INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Mechanism of Physiological Function of Sphingosine-1-Phosphate: Extracellular Action and Demonstration of Alleged Receptor

by

Soichiro Yamamura

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington
1998

Approved by 
(Chairperson of the Supervisory Committee)

Program Authorized
to Offer Degree Department of Pathobiology

Date 8/13/98

UMI Number: 9907966

UMI Microform 9907966
Copyright 1998, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized copying under Title 17, United States Code.

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O., Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microfilm and/or (b) printed copies of the manuscript made from microfilm"

Signature 

Date 8/13/98

University of Washington

Abstract

Mechanism of Physiological Function of Sphingosine-1-Phosphate: Extracellular Action and Demonstration of Alleged Receptor

by

Soichiro Yamamura

Chairperson of the Supervisory Committee:
Professor Sen-itiroh Hakomori
Department of Pathobiology

Sphingosine-1-phosphate (Sph-1-P), the first product of sphingosine (Sph) catabolism, causes changes of cell physiology when exogenously added to culture medium. Cellular response varies depending on type of cell. For example: (i) 10-20 μM exogenous Sph-1-P causes shape change and aggregation of platelets. (ii) 10-100 nM exogenous Sph-1-P inhibits motility of B16 mouse melanoma and various other tumor cells. (iii) 5-10 μM exogenous Sph-1-P stimulates proliferation of Swiss 3T3 fibroblasts.

Platelets lack Sph-1-P lyase activity, possess persistently active Sph kinase, and therefore abundantly store Sph-1-P. Exogenous Sph-1-P promptly activates platelets, even though the quantity of Sph-1-P transported into cells is negligible as compared to the high level of stored intracellular Sph-1-P. This suggests that the initial action of Sph-1-P is based on binding of this agonist at a cell surface receptor. This concept was applied to study the initial effects of exogenous Sph-

1-P causing two processes: platelet activation and melanoma cell motility inhibition.

Contact of cell surface with immobilized Sph-1-P whose ω-carbon is covalently linked to porous glass beads resulted in activation of platelets, and motility inhibition of melanoma cells. Binding assays with radiolabeled Sph-1-P revealed the presence of specific Sph-1-P binding sites on platelets and melanoma cells. These results suggest that Sph-1-P acts extracellularly on these cells through surface binding sites.

Sph-1-P inhibits cell motility at very low concentrations (10-100 nM) which may be physiologically relevant. Internally stored Sph-1-P was released from platelets and other cells upon stimulation, and concentration of Sph-1-P in human plasma was high enough to control cell motility. Thus, Sph-1-P may act as an intercellular (cell-to-cell) modulator or messenger in regulation of cell motility.

TABLE OF CONTENTS

List of Figures.....ii

List of Tables.....iv

Chapter 1: Introduction: Structural and Metabolic Features, and General Function Role of Sphingolipid.....1

Chapter 2: Sphingosine-1-Phosphate Induces Platelet Activation through Binding to its Receptor, which is Shared with Lysophosphatidic Acid.....14
Introduction.....14
Materials and Methods.....15
Results.....22
Discussion.....27

Chapter 3: Sphingosine-1-phosphate Regulates Melanoma Cell Motility through a Receptor-coupled Extracellular Action.....50
Introduction.....50
Materials and Methods.....51
Results.....56
Discussion.....63

Chapter 4: Concluding Remarks.....81

Bibliography.....86

Appendix: Mechanism for inhibition of haptotactic motility by sphingosine-1-phosphate.....101

LIST OF FIGURES

Figure 1.1(A) Conformational structure of a typical GSL antigen. (B) Structure and organization of glycosphigolipid antigens at the cell surface.....10

Figure 1.2. Biosynthesis of sphingolipids from serine to glucosylceramide.....11

Figure 1.3. Biosynthesis of Sph-1-P and its degradation.....12

Figure 2.1. Structure of the conjugates of Sph-1-P, 1, and Sph, 2, with controlled pore glass (CPG) beads used in this study.....32

Figure 2.2. Scheme for preparation of Sph-1-P and Sph immobilized on controlled pore glass beads.....33

Figure 2.3. Synthesis of [3-^{3}H]Sph-1-P.....34

Figure 2.4. Sph-1-P induced human platelet aggregation (A) and shape change (B).....35

Figure 2.5. Concentration-dependent platelet aggregation and shape change induced by Sph-1-P.....36

Figure 2.6. Scanning electron microscopy of shape-changed platelets.....37

Figure 2.7. Sph-1-P-induced changes in platelet intracellular Ca^{2+} concentration.....38

Figure 2.8. Platelet activation by exogenously-added Sph-1-P, but not Sph, in spite of extensive conversion of added Sph into Sph-1-P.....39

Figure 2.9. Platelet activation induced by Sph-1-P immobilized on a solid support.....40

Figure 2.10. Platelet intracellular Ca^{2+} mobilization induced by Sph-1-P immobilized on a solid support.....41

Figure 2.11. Specific binding of [^{3}H]Sph-1-P to platelets.....42

Figure 2.12. Effect of various lipids on [^{3}H]Sph-1-P binding to platelets.....43

Figure 2.13. Desensitization of platelet aggregation response to LPA by prior addition of Sph-1-P.....44

Figure 2.14. TPA-induced Sph-1-P release.....45

Figure 2.15. Thrombin-, TPA-, or 1-oleoyl-2-acetyl-glycerol-induced Sph-1-P release from platelets and its inhibition by staurosporine.....46

Figure 2.16 Sph-1-P as an intracellular second messenger.....47

Figure 2.17 Sph-1-P as an intercellular messenger.....48

Figure 3.1. Effect of immobilized sphingolipids on F10 cell motility.....69

Figure 3.2. Specific binding of [^{3}H]Sph-1-P to F10 cell surface.....70

Figure 3.3. Effect of various lipids on [^{3}H]Sph-1-P binding to F10 cells.....71

Figure 3.4. Effect of Sph-1-P, DMS-1-P and LPA on F10 cell motility.....72

Figure 3.5. Effect of suramin on inhibition of cell motility by 0.1 μM Sph-1-P.....73

Figure 3.6. Cell motility inhibition by Sph-1-P is not sensitive to pertussis toxin treatment.....74

Figure 3.7. Identification of the Sph-1-P binding protein.....75

Figure 3.8. Metabolism of [3-^{3}H]Sph in A31 cells and the release of ^{3}H-labeled Sph-1-P into the medium containing 10 % BCS.....76

Figure 3.9. Metabolism of [3-^{3}H]Sph in F10 cells and release of ^{3}H-labeled Sph-1-P into the medium containing 10 % FCS.....77

Figure. 3.10. Determination of the Sph-1-P levels in human plasma and serum.....78

Figure 3.11. Sph-1-P as an intercellular messenger which regulates cell motility.....79

LIST OF TABLES

Table I. Function of sphingolipids and glycosphingolipids.....13

Table II. Sph-1-P lyase activity.....49

Table III. Sph-1-P levels in human platelets, neutrophils, and erythrocytes.....80

ACKNOWLEDGMENTS

I wish to express my gratitude to Professor Sen-itiroh Hakomori for his continued support over the years, and my sincere appreciation to Dr. Yasuyuki Igarashi for all his guidance. I appreciated all the assistance I received from Drs. William G. Carter, John A. Glomset, Ted Kuo,George M. Martin and Eline W. Raines. I also would like to thank the members of The Biomembrane Institute and Dr. Hakomori's lab.

Chapter 1

Introduction: Structural and Metabolic Features, and General Functional Role of Sphingolipids

Glycerophospholipids are the major structural components of cell membrane, although some of their metabolites (*e.g.* diacylglycerol (DAG), lyso-phosphatidyl derivatives, inositol triphosphate, arachidonic acid and its derivatives) are functionally important as second messengers in signal transduction leading to cell proliferation, differentiation, and apoptosis (Nishizuka, 1986; Berridge, 1988; Exton, 1990). In contrast, sphingolipids and glycosphingolipids (GSLs) are classically considered as structural components of cell surface membrane, and carbohydrate residues of GSLs are known as cell type-specific antigens (Hakomori, 1981a), and receptors for toxins and microbial infection (Karlsson, 1989).

The idea that GSLs, particularly gangliosides, act as modulators of signal transduction through interaction with growth factor receptors or protein kinase C (PKC) originally suggested in the early 1980s (Bremer and Hakomori, 1982; 1984; Kreutter, 1987). In association with this line of study, sphingosine (Sph), the backbone lipid component of all sphingolipids (see below), was found as an inhibitor of PKC, in contrast to the stimulatory effect of DAG (Hannun et al., 1986). The effect of stereoisomers and derivatives of Sph and D-erythro-N,N-dimethyl-Sph (DMS) on PKC is stronger than that of other stereoisomers (Igarashi et al., 1989). In Swiss 3T3 cells, exogenous Sph stimulates cell proliferation, probably through its conversion to Sph-1-phosphate (Sph-1-P) (Zhang et al., 1990). A series of subsequent studies led to the concept that Sph derivatives, including Sph-1-P, play essential roles as a second messenger in signal transduction (Zhang et al. 1991; Olivera and Spiegel, 1993).

Ceramide (Cer) has been also shown to be a second messenger in signal transduction. Extracellular agents such as interleukin-1, tumor necrosis factor (TNF) and Fas ligand are capable of activating sphingomyelinase, and generate Cer by sphingomyelin hydrolysis (Hannun, 1994; Kolesnick and Golde, 1994). Cer has been shown to mediate inflammatory and apoptotic responses to TNFα. It activates a specific type of protein kinase (ceramide-activated kinase), initiating signaling of mitogen-activated protein kinase cascade in the inflammatory response (Yao et al., 1995). TNFα induces hydrolysis of sphingomyelin to Cer, which causes apoptosis. Cer has been also shown to induce Fas-mediated apoptosis (Cifone et al., 1993). Cer has been recently claimed to function as a "biostat" controlling homeostasis of intercellular environment during cellular response to stress, rather than as a second messenger, because it does not function solely in signal transduction, and duration of Cer change in response to cellular stimulus is much longer than that observed for typical second messengers such as IP_3, PIP_3 and cyclic AMP (Hannun, 1996).

Although interpretations vary, sphingolipid catabolites also function as modulators of signal transduction or as second messengers, in analogy to phospholipid catabolites (for reviews see Hannun and Bell, 1989; Hakomori, 1990; Liscovitch and Cantley, 1994; Spiegel and Milstien, 1995; Hakomori and Igarashi, 1995). Functions of sphingolipids and GSLs are summarized in Table I.

Structural and functional features of sphingolipids

The characteristic structural feature of sphingolipids is the presence of Sph and sphingoid bases. The common structure for Sph and other sphingoid bases (dihydro-Sph, phyto-Sph, eicosa-Sph, etc.) is aliphatic amino alcohol (1,3-dihydroxy-2-amino-alkane or -alkene) having D-erythro configuration as head group, and with 4,5-trans-olefinic structure for alkene. The amino group of Sph

is acylated with fatty acid with or without α-hydroxyl group, and with varying chain length, to form Cer. The primary hydroxyl group (1-hydroxy) is glycosylated to form GSLs or linked to the phosphate group of phosphorylcholine to form phosphosphingolipid, *i.e.* sphingomyelin.

Conformational structure of a typical GSL antigen is shown in Figure 1.1A. Note that the axes of the two hydrophobic tails, *i.e.* Cer, are oriented perpendicular to the carbohydrate chain axis. Such GSLs, and perhaps sphingolipids in general, have a strong ability to self-assemble to form microdomains separated from glycoprotein clusters at the cell surface membrane (Figure 1.1B) (Iwabuchi et al., 1998).

Sphingolipids are essential structural components of animal and plant cell membranes, and for a specific bacteria (Sphingobacterium). They are considered to provide greater stability and rigidity of the membrane than glycerides or glycerophospholipids. Many recent studies indicate that sphingolipids in membrane form clustered microdomains which are functionally distinct from glycerophospholipids. Stability and rigidity of sphingolipid-enriched microdomain are conferred by Cer, *i.e.* the presence of long-chain, saturated aliphatic residue which form an "ordered liquid phase" with restricted fluidity in membrane. In contrast, "disordered liquid phase" is associated with high fluidity caused by a predominance of lipids with unsaturated fatty acids (Schroeder et al. 1994; Ahmed et al. 1997; for review see Brown & London 1997). Microdomains enriched in glycosphingolipids are closely associated with various protein kinases involved in cell adhesion coupled with signal transduction (Iwabuchi et al. 1998).

Cellular morphology depends highly on the state of organization of sphingolipids in membrane. Myelin sheath, a characteristic component of neuronal cells, consists of multilayered plasma membrane. The multilayered

structure may be based on interaction of GalCer and sulfatide (3-O-sulfated GalCer), which are abundant in myelin sheath membrane (Stewart and Boggs, 1993a). Polarity of epithelial cells has been explained by the distinct distribution patterns of sphingolipids vs. glycerophospholipids. I.e. the apical surface is enriched in sphingolipids, while the basolateral surface is enriched in glycerophospholipids (Simons and van Meer, 1988; Brown and Rose, 1992).

Sphingomyelin is a major source of Cer, Sph, and their derivatives, which play essential roles in signal transduction, as described above and in Table I. Various functional roles of GSLs as ligand receptors and as modulators of receptor function are also listed in Table I. I give references for each functional role of GSLs, but will not explain them in detail, since these are not within the topic of my thesis.

Metabolic features of sphingolipids

I will describe briefly the biosynthesis and biodegradation of sphingolipids in general (see Figure 1.2). Biosynthesis of sphingosine and other sphingoid bases is initiated by condensation of palmitoyl-CoA with serine, catalyzed by the serine: palmitoyltransferase, leading to formation of 3-ketosphinganine (Figure 1.2). 3-Ketosphinganine is reduced to sphinganine by a NADPH-dependent oxidoreductase. Sphinganine thus formed is converted to dihydroceramide (N-acyl-sphinganine) by transfer of fatty acid from fatty acyl-CoA: sphingoid base N-acyltransferase (dihydroceramide synthase). The enzyme appears to be present in multiple forms, having different substrate specificity and tissue distribution. Subsequently, unsaturated bond (double bond) is introduced at the 4,5 position of dihydrosphingosine moiety in dihydroceramide by a FAD-dependent oxidoreductase, leading to formation of Cer (Rother et al., 1992). It should be noted that double bond introduction does not occur in sphingosine, but

does occur in dihydroceramide. In rodent liver, all the enzymes involved in *de novo* biosynthesis of Cer occur in the smooth endoplasmic reticulum and appear to expose their active sites at the cytosolic face (Mandon et al., 1992), the only exception being dihydroceramide oxidoreductase, whose topology has not yet been established. The acyl-CoA:sphingoid base acyl transferase, which in the *de novo* synthesis pathway uses sphinganine as acceptor, is also capable of promoting the acylation of Sph, if it is available, both *in vitro* and *in vivo*.

A small amount of free Sph (C-18 and C-20 species), as well as sphinganine, can be detected in cells. The origin of free Sph is catabolic: Cer is split into Sph and fatty acid by ceramidase. A further metabolic engagement of sphingoid bases suggested for C18-Sph is N-methylation, with the formation of the monomethyl- and dimethyl derivative (Igarashi and Hakomori, 1989). Sph-N-methyltransferase seems to be a membrane-bound (possibly microsomal) enzyme, although it was not further characterized. DMS is detectable in small quantity in HL60 cells by systematic HPLC and mass spectrometry (Mano et al., 1997).

The addition of carbohydrate groups from UDP-sugar to Cer forms complex GSLs. Glucosyl-Cer is biosynthesized from Cer and UDP-glucose by a Cer glucosyl transferase (Matsuo et al., 1992). This enzyme resides on the cytoplasmic face of membranes in the cis-Golgi and, possibly, a pre-Golgi compartment. The major metabolic involvement of glucosyl-Cer is to be further glycosylation, yielding more complex glycosphingolipids, like gangliosides. This process takes place in the luminal face of the cis-Golgi stacks (Burger et al., 1996) and requires translocation of glucosyl-Cer from the cytoplasmic to the luminal face of the Golgi membrane.

The main pathway of sphingomyelin biosynthesis in mammalian cells is based on the transfer of phosphocholine from phosphatidylcholine to Cer,

catalyzed by sphingomyelin synthase. This enzyme appears to be located mainly in the cis- and medial Golgi membrane, with the active site exposed on the luminal leaflet (Lipsky and Pagano, 1983). An alternative mechanism for sphingomyelin biosynthesis consists of the transfer of phosphocholine from CDP-choline to Cer.

Sphingosine-1-phosphate and its physiological function

Sphingosine-1-phosphate (Sph-1-P) is the initial product of Sph catabolism by Sph kinase (Stoffel et al., 1973), and is cleaved by Sph-1-P lyase to phosphoethanolamine and palmitaldehyde (Stoffel, 1970) (Figure 1.3).

Nearly 20 years later, this catabolite is in the forefront of cell biological research because of its essential functional role in signal transduction. This is based on four lines of study:

i. Exogenous addition of an enzymatically converted product of Sph, presumably Sph-1-P, caused high Ca^{2+} mobilization in smooth muscle cell, including the inositol 1,4,5-triphosphate-sensitive Ca^{2+} pool (Ghosh et al., 1990). Sph-1-P generated in endoplasmic reticulum membrane caused release of stored Ca^{2+} (Ghosh et al., 1994).

ii. Exogenous Sph causes mitogenesis of Swiss 3T3 cells through a PKC-independent pathway (Zhang et al., 1990). This effect was shown to be due to conversion of Sph to Sph-1-P, which is responsible for cellular proliferation (Zhang et al., 1991). This catabolite also causes increased phosphatidic acid level through phospholipase D activation (Desai et al., 1992), and causes Ca^{2+} mobilization independent from inositol triphosphate pathway (Mattie et al., 1994). Sph kinase, a key enzyme controlling Sph-1-P level, is up-regulated upon proliferative stimulation of Swiss 3T3 cells by platelet-derived growth factor (PDGF) (Olivera and Spiegel, 1993), thus establishing the idea that Sph-1-P may

act as a second messenger following PDGF stimulation. Similar enhancement of Sph kinase activity was observed following TPA stimulation of Balb/c 3T3 A31 cells (Mazurek et al., 1994). Later, Ca^{2+} mobilization following stimulation of IgE receptor (FcεRI antigen) was found to be mediated by Sph kinase, resulting in enhanced Sph-1-P level (Choi et al., 1996). Further studies on the mechanism of Sph-1-P as a basis of mitogenesis indicate that this catabolite enhances DNA-binding activity of AP-1 (Su et al., 1994), and that Sph-1-P-dependent signaling pathway involves the pertussis toxin-sensitive GTP-binding protein (Goodemote et al., 1995; Wu et al., 1995). Recently, Sph kinase was extensively purified and characterized from rat kidney (Olivera et al., 1998).

iii. Distinct from lines of study (i) and (ii), the effect of Sph-1-P on platelets to cause shape change and aggregation reaction is novel, since Sph-1-P is present at unusually high level in these anuclear cells even at resting state. Exogenous addition of Sph-1-P (10-20 μM range) causes immediate shape change and aggregation, similar to the effect of weak platelet agonists such as adenosine diphosphate and epinephrine. Subthreshold concentrations of Sph-1-P and these weak agonists activate platelets synergistically (Yatomi et al., 1995). Sph-1-P mobilized intracellular Ca^{2+}, and the dose response for Ca^{2+} release was correlated closely with the concentration required for induction of shape change. How exogenous Sph-1-P causes platelet activation is an enigma, since the intracellular presence of this catabolite may not be essential for this process. To solve this enigma, the hypothesis was proposed that Sph-1-P acts only extracellularly, and does not increase Sph-1-P level intracellularly.

iv. Another remarkable effect of Sph-1-P, distinguishable from mitogenesis or platelet activation, is that it inhibits haptotactic and chemotactic motility of mouse melanoma B16, human arterial smooth muscle cells and many other types of cells at very low concentrations (10-100 nM). No other Sph-

related compounds are known to inhibit cell motility. Sph-1-P does not affect proliferation of mouse melanoma B16 or human arterial smooth muscle cells (Sadahira et al., 1992, Bornfeldt et al., 1995), but does inhibit integrin-dependent motility of mouse melanoma B16 cells (Sadahira et al., 1994). It also inhibits pseudopodium formation by blocking polymerization and reorganization of actin filaments in newly formed pseudopodia, and reduced F-actin by ~ 25% in B16 F1 cells. A pyrene-labeled actin nucleation assay revealed that Sph-1-P inhibits actin nucleation mediated by F1 cell plasma membranes (Yamamura et al., 1996).

Although phenotypic changes caused by Sph-1-P vary extensively, the initial common step is its effect at the cell surface. Two mechanisms for the initial effect of Sph-1-P can be considered: (a) Sph-1-P enters the cell without a specific receptor and affects signal transduction. (b) Sph-1-P binds initially to a specific cell surface receptor which triggers signal transduction leading to various phenotypic changes. Sph-1-P is released to varying degrees from various types of cells, and may act as an extracellular messenger to modulate cellular functions. Metabolism of Sph-1-P in platelets (Chapter 2) also suggests existence of a Sph-1-P receptor. I therefore studied the possibility of existence of a Sph-1-P receptor on platelets and B16 melanoma cells.

In Chapter 2, I describe the metabolism of Sph-1-P and mechanism by which Sph-1-P induces activation in anuclear human platelets. Platelets lack Sph-1-P lyase activity, possess persistently active Sph kinase, and abundantly store Sph-1-P. Although exogenous Sph-1-P activated platelets, intracellular Sph-1-P formed from exogenously-added Sph by cytosolic Sph kinase did not. Contact of platelet surface with immobilized Sph-1-P whose ω-carbon is covalently linked to glass beads resulted in platelet activation. Finally, I found specific binding sites for radiolabeled

Sph-1-P on the platelet surface, and found that internally stored Sph-1-P is released from platelets upon platelet stimulation. These observations suggest that a binding receptor for Sph-1-P is present at the platelet plasma membrane.

In Chapter 3, the concept of the extracellular action of Sph-1-P in platelet activation was applied to cell motility inhibition, and the possibility of a physiological role of Sph-1-P as a regulator of cell motility, given that Sph-1-P at very low concentrations (10-100 nM) inhibits cell motility. I identified and characterized specific cell surface binding sites for Sph-1-P in mouse melanoma cells, demonstrating that Sph-1-P may act through the binding protein (receptor) in the cell motility regulation. Pertussis toxin did not abolish the inhibitory effect of Sph-1-P on cell motility, suggesting that the signaling pathways do not involve a pertussis toxin-sensitive GTP-binding protein (G protein). Sph-1-P itself was released into culture medium from cells upon serum stimulation, and concentration of Sph-1-P in human plasma was high enough to control cell motility. These results suggest that Sph-1-P can act as an intercellular modulator or messenger in regulation of cell motility. The results in these Chapters do not exclude mechanism (a) Sph-1-P enters the cells without a specific receptor and affects signal transduction.

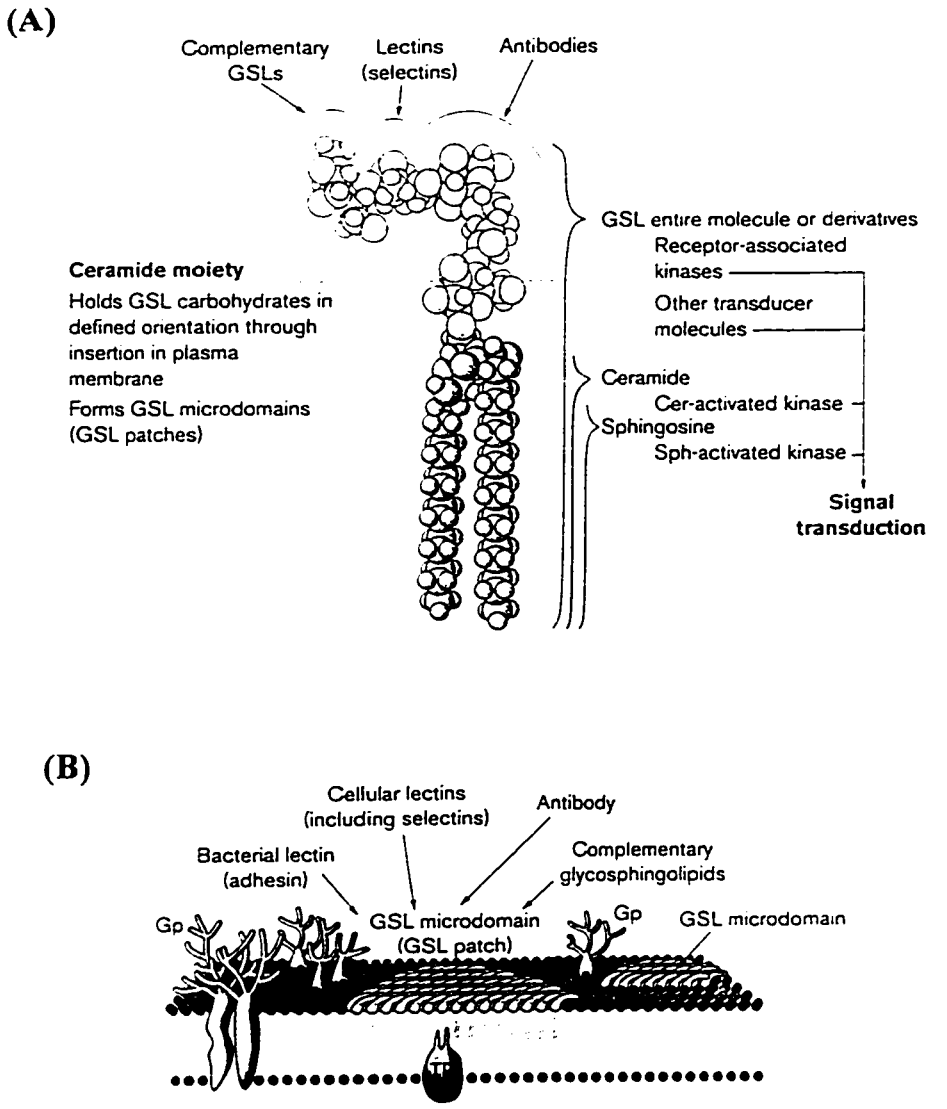
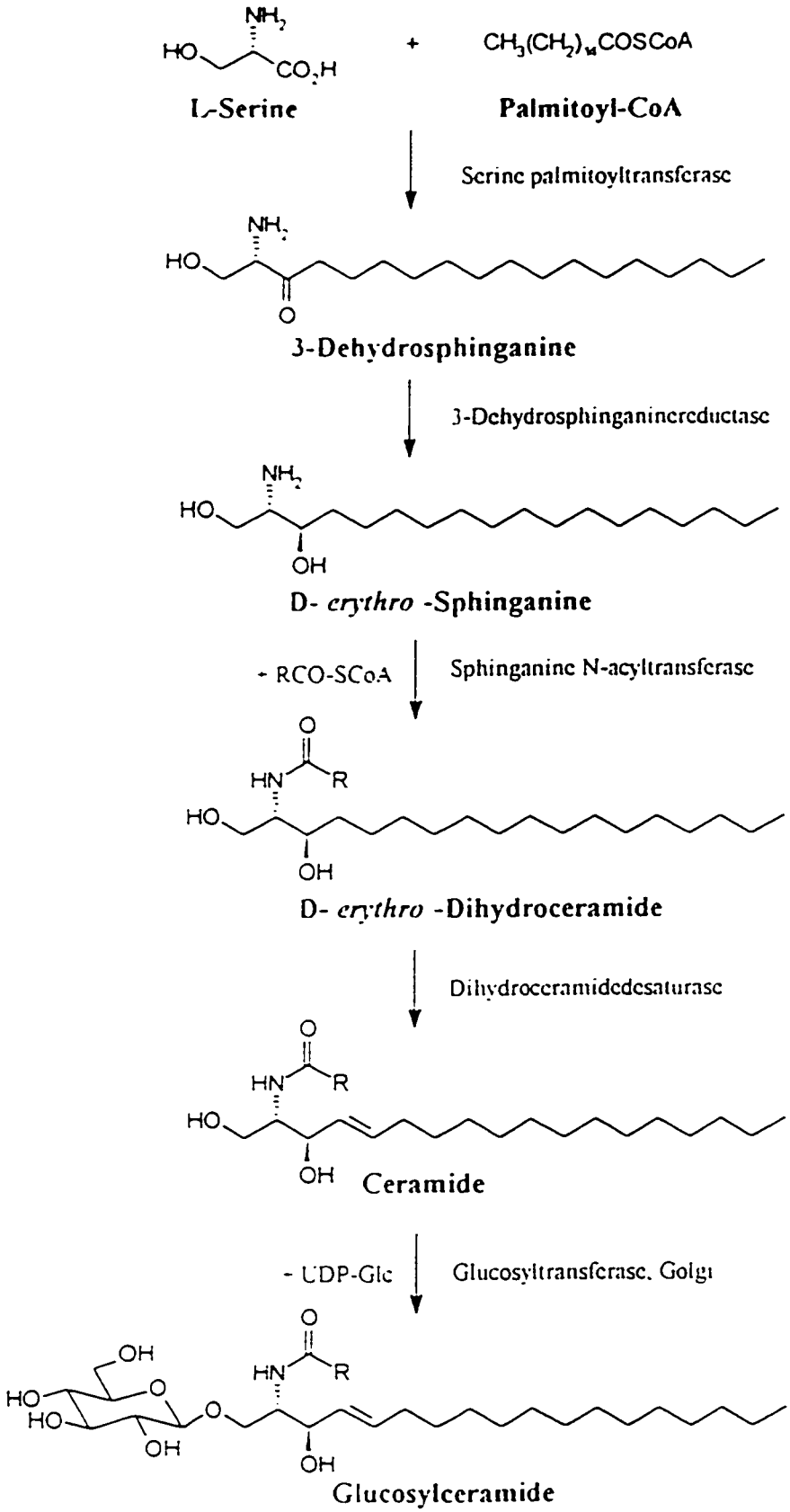


Fig. 1.1(A) Conformational structure of a typical GSL antigen. The axes of the two hydrophobic tails, *i.e.* Cer, are oriented perpendicular to the carbohydrate chain axis. **(B) Structure and organization of glycosphigolipid antigens at the cell surface.** GSLs inserted in the plasma membrane tend to form GSL microdomains. Glycoproteins (Gp) are arranged in clusters that are separated from GSL microdomains. Transducer molecules (TD) such as Src, Ras and Rho are sometimes associated with GSL microdomains.

Fig. 1.2. Biosynthesis of sphingolipids from serine to glucosylceramide.


L-Serine
+
Palmitoyl-CoA
Serine palmitoyltransferase
3-Dehydrosphinganine
3-Dehydrosphinganinereductase
D- erythro -Sphinganine
+ RCO-SCoA
Sphinganine N-acyltransferase
D- erythro -Dihydroceramide
Dihydroceramidedesaturase
Ceramide
+ UDP-Glc
Glucosyltransferase, Golgi
Glucosylceramide

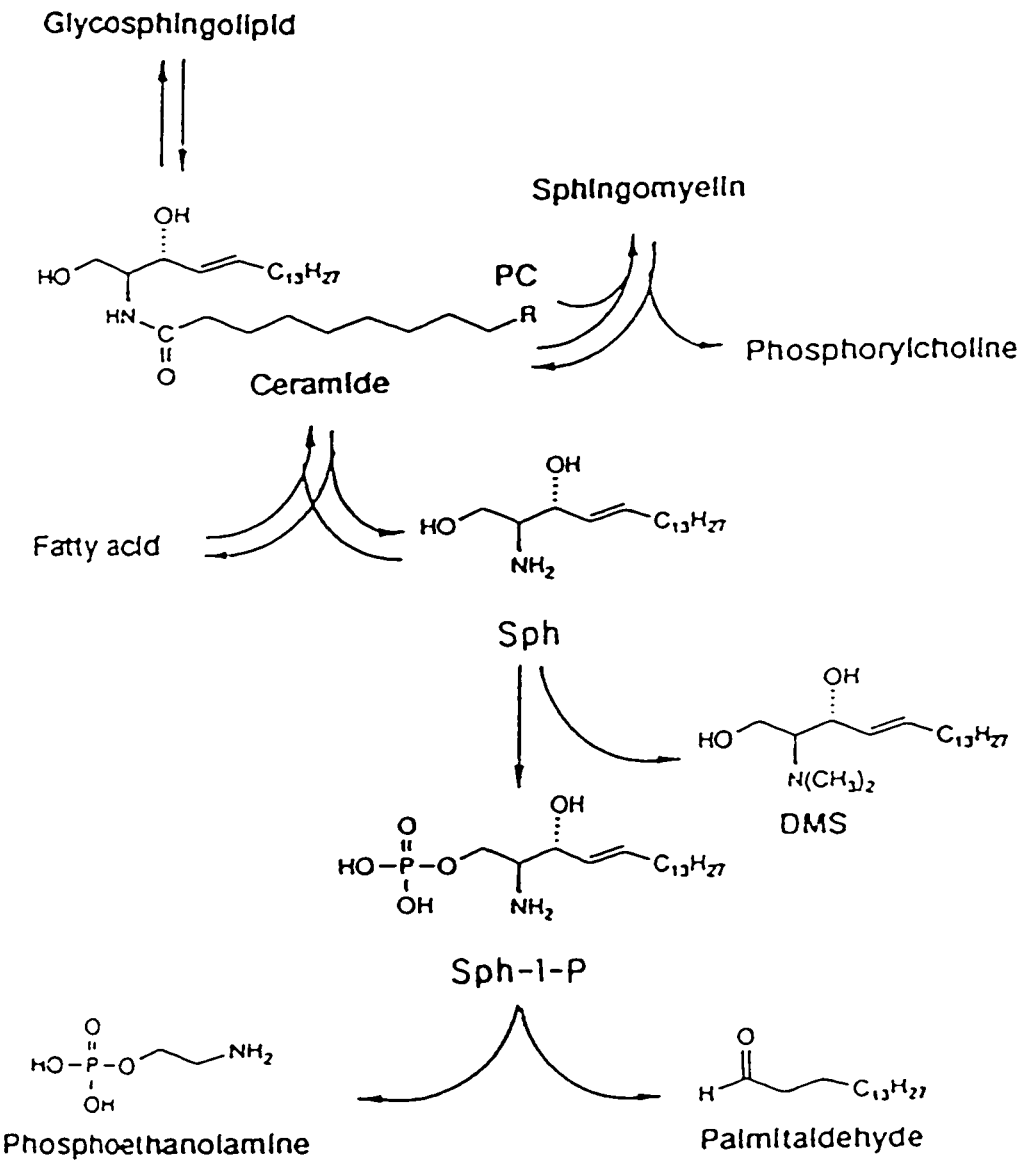


Fig. 1.3. Biosynthesis of Sph-1-P and its degradation.

Table I. Function of sphingolipids and glycosphingolipids

Sphingomyelin (phosphosphingolipid)

- Structural component → rigid clustered structure with cholesterol, forming microdomain in plasma membrane (Brown and London, 1997)
- Yields ceramide, sphingosine, and Sph-1-P, functioning as second messengers or modulators in signal transduction (Hannun and Linardic, 1993; Igarashi, 1997; Merrill and Stevens, 1989; Spiegel *et al.*, 1996; Spiegel and Milstien, 1995)

Glycosphingolipids

- Ligand receptors
  - Antigens (histo-blood group ABH, PP_1P^kp, Le^a/Le^b, I/i, heterophile Forssman (Hakomori, 1981a; Marcus *et al.*, 1981)
  - Oncodevelopmental and tumor-associated antigens (Hakomori, 1984; Hakomori, 1989)
  - Galectin/ selectin ligands (Gabius and Gabius, 1997; Handa *et al.*, 1997)
  - Cell-cell adhesion through GSL-GSL interaction (Eggens et al., 1989; Kojima et al., 1991; Stewart and Boggs, 1993; Boubelik et al., 1998)
  - Bacterial adhesion → infection (Karlsson, 1989)
  - Toxin receptors (Hakomori, 1981b)
- Modulators of receptor function
  - Growth factor receptors (Hakomori, 1990; Hakomori and Igarashi, 1995)
  - Integrin receptor modulators (Zheng *et al.*, 1993)

Chapter 2

Sphingosine-1-Phosphate Induces Platelet Activation through Binding to its Receptor, which is Shared with Lysophosphatidic Acid

Introduction

Platelets, novel anuclear cells, are highly sensitive to various environmental stimuli. Their response (*e.g.* shape change, Ca^{2+} mobilization, "secretory response") varies depending on strength and quality of stimulus. Platelets play a central role in blood coagulation, maintenance of homeostasis, and vascular and hematopoietic functions. A feature of sphingolipid metabolism of platelets is the presence of an unusually high level of Sph-1-P, in contrast to the extremely low level found in various other types of cells (fibroblasts, epithelial cells, blood cells, etc.). Synthesis of Sph-1-P, the first step in catabolism of Sph, is catalyzed by Sph kinase. Thus, both Sph kinase and Sph-1-P lyase (the enzyme which cleaves Sph-1-P to produce palmital and phosphoethanolamine) are rate-limiting factors in Sph catabolism (Stoffel and Bister 1973) (see "Metabolic features of sphingolipids" in Chapter 1). Activity level of Sph kinase in platelets appears to be normal. The high level of Sph-1-P in platelets appears to be ascribable to the fact that platelets have a minimal (undetectable) level of Sph-1-P lyase.

Intracellular Sph-1-P level is transiently enhanced after stimulation of specific types of cells by agonists (*e.g.* Swiss 3T3 cells by PDGF; rat mast cells by IgE), with consequent phenotypic changes (see Chapter 1). These observations lead to the concept that Sph-1-P may act as a second messenger in response to cell stimuli, although the exact mechanism remains to be explored. This concept is difficult to apply to platelets, since a high level of intracellular Sph-1-P is present in resting platelets (see Chapter 1).

I found that exogenous Sph-1-P induces mild platelet activation in terms of shape change and intracellular Ca^{2+} mobilization. Since this may not due to enhancement of intracellular Sph-1-P level, a plausible possibility is that Sph-1-P may bind to a specific site at the platelet surface membrane, which induces Ca^{2+} mobilization and other events leading to signal transduction. The pre-existing high level of intracellular Sph-1-P may have no function in signal transduction. This mechanism appears to be different from the effect of Sph-1-P on Swiss 3T3 cells or other types of cells.

I utilized several approaches to solve this problem, as follows:

i. Application of immobilized Sph-1-P whose head group is exposed to contact the cell surface and whose ω-carbon is linked covalently to amino group affixed to porous glass surface. Similarly immobilized Sph (without phosphate) was used as control. Contact of head group of Sph-1-P, but not of Sph, caused platelet stimulation of terms of shape change and intracellular Ca^{2+} mobilization.

ii. Detection of specific binding site at platelet surface membrane for ^{3}H-labeled Sph-1-P.

iii. Binding of ^{3}H-labeled Sph-1-P to platelet surface and Sph-1-P-dependent platelet aggregation were inhibited by lysophosphatidic acid (LPA) whose molecular conformation is similar to that of Sph-1-P.

iv. Synthesis of Sph-1-P, absence of Sph-1-P lyase (aldolase), and unusual accumulation of Sph-1-P in platelets were demonstrated by biochemical/ metabolic analysis employing ^{3}H-Sph. Release of Sph-1-P from platelets upon activation was also demonstrated.

Materials and Methods

Materials

Sph-1-P was synthesized chemically (Sadahira et al, 1994). Sph, was synthesized as previously described (Igarashi et al, 1989; Toyokuni et al, 1991). C2-ceramide (C2-Cer) was synthesized according to the literature (Vunnam and Radin, 1979). These sphingolipids were dissolved in ethanol/water (1:1), and stored at -20°C until required. All other reagents were purchased from Sigma (St. Louis, MO).

Preparation of platelets

Citrated venous blood was obtained from healthy adult volunteers. The blood was anticoagulated with 3.8% sodium citrate (9 vol of blood for 1 vol of sodium citrate) and centrifuged at 120 g for 10 min. The platelets obtained were washed and finally resuspended in a buffer containing 138 mM NaCl, 3.3 mM NaH_2PO_4, 2.9 mM KCl, 1.0 mM $MgCl_2$, 1 mg/ml of glucose and 20 mM HEPES (pH 7.4). The suspensions were adjusted to 3 x 10^8/ml and supplemented with 1 mM $CaCl_2$ unless otherwise stated. Bovine serum albumin (fatty acid-free) (1%) was added when indicated. For shape change studies, the platelets, after final centrifugation, were left for at least 60 min at 37°C because this resulted in greater response. All experiments using intact platelet suspensions were performed at 37°C.

For Sph-1-P lyase activity assay, outdated platelet concentrates obtained from Oregon Red Cross (Portland, OR) were used.

Sph-1-P extraction from platelets

Sph-1-P was extracted from platelets suspended in 0.5 ml of the buffer described above. 3 ml of ice-cold chloroform/methanol (1:2) was added to the samples, followed by thorough mixing and sonication for 30 min. Phases were

separated by adding 2 ml chloroform, 2 ml 1 M KCl, and 100 μl 7N NH_4OH. The alkaline upper phases were transferred to new tubes, to which 3 ml chloroform and 200 μl concentrated HCl were added. The lower chloroform phases, formed under these new acidic conditions, were evaporated under N_2 and assayed for Sph-1-P as described below. The initial lower chloroform phases were dried and used for a quantitative assay of Sph.

Quantitative Measurements of Sph and Sph-1-P

Sph (Ohta et al., 1994) and Sph-1-P (Yatomi et al., 1995) were quantitatively measured by N-acetylation with [^{3}H]acetic anhydride into [^{3}H]C2-Cer (N-[^{3}H]acetylated Sph) and [^{3}H]C2-Cer-1-P (N-[^{3}H]acetylated Sph-1-P), respectively, as described previously. Dried samples were dissolved in 40 μl 0.008N NaOH in methanol/10 mM solution of [^{3}H]acetic anhydride (1:1) by sonication. Acetylation reactions were allowed to proceed at 37°C for 2 h. The remaining anhydride was hydrolyzed by addition of 0.2 ml of 0.2N NaOH in methanol. Following a 1-h incubation at room temperature, the C2-Cer or C2-Cer-1-P formed was extracted by addition of 0.78 ml methanol, 0.98 ml chloroform, 0.9 ml 1 M KCl and 0.78 ml methanol, 0.98 ml chloroform, 0.9 ml 1 M KCl and 20 μl concentrated HCl, respectively. The resultant lower chloroform phase was washed twice by 1 ml of chloroform/methanol/water (3:48:47) for C2-Cer, and 1 ml of chloroform/methanol/water (3:48:47) plus 10 μl concentrated HCl for C2-Cer-1-P. Samples from the chloroform phase were evaporated under N_2 and then resuspended in small volumes of chloroform/methanol (2:1). Portions of the lipids obtained, with appropriate standards, were applied to silica gel 60 HPTLC plates (Merck, Darmstadt, Germany) and the plates were developed in chloroform/methanol/7N

NH_4OH/water (80:20:0.5:0.5) or butanol/acetic acid/water (3:1:1) for C2-Cer or C2-Cer-1-P, respectively. After enhancer (Resolution TLC; E.M. corp., Chestnut Hill, MA) treatment of the plate, autoradiography was conducted with Kodak X-Omat film (Eastman Kodak, Rochester, NY). The silica gel areas which contained the radiolabeled sphingolipids were scraped off and the radioactivity was counted with a liquid scintillator.

Platelet aggregation and shape change

Platelet aggregation and shape change were determined turbidometrically (Zucker, 1989), using Platelet Ionized Calcium Aggregometer (ChronoLog, Havertown, PA) with stirring at 1,000 rpm. Calibration was performed with zero light transmission defined for platelet suspension and 100% transmission for the buffer. When aggregation was measured, human fibrinogen (500 μg/ml) was added to platelet suspensions shortly before the addition of stimuli.

Shape change was observed by adding 5 mmol/l EDTA (instead of Ca^{2+}) before administration of stimuli to prevent aggregation and indicated by a decrease in light transmission (Zucker, 1989).

Measurement of intracellular Ca^{2+} concentration

Platelet-rich plasma was incubated at 37 °C with 3 μM-fura2-AM for 30 min. The platelets were then washed twice and resuspended in Hepes-Tyrode's buffer. Intracellular Ca^{2+} concentration was measured by a Model LS 50B Luminescence Spectrometer (Perkin-Elmer Ltd., Beaconsfield, England) with a program for measurement of intracellular Ca^{2+} concentration (The Intracellular Biochemistry Application). An excitation and emission wavelengths were 340/380 nm and 500 nm, respectively. Values of peak increases after addition of an agent were quantified.

Sph-1-P lyase activity assay

Sph-1-P lyase activity was measured as described previously (Van Veldhoven and Mannaerts, 1991) except that [3-^{3}H]Sph-1-P instead of [4,5-^{3}H]dihydrosphingosine-1-phosphate was used as a substrate. The reaction products were applied to silica gel 60 HPTLC plates (Merck, Darmstadt, Germany), and the plates were developed in chloroform/ methanol/ acetic acid (50: 50: 1). In this polar solvent, all radioactive lyase metabolites (palmitaldehyde, palmitic acid, and palmitol) ran closely together near the front, and the whole region was scraped off and counted by liquid scintilation counting.

Preparation and application of Sph-1-P or Sph immobilized on a solid support

To test the extracellular effect of Sph-1-P and Sph on platelet function, it is necessary to immobilize these compounds through their ω-carbon. Considerable effort was made to synthesize such structures, in which head groups are accessible for contact with the platelet surface. Structures designated as conjugates **1** and **2** in Figure 2.1 were synthesized by creation of ω-carboxyl group in Sph-1-P or Sph, covalently linked to the surface of controlled pore glass beads (CPG Inc., Fairfield, NJ) through an alkyl amine. The synthesis of **1** is summarized in Figure 2.2. Basically, the allylic alcohol **3**, readily obtained by our approach (Ruan et al., 1995), was protected to afford **4**. Treatment **4** with the Jones reagent, followed by esterification, yielded **5**. Sequential selective cleavage of the acetal moiety in **5**, phosphorylation of the resulting primary hydroxy group of **6**, and reductive removal of the p-nitrobenzyl group provided the desired protected ω-carboxyl Sph-1-P, **8**. The condensation of the carboxilic acid **8** to the amino groups of CPG followed by capping of the remaining free

amino groups gave **9**. Final deprotection of **9** yielded the desired conjugate **1**. The preparation of **2** was accomplished as described previously (Ruan et al., 1995). Quantitative Kaiser assay indicated that the substitution level of **1** and **2** used in our studies was 22 μmol and 39 μmol/g, respectively. Direct acetylation of the amino groups on the CPG beads gave the control sample for this studies. The beads were added into platelet suspensions under constant stirring at 1,000 rpm, with the Sph-1-P or Sph concentration at 40 μM, followed by scanning electron microscopy (Yatomi et al., 1995) and measurement of intracellular Ca^{2+} concentration.

Synthesis of [3-^{3}H]Sph-1-P

[3-^{3}H]Sph (22.0 Ci/mmol) was purchased from DuPont-New England Nuclear (Boston, MA). [3-^{3}H]Sph-1-P was prepared enzymatically from [3-^{3}H]Sph with crude Sph kinase obtained from BALB/3T3 clone A31 cells (Mazurek et al, 1994) (Figure 2.3).

Briefly, A31 cells were washed three times with cold PBS and harvested in lysis buffer (20 mM Tris (pH 7.5), 0.5 mM EDTA, 0.5 mM EGTA, 1 mM Na_3VO_4, 10 mM NaF, 1 mM PMSF, 10 mM β-mercaptoethanol, 1% glycerol, 25 μg/ml leupeptin, 25 μg/ml aprotinin, 25 μg/ml trypsin inhibitor, 10 mM benzamidine and 0.1% Triton X-100). Harvested cells were homogenized by a Dounce homogenizer, lysates were incubated on ice for another 30 min with agitation, and the homogenate was centrifuged at 12,000 x g for 2 min. The supernatant was collected and loaded onto DEAE cellulose (DE52) anion exchange column pre-equilibrated with column buffer (20 mM Tris (pH 7.5), 0.5 mM EDTA, 10 mM β-mercaptoethanol, 1 mM PMSF). The column was washed with 10 volumes of column buffer. The Sph kinase fraction was eluted with two

volumes of column buffer containing 200 mM NaCl and leupeptin and aprotinin (20 μg/ml).

[3-^{3}H]Sph (0.5 μCi) was dried and resuspended in 20 μl of reaction buffer (255 mM n-octyl β-D-glucopyranoside, 200 mM MOPS (pH 7.2), 10% glycerol, 10 mM β-mercaptoethanol, 20 mM $MgCl_2$). Five μl of ATP- $MgCl_2$ solution (10 mM ATP, 50 mM $MgCl_2$) and 10 μl of crude Sph kinase fraction was added to the reaction buffer containing [3-^{3}H]Sph every 4 h while this reaction mixture was incubated at 37 °C for 20 h . The reaction mixture was applied to silica gel 60 HPTLC plates (Merck, Darmstadt, Germany) and the plates were developed in butanol/acetic acid/water (3:1:1). Autoradiography was conducted, the silica gel areas which contained [3-^{3}H]Sph-1-P were scraped off, and [3-^{3}H]Sph-1-P was eluted with chloroform/methanol (2:1) from the gel. The solution was dried and resolubilized in ethanol/water (1:1).

Binding assay

Binding of Sph-1-P to the surface of platelets was assayed as described previously (Brown and Goldstein, 1974) with modifications. The suspension of platelets ($1x10^8$ cells) in PBS (without Ca^{2+} and Mg^{2+}) were incubated with [3-^{3}H]Sph-1-P at 4°C for 1 h. The platelets were washed three times with 1 ml of 1 mg/ml bovine serum albumin, 150 mM NaCl and 50 mM Tris-HCl, pH 7.5. Radioactivity of the platelets was counted with a liquid scintillator. The amount of specifically bound [3-^{3}H]Sph-1-P was determined by subtracting the radioactivity bound in the presence of unlabeled Sph-1-P (50 μM) from the radioactivity bound in the absence of unlabeled Sph-1-P. Scatchard analysis was conducted according to the literature (Scatchard, 1949). The effect of treatment with a protease (type XXV) (Sigma Chemical Co., St. Louis, MO) on the binding

was also examined. Platelets ($1x10^8$ cells) were incubated in PBS (without Ca^{2+} and Mg^{2+}) that contained 0.02% of protease (type XXV) at room temperature for 30 min. Binding assay was performed as described above.

Sph-1-P release from platelets labeled with [^{3}H]Sph

[^{3}H]Sph-labeled platelet suspensions, to which 1% bovine serum albumin had been added, were centrifuged for 15 s at 12,000 x g. Lipids were then extracted from the resultant medium supernatant and cell pellet and analyzed as described above. Albumin was included in the medium to prevent released Sph-1-P, a lipophilic molecule, from being non-specifically attached to the plasma membrane surface and consequently underestimated. Percent Sph-1-P release into medium was calculated as 100 x ([^{3}H]Sph-1-P in medium)/ (total [^{3}H]Sph-1-P in medium plus cell pellet).

Results

Effects of Sph-1-P on functional responses of human platelets

Human platelets showed reversible aggregation reaction when treated with 20-40 μM Sph-1-P (Figure 2.4A). I next measured platelet shape change. The decrease of light transmission of a stirred platelet suspension monitored the change from discoid to spheroid shape (Zucker, 1989). Administration of Sph-1-P resulted in rapid shape change (Figure 2.4B) in a concentration-dependent manner (Figure 2.5). I also determined platelet shape change by scanning electron microscopy. Under the conditions more than 90% of platelets were discoid and lacked pseudopodia before stimulation (Figure 2.6A). After exposure to Sph-1-P, nearly all platelets changed to spheroid shape with numerous pseudopodia (Figure 2.6B). These positive effects of Sph-1-P on

platelet functional responses were specific; Cer (type III), C2-Cer, C8-Cer, DMS, and sphingomyelin elicit no aggregation or shape change.

Sph-1-P-induced intracellular Ca^{2+} mobilization in platelets

It is well known that intracellular Ca^{2+} mobilization plays a central role in eliciting platelet functional responses (Siess, 1989), and that Sph-1-P causes an increase in intracellular Ca^{2+} concentration in other type of cells (Zhang et al., 1991; Mattie et al., 1994; Ghosh et al., 1994). Accordingly, we studied the effect of Sph-1-P on Ca^{2+} levels in human platelets. Sph-1-P induced a prompt increase in intracellular Ca^{2+} concentration with the peak value obtained within 20 seconds of addition (Figure 2.7A). The increase induced by 40 μM Sph-1-P was 124 ± 10 nM (mean ± SD, n = 3) in the presence of extracellular Ca^{2+}, and 68 ± 7 nM (mean ± SD, n = 3) when 1 mM EGTA, instead of Ca^{2+} was added to chelate extracellular Ca^{2+}. The concentration dependency for Ca^{2+} release induced by Sph-1-P (Figure 2.7B) correlated closely with the concentration required for shape change (Figure 2.5).

Absence of Sph-1-P lyase activity in platelets

When [^{3}H]Sph was added into platelet homogenates under established assay conditions for Sph-1-P lyase (Van Veldhoven and Mannaerts, 1991), no lyase products (palmitaldehyde, palmitic acid and palmitol) were formed. Under the same conditions, high lyase activity was detected in bovine (Table II) and rat livers (~12.8 pmol/min.mg of protein) as reported previously (see Chapter 1). Low but significant Sph-1-P lyase activity was detected in CMK and K562 cells (Table II). CMK is a human megakaryoblastic cell line (Sato et al., 1989); K562 is a human erythroleukemia cell line capable of megakaryocytic differentiation

(Cheng et al., 1994). My results clearly shows the absence of Sph-1-P lyase in platelets. It is likely that megakaryocytes lose Sph-1-P lyase activity during late-stage differentiation into platelets.

Determination of Sph and Sph-1-P levels in platelets

I determined the mass levels of Sph-1-P and Sph extracted from platelets. The content of Sph-1-P and Sph was calculated as 1.41 ± 0.04 nmol and 374 ± 61 pmol (mean $\pm$ SD, n=3 and 5), respectively, per 10^9 platelets. The cellular content of Sph-1-P is much higher than that of Sph.

I measured the time course of mass change of Sph-1-P and Sph in platelets incubated with non-radioactive Sph, the substrate of Sph kinase. Exogenous Sph was rapidly converted into Sph-1-P (Figure 2.8A). Consistent with our previous finding (Yatomi et al., 1995), exogenous Sph-1-P induced platelet shape change and intracellular Ca^{2+} mobilization (Figure 2.8B). These positive reactions induced by Sph-1-P were not mimicked by Sph (Figure 2.8B). Sph-1-P (5 μM), but not 10 μM Sph, activated platelets (Figure 2.8B) despite the fact that over 5 μM Sph-1-P was rapidly formed in platelets incubated with 10 μM Sph (Figure 2.8A) by cytoplasmic Sph kinase (Ghosh et al., 1994; Stoffel et al., 1973).

Platelet activation induced by Sph-1-P immobilized on a solid support

The finding that only exogenous Sph-1-P activates platelets raises the possibility that Sph-1-P acts on these cells from outside. To test this possibility, I applied Sph-1-P immobilized on glass beads to platelets. Upon the addition of the immobilized Sph-1-P, platelets underwent shape change and aggregate formation, as determined by scanning microscopy (Figures 2.9C and D), and intracellular Ca^{2+} mobilization (Figure 2.10). Sph-bound beads (Figures 2.8A

and B, and Figure 2.10) or control beads (data not shown) had no effect on platelets. The fact that immobilized Sph-1-P mimics free Sph-1-P in terms of activating platelets strongly suggests that the site of Sph-1-P action is not inside platelet but on the surface. This hypothesis would explain why exogenous addition of Sph-1-P but not Sph results in platelet activation in spite of rapid Sph conversion into Sph-1-P in platelets.

[^{3}H]Sph-1-P binding studies

To examine hypothesis that exogenous Sph-1-P acts on platelets via interaction with a plasma membrane receptor, I assayed the binding of Sph-1-P to platelets with [3-^{3}H]Sph-1-P. Figure 2.11A shows the time course of specific binding of [3-^{3}H]Sph-1-P at 4°C. The specific binding depended on time, reached equilibrium within 1 hour, and remained at equilibrium for at least 1 hour. Figure 2.11B shows dependence of [3-^{3}H]Sph-1-P concentration on specific binding of [3-^{3}H]Sph-1-P to Platelets. The specific binding was saturated around 2 nM of the ligand. Scatchard analysis of these data revealed two binding sites for Sph-1-P (Figure 2.11C). The Kd values were estimated to be 110 nM and 2.6 μM, and the numbers of the binding site were ca. 200/cell and 2400/cell, respectively.

I examined effect of a protease on the specific binding of [3-^{3}H]Sph-1-P. Treatment of platelets with a protease (type XXV) disrupted the specific binding of Sph-1-P to platelets almost completely (n=2), indicating that the binding sites are proteins which are located on the surface of the platelets. These results indicate the presence of a cell surface receptor for Sph-1-P.

To determine the structural characteristic of Sph-1-P required for its binding to the surface of platelets, I conducted competition binding experiments with unlabeled lipids which have similar structures to Sph-1-P (Figure 2.12).

Unlabeled Sph-1-P reduced the binding of [3-^{3}H]Sph-1-P to platelets, and this effect increased as the concentration of unlabeled Sph-1-P increased, reaching non-specific binding of ca. 30% of the total binding at 50 μM of unlabeled Sph-1-P. Sph and C2-Cer did not inhibit the binding of [3-^{3}H]Sph-1-P. Sphingosylphosphorylcholine (SPC) only slightly inhibited the binding of [3-^{3}H]Sph-1-P. DMS-1-P (50 μM) inhibited ca. 50%, suggesting that the phosphate group is required for the specific binding of Sph-1-P besides the hydrophobic portion of the molecule. LPA which has a very similar structure to Sph-1-P reduced the binding of [3-^{3}H]Sph-1-P as much as unlabeled Sph-1-P did.

Inhibition of LPA-induced platelet aggregation by Sph-1-P

LPA is a well-established platelet agonist capable of inducing strong, irreversible aggregation response (Benton et al., 1982). Sph-1-P is also a platelet-aggregation agent, although its effect is weaker (Yatomi et al., 1995). I found that LPA-induced aggregation is inhibited (desensitized) by Sph-1-P (Figure 2.13). This observation may be related to the fact that Sph-1-P and LPA share a platelet surface receptor, although we cannot completely rule out the desensitization mechanism resulting from the modification of intracellular signaling pathways involved. LPA (5 μM) induced marked platelet aggregation, which was comparable with that induced by 5 μg/ml of collagen (Siess, 1989). Prior addition of subthreshold concentrations of Sph-1-P inhibited the platelet aggregation response to LPA. In contrast, Sph-1-P did not affect the response to collagen.

Sph-1-P release from activated platelets

In view of the fact that Sph-1-P activates platelets from outside but is abundantly stored inside, we examined the possibility that Sph-1-P is released from platelets and acts intercellularly as a local mediator. When platelets were stimulated with 1 μM TPA, which can act as a substitute for diacylglycerol and directly activates protein kinase C (Siess, 1989; Nishizuka, 1984), 8, 35, 53, 57 and 60% of stored Sph-1-P was released extracellularly at 2, 5, 20, 60 and 120 min after challenge, respectively. This protein kinase C activator was also found to release Sph-1-P in a dose dependent manner (Figure 2.14) Under the same conditions, thrombin, which produces diacylglycerol and activates protein kinase C as a result of phosphatidylinositol-4,5-bisphosphate hydrolysis (Siess, 1989; Nishizuka, 1984), and 1-oleoyl-2-acetyl-glycerol, a membrane-permeable diacylglycerol (Nishizuka, 1984), also caused marked Sph-1-P release (Figure 2.14). The Sph-1-P release induced by these protein kinase C activators was inhibited by staurosporine (Figure 2.15), an inhibitor of protein kinases, including protein kinase C. These finding indicate that Sph-1-P can be released from activated platelets and suggest that protein kinase C activation may be the mechanism involved.

Discussion

Sph-1-P was discovered over 25 years ago as an intermediate catabolite during studies of Sph catabolism by Wilhelm Stoffel and his colleagues (Stoffel, 1970). This is a counterpart of Sph anabolism (in which 3-keto-Sph is formed from serine and palmitoyl CoA) as formulated by Esmond Snell around the same time (Snell et al., 1970).

Whereas Sph kinase (for Sph-1-P synthesis) and Sph-1-P aldolase (lyase) are rate-limiting enzymes for Sph catabolism, and Sph-1-P level is known to be high in platelets (Stoffel et al., 1970; Stoffel, 1973), the pattern of Sph

catabolism in various types of cells and tissues are not available. My initial studies of Sph catabolism indicated that Sph-1-P lyase (and also phosphatase (data not shown)) activity is completely absent in platelets, in striking contrast to its consistent presence in megakaryocytes and other types of cells. Since megakaryocytes are precursors of platelets, this enzyme is down-regulated and its activity ceases in association with platelet differentiation. I found that platelets contain a high level of Sph-1-P (0.21 mol% Sph-1-P/phospholipid), in contrast to its low level in various other cells such as melanoma cells (~0.02 mol% Sph-1-P/phospholipid) and Balb/c 3T3 A31 cells (~0.03 mol% Sph-1-P/phospholipid). Thus, I established and confirmed the idea that the novelty of Sph metabolism in platelets as compared to other nucleated cells is the accumulation of Sph-1-P resulting from absence of Sph-1-P lyase and phosphatase activity.

The idea that exogenous Sph-1-P enters cells and cause altered signal transduction leading to shape change and Ca^{2+} mobilization is highly unlikely, since platelets contain a constant high level of intracellular Sph-1-P. However, one possible explanation is that intracellular Sph-1-P is stored in a specific locus and not used for signal transduction. Therefore, two mechanisms for the initial effect of Sph-1-P can be considered: (a) Sph-1-P enters the platelet without a specific receptor and affects signal transduction. (b) Sph-1-P binds initially to a specific cell surface receptor which triggers signal transduction leading to phenotypic changes. There may be those two mechanisms for platelet activation by Sph-1-P. My studies have been focused on mechanism (b), which is supported by various observations as described under "Results" and summarized below.

1. [3-^{3}H]Sph-1-P binds to platelet surface membrane in a time- and dose-dependent manner. This binding is inhibited by cold Sph-1-P but not by Sph or Cer. Scatchard plot analysis of specific binding of [3-^{3}H]Sph-1-P to

platelet surface indicates the two binding sites with dissociation constant of 110 nM and 2.6 μM. A large molar excess of unlabeled Sph-1-P was required for the replacement of [^{3}H]Sph-1-P, which casts a doubt on this experiment. However, lipids are naturally highly lipophilic and promote high levels of nonspecific binding (Balsa et al., 1996).

2. Sph-1-P or Sph immobilized through their ω-carbon to porous glass beads were synthesized by novel multi-step processes including carboxylation at ω-carbon. Head groups of immobilized Sph-1-P caused shape change and aggregation when they touched the platelet surface. Head groups of immobilized Sph had no effect.

3. Binding of [3-^{3}H]Sph-1-P to platelet surface was inhibited by not only Sph-1-P but also LPA. LPA causes platelet activation, similar to but stronger than that caused by Sph-1-P, including release of LPA and Sph-1-P from platelets (Benton et al., 1982; Eichholtz et al., 1993). LPA-induced aggregation is inhibited (desensitized) by Sph-1-P. This may be due to the fact that Sph-1-P and LPA have similar conformational structure and therefore may share the same receptor (Durieux et al., 1993; Durieux and Lynch, 1993).

These results shows the existance of binding sites on platelet surface but I do not know that the binding site is a functional receptor. The results here indicate possibility that Sph-1-P activates platelets through the binding sites, and do not exclude mechanism (a) Sph-1-P enters the platelet without a specific receptor and affects signal transduction. Cloning and characterization of a putative Sph-1-P receptor is essential.

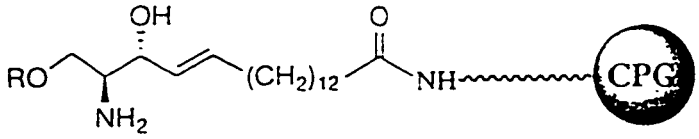
These findings are not consistent with several recent studies reporting lack of cross-desensitization between Sph-1-P and LPA in various nucleated cells (Jalink et al., 1995; van Koppen et al., 1996; Postma et al., 1996). In this

context, it is noted that platelets are shown to possess two different levels of binding sites, a high affinity site (Kd, 110 nM) and a low affinity site (Kd, 2.6 μM), and one can assume that the low affinity site might be functional for platelets, judging from the fact that μM order of Sph-1-P concentrations is needed to induce platelet shape change and aggregation. Although the molecular mechanism of our observation in platelets remains to be solved, one possibility is that platelets possess unique receptor(s) for lysophospholipids including Sph-1-P and LPA, leading to platelet aggregation.

In connection with the concept of Sph-1-P receptor as described above, it should be noted that signaling pathways of Sph-1-P are regulated by heterotrimeric GTP-binding proteins (Wu et al., 1995; Goodemote et al., 1995; van Koppen et al., 1996; Okajima et al., 1996), whose activation is receptor-dependent (Neer, 1995). Only exogenously (not intracellularly) added Sph-1-P induces biological responses in guinea-pig atrial myocytes (van Koppen et al., 1996) and N1E neuronal cells (Postma et al., 1996). Furthermore, an orphan receptor, endothelial differentiation gene product (EDG-1), was found recently to bind to Sph-1-P with high affinity (Lee et al., 1998). In addition to intracellular actions after passing through the plasma membrane, activation of plasma membrane receptor(s) may be a critical mechanism by which Sph-1-P exerts biological responses in various cells.

Platelets store a variety of biologically active molecules which are secreted upon stimulation (Hawiger, 1989). The secreted molecules interact with other platelets, plasma proteins, and the vessel wall. Sph-1-P is released from platelets (Figures 2.14 and 2.15), as expected given that Sph-1-P activates platelets from outside but is abundantly stored inside. Sph-1-P release may be mediated by PKC, which is also highly expressed in platelets (Kikkawa et al., 1982). I propose that Sph-1-P be added to the list of bioactive molecules stored

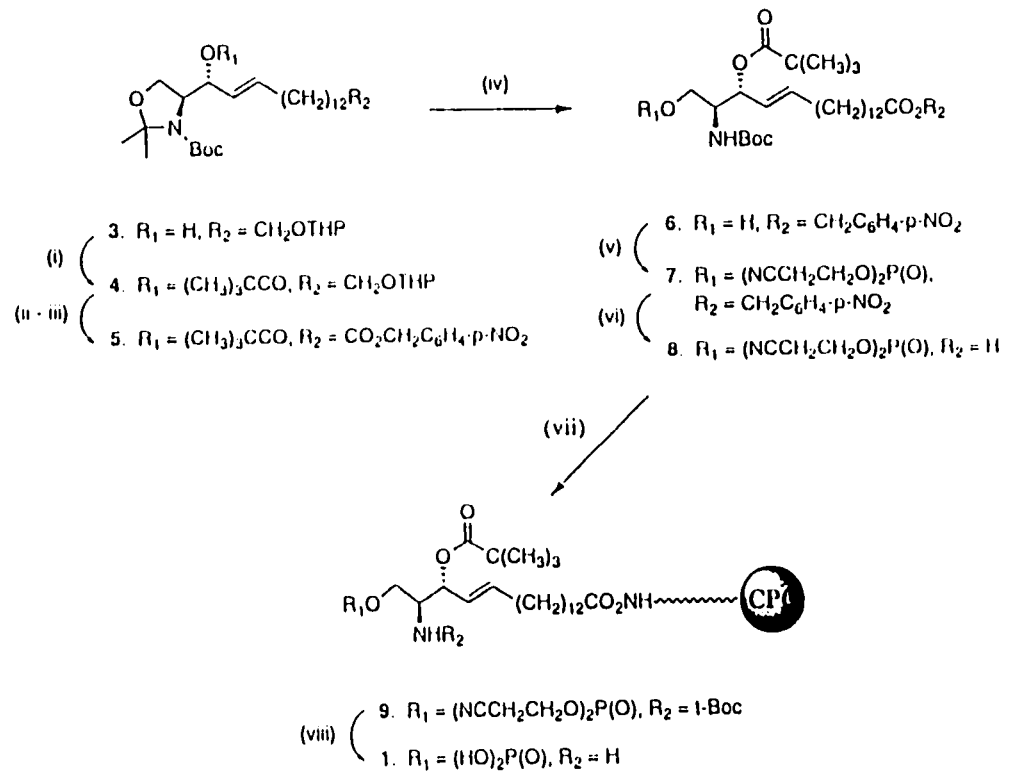
in platelets and released from them upon stimulation, even though at present I do not have direct evidence for platelet activation caused by released Sph-1-P; the fact that not only Sph-1-P but also more potent lipid mediators such as thromboxane A2 are released from activated platelets makes such experiments difficult. Previous studies, mainly on Swiss 3T3 cells, suggest that Sph-1-P is an intracellular second messenger in cell growth regulation: Sph kinase, which produces Sph-1-P in the cells, was found to be regulated by platelet-derived growth factor (PDGF) (Spiegel et al., 1996; Spiegel and Milstien, 1995) (Figure 2.16). However, as shown here, Sph-1-P can act extracellularly as a local mediator through its discharge from cells to regulate cellular functions in an autocrine or paracrine fashion (Figure 2.17).



1. R = $(HO)_2P(O)$
2. R = H

Fig. 2.1. Structure of the conjugates of Sph-1-P, 1, and Sph, 2, with controlled pore glass (CPG) beads used in this study.

Fig. 2.2. Scheme for preparation of Sph-1-P and Sph immobilized on controlled pore glass beads.



Reagents: (i) pivaloyl chloride, pyridine; (ii) Jones reagent, acetone; (iii) p-$NO_2C_6H_4CH_2Cl$, NaI, $NaHCO_3$, DMF; (iv) p-$TsOH$, CH_3OH; (v) (a) $(NCCH_2CH_2O)_2PN(i\text{-}Pr)_2$, tetrazole, CH_2Cl_2, (b) MCPBA, CH_2Cl_2; (vi) zinc dust, $HOAc/H_2O$ (9:1); (vii) (a) CPG, HOBt, DIC, CH_2Cl_2, DMF, (b) HOAc; (viii) (a) NaOH, EtOH, dioxane, (b) TFA, CH_2Cl_2.

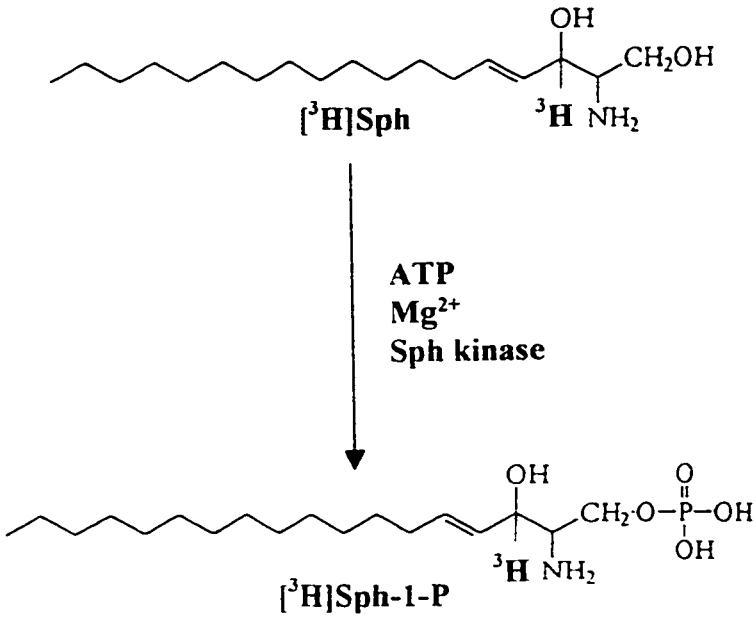


Fig. 2.3. Synthesis of [3-^{3}H]Sph-1-P. [3-^{3}H]Sph-1-P was prepared enzymatically from [3-^{3}H]Sph with crude Sph kinase obtained from BALB/3T3 clone A31 cells.

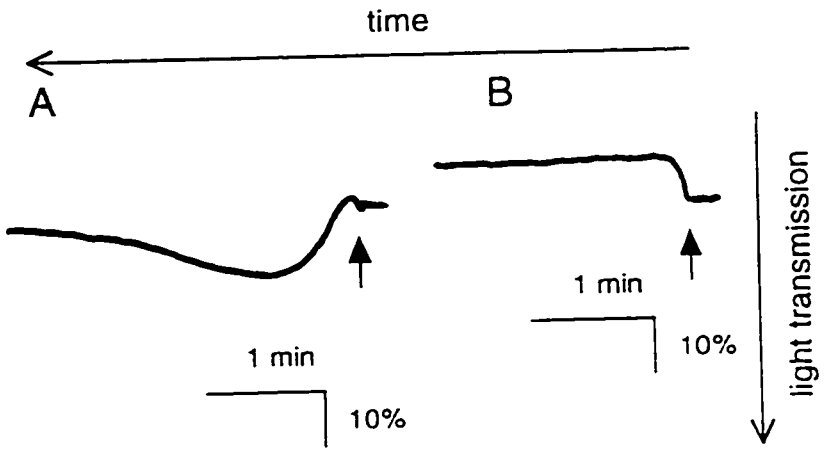


Fig. 2.4. Sph-1-P induced human platelet aggregation (A) and shape change (B). Human platelets were challenged with 40 μM Sph-1-P as indicated by arrows. aggregation (A) and shape change (B) were observed as described in "Materials and Methods".

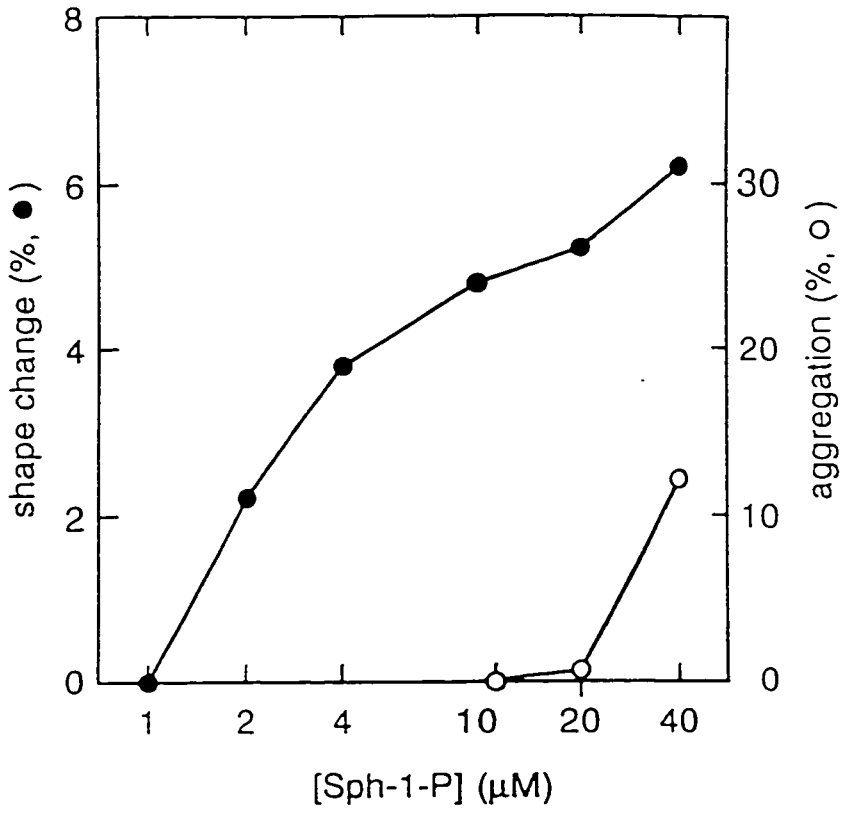


Fig. 2.5. Concentration-dependent platelet aggregation and shape change induced by Sph-1-P. Platelet aggregation (O) and shape change (●) induced by various concentrations of Sph-1-P are shown. Results are expressed as means of two or three determinations.

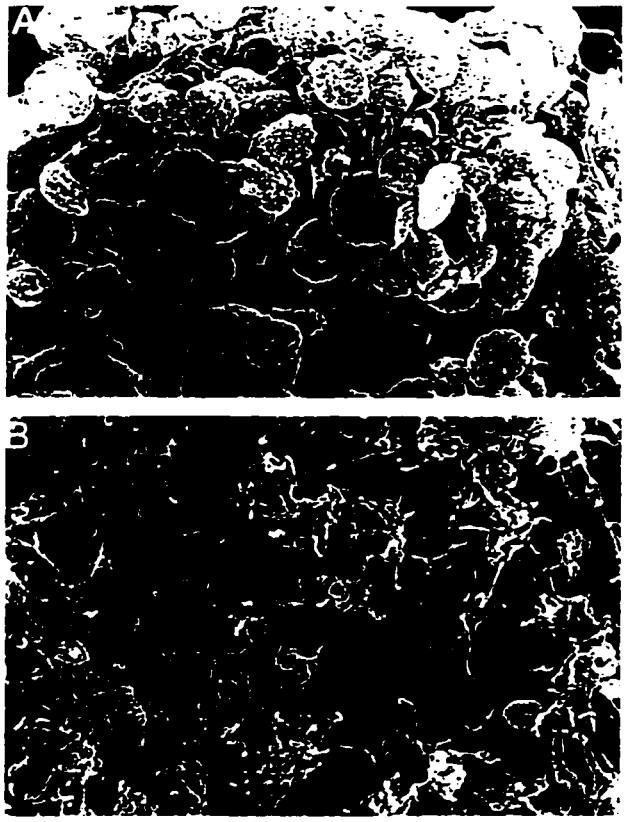


Fig. 2.6. Scanning electron microscopy of shape-changed platelets. Platelets were treated without (A) or with (B) 20 μM Sph-1-P for 1 min. Scanning electron microscopy was performed to determine the cell shape. Original magnification x 5,000.

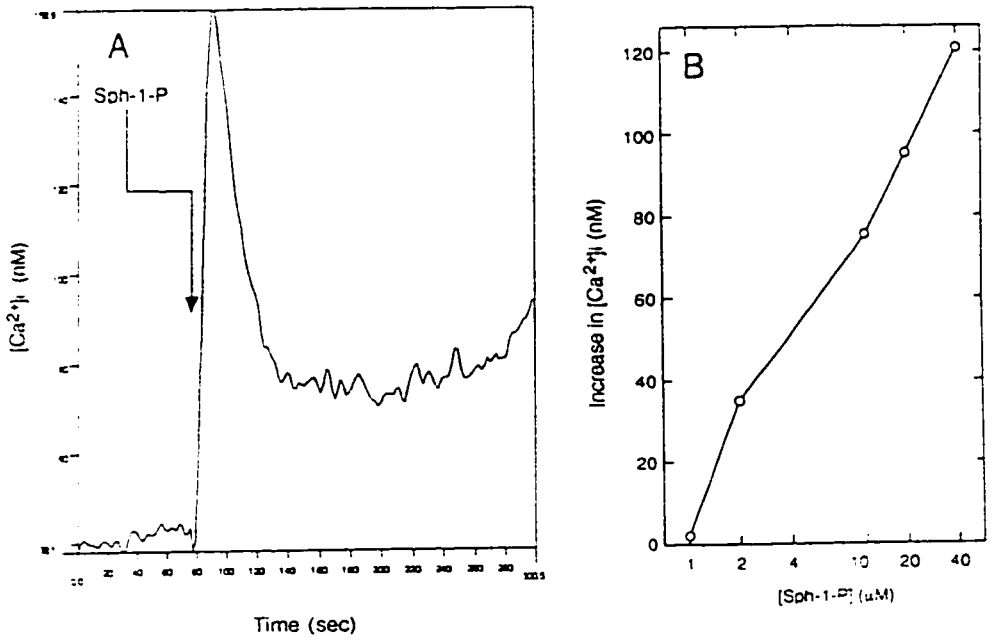


Fig. 2.7. Sph-1-P-induced changes in platelet intracellular Ca^{2+} concentration. (A) Fura2-loaded platelets were challenged with 40 mM Sph-1-P, as indicated by an arrow. (B) Fura2-loaded platelets were stimulated with various concentration of Sph-1-P. The results are the values of peak increases after Sph-1-P addition and are from a single experiment representive of three.

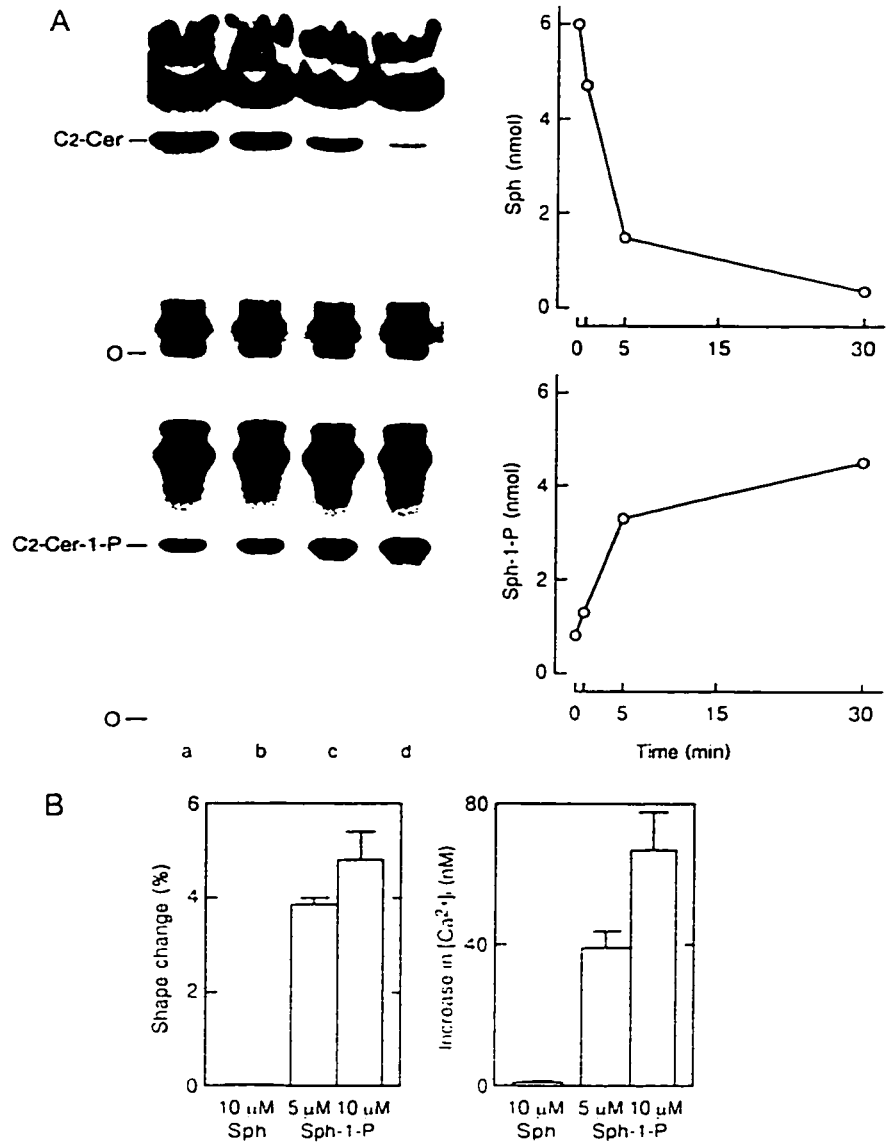


Fig. 2.8. Platelet activation by exogenously-added Sph-1-P, but not Sph, in spite of extensive conversion of added Sph into Sph-1-P. (A) Platelet suspensions (0.5 ml) were incubated with 10 μM (5 nmol) Sph for indicated period. Sph and Sph-1-P were quantitatively measured by N-acylation with [^{3}H]acetic anhydride. (B) Platelets were challenged with Sph (10 μM) or Sph-1-P (5 or 10 μM), and shape change reaction (left) and intracellular Ca^{2+} concentration (right) were monitored. Columns and error bars represent means ± S.D. (n = 3).

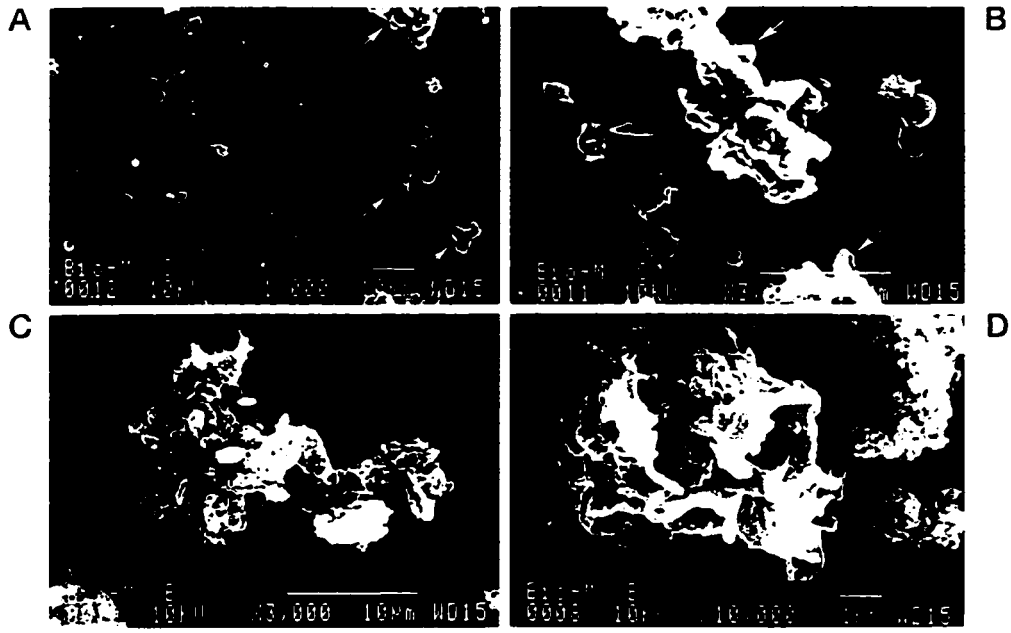


Fig. 2.9. Platelet activation induced by Sph-1-P immobilized on a solid support. A and B, platelets were treated for 1 min with Sph-bound glass beads and examined by scanning electron microscopy. C and D, platelets were treated for 1 min with Sph-1-P bound beads and examined by scanning electron microscopy. Bars: A, B and C, 10 μm; D, 1 μm.

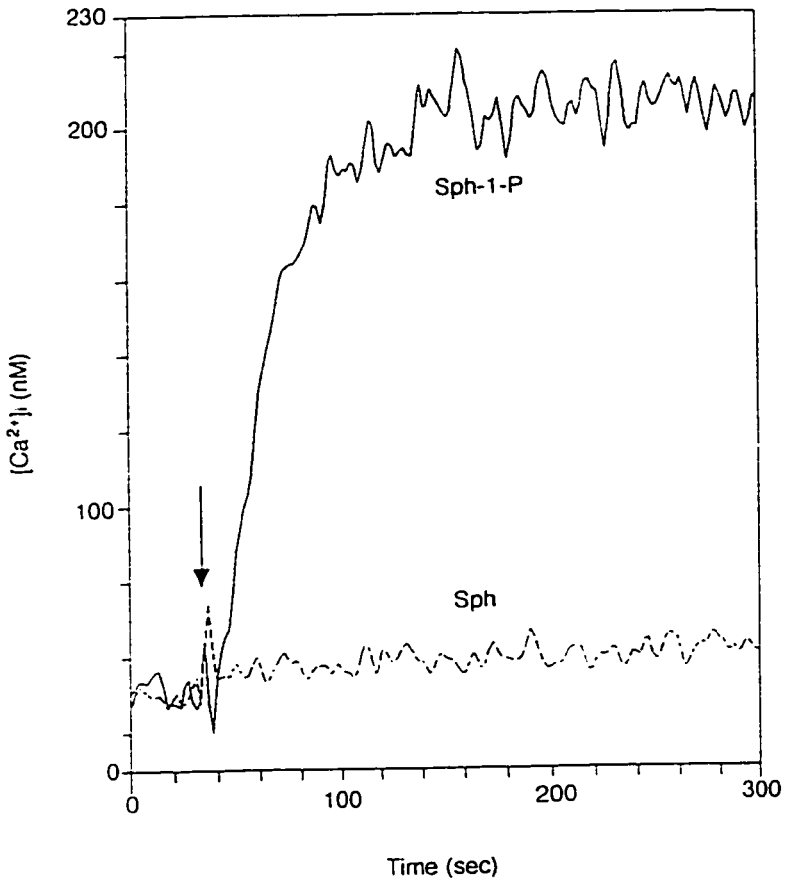


Fig. 2.10. Platelet intracellular Ca^{2+} mobilization induced by Sph-1-P immobilized on a solid support. Platelets were challenged with Sph- or Sph-1-P-bound glass beads (arrow), and intracellular Ca^{2+} concentration was monitored.

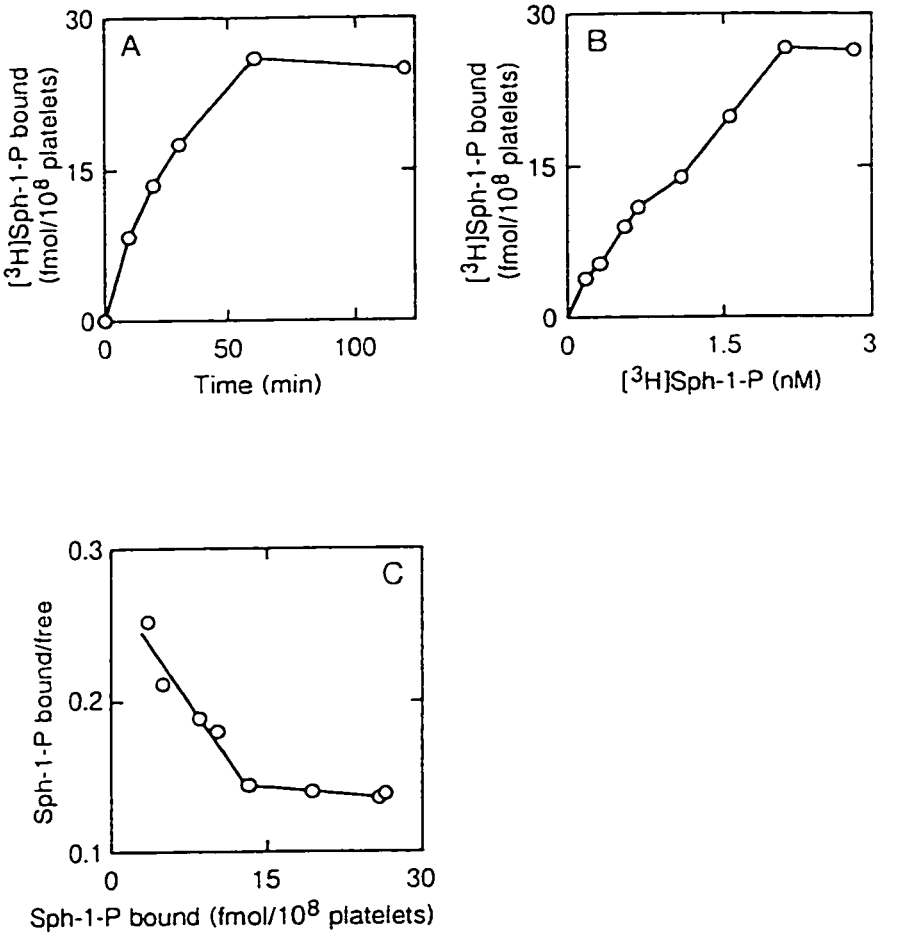


Fig. 2.11. Specific binding of [3-^{3}H]Sph-1-P to Platelets. The specific binding was determined as described in "Materials and Methods". Data are representative of two independent experiments. (A) Time course of specific binding of [3-^{3}H]Sph-1-P to platelets. Platelets (1 x10^8 cells) were incubated with 2.1 nM [3-^{3}H]Sph-1-P at 4 °C in the presence and the absence of 50 μM unlabeled Sph-1-P. (B) Dependence of [3-^{3}H]Sph-1-P concentration on specific binding of [3-^{3}H]Sph-1-P to Platelets. Platelets (1x10^8 cells) were incubated with various concentrations of [3-^{3}H]Sph-1-P at 4 °C for 1 hour in the presence and the absence unlabeled Sph-1-P (50 μM). (C) Scatchard analysis of specific binding of [3-^{3}H]Sph-1-P to Platelets. The data in panel B were plotted by the method of Scatchard.

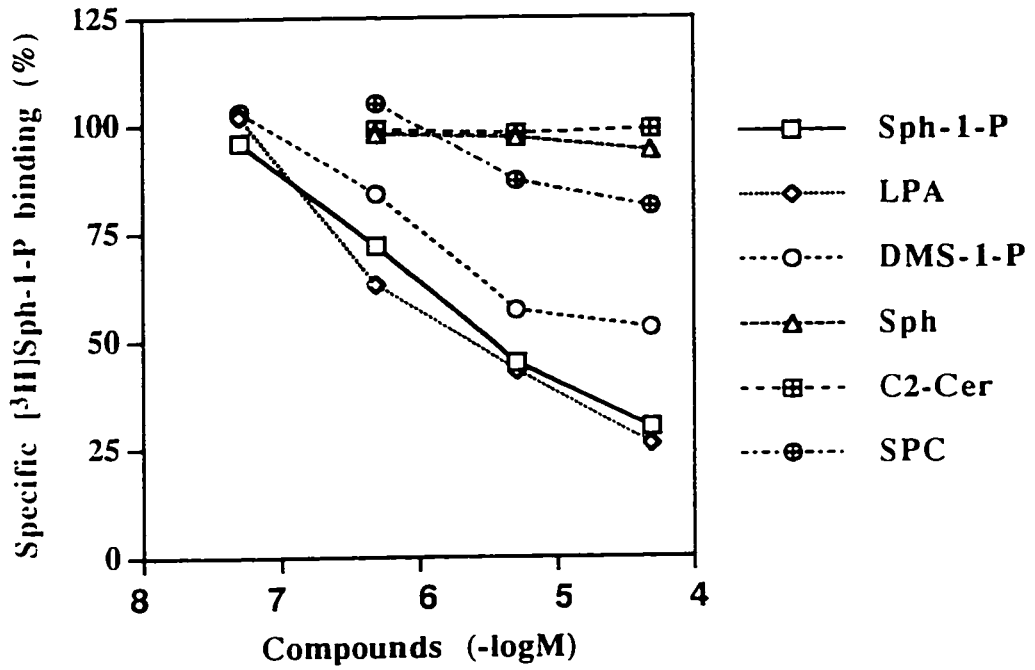


Fig. 2.12. Effect of various lipids on [3-^{3}H]Sph-1-P binding to platelets. Platelets ($1x10^8$ cells) were incubated with 2.1 nM [3-^{3}H]Sph-1-P at 4 °C for 1 hour in the presence and the absence of various lipids (50 μM). Data are representative of two independent experiments.

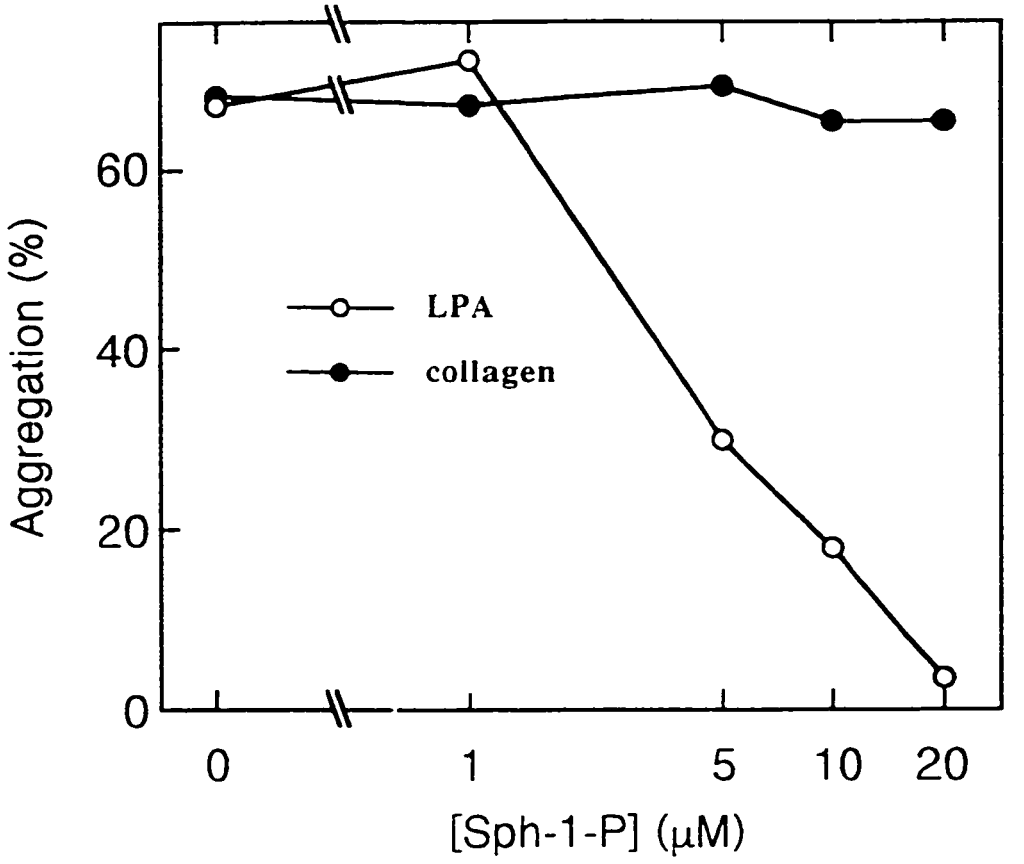


Fig. 2.13. Desensitization of platelet aggregation response to LPA by prior addition of Sph-1-P. Platelets were preincubated with various concentrations of Sph-1-P for 1 min and then challenged with 5 μM LPA or 5 μg/ml of collagen. Platelet aggregation was analyzed as described in "Materials and Methods". Under the conditions employed, Sph-1-P, at concentrations below 20 μM did not induce platelet aggregation. Data are representative of two independent experiments.

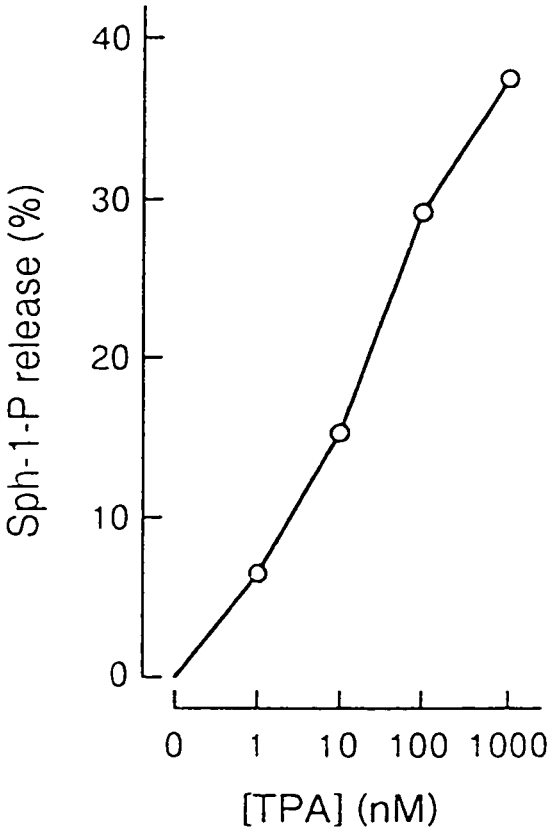


Fig. 2.14. TPA-induced Sph-1-P release. Platelet labeled with [^{3}H]Sph were stimulated with various concentrations of TPA for 5 min, and the percent Sph-1-P release into the medium was determined as described under "Materials and Methods".

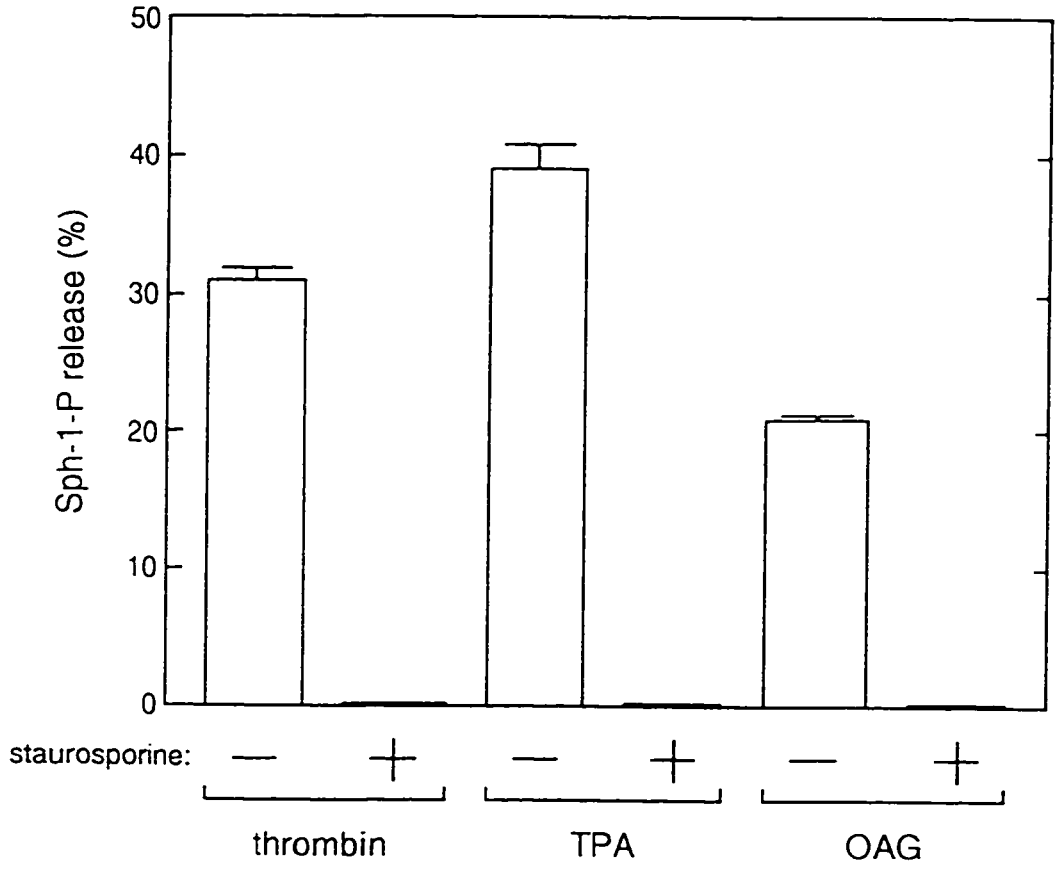


Fig. 2.15. Thrombin-, TPA-, or 1-oleoyl-2-acetyl-glycerol-induced Sph-1-P release from platelets and its inhibition by staurosporine. The percentage of Sph-1-P release into the medium was measured in platelets treated with 0.5 unit/ml of thrombin, 1 μM TPA or 25 μg/ml of 1-oleoyl-2-acetyl-glycerol for 5 min in the absence (-) or presence (+) of 1 μM staurosporine. Data are the means (bars, S.D.) of three determinations.

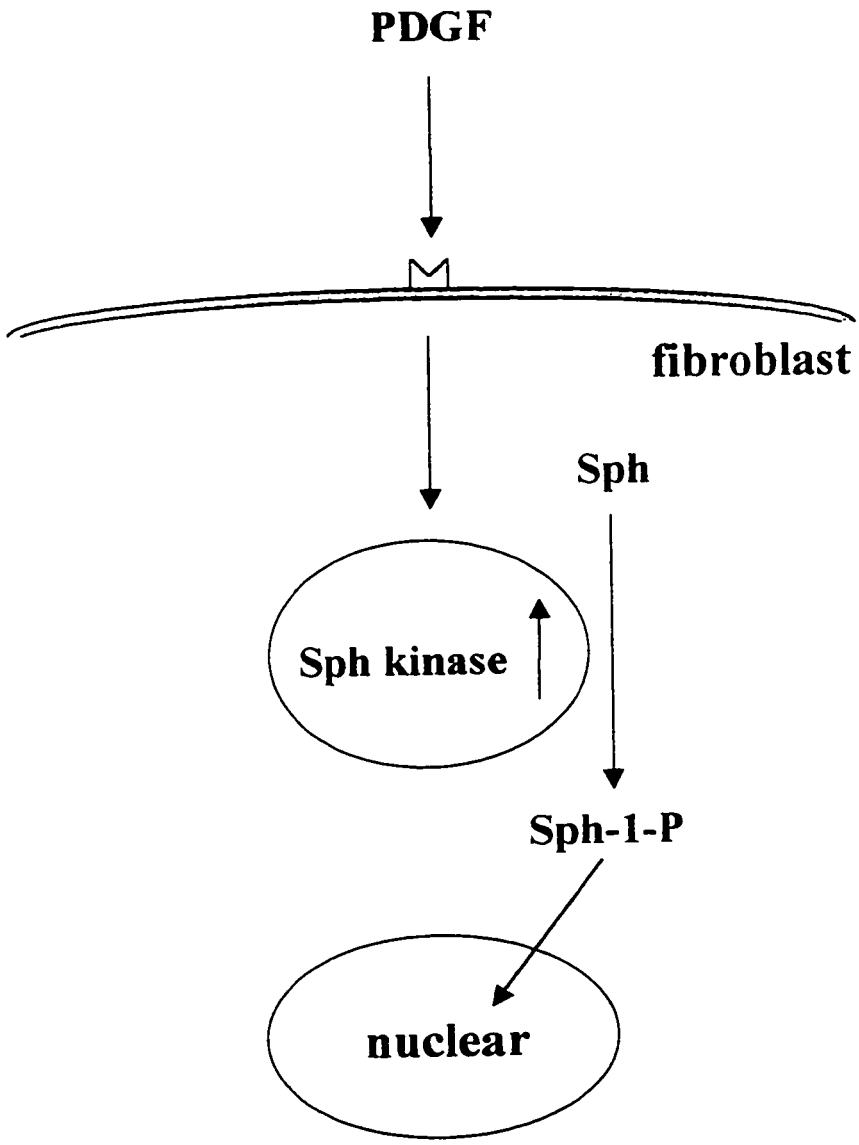


Fig. 2.16. Sph-1-P as an intracellular second messenger. Sph kinase activity is stimulated and Sph-1-P level is transiently increased by PDGF.

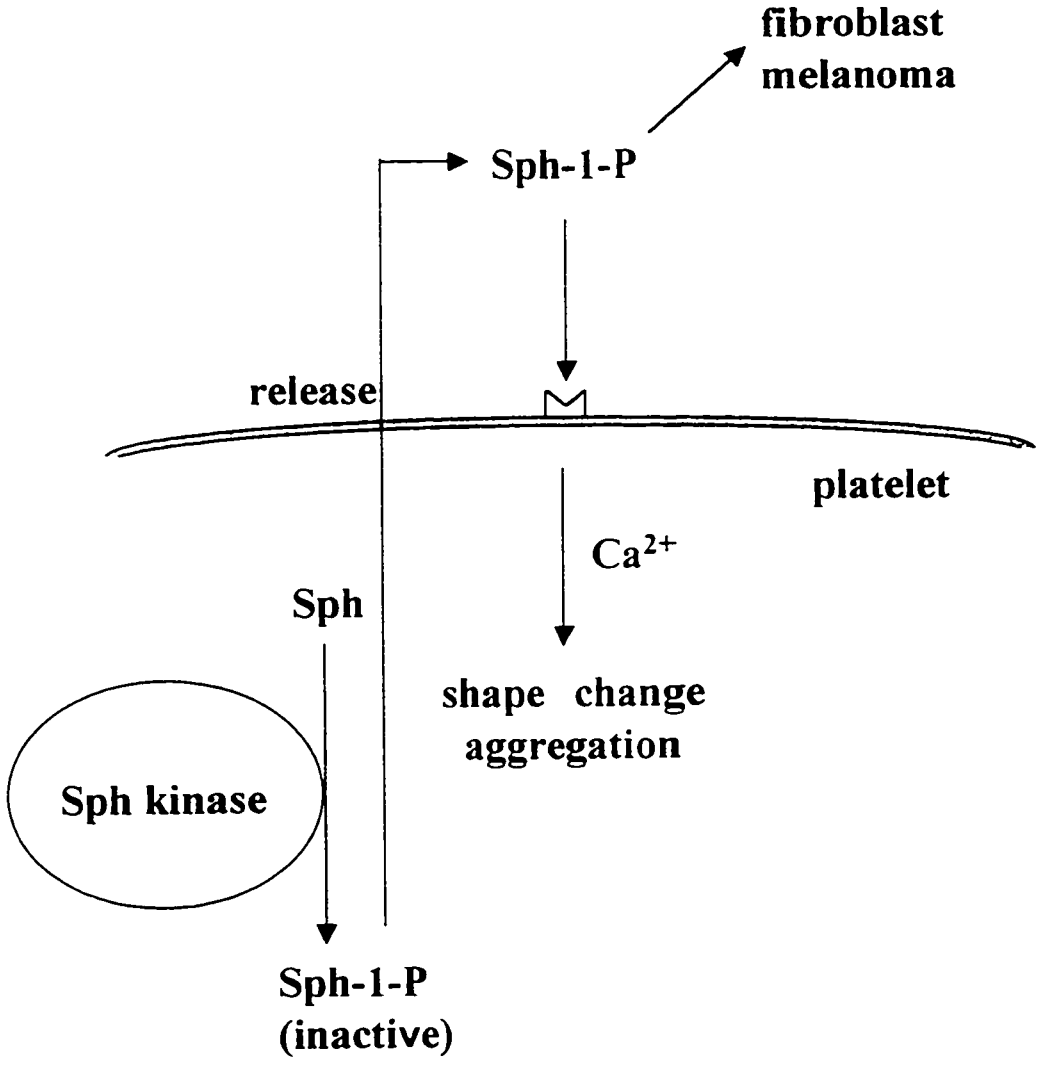


Fig. 2.17. Sph-1-P as an intercellular messenger. Sph-1-P may act intercellularly (cell to cell) as a local mediator through its discharge from platelets to regulate cellular functions in an autocrine or paracrine fashion.

Table II. Measurement of Sph-1-P lyase activity.

Cells or tissue	Sph-1-P lyase activity (pmol/ min. mg of protein)
Platelets	not detectable
CMK	0.253 ± 0.036
K562	0.539 ± 0.031
Bovine liver	11.2 ± 0.2

Chapter 3

Sphingosine-1-phosphate Regulates Melanoma Cell Motility through a Receptor-coupled Extracellular Action

Introduction

In Chapter 2, I demonstrated that Sph-1-P acts on platelets extracellularly to activate them through a suggestive receptor. The concentration of Sph-1-P alone required for activation was high (20-40 μM), throwing doubt on the physiological relevance.

In contrast, Sph-1-P inhibits chemotactic and haptotactic motility of mouse melanoma B16, Balb/c 3T3, human arterial smooth muscle cells, and many other types of cells at very low concentrations (10-100 nM) (Sadahira et al., 1992; Sadahira et al., 1994; Bornfeldt et al., 1995). No other Sph-related compounds are known to inhibit cell motility. Thus, susceptibility of motility of various types of cells to Sph-1-P is much higher than platelet activation susceptibility to Sph-1-P, suggesting a physiological role.

Two possible mechanisms for Sph-1-P action can be considered: (i) Sph-1-P enters cells rapidly and binds to target molecules which control motility; or (ii) receptors exist, analogous to those on the platelet surface. Sph-1-P binds to the receptors, which quickly mediate signaling to inhibit motility.

In this chapter, the concept of the extracellular action of Sph-1-P in platelet activation was applied to cell motility inhibition and demonstrated an alleged receptor having physiological relevance for extracellular action of Sph-1-P on B16 cells. The target molecule affected immediately after addition of Sph-1-P appears to be small G-protein Rho. A proposed molecular mechanism following the effect of Sph-1-P on Rho A is summarized in the Appendix.

Materials and Methods

Materials

Sph-1-P was synthesized chemically (Sadahira et al, 1994). Sph, N,N-dimethylsphingosine (DMS), and N,N,N-trimethylsphingosine (TMS) were synthesized as described previously (Igarashi et al, 1989; Toyokuni et al, 1991). C2-ceramide (C2-Cer) was synthesized as described (Vunnam and Radin, 1979). N,N-dimethylsphingosine-1-phosphate (DMS-1-P) was synthesized in our lab (F. Ruan, S. Hakomori and Y. Igarashi, unpublished data). Sphingosylphosphorylcholine (SPC) was purchased from Sigma Chemical Co. (St. Louis, MO) and purified by HPLC. [3-^{3}H]Sph-1-P was prepared enzymatically from [3-^{3}H]Sph with crude Sph kinase obtained from BALB/3T3 clone A31 cells (Mazurek et al, 1994) as described in Chapter 2. These sphingolipids were dissolved in ethanol/water (1:1), and stored at -20°C until required. IODO-BEADS were purchased from Pierce (Rockford, IL). Suramin and [^{32}P]NAD were obtained from ICN Pharmaceuticals (Costa Mesa, CA). All other reagents were purchased from Sigma (St. Louis, MO).

Cell culture

Mouse melanoma cells B16 F10 were obtained from Dr. I. J. Fidler (M. D. Anderson Cancer Center, University of Texas, Houston, TX) and cultured in DMEM supplemented with 10% fetal calf serum (FCS) (HyClone, Logan, UT), 2 mM L-glutamine, 2 mM pyruvic acid, 4.5 mg/ml D-glucose, 100 units/ml penicillin G, and 100 μg/ml streptomycin. A31 cells were obtained from the American Type Culture Collection and cultured in the same medium, except that bovine calf serum (BCS) (HyClone) was used.

Preparation and application of Sph-1-P or Sph immobilized on glass beads

Sph-1-P or Sph, each with a free ω-carboxyl group, was synthesized. The ω-carboxyl group of the lipid was covalently linked to a long chain alkyl amine on the surface of controlled-pore glass beads (CPG Inc., Fairfield, NJ, USA) through amide bonding (see Chapter 1). Cells were briefly trypsinized, washed, and the glass beads were added into the cell suspensions. The cells were cultured with the beads for 2 days. Viability of cells which adhered to the beads was checked by trypan blue exclusion assay. The interaction between cells and the beads was studied with scanning electron microscopy (Yatomi et al., 1995) or light microscopy.

Cell motility assay

Haptotactic cell motility of F10 cells was assayed using Transwell compartments (Costar, Cambridge, MA) with 6.5 mm diameter polycarbonate filters (8 μm pore size) (Sadahira et al., 1994). Briefly, the lower surface of the filter was coated with human plasma fibronectin (5 μg/filter) (Boehringer Mannheim Biochemicals, Indianapolis, IN) and dried at 37°C overnight. The cells were harvested after brief exposure to 0.05% trypsin and 0.02% EDTA, and resuspended to a concentration of 4×10^5 cells/ml in DMEM supplemented with 0.1% BSA. Cells (100 μl of suspension) were seeded in the upper compartment. DMEM (0.6 ml) with 0.1% BSA was placed in the lower compartment, which was then connected to the upper compartment. After incubation for 18 h (unless otherwise indicated) at 37°C, cells on the lower surface of the filter were stained and counted.

Binding assay

Binding of Sph-1-P to the cell surface was assayed as described previously (Brown and Goldstein, 1974) with modifications. Briefly, F10 cells were harvested with 0.02% EDTA in PBS. Cells were resuspended (1.2×10^6 or 4×10^5 cells/ml) in DMEM and incubated with [3-^{3}H]Sph-1-P at 4°C for the indicated times, washed three times with 1 ml of 2 mg/ml BSA, 150 mM NaCl and 50 mM Tris-HCl, pH 7.5, and radioactivity was counted with a scintillation counter. The amount of specifically bound [3-^{3}H]Sph-1-P was determined by the difference of radioactivity bound in the presence vs. absence of unlabeled Sph-1-P (50 μM). Scatchard plot analysis was conducted as described (Scatchard, 1949). For protease experiments, F10 cells (1×10^7 cells) were incubated at room temperature for 10 min in 5 ml of PBS containing 0.02% protease (type XXV), and binding assays were performed as described above.

Preparation of cell membranes

F10 cells were washed twice with PBS and scraped into buffer A (10 mM HEPES (pH 7), 5 mM $MgCl_2$, 1 mM EDTA, 0.5 mM PMSF) (Watanabe et al., 1991). Cell suspensions were homogenized in a Dounce homogenizer in buffer A at 4°C and centrifuged at 1000 g for 15 min at 4°C, and the pellet was resuspended in buffer A.

ADP-ribosylation of membranes by pertussis toxin

F10 cells were cultured overnight in DMEM supplemented with 0.25% BSA in the presence or absence of 100 ng/ml pertussis toxin (PTX, Sigma). Membranes were prepared as described above from PTX-treated and control cells. A reaction mixture containing cell membranes (50 μg total protein), 100 mM Tris-HCl (pH 7.5), 20 μg/ ml preactivated PTX, 10 mM thymidine, 1 mM

ATP, 0.1 mM GTP, 1 mM $MgCl_2$, 0.5 mM EDTA and 20 μM [^{32}P]NAD was incubated at 30 °C for 30 min (Watanabe et al.). Aliquots were loaded and run on 10% SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The gel was dried and subjected to autoradiography.

Iodination of cell surface proteins

$Na^{125}I$ (100 μCi) and three Iodo-Beads (Pierce, Rockford, IL) were added to F10 cells (1 x 10^6 cells) in 0.5 ml of 50 mM Tris-HCl, pH 7.5, and incubated at 4°C for 15 min. The reaction was stopped by removing the Iodo-Beads. The cells were washed 3 times with 0.5 ml of 50 mM Tris-HCl, pH 7.5, then sonicated in 0.5 ml of buffer B (50 mM Tris-HCl, pH 7.5, 0.5% CHAPS, 1 mM EDTA, 1 mM PMSF, and 10 μg/ml aprotinin) at 4°C for 10 min. The lysate was centrifuged at 4°C for 30 min. Supernatant (50 μl) was incubated with 100 μM Sph-1-P, Sph, or no compound at room temperature. Another 50 μl of supernatant was heated at 90°C for 3 min. These four samples were incubated with 1 mg immobilized Sph-1-P on glass beads at room temperature for 30 min. The beads were washed 3 times with buffer B and then incubated for 30 min with 100 μl of 100 μM Sph-1-P in buffer B at room temperature. The eluted proteins were analyzed on 10% SDS-PAGE.

Metabolism of [3-^{3}H]Sph

Cells in a 60-mm diameter dish were incubated with 1.5 μM [3-^{3}H]Sph (0.3 μCi) in 2.5 ml DMEM supplemented with 10% calf serum. The culture medium was collected, and cells were washed with 0.5 ml of fresh medium which was then combined with the recovered medium. Cells were washed twice with 2.5 ml PBS, and scraped twice with 0.5 ml ice-cold methanol. 1 ml

chloroform, 10 μl concentrated HCl and 1 ml methanol were added to the cell suspension (final ratio of chloroform:methanol: concentrated HCl was 100:200:1). For extraction, 7.5 ml of this same solution was added to 2 ml of medium. Lipids were extracted from the cell suspension or medium as described (Bligh and Dyer, 1959). The lower chloroform phase sample was dried and resuspended in a small volume of chloroform/methanol (2:1). A portion of the lower phase lipids was applied to silica gel 60 HPTLC plates (Merck, Darmstadt, Germany), and the plate was developed with butanol:acetic acid:water (3:1:1). After enhancer (Resolution TLC; E.M. Corp., Chestnut Hill, MA) treatment of the plate, autoradiography was performed with Kodak X-Omat film (Eastman Kodak, Rochester, NY) at -80°C for 3 to 7 days. The silica gel areas which contained radiolabeled sphingolipids were scraped off, and radioactivity was counted with a scintillation counter (Beckman).

Preparation of platelets, neutrophils, erythrocytes, plasma and serum

Platelets were prepared as described in Chapter 2. Neutrophils were isolated by dextran sedimentation and centrifugation on a Ficoll-Paque (Pharmacia LKB, Uppsala, Sweden) cushion, followed by hypotonic lysis to remove contaminating erythrocytes. Erythrocytes were washed in normal saline, and buffy coat was carefully removed. The washed cells were resuspended in a buffer composed of 138 mM NaCl, 3.3 mM NaH_2PO_4, 2.9 mM KCl, 1.0 mM $MgCl_2$, 1 mg/ml glucose, and 20 mM Hepes (pH 7.4).

For preparation of plasma, venous blood was mixed with 15% volume of ACD anticoagulant (Mustard et al., 1989), centrifuged at 2,000 x g for 15 min, and resultant plasma supernatant was collected. For preparation of serum, venous blood was placed directly into glass tubes and left for 60 min at room

temperature to allow blood clots to form. 15% volume of ACD was added to samples, and serum was separated by centrifugation at 2,000 x g for 15 min.

Quantitative measurement of Sph-1-P

Sph-1-P was extracted from 0.5 - 1.0 ml serum and plasma, 1-2 x 10^9 platelets, 5-10 x 10^7 neutrophils and 1-5 x 10^9 erythrocytes, and quantitatively measured as described in Chapter 2.

Results

Effects of immobilized Sph-1-P on cell motility

To investigate if Sph-1-P acts on cells from the outside, I first utilized Sph-1-P immobilized on controlled-pore glass beads, as described in Chapter 2. Chemical synthesis of conjugate through the ω-carboxyl group of Sph-1-P and Sph required at least 10 steps, and control beads contained N-acetylated amino group (see Chapter 2).

Cell motility assays were conducted with the immobilized sphingolipids. Figure 3.1 shows the effect of the immobilized Sph-1-P on cell motility of F10 cells. As the total amount of the immobilized Sph-1-P increased, the inhibitory effect on cell motility increased. At 0.2 mg/filter (4.4 nmol Sph-1-P/filter) of the beads, the immobilized Sph-1-P inhibited the cell motility almost completely, but the control beads did not. The immobilized Sph beads at 0.2 mg/filter (7.8 nmol/filter) inhibited the cell motility by ca. 35% of that seen with control beads. This may be due to concentrated Sph on the surface of the beads since 10 μM Sph has been shown to inhibit cell motility completely (Sadahira et al., 1992). Upon analysis of the medium by thin layer chromatography (TLC), we observed

no release of Sph-1-P or Sph from the beads during the cell motility assay. These results suggest that Sph-1-P acts on these cells from the outside.

Upon the addition of immobilized Sph-1-P to a suspension of F10 cells, cells adhered to the beads instantly and retained their round shapes even after 2 days in culture. This was in contrast to a percentage of cells which did not come into contact with the immobilized Sph-1-P and which then adhered normally to the dish (data not shown). Upon the addition of immobilized Sph, only a few F10 cells adhered to the beads, motility was hardly affected, and cells became round, and remained round as long as beads were attached. Control beads did not interact with the cells, and did not cause any change in motility or shape.

On the other hand, K562 cells, previously shown to have no response to Sph-1-P (Durieux et al., 1993), did not interact with immobilized Sph-1-P (data not shown). I examined viability of the round-shaped cells on immobilized Sph-1-P or Sph after 2 days of culture, and found viability to be >95% for each culture.

Specific binding of Sph-1-P to the surface of F10 cells

I assayed the binding of Sph-1-P to the F10 cell surface using [3-^{3}H]Sph-1-P. Figure 3.2A shows the time course of specific binding of [^{3}H]Sph-1-P at 4°C. The specific binding depended on time, reached equilibrium within 1 hour, and remained at equilibrium for at least 1 hour. Excess amounts of unlabeled Sph-1-P reduced the binding of [^{3}H]Sph-1-P to F10 cells (Figure 3.2B). This effect increased with the concentration of unlabeled Sph-1-P, reaching a constant at 50 μM where unlabeled Sph-1-P reduced the binding by ca. 75%. A large molar excess of unlabeled Sph-1-P was required for the replacement of [^{3}H]Sph-1-P because lipids are naturally highly lipophilic and promote high levels of nonspecific binding (Balsa et al., 1996). The specific binding was

saturated around 3 nM of the ligand (Figure 3.2C). Scatchard analysis of these data revealed a single class of affinity binding sites for Sph-1-P (Figure 3.2D). The Kd value was estimated to be ~60 nM, and the number of the binding sites was $4x10^4$ - $1x10^5$/cell. K562 cells showed only 5-10% of specific binding of [^{3}H]Sph-1-P, as compared with that of F10 cells. I also assayed for binding of [^{3}H]Sph to F10 cells, and found no specific binding.

Figure 3.3 shows results of competition binding experiments with unlabeled lipids which have similar structures to Sph-1-P. Sph, C2-Cer, TMS and Sphingosylphosphorylcholine (SPC) only slightly inhibited the binding of [^{3}H]Sph-1-P. SPC has been reported to activate K^+ conductance in guinea pig atrial myocytes through the same putative receptor as that of Sph-1-P (Bunemann et al., 1996), however, our result indicates that their receptor may be distinct in F10 cells. DMS inhibited the binding by ca. 35%, and DMS-1-P by ca. 55%, suggesting that the phosphate group is required for the specific binding. LPA, a structurally similar compound to Sph-1-P, remarkably reduced the binding (by ca. 85%), suggesting the possibility that LPA shares the binding site with Sph-1-P. I also examined the effect of a protease on the specific binding. Treatment of F10 cells with a non-specific protease (type XXV) remarkably reduced the specific binding of Sph-1-P by 81.2 ± 8.5% (n=3), indicating that the binding sites are proteins which are located on the surface of the cells. During these binding assays, [^{3}H]Sph-1-P was found not to be converted to any other compound by analysis with TLC. Taken together, these results indicate the presence of specific cell surface binding sites for Sph-1-P.

Motility inhibition and cell surface binding

Because DMS-1-P and LPA reduced the binding of Sph-1-P in the competition binding experiments (Figure 3.3), I tested the effects of these

compounds on cell motility (Figure 3.4). LPA (1 μM) had a strong inhibitory effect on F10 cell motility, inhibiting by ca. 85% and ca. 70% for 18 h (Figure 2.4A) and 8 h incubations (Figure 3.4B), respectively. The maximum effect of Sph-1-P appeared at 10 - 100 nM where it inhibited cell motility by 80 - 90% (Figures 3.4 A and B). In the competition binding experiments, LPA inhibited binding of [^{3}H]Sph-1-P as effectively as Sph-1-P (Figure 3.3). As there is no significant difference in dose dependence in cell motility inhibition, these results together suggest that LPA may be able to bind to the identical sites to those of Sph-1-P.

DMS-1-P at 0.1 μM and 1 μM inhibited the cell motility by ca. 40% and 60%, respectively (Figure 3.4A), presumably by binding to the Sph-1-P binding site. SPC did not inhibit F10 cell motility at low concentrations. At 0.01 μM SPC, cells migrated were 117.0 ± 18.7% of control (n=3), indicating that SPC binding sites are most likely distinct from those of Sph-1-P. Sph, C2-Cer, TMS and DMS have been shown to have no significant effect on cell motility (Sadahira et al., 1992; Sadahira et al., 1994; Bornfeldt et al., 1995). These results indicate that the binding seen in Figure 2.4 well associates with inhibition of cell motility.

Effects of suramin and pertussis toxin on the inhibitory effect of Sph-1-P on cell motility

I tested the effect of suramin, a known inhibitor of ligand-receptor interactions (Postma et al., 1996), on the inhibition of cell motility by 100 nM Sph-1-P. Suramin reduced the inhibitory effect by ca. 40% of the control (Figure 2.5). This finding further indicates the presence of a cell surface receptor for Sph-1-P.

A PTX-sensitive G protein has been shown to mediate Sph-1-P-induced signaling pathways for cell proliferation and intracellular Ca^{2-} mobilization (Goodemote et al., 1995; van Koppen et al., 1996). To study mediation of the G protein, I assayed haptotactic cell motility of F10 cells in the presence of 1-100 ng/ml PTX with a 30 min preincubation. Although [^{32}P]ADP-ribosylation experiments indicated that greater than 90% of total PTX-sensitive G proteins (possibly Gi and Go based on their molecular weights) were ADP-ribosylated (Figure 3.6A), PTX did not abolish the inhibitory effect of Sph-1-P on cell motility (Figure 3.6B). This suggests that PTX-sensitive G proteins are not involved in the signaling pathways of Sph-1-P inhibition of cell motility.

Identification of a cell surface binding protein for Sph-1-P

To identify the cell surface Sph-1-P binding protein, I labeled the surface of F10 cells with ^{125}I. The iodinated proteins were applied to the Sph-1-P immobilized on beads in the presence of Sph-1-P, Sph, or no compound. The bound proteins were eluted with Sph-1-P, and analyzed on SDS-PAGE. Two radioactive bands with molecular masses of ca. 41 and 79 kDa were observed, but after denaturing by heating no band was apparent (Figure 3.7, lane 1). The intensity of the 79-kDa band was reduced by the addition of Sph-1-P (Figure 3.7, lane 3) but not by that of Sph (Figure 3.7, lane 2) in comparison with the control (Figure 3.7, lane 4). The intensity of the 41-kDa band was less sensitive to the addition of Sph-1-P (Figure 3.7, lanes 2). ^{125}I-labeling of the surface of K562 cells, which showed only a small amount of the specific binding, revealed no 79-kDa or 41 kDa protein that binds to Sph-1-P. These results suggest that the 79-kDa, and less likely the 41-kDa, is a specific Sph-1-P binding protein on the cell surface.

Release of Sph-1-P from cells

Having demonstrated the presence of a cell surface binding protein (or a receptor) for Sph-1-P, I raised the possibility that Sph-1-P acts as an intercellular modulator or messenger. To test this possibility, I examined if Sph-1-P is released from the cells into the culture medium upon stimulation. Figure 1.8 shows the metabolism of [3-^{3}H]Sph in A31 cells and the subsequent release of ^{3}H-labeled Sph-1-P into the medium in the presence of serum. Exogenously-added [^{3}H]Sph was rapidly taken up within 3 min, reaching a plateau at 8 min. [^{3}H]Sph was quickly phosphorylated to ^{3}H-labeled Sph-1-P (Figures 3.8A and C) with ceramide and sphingomyelin (SM) also being formed. The level of Sph-1-P peaked at 8 min and then decreased. Analysis of lipids in the medium showed that Sph-1-P was released from the cells, but ceramides or other lipids produced in the cells were not (Figure 3.8B). At 3 min, Sph-1-P was slightly detectable in the medium (Figure 3.8C). At 60 min, 28.5$\pm$3.8% (n=3) of Sph-1-P was in the medium and decreased thereafter. The decrease of Sph-1-P in the medium can be ascribed to internalization or binding to the cells. At 180 min, the cell viability was ca. 98%. Sph-1-P was scarcely released from the cells into the medium which contained only 0.1% BSA instead of serum.

F10 cells showed a similar metabolism of [3-^{3}H]Sph to that in A31 cells (Figure 3.9A). Exogenously-added [^{3}H]Sph was rapidly incorporated into the cells, and ^{3}H-labeled Sph-1-P was quickly formed. The maximum level of Sph-1-P appeared at 18 min. The analysis of the medium (Figure 3.9B) showed that a small amount of Sph-1-P was released into the media (ca. 2.5% at 60 min) in the presence of serum.

Sph-1-P levels in plasma and serum

I examined if Sph-1-P levels in plasma and serum are high enough to regulate cell motility. When plasma or serum extracts were N-acetylated with [^{3}H]acetic anhydride for Sph-1-P measurements, one clear band was detected, on TLC (Figure 3.10A). The band (acylation product in plasma and serum) coincided with FAB-MS-identified C_2-Cer-1-P in TLC mobility with three different solvent systems, i.e., butanol/acetic acid/water (3:1:1) (Figure 3.10A), chloroform/methanol/7N NH_4OH/water (80:20:0.5:0.5) (data not shown) and chloroform/methanol/acetic acid/water (65:43:1:3) (data not shown). These results indicates that Sph-1-P was clearly detected and identified as a normal constituent of plasma and serum. Human plasma, which does not include platelet discharge, contained about 190 pmol/ml (190 nM) of Sph-1-P, while clotted blood serum, into which the contents of platelets should be released, contained about 480 pmol/ml (480 nM) of Sph-1-P. When the Sph-1-P levels in paired plasma and serum samples obtained from 6 healthy adults were measured, the serum Sph-1-P/plasma Sph-1-P ratio was found to range from 1.36 to 4.05, the average being 2.65 $\pm$ 1.26 (mean $\pm$ SD) (Figure 3.10B).

Determination of Sph-1-P level in blood cells

Next, I analyzed blood cells for Sph-1-P. Platelets lack Sph-1-P lyase activity, possess persistently active Sph kinase (Chapter 2). In agreement with this, platelets contain 1.4 nmol of Sph-1-P/10^9 platelets. When a comparison was made using the mol% Sph-1-P/phospholipid value, the Sph-1-P level in platelets was found to be over 10 times higher than that in neutrophils or erythrocytes (Table III). Since Sph-1-P is most abundantly stored in platelets, compared with in other blood cells, and can be released into the medium upon the stimulation of platelets (Chapter 2), it is most likely that the source of discharged Sph-1-P during the clotting process is platelets.

Discussion

Sph-1-P, the first product of Sph catabolism, causes changes of cell physiology when exogenously added to culture medium. Cellular response varies depending on type of cell. So far, two types of cell physiological change caused by high concentrations (μM order) of exogenous Sph-1-P are known:

(i) Platelets. 10-20 μM exogenous Sph-1-P, but not Sph, Cer, or any other sphingolipid tested, causes shape change, aggregation, and Ca^{2+} mobilization in platelets, even though platelets contain a high level of endogenous Sph-1-P (see Chapter 2).

(ii) Fibroblasts. 5-10 μM exogenous Sph-1-P causes intracellular Ca^{2+} mobilization in some types of cells (*e.g.* Swiss 3T3), leading to strong mitogenesis. Transiently enhanced level of Sph-1-P in Swiss 3T3 cells is also caused by mitogenic stimuli (*e.g.* PDGF, TPA), with associated increase in Ca^{2+} mobilization leading to cell proliferation. Thus, Sph-1-P is implicated as a second messenger following primary stimuli of cells.

In contrast to cellular responses (i) and (ii) described above, cell motility response to Sph-1-P is unique in that very low concentrations (10-100 nM) of exogenous Sph-1-P are sufficient to inhibit both chemotactic and haptotactic motility of various types of cells (*e.g.* B16 mouse melanoma, Balb/c 3T3 fibroblast, human melanoma, fibrosarcoma, osteosarcoma). Under this condition, Sph-1-P does not affect mitogenesis of these cells (Sadahira et al., 1992; Sadahira et al., 1994). Such low Sph-1-P concentrations causing motility change may be physiologically relevant. In this chapter, I demonstrate the physiological

relevance of extracellular action of Sph-1-P on mouse B16 melanoma cells through its receptor, in analogy to its effect on platelets.

The motility-inhibitory effect of Sph-1-P on B16 melanoma or Balb/c 3T3 cells could be based on either of the following mechanisms: (a) Sph-1-P enters the cell without a specific receptor, and binds to a target molecule or molecules that control cell motility; (b) Sph-1-P binds initially to a specific cell surface receptor which triggers signaling, causing a functional change of specific molecule(s) that control cell motility. Endogenous Sph-1-P level in these cells is low, and there is no strong evidence against the hypothesis that exogenously added Sph-1-P is transported into cells and affects signal transduction to inhibit cell motility. However, the following observations, analogous to those in platelets, suggest the presence of an Sph-1-P receptor that mediates initial interaction of Sph-1-P with the cell, *i.e.* mechanism (b) above:

1. Contact of melanoma cell surface with Sph-1-P immobilized on porous glass beads through ω-carbon inhibited the cell motility. Control experiments with immobilized Sph or N-acetylated beads did not cause motility inhibition. These results indicate that contact with head group of Sph-1-P induces motility inhibition. A possibility that the cells hydrolyze ω-carboxyl linked amide group affixed to glass beads is highly unlikely, since the presence of extracellular "amidase" is unknown.

2. Binding studies with [^{3}H]Sph-1-P showed the presence of specific binding sites for Sph-1-P on the surface of F10 cells. Competition experiments with various Sph-1-P analogs indicated that the head phosphate group and core structure of Sph are both required for the specific binding. Cell motility assays with various compounds revealed that the binding is associated with inhibition of cell motility. Treatment with a protease disrupted the specific binding,

suggesting that the binding sites are proteins which are located on the surface of the cells.

3. Suramin, an inhibitor of ligand-receptor interactions (Postma et al., 1996), reduced the inhibitory effect of Sph-1-P on cell motility. Suramin may be toxic to the cells, which lowers the validity of this experiment.

4. Two cell surface labeled proteins with Mr 41 and 79 kDa were demonstrated as the Sph-1-P binding site by labeling with ^{125}I followed by affinity chromatography using Sph-1-P-immobilized beads. The intensity of the 79-kDa band was reduced by the addition of Sph-1-P, whereas that of the 41-kDa band was less sensitive to Sph-1-P addition. This suggests that the 79-kDa protein might be a specific cell surface receptor for Sph-1-P and that Sph-1-P may regulate cell motility through this specific cell surface binding site.

5. The binding site for Sph-1-P appears to be shared with LPA, since addition of LPA abolished Sph-1-P binding to the melanoma cell surface, in analogy to the process in platelets.

Recently, G-protein-coupled orphan receptor EDG-1 was identified as a high affinity Sph-1-P receptor (Lee et al., 1998). This finding supports the existence of a Sph-1-P receptor, although the receptor on mouse melanoma is not identical to EDG-1 (Hla, T., personal communication).

LPA has structural similarity to Sph-1-P (Durieux et al., 1993) and elicits various biological effects on such phenomena as cell proliferation and differentiation, through its receptor, coupled with PTX-sensitive G-proteins (Moolenaar, 1995). The cloning of a LPA receptor of neocortical neuroblast cell line was reported (Hecht et al., 1996). In our present study, in competition binding experiments, LPA abolished binding of [^{3}H]Sph-1-P as effectively as did

Sph-1-P itself (Figure 3.3) suggesting that LPA and Sph-1-P share the same binding site. Our observation is inconsistent with recent findings that an LPA receptor is completely distinct from that of Sph-1-P in various cells (Jalink et al., 1995; van Koppen et al., 1996; Postma et al., 1996). However, cross-desensitization between Sph-1-P and LPA was observed in *Xenopus* oocytes (Duriex and Lynch, 1993), atrial myocytes (Bunemann et al., 1996), human platelets (Chapter 1, Yatomi et al., 1997a) and FRTL-5 thyroid cells (Okajima et al., 1997). These observations strongly suggest the existence of cell-type specific receptors shared by Sph-1-P and LPA in some types of cells, whereas in other types of cells, receptors for these two bioactive lipids are distinct. Regarding cell motility, LPA has been also reported to enhance cancer cell motility and invasiveness, through Rho-mediated protein tyrosine phosphorylation (Imamura et al, 1993; 1996), and through actin polymerization by a PTX-sensitive mechanism (Ha et al, 1994). As LPA was found to be a weaker inhibitor of cell motility compared to Sph-1-P, I cannot deny the possibility that LPA may act on both Sph-1-P binding sites and its own receptor, which have distinct or opposite effects on F10 cell motility. Recently, interestingly, the cloning of an LPA receptor of neocortical neuroblast cell lines was reported, showing its molecular mass to be 41-42 kDa (Hecht et al., 1996). The 41-kDa protein which is less sensitive to Sph-1-P (Figure 3.7) has possibility of an LPA receptor.

Many known receptors transduce signals through G proteins sensitive to PTX, which ADP-ribosylates and inactivates Gα proteins. Sph-1-P induces signaling pathways of cell proliferation or Ca^{2+} mobilization that are mediated by PTX-sensitive G proteins (Goodemote et al., 1995; Wu et al., 1995; van Koppen et al., 1996). In FRTL-5 thyroid cells, Sph-1-P-induced responses were reported to be mediated by both PTX-sensitive and insensitive G proteins (Okajima et al.,

1997). In the present study, PTX treatment did not reduce the inhibitory effect of Sph-1-P on F10 cell motility, indicating that PTX-sensitive G proteins are not involved in the signaling pathways of cell motility regulation. In this regard, the signaling pathways of Sph-1-P in motility inhibition of F10 cells are unique among the reported signaling mechanisms induced by Sph-1-P in other systems.

Recent studies showed that focal adhesion kinase (FAK) plays an important role in cell migration, as it is localized to focal adhesion sites and is phosphorylated and activated by integrin/ligand interactions (Zachary and Rozengurt, 1992; Ilic et al., 1995). The small G protein Rho also has been implicated in the regulation of actin stress fiber formation, motility, adhesion, and morphology of cells (Takaishi et al., 1993; Nobes and Hall, 1995). Sph-1-P was found to induce Rho-dependent neurite retraction in neuronal cells (Postma et al., 1996). It is possible that Sph-1-P controls cell motility through its receptor by affecting these factors and regulating actin filament formation. In a follow-up study, the target molecule of Sph-1-P was found to be small G-protein Rho. A proposed molecular mechanism following the effect of Sph-1-P on Rho A is summarized in the Appendix.

Very low concentrations (10 -100 nM) of exogenous Sph-1-P are capable of controlling F10 cell motility, suggesting a physiological role of Sph-1-P. I found that Sph-1-P was specifically released from F10 melanoma and A31 fibroblast cells into culture medium in response to serum stimulation. If 30% of cellular Sph-1-P is released (Figure 3.8), the locally-concentrated Sph-1-P in the extracellular environment could be enough to regulate cell motility. This implies that Sph-1-P released from the cells might regulate cell motility in an autocrine or a paracrine fashion, through the proposed receptor mechanisms. I also found that Sph-1-P stored in platelets is released upon agonist stimulation (Chapter 2), and that human plasma contains 50-350 nM Sph-1-P (Figure 3.10B). Major portions

of plasma Sph-1-P seemed to originate from activated platelets, although the possibility remains that cells other than platelets may release Sph-1-P into the extracellular environment. Concentration of Sph-1-P in plasma is high enough to control cell motility, suggesting that Sph-1-P plays a role under physiological conditions.

Here, I demonstrate and characterize the specific binding protein for Sph-1-P on the cell surface. It is still unknown that this binding protein is a functional receptor. Purification and cloning of this protein is required to examine if this is a Sph-1-P receptor. I further illustrate that Sph-1-P regulates cell motility, at least partly, by binding to the cell surface receptor which triggers a PTX-sensitive G protein-independent signaling mechanism. Sph-1-P is specifically released from the cells into the extracellular environment, and human plasma contains high concentrations of Sph-1-P. These results suggest a physiological role of Sph-1-P as an autocrine or a paracrine intercellular (cell to cell) modulator or messenger which regulates cell motility (Figure 3.11).

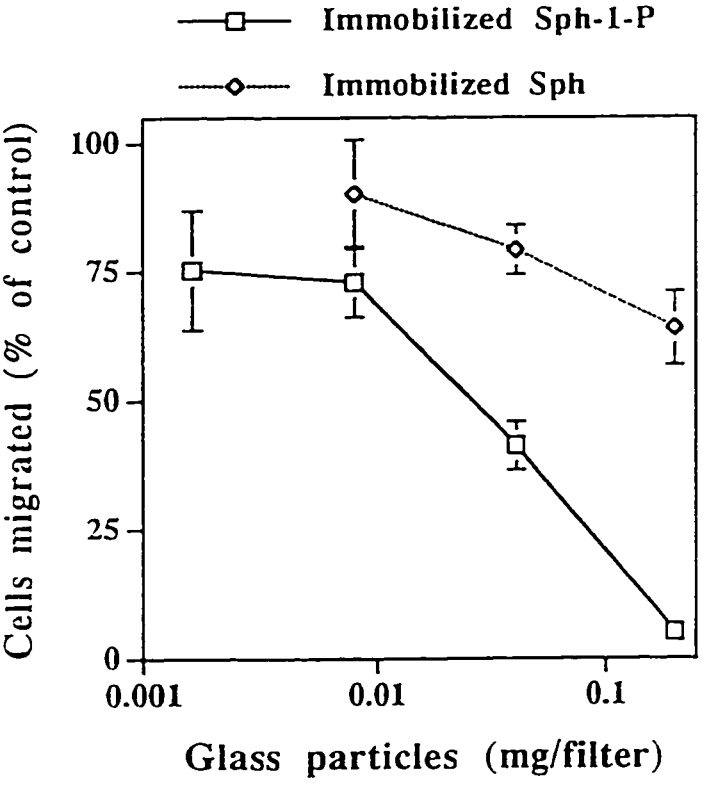


Fig. 3.1. Effect of immobilized sphingolipids on F10 cell motility. Haptotactic cell motility of F10 cells was assayed using Transwell chambers as described under "Materials and Methods". Glass beads with immobilized sphingolipids were added to the upper compartment with the cells. Data represent% of motility assayed with control beads (mean ± S.D. (n = 3)).

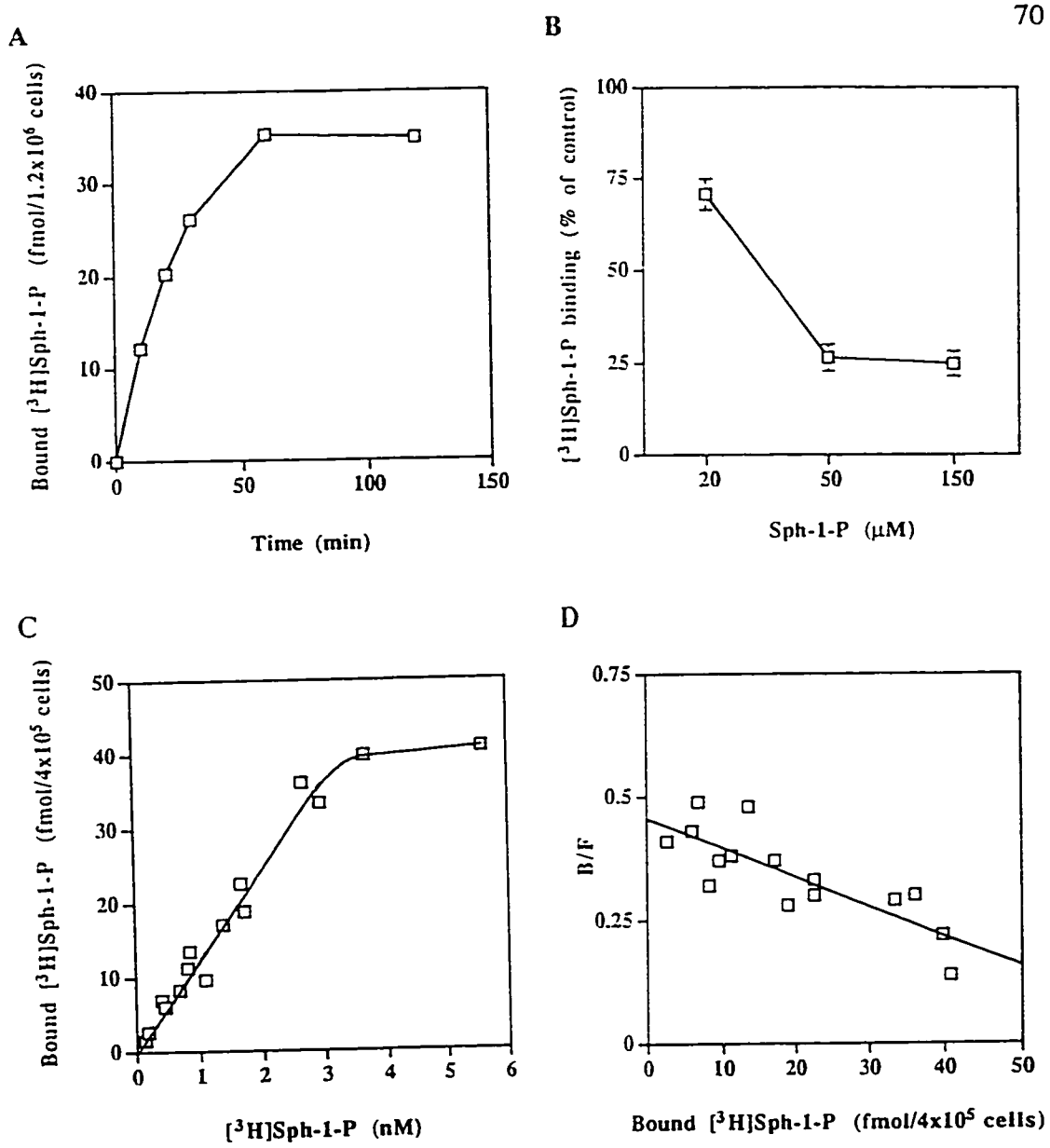


Fig. 3.2. Specific binding of [^{3}H]Sph-1-P to F10 cell surface. The specific binding was determined as described under "Materials and Methods". (A) Time course of specific binding of [^{3}H]Sph-1-P. F10 cells ($1.2x10^6$ cells) were incubated with 2.7 nM [^{3}H]Sph-1-P at 4 °C for the indicated time. (B) Effect of unlabeled Sph-1-P on [^{3}H]Sph-1-P binding. F10 cells ($1.2x10^6$ cells) were incubated with 2.7 nM [^{3}H]Sph-1-P at 4 °C for 1 hour in the presence of the indicated concentrations of unlabeled Sph-1-P. (C) Dose curve of specific binding of [^{3}H]Sph-1-P. F10 cells ($4x10^5$ cells) were incubated with various concentrations of [^{3}H]Sph-1-P at 4 °C for 1 hour. (D) Scatchard analysis of specific binding of [^{3}H]Sph-1-P to F10 cells using the data in panel C. Data are typical from three independent experiments.

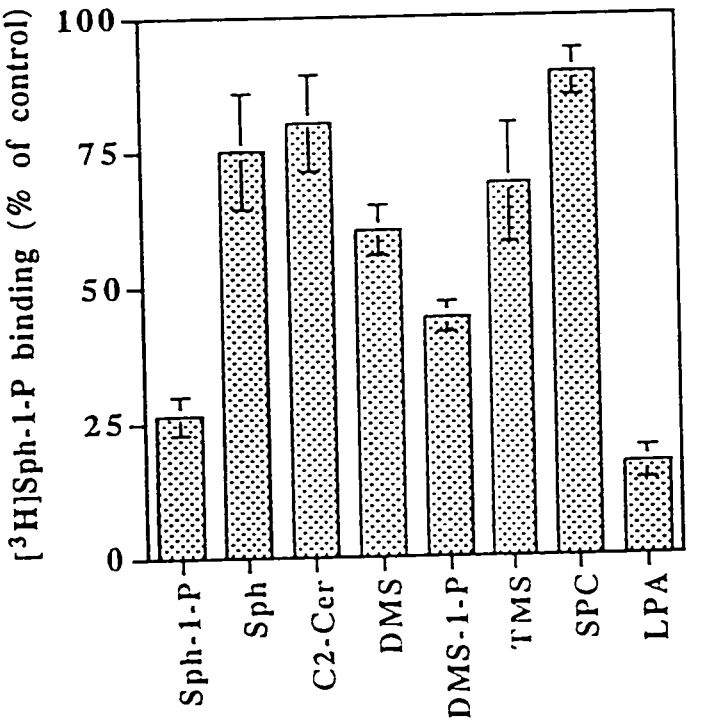
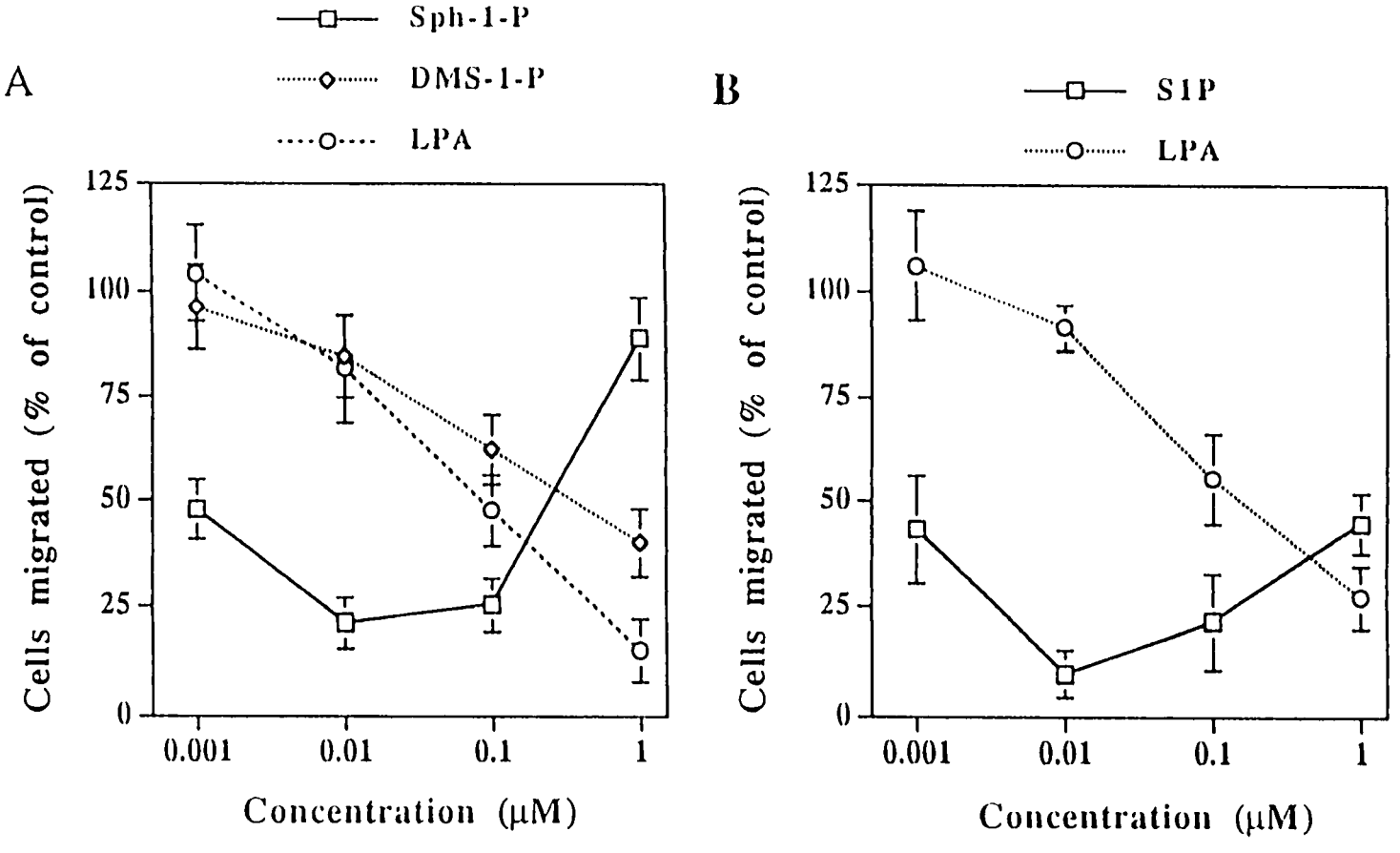


Fig. 3.3. Effect of various lipids on [^{3}H]Sph-1-P binding to F10 cells. F10 cells ($1.2x10^6$ cells) were incubated with 2.7 nM [^{3}H]Sph-1-P at 4 °C for 1 hour in the presence and the absence of various lipids (50 μM). The binding was determined as described under "Materials and Methods". Data represent mean ± S.D. (n = 3).

Fig. 3.4. Effect of Sph-1-P, DMS-1-P and LPA on F10 cell motility.
Haptotactic cell motility of F10 cells was assayed using Transwell chambers as described under "Materials and Methods". The test compound was added in the lower chamber. Data represent mean $\pm$ S.D. (n = 3). Incubation time : (A) 18 h; (B) 8 h.


A
Sph-1-P
DMS-1-P
LPA
Cells migrated (% of control)
125
100
75
50
25
0
0.001
0.01
0.1
1
Concentration (μM)
B
S1P
LPA
Cells migrated (% of control)
125
100
75
50
25
0
0.001
0.01
0.1
1
Concentration (μM)

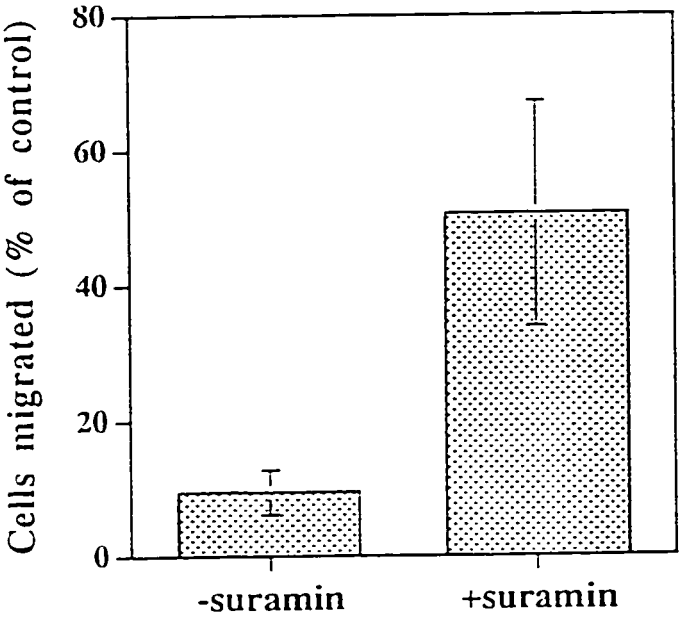
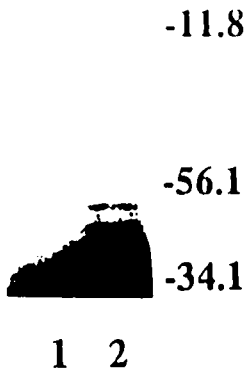
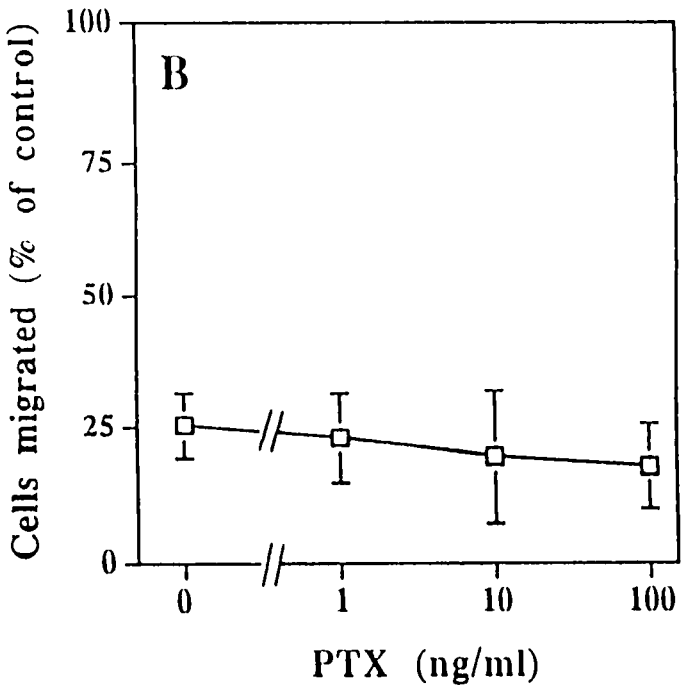


Fig. 3.5. Effect of suramin on inhibition of cell motility by 0.1 μM Sph-1-P. F10 cells were preincubated with 1 mg/ml suramin for 30 min or untreated, and haptotactic cell motility was assayed with or without 1 mg/ml suramin in both upper and lower chambers. Data represent mean ± S.D. (n = 3).

Fig. 3.6. Cell motility inhibition by Sph-1-P is not sensitive to pertussis toxin (PTX) treatment. (A) PTX-induced [^{32}P]ADP-ribosylation of membrane proteins from F10 cells. F10 cells were incubated with or without 100 ng/ml of PTX overnight and [^{32}P]ADP-ribosylation of membrane proteins by PTX was performed, followed by analysis on 10% SDS-PAGE. Lane 1, membrane from PTX-treated cells (12.5 μg); Lane 2, membrane from control cells (12.5 μg). Unreacted [^{32}P]NAD is seen in the low molecular weight regions. Data is typical from three independent experiments. Values on the figure indicate molecular weight (kDa). (B) Effect of PTX on inhibition of cell motility by 0.1 μM Sph-1-P. F10 cells were preincubated with the indicated concentrations of PTX for 30 min, and haptotactic cell motility was assayed with the indicated concentration of PTX in both upper and lower chambers. Data represent mean ± S.D. (n = 3).

A


-11.8
-56.1
-34.1
1 2


B
Cells migrated (% of control)
100
75
50
25
0
0
1
10
100
PTX (ng/ml)

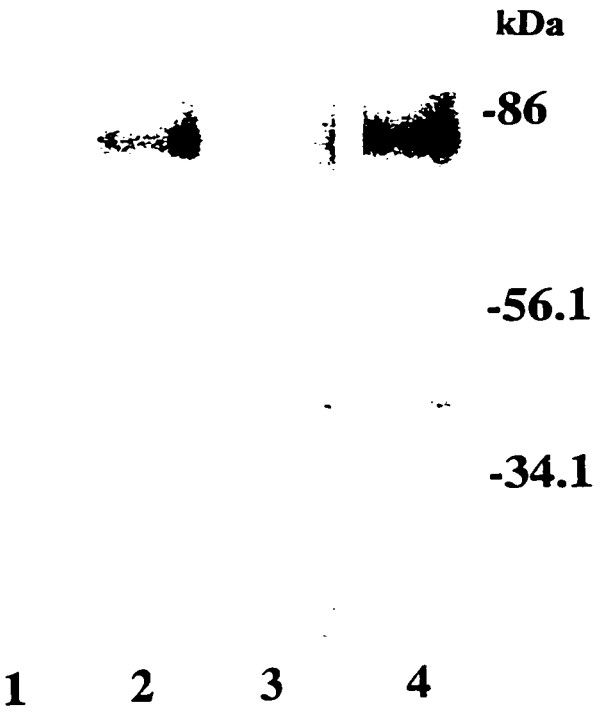


Fig. 3.7. Identification of the Sph-1-P binding protein. The surface proteins of F10 cells were labeled with ^{125}I. The cells were lysed, and proteins were denatured by heating (lane 1) or were incubated with 100 μM Sph (lane 2), 100 μM Sph-1-P (lane 3) or no compound (lane 4). The proteins were applied to the Sph-1-P immobilized on glass beads. The bound proteins on the beads were eluted with 100 μM Sph-1-P, and were analyzed on 10% SDS-PAGE. Values on the figure indicate molecular weight (kDa). Data are representative of three independent experiments.

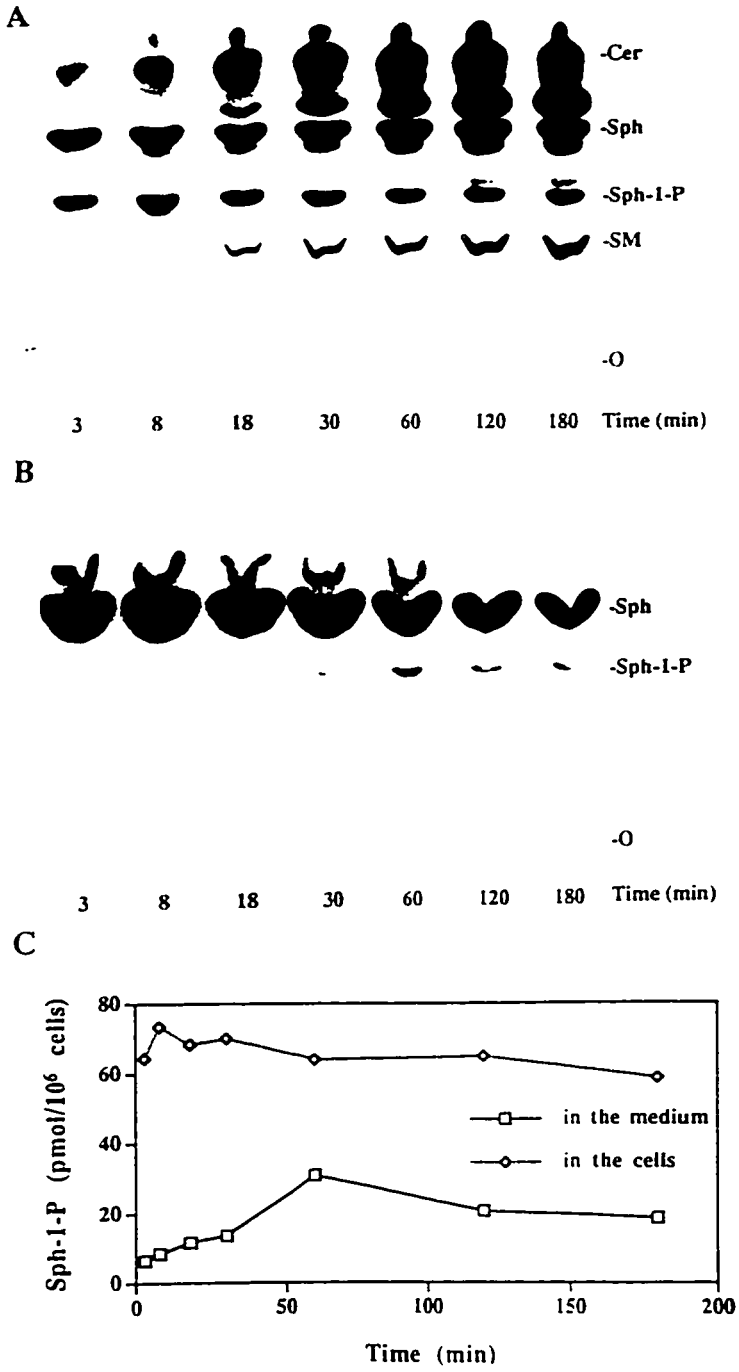


Fig. 3.8. Metabolism of [3-^{3}H]Sph in A31 cells and the release of ^{3}H-labeled Sph-1-P into the medium containing 10% BCS. Confluent cultures of A31 cells were incubated with 1.5 μM [^{3}H]Sph for the indicated period. The lipids in the cells (A) and the medium (B) were extracted and analyzed by TLC as described under "Materials and Methods". (C) Radioactive bands of Sph-1-P on the TLC plate shown in panels A and B were scraped and quantified. Data are representative of two independent experiments.

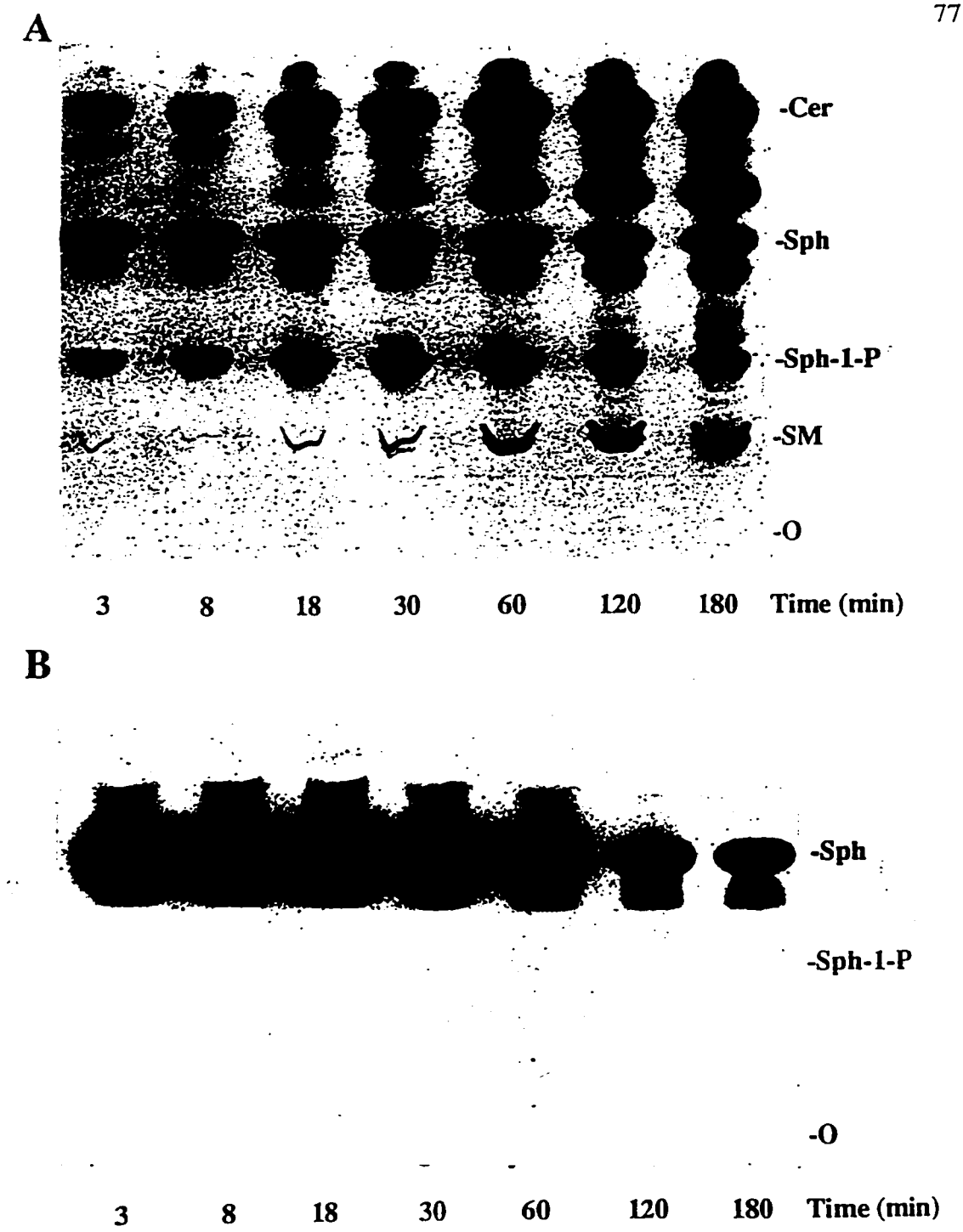


Fig. 3.9. Metabolism of [3-^{3}H]Sph in F10 cells and release of ^{3}H-labeled Sph-1-P into the medium containing 10% FCS. F10 cells ($9x10^6$ cells) were incubated with 1.5 μM [^{3}H]Sph for the indicated period. The lipids in the cells (A) and the medium (B) were extracted and analyzed by TLC as described under "Materials and Methods". Data are representative of two independent experiments.

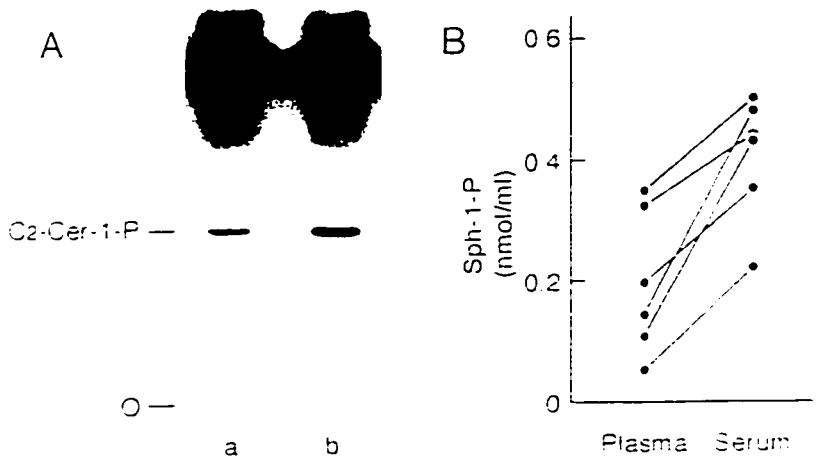


Fig. 3.10. Determination of the Sph-1-P levels in human plasma and serum. (A) Sph-1-P extracted from human plasma (lane a) and serum (lane b) was N-acylated with [^{3}H]acetic anhydride into [^{3}H]C2-Cer-1-P. (B) Determination of the Sph-1-P levels in paired plasma and serum samples obtained from 6 healthy adults.

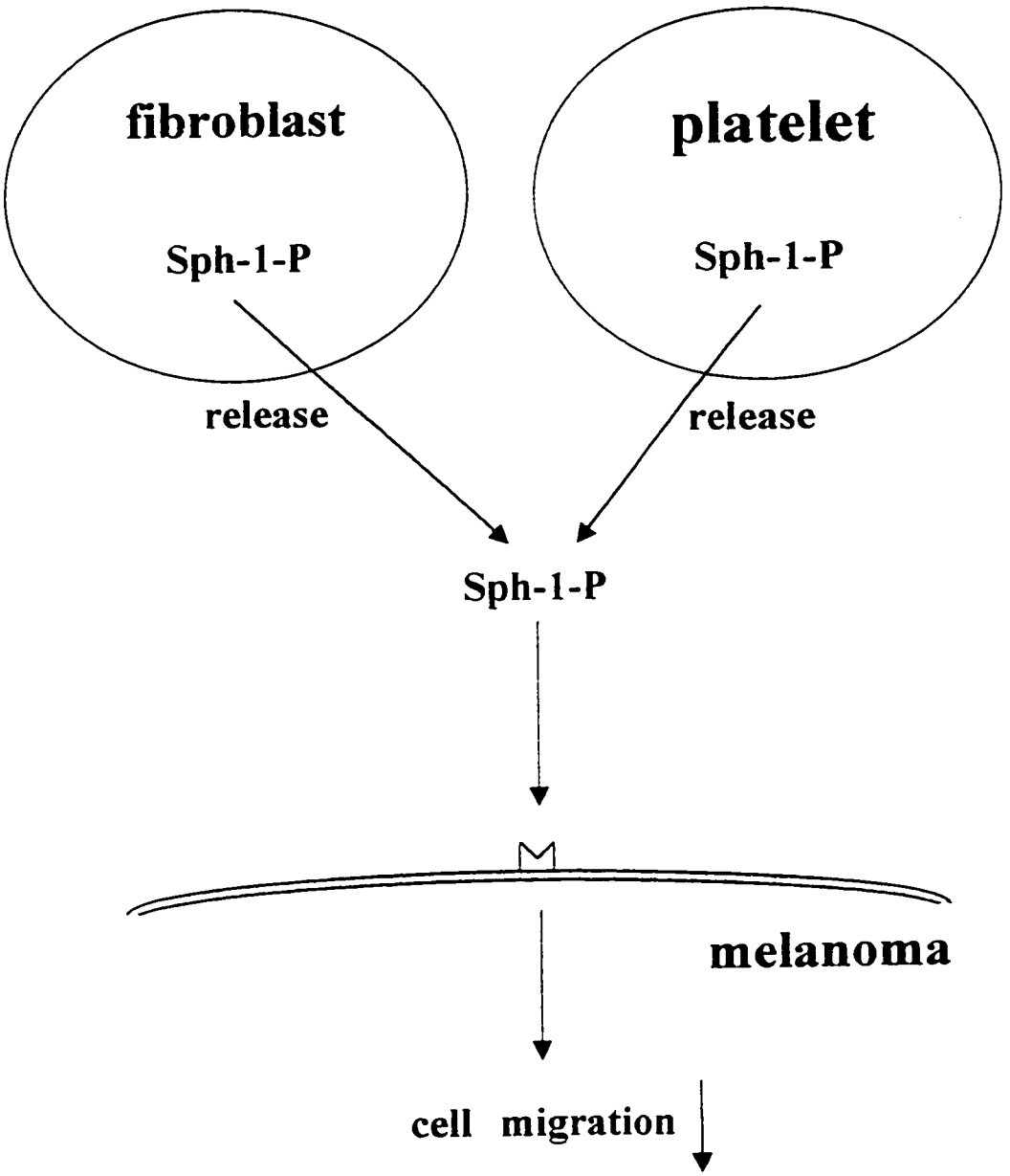


Fig. 3.11. Sph-1-P as an intercellular messenger which regulates cell motility. Sph-1-P released from platelets or other cells may regulate cell motility.

Table III. Sph-1-P levels in human platelets, neutrophils, and erythrocytes. Sph-1-P extracted from blood platelets, neutrophils and erythrocytes was measured and adjusted as to the phospholipid level. Values are means ± SD ($n=3$).

	Cell count	Sph-1-P (pmol)	mol% Sph-1-P/ phospholipid
Platelets	10^8	141.0 ± 4.0	0.211 ± 0.014
Neutrophils	10^7	15.2 ± 2.8	0.018 ± 0.003
Erythrocytes	10^8	7.2 ± 1.7	0.019 ± 0.002

Chapter 4

Concluding Remarks

Sphingolipids and their metabolites have been demonstrated to function as second messengers or modulators in signal transduction. Sphingosine-1-phosphate (Sph-1-P), formed from sphingosine (Sph) by Sph kinase, is the initial product of Sph catabolism, and is cleaved by Sph-1-P lyase to phosphoethanolamine and palmitaldehyde. Sph-1-P was reported to act as a lipid second messenger, leading to mitogenesis in certain cell types (*e.g.* Swiss 3T3 fibroblasts).

Sph-1-P causes changes of cell physiology when added exogenously to culture medium. Cellular response to Sph-1-P varies extensively depending on type of cell. So far, three types of cell physiological change have been reported: (i) 10-20 μM exogenous Sph-1-P, but not Sph or ceramide, causes shape change, aggregation, and Ca^{2+} mobilization in platelets, even though platelets have a high level of stored Sph-1-P. The effect is synergistic with weak platelet agonists such as epinephrine and ADP. (ii) A very low concentration (10-100 nM) of Sph-1-P, but not Sph or ceramide, inhibits chemotactic and haptotactic motility of various types of cells (*e.g.* B16 mouse melanoma, Balb/c 3T3 fibroblast, arterial smooth muscle, human melanoma, fibrosarcoma, osteosarcoma, etc.), but does not affect mitogenesis of these cells. (iii) Addition of 5-10 μM exogenous Sph-1-P to certain types of cells (*e.g.* Swiss 3T3) causes intracellular Ca^{2+} mobilization, leading to strong mitogenesis. Transient enhanced level of Sph-1-P in Swiss 3T3 cells is also caused by mitogenic stimuli (*e.g.* PDGF, TPA), with subsequent increase in Ca^{2+} mobilization leading to cell proliferation. Thus, Sph-1-P is implicated as a second messenger following primary stimuli of cells.

Although phenotypic changes caused by Sph-1-P vary extensively as described above, the initial common step is its effect at the cell surface, presumably through a specific receptor. Interestingly, Sph-1-P is released from various types of cells after cell stimulation. Sph-1-P may act as an extracellular/ intercellular messenger to modulate cellular functions. This thesis is focused primarily on the possible existence of a Sph-1-P receptor at the surface of platelets and B16 melanoma cells.

In Chapter 2, studies of Sph-1-P metabolism and the mechanism by which Sph-1-P induces activation of anuclear human platelets are described. Platelets display persistently active Sph kinase and abundantly store Sph-1-P because of lack of Sph-1-P lyase activity. Although exogenously-added Sph-1-P promptly activated platelets, intracellular Sph-1-P, formed from exogenously-added Sph by cytosolic Sph kinase, had a negligible effect. This suggests that Sph-1-P acts extracellularly through its specific receptor. However, one possible explanation is that intracellular Sph-1-P is stored in a specific locus and not used for signal transduction. Therefore, two mechanisms for the initial effect of Sph-1-P can be considered: (a) Sph-1-P enters the platelet without a specific receptor and affects signal transduction. (b) Sph-1-P binds initially to a specific cell surface receptor which triggers signal transduction leading to phenotypic changes. I focused on mechanism (b), and to test this possibility, two approaches were employed.

In the first approach, a specific probe was synthesized, *i.e.* an immobilized conjugate of Sph-1-P with controlled-pore glass beads covalently linked through the ω-carbon of Sph residue. For this purpose, Sph-1-P compound having ω-carboxyl residue was synthesized and covalently coupled to a long chain alkyl amine on the bead surface. This chemical synthesis required at least 10 reaction steps. Contact of platelet surfaces with immobilized Sph-1-P resulted in platelet activation in terms of shape change and aggregation. However, similar synthetic

Sph residue, without phosphate group, affixed through ω-carbon to beads, did not activate platelets. This experiment suggests that Sph-1-P touches the cell surface and induces activation, *i.e.* Sph-1-P acts extracellularly.

In a second approach, I detected specific binding sites for [^{3}H]Sph-1-P on the platelet surface, suggesting extracellular effects of Sph-1-P on plasma membrane receptors. This specific Sph-1-P binding was inhibited not by other sphingolipids but by lysophosphatidic acid (LPA), and platelet aggregation response to LPA was specifically desensitized by prior addition of Sph-1-P. Finally but interestingly, internally stored Sph-1-P was released extracellularly upon platelet activation. These results suggest that Sph-1-P acts not intracellularly but intercellularly, following release from activated platelets.

Chapter 3 presents evidence for a physiological role of Sph-1-P as an extracellular regulator of cell motility. A very low concentration (10-100 nM) of Sph-1-P, which may be physiologically relevant, is capable of inhibiting cell motility. I attempted to identify and characterize specific cell surface binding sites for Sph-1-P in mouse melanoma F10 cells, using an approach similar to that used to study platelet activation, *i.e.* immobilized Sph-1-P covalently linked to glass beads. Whenever F10 cells were touched with Sph-1-P residues affixed on beads, cell motility stopped. In this process, Sph-1-P residue cannot be internalized. Sph residue without phosphate group, affixed to beads, does not stop cell motility.

Binding assays with [^{3}H]Sph-1-P revealed the presence of specific cell surface binding sites for Sph-1-P in F10 cells. Scatchard analysis demonstrated a single class of binding sites for Sph-1-P with Kd ~60 nM. The binding of [^{3}H]Sph-1-P to F10 cells was inhibited by addition of excess unlabeled Sph-1-P but not other natural sphingolipids, and was abolished by protease treatment. In

order to identify cell surface binding site, cell surface proteins were labeled with ^{125}I iodination followed by affinity chromatography using Sph-1-P immobilized on glass beads. Two bands, with 41 and 79 kDa, were thus detected for the first time as Sph-1-P binding proteins expressed at the melanoma cell surface, although the 41-kDa protein was less specific to Sph-1-P. Binding of Sph-1-P to melanoma cells was also inhibited by LPA, whose molecular configuration is similar to that of Sph-1-P, although a higher concentration of LPA (1-10 μM) was required to ensure motility inhibition. Pertussis toxin (PTX) treatment did not abolish the motility inhibition by Sph-1-P, suggesting that no PTX-sensitive G-protein is involved in the signaling. These results strongly suggest that Sph-1-P regulates cell motility through specific binding to the cell surface receptor protein, independent of PTX-sensitive G-protein. Sph-1-P itself was released into the culture medium from the cells upon serum stimulation, and concentration of Sph-1-P in plasma was also high enough to control cell motility. Collectively, the results suggest that Sph-1-P can function as an intercellular (cell to cell) messenger in regulating haptotactic and chemotactic cell motility.

Recently, Sph-1-P receptor was alleged to be G-protein-coupled orphan receptor EDG-1 (Lee et al., 1998), to which Sph-1-P binds with high affinity. This finding supports the existence of a Sph-1-P receptor, although the receptor on mouse melanoma is not identical to EDG-1 (T. Hla, personal communication).

The initial step of physiological function of Sph-1-P is its interaction at a cell surface binding site, which is alleged to be a specific receptor shared with LPA It is still unknown that this binding site is a functional receptor. Purification and cloning of this binding site is required to examine if this is a functional Sph-1-P receptor. All the signal transduction events induced by Sph-1-P occur subsequent to its binding to the receptor. Consequences are

mitogenesis (in certain cell types), motility inhibition (often observed in various cell types, particularly tumor cells), and various functional changes as seen in platelet activation. Events leading to cell motility inhibition in B16 melanoma appear to be mediated initially by Rho activation (see Appendix), while the initial event associated with platelet activation is induction of Ca^{2+} influx. The exact mechanism of these events is far from being known. Nevertheless, my studies show that Sph-1-P acts on cells extracellularly to elicit functional changes. Further characterization of the receptor and consequent changes of signal transduction are important topics for future studies.

Bibliography

Ahmed, S. N., London, E. and Brown, D. (1997) On the origin of sphingolipid/cholesterol-rich detergent-insoluble cell membranes: Physiological concentrations of cholesterol and sphingolipid induce formation of a detergent-insoluble, liquid-ordered lipid phase in model membranes. Biochemistry 36, 10944-10953.

Balsa, D., Merlos, M., Gial, M., Fernando, R., Garcia-Rafanell, J. and Forn, J. (1996) Characterization of [^{3}H]apafant binding to PAF receptor on rabbit platelet membranes: a comparison of a microplate filtration system and a standard method. J. Pharmacol. Toxicol. Methods 36, 53-62.

Benton, A. M., Gerrard, J. M., Michiel, T. and Kindom, S. E. (1982) Are lysophosphatidic acids or phosphatidic acids involved in stimulus activation coupling in platelets? Blood 60, 642-649.

Berridge, M. J., (1988) Inositol lipids and calcium signalling. Proc. Royal Soc. London (Biol) 234, 359-378.

Bligh, E. G. and Dyer, W. J. (1959) A rapid method of total lipid extraction and purification. Can. J. Biochem. Physiol. 37, 911-917.

Bornfeldt, K. E., Graves, L. M., Raines, E. W., Igarashi, Y., Wayman, G., Yamamura, S., Yatomi, Y., Sidhu, J. S., Krebs, E. G., Hakomori, S. and Ross, R. (1995) Sphingosine-1-phosphate inhibits PDGF-induced chemotaxis of human arterial smooth muscle cells: spatial and temporal modulation of PDGF chemotactic signal transduction. J. Cell Biol. 130, 193-206.

Boubelik, M., Floryk, D., Bohata, J., Draberova, L., Macak, J., Smid, F. and Draber, P. (1998) Le^x glycosphingolipids-mediated cell aggregation. Glycobiology 8, 139-146.

Bremer, E. G. and Hakomori, S. (1982) GM_3 ganglioside induces hamster fibroblast growth inhibition in chemically-defined medium: Ganglioside may regulate growth factor receptor function. Biochem. Biophys. Res. Commun. 106. 711-718.

Bremer, E. G. and Hakomori, S. (1984) Gangliosides as receptor modulators. Adv. Exp. Med. Biol. 174, 381-394.

Brown, M. S. and Goldstein J. L. (1974) Familial hypercholesterolemia: defective binding of lipoproteins to cultured fibroblasts associated with impaired regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity. Proc. Nat. Acad. Sci. USA 71, 788-792.

Brown, D. A. and London, E. (1997). Structure of detergent-resistant membrane domains: Does phase separation occur in biological membranes? Biochem. Biophys. Res. Commun. 240, 1-7.

Brown, D. A. and Rose, J. K. (1992) Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. Cell 68, 533-544.

Cheng, T., Wang, Y., and Dai, W. (1994) Transcription factor egr-1 is involved in phorbol 12-myristate 13-acetate-induced megakaryocytic differentiation of K562 cells. J. Biol. Chem. 269, 30848-30853.

Choi, O. H., Kim, J. and Kinet, J. (1996) Calcium mobilization via sphingosine kinase in signaling by the FcεRI receptor. Nature 380, 634-636.

Cifone, M. C., De Maria, R., Roncaioli, P., Rippo, M. R., Azuma, M., Lewis, L. L., Santoni, A., and Testi, R. (1993) Apoptotic signaling through CD95 (Fas/Apo-1) activates an acidic sphingomyelinase. J. Exp. Med. 177, 1547-1552.

Desai, N. N., Zhang, H., Olivera, A., Mattice, M. E. and Spiegel, S. (1992) Sphingosine-1-phosphate, a metabolite of sphingosine, increases phosphatidic acid levels by phospholipase D activation. J. Biol. Chem. 267, 23122-23128.

Durieux, M. E., Carlisle, S. J., Salafranca, M. N. and Lynch, K. R. (1993) Response to sphingosine-1-phosphate in X. laevis oocytes: similarities with lysophosphatidic acid signaling. Am. J. Physiol. 264, C1360-1364.

Durieux, M. E. and Lynch, K. R. (1993) Signalling properties of lysophosphatidic acid. Trends Pharmacol. Sci. 14, 249-254.

Eggens, I., Fenderson, B. A., Toyokuni, T., Dean, B., Stroud, M. R. and Hakomori, S. (1989) Specific interaction between Le^x and Le^x determinants: A possible basis for cell recognition in preimplantation embryos and embryonal carcinoma cells. J. Biol. Chem. 264, 9476-9484.

Eichholtz, T., Jalink, K., Fahrenfort, I. and Moolenaar, W. H. (1993) The bioactive phospholipid phosphatidic acid is released from activated platelets. Biochem J. 291, 677-680.

Exton, J. H. (1990) Signaling through phosphatidylcholine breakdown. J. Biol. Chem. 265, 1-4.

Gabius, H. J. and Gabius, S. (1997). Glycosciences: Status and perspectives. London: Chapman & Hall.

Ghosh, T. K., Bian, J. and Gill, D. L. (1990) Intracellular calcium release mediated by sphingosine derivatives generated in cells. Science 248, 1653-1656. l

Ghosh, T. K., Bian, J., and Gill, D. L. (1994) Sphingosine-1-phosphate generated in the endoplasmic reticulum membrane activates release of stored calcium. J. Biol. Chem. 269, 22628-22635.

Goodemote, K. A., Mattie, M. E., Berger, A. and Spiegel, S. (1995) Involvement of a pertussis toxin-sensitive G protein in the mitogenic signaling pathways of sphingosine-1-phosphate. J. Biol. Chem. 270, 10272-10277.

Ha, K.- S., Yeo, E.- J. and Exton, J. H. (1994) Lysophosphatidic acid activation of phosphatidylcholine-hydrolyzing phospholipase D and actin polymerization by a pertussis toxin-sensitive mechanism. Biochem. J. 303, 55-59.

Hakomori, S. (1981a). Blood group ABH and Ii antigens of human erythrocytes: Chemistry, polymorphism, and their developmental change. Semin. Hematol. 18, 39-62.

Hakomori, S. (1981b). Glycosphingolipids in cellular interaction, differentiation, and oncogenesis. Annu. Rev. Biochem. 50, 733-764.

Hakomori, S. (1984). Tumor associated carbohydrate antigens. Annu. Rev. Immunol. 2, 103-126.

Hakomori, S. (1989). Aberrant glycosylation in tumors and tumor-associated carbohydrate antigens. Adv. Cancer Res. 52, 257-331.

Hakomori, S. (1990) Bifunctional role of glycosphingolipids. Modulators for transmembrane signaling and mediators for cellular interactions. J. Biol. Chem. 265, 18713-18716.

Hakomori, S. and Igarashi, Y. (1993) Gangliosides and glycosphingolipids as modulators of cell growth, adhesion, and transmembrane signaling. Adv. Lipid. Res. 25, 147-162

Hakomori, S. and Igarashi, Y. (1995) Functional role of glycosphingolipids in cell recognition and signaling. J. Biochem. 118, 1091-1103.

Hall, A. (1998) Rho GTPase and actin cytoskeleton. Science 279, 509-514.

Handa, K., Stroud, M. R. and Hakomori, S. (1997). Sialosyl-fucosyl poly-LacNAc without the sialosyl-Le^x epitope as the physiological myeloid cell

ligand in E-selectin-dependent adhesion: Studies under static and dynamic flow conditions. Biochemistry 36, 12412-12420.

Hannun, Y. A. (1994) The sphingomyelin cycle and the second messenger function of ceramide. J. Biol. Chem. 269, 3125-3128.

Hannun, Y. A. (1996) Functions of ceramide in coordinating cellular responses to stress. Science, 274, 1855-1859.

Hannun, Y. A. and Bell, R. M. (1989) Functions of sphingolipids and sphingolipid breakdown products in cellular regulation. Science 243, 500-507.

Hannun, Y. A. and Linardic, C. M. (1993). Sphingolipid breakdown products: Anti-proliferative and tumor-suppresser lipids. Biochim. Biophys. Acta 1154, 223-236.

Hannun, Y. A., Loomis, C. R., Merrill, A. H. J., and Bell, R. M. (1986) Sphingosine inhibition of protein kinase C activity and of phorbol dibutyrate binding *in vitro* and in human platelets. J. Biol. Chem. 261, 12604-12609.

Hawiger, J. (1989) Platelet secretory pathways: an overview. Methods Enzymol. 169, 191-195.

Hecht, J. H., Weiner, J. A., Post, S. R. and Chun, J. (1996) Ventricular zone gene-1 (vzg-1) encodes a lysophosphatidic acid receptor expressed in neurogenic regions of the developing cerebral cortex. J. Cell Biol. 135, 1071-1083.

Igarashi, Y. (1997). Functional roles of sphingosine, sphingosine-1-phosphate, and methylsphingosines: In regard to membrane sphingolipid signaling pathways. J. Biochem.(Tokyo) 122, 1080-1087.

Igarashi, Y., Hakomori, S., Toyokuni, T., Dean, B., Fujita, S., Sugimoto, M., Ogawa, T., El-Ghendy, K. and Racker, E. (1989) Effect of chemically well-

defined sphingosine and its N-methyl derivatives on protein kinase C and src kinase activities. Biochemistry 28, 6796-6800.

Igarashi, Y., Kitamura, K., Toyokuni, T., Dean, B., Fenderson, B., Ogawa, T. and Hakomori, H. (1990) A specific enhancing effect of N,N-dimethylsphingosine on epidermal growth factor receptor autophosphorylation. Demonstration of its endogenous occurrence (and the virtual absence of unsubstituted sphingosine) in human epidermoid carcinoma A431 cells. J. Biol. Chem. 265, 5385-5389.

Ilic, D., Furuta, Y., Kanazawa, S., Takeda, N., Sobue, K., Nakatsuji, N., Nomura, S., Fujimoto, J., Okada, M., Yamamoto, T. and Aizawa, S. (1995) Reduced cell motility and enhanced focal adhesion contact formation in cells from FAK-deficient mice Nature 377, 539-544.

Imamura, F., Horai, T., Mukai, M., Shinkai, K., Sawada, M. and Akedo, H. (1993) Induction of in vitro tumor cell invasion of cellular monolayers by lysophosphatidic acid or phospholipase D. Biochem. Biophys. Res. Commun. 193, 497-503.

Imamura, F., Shinkai, K., Mukai, M., Yoshioka, K., Komagome, R., Iwasaki, T. and Akedo, H. (1996) rho-Mediated protein tyrosine phosphorylation in lysophosphatidic-acid -induced tumor cell invasion. Int. J. Cancer 65, 627-632.

Iwabuchi, K., Yamamura, S., Prinetti, A., Handa, K. and Hakomori, S. (1998) GM3-enriched microdomain involved in cell adhesion and signal transduction through carbohydrate-carbohydrate interaction in mouse melanoma B16 cells. J. Biol. Chem. 273, 9130-9138.

Jalink, K., Hengeveld, T., Mulder, S., Postma, F. R., Simon, M.-F., Chap, H., van der Marel, G. A., van Boom, J. H., van Blitterswijik, W. J. and Moolenaar, W. H. (1995) Lysophosphatidic acid-induced Ca^{2+} mobilization in human A431 cells: structure-activity analysis. Biochem. J. 307, 609-616.

Karlsson, K.-A. (1989) Animal glycosphingolipids as membrane attachment sites for bacteria. Annu. Rev. Biochem. 58, 309-350.

Kikkawa, U., Takai, Y., Minakuchi, R., Inohara, S. and Nishizuka, Y. (1982) Calcium-activated, phospholipid-dependent protein kinase from rat brain. Subcellular distribution, purification, and properties. J. Biol. Chem. 257, 13341-13348.

Kindman, L. A., Kim, S., McDonald, T. V. and Gardner, P. (1994) Characterization of a novel intracellular sphingolipid-gated Ca^{2+}-permeable channel from rat basophilic leukemia cells. J. Biol. Chem. 269, 13088-13091.

Kojima, N. and Hakomori, S. (1991) Cell adhesion, spreading, and motility of GM3-expressing cells based on glycolipid-glycolipid interaction. 266, 17552-17558.

Kolesnick, R. and Golde, D. W. (1994) The sphingomyelin pathway in tumor necrosis factor and interleukin-1 signaling. Cell 77, 325-328.

Kreutter, D., Kim, J. Y., Goldenring, J. R., Rasmussen, H., Ukomadu, C., DeLorenzo, R. J. and Yu, R., K. (1987) Regulation of protein kinase C activity by gangliosides. J. Biol. Chem. 262, 1633-37, 1987

Lee, M-J, Brocklyn, J. R. V., Thangada, S., Liu, C. H., Hand, A. R., Menzeleev, R., Spiegel, S. and Hla, T. (1998) Sphingosine-1-phosphate as a ligand for the G protein-coupled receptor EDG-1. Science 279, 1552-1555.

Liscovitch, M. and Cantley, L. C. (1994) Lipid second messengers. Cell 77, 329-334

Mano, N., Oda, Y., Yamamda, K., Asakawa, N.,and Katayama, K. (1997) Simultaneous quantitative determination method for sphingolipid metabolites by liquid chromatography/ionspray ionization tandem mass spectroscopy. Anal. Biochem. 244, 291-300.

Mao, C., Kim, S. H., Almenoff, J. S., Rudner, X. L., Kearney, D. M. and Kindman, L. A. (1996) Molecular cloning and characterization of SCaMPER, a sphingolipid Ca^{2+} release-mediating protein from endoplasmic reticulum. Proc. Natl. Acad. Sci. USA 93, 1993-1996.

Marcus, D. M., Kundu, S. K. and Suzuki, A. (1981). The P blood group system: Recent progress in immunochemistry and genetics. Semin.Hematol. 18, 63-71.

Mattie, M., Brooker, G., and Spiegel, S. (1994) Sphingosine-1-phosphate, a putative second messenger, mobilizes calcium from internal stores via an inositol triphosphate-independent pathway. J. Biol. Chem. 269, 3181-3188.

Mazurek, N., Megidish, T., Hakomori, S. and Igarashi, Y. (1994) Regulatory effect of phorbol esters on sphingosine kinase in Balb/c 3T3 fibroblasts (variant A 31): Demonstration of cell type-specific response. Biochem. Biophys. Res. Commun. 198, 1-9.

Merrill, A. H. J. and Stevens, V. L. (1989). Modulation of protein kinase C and diverse cell functions by sphingosine -- a pharmacologically interesting compound linking sphingolipids and signal transduction. Biochim. Biophys. Acta 1010, 131-139.

Moolenaar, W. H. (1995) Lysophosphatidic acid, multifunctional phospholipid messenger. J. Biol. Chem. 270, 12949-12952.

Mustard, J. F., Kinlough-Rathbone, R. L. and Packham, M. A. (1989) Isolation of human platelets from plasma by centrifugation and washing. Methods Enzymol. 169, 3-11.

Neer, E. J. (1995) Heterotrimeric G proteins: organization of transmembrane signals. Cell 80, 249-257.

Nishizuka, Y. (1984) The role of protein kinase C in cell surface signal transduction and tumor promotion. Nature 308, 693-698.

Nishizuka, Y. (1986) Studies and perspectives of protein kinase C. Science 233, 305-312.

Nobes, C. D. and Hall, A. (1995) Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. Cell 81, 53-62.

Ohta, H., Ruan, F., Hakomori, S. and Igarashi, Y. (1994) Quantification of free sphingosine in cultured cells by acylation with radioactive acetic anhydride. Anal. Biochem. 222, 489-494.

Okajima, F., Tomura, H., Sho, K., Nochi, H., Tamoto, K. and Kondo, Y. (1996) Involvement of pertussis toxin-sensitive GTP-binding proteins in sphingosine-1-phosphate-induced activation of phospholipase C-Ca^{2+} system in HL60 leukemia cells. FEBS Lett. 379, 260-264.

Okajima, F., Tomura, H., Sho, K., Kimura, T., Sato, K., Im. D-S., Akbar, M. and Kondo, Y. (1997) Sphingosine-1-phosphate stimulates hydrogen peroxide generation through activation of phospholipase C-Ca^{2+} system in FRTL-5 thyroid cells: possible involvement of guanosine triphosphate-binding proteins in the lipid signaling. Endocrinology 138, 220-229.

Olivera, A., Kohama, T., Tu, Z. X., Milstien, S. and Spiegel, S. (1998) Purification and characterization of rat kidney sphingosine kinase. J. Biol. Chem. 273, 12576-83.

Olivera, A., and Spiegel, S. (1993) Sphingosine-1-phosphate as second messenger in cell proliferation induced by PDGF and FCS mitogens. Nature 365, 557-560.

Olivera, A., Zhang, H., Carlson, R. O., Mattie, M. E., Schmidt, R. R. and Spiegel, S. (1994) Stereospecificity of sphingosine-induced intracellular calcium mobilization and cellular proliferation. J. Biol. Chem. 269, 17924-17930.

Postma, F., R., Jalink, K., Hengeveld, T. and Moolenaar, W., H. (1996) Sphingosine-1-phosphate rapidly induces Rho-dependent neurite retraction: action through a specific cell surface receptor. EMBO J. 15, 2388-2395.

Ruan, F., Yamamura, S., Hakomori, S. and Igarashi, Y. (1995) Synthesis of Sphingosine conjugate with controlled pore glass beads. Tetrahedron Lett. 36, 6615-6618.

Sadahira, Y., Ruan, F., Hakomori, S. and Igarashi, Y. (1992) Sphingosine-1-phosphate, a specific endogenous signaling molecule controlling cell motility and tumor cell invasiveness. Proc. Natl. Acad. Sci. USA 89, 9686-9690.

Sadahira, Y., Zheng, M., Ruan, F Hakomori, S. and Igarashi, Y. (1994) Sphingosine-1-phosphate inhibits extracellular matrix protein-induced haptotactic motility but not adhesion of B16 mouse melanoma cells. FEBS Lett. 340, 99-103.

Sato, T., Fuse, A., Eguchi, M., Hayashi, Y., Ryo, R., Adachi, M., Kishimoto, Y., Terayama, M., Mizoguchi, H., Shima, Y., Komori, I., Sunami, S., Okimoto, Y. and Nakajima, H. (1989) Establishment of a human leukaemic cell line (CMK) with megakaryocytic characteristics from a Down's syndrome patient with acute megakaryoblastic leukemia. BR. J. Haematol. 72, 184-190.

Scatchard, G. (1949) The attraction of proteins for small molecules and ions. Ann. N. Y. Acad. Sci. 51, 600-672.

Schroeder, R., London, E. and Brown, D. (1994) Interactions between saturated acyl chains confer detergent resistance on lipids and glycosylphosphatidylinositol (GPI)-anchored proteins: GPI-anchored proteins in liposomes and cells show similar behavior. Proc. Natl. Sci. USA 91, 12130-12134.

Siess, W. (1989) Molecular mechanisms of platelet activation. Physiol. Rev. 69, 58-178.

Simons, K. and van Meer, G. (1988) Lipid sorting in epithelial cells. Biochemistry, 27, 6197-6202.

Snell, E. E., Dimari, S. J. and Brady, R. N. (1970) Biosynthesis of sphingosine and dihydrosphingosine by cell-free systems from *Hansenula ciferri*. Chem. Phys. Lipids 5, 116-138.

Spiegel, S., Foster, D. and Kolesnick, R. (1996) Signal transduction through lipid second messengers. Curr. Opin. Cell Biol. 8, 159-167.

Spiegel, S. and Milstien, S. (1995) Sphingolipid metabolites: members of a new class of lipid second messengers. J. Membrane Biol. 146, 225-237.

Stewart, R. J. and Boggs. J. M. (1993a) Exposure of galactosylceramide to galactose oxidase in liposomes: dependence on lipid environment and ceramide composition. Biochemistry 32, 5605-5614.

Stewart, R. J. and Boggs. J. M. (1993b) A carbohydrate-carbohydrate interaction between galactosylceramide-containing liposomes and cerebroside sulfate-containing liposomes: Dependence on the glycolipid ceramide composition. Biochemistry 32, 10666-10674.

Stoffel, W. (1970) Studies on the biosynthesis and degradation of sphingosine bases. Chem. Phys. Lipids 5, 139-158.

Stoffel, W. (1973) Sphingosine metabolism and its link to phospholipid biosynthesis. Mol. Cell. Biochem. 1, 147-155.

Stoffel, W., Assmann, G., and Binczek, E. (1970) Enzymatic synthesis of 1-phosphate esters of 4t-sphingenine (sphingosine), sphinganine (dihydrosphingosine), 4-hydroxysphinganine (phytosphingosine) and 3-dehydrosphinganine by erythrocytes. Hoppe-Seyler's Z. Physiol. Chem. 351, 635-642.

Stoffel, W., Bauer, E. and Stahl, J. (1974) The metabolism of sphingosine bases in tetrahymena pyriformis. Sphingosine kinase and sphingosine-1-phosphate lyase. Hoppe-Seyler's Z. Physiol. Chem. 355, 61-74.

Stoffel, W. and Bister, K. (1973) Stereospecificities in the metabolic reactions of the four isomeric sphingosines (dihydrosphingosines) in rat liver. Hoppe-Seyler's Z. Physiol. Chem. 354, 169-181.

Stoffel, W., Heimann, G. and Hellenbroich, B. (1973) Sphingosine kinase in blood platelets. Hoppe-Seyler's Z. Physiol. Chem. 354, 562-566.

Stoffel, W., Sticht, G. and LeKim, D. (1968) Metabolism of sphingosine bases. IX. Degradation in vitro of dihydrosphingosine and dihydrosphingosine phosphate to palmitaldehyde and ethanolamine phosphate. Hoppe-Seyler's Z. Physiol. Chem. 349, 1745-1748.

Su, Y., Rothental, D., Smulson, M., and Spiegel, S. (1994) Sphingosine-1-phosphate, a novel signaling molecule, stimulates DNA binding activity of AP-1 in quiescent Swiss 3T3 fibroblasts. J. Biol. Chem. 269, 16512-16517.

Takaishi, K., Kikuchi, A., Kuroda, S., Kotani, K., Sasaki, T. and Takai, Y. (1993) Involvement of rho p21 and its inhibitory GDP/GTP exchange protein (rho GDI) in cell motility. Mol. Cell. Biol. 13, 72-79.

Toyokuni, T., Nisar, M., Dean, B. and Hakomori, S. (1991) A facile and regiospecific triation of sphingosine: synthesis of (2S,3R,4E)-2-amino-4-octadecene-1,3-diol-1-^{3}H. J. Labelled Comp. Radiopharm. 29, 567-574.

van der Bend, R., L., Brunner, J., Jalink, K., van Corven, E. J., Moolenaar, W. H. and Van Blitterswijk, W. J. (1992) Identification of a putative membrane receptor for the bioactive phospholipid, lysophosphatidic acid. EMBO J. 11, 2495-2501.

van-Koppen, C., Meyer-zu-Heringdorf, M., Laser, K. T., Zhang, C., Jakobs, K. H., Bunemann, M. and Pott, L. (1996) Activation of a high affinity Gi

protein-coupled plasma membrane receptor by sphingosine-1-phosphate. J. Biol. Chem. 271, 2082-2087.

Van Veldhoven, P. P., and Mannaerts, G. P. (1991) Subcellular localization and membrane topology of sphingosine-1-phosphate lyase in rat liver. J. Biol. Chem. 266, 12502-12507.

Van Veldhoven, P. P., and Mannaerts, G. P. (1993) Sphingosine-phosphate lyase. Advances Lipid Res. 26, 69-98.

Vunnam, R. R. and Radin, N. S. (1979) Short chain ceramides as substrates for glucocerebroside synthetase. Differences between liver and brain enzymes. Biochim. Biophys. Acta 573, 73-82.

Wu, J., Spiegel, S., and Sturgill, T. W. (1995) Sphingosine-1-phosphate rapidly activates the mitogen-activated protein kinase pathway by a G protein-dependent mechanism. J. Biol. Chem. 270, 11484-11488.

Yamamura, S., Sadahira, Y., Ruan, F., Hakomori, S. and Igarashi, Y. (1996) Sphingosine-1-phosphate inhibits actin nucleation and pseudopodium formation to control cell motility of mouse melanoma cells. FEBS Lett. 382, 193-197.

Yamamura, S., Yatomi, Y., Ruan, F., Sweeney, E. A., Hakomori, S. and Igarashi, Y. (1997) Sphingosine-1-phosphate regulates melanoma cell motility through a receptor-coupled extracellular action and in a pertussis toxin-sensitive manner. Biochemistry 36, 10751-10759.

Yao, B., Zhang, Y., Delikat, S., Mathias, S., Basu, S. and Kolesnick, R. (1995) Ceramide-activated protein kinase is a Raf kinase. Nature 378, 307-310.

Yatomi, Y., Ozaki, Y., Satoh, K. and Kume, S. (1994) Synthesis of phosphatidylinositol 3,4-bisphosphate is regulated by protein-tyrosine phosphorylation but the p85α subunit of phosphatidylinositol 3-kinase may

not be a target for tyrosine kinases in thrombin-stimulated human platelets. Biochim. Biophys. Acta. 1212, 337.

Yatomi, Y., Ruan, F., Ohta, H., Welch, R. J., Hakomori, S. and Igarashi, Y. (1995) Quantitative measurement of sphingosine-1-phosphate in biological samples by acylation with radioactive acetic anhydride. Anal. Biochem. 230, 315-320.

Yatomi, Y., Ruan, F., Hakomori, S. and Igarashi, Y. (1995) Sphingosine-1-phosphate: a platelet-activating sphingolipid released from agonist-stimulated human platelets. Blood 86, 193-202.

Yatomi, Y., Yamamura, S., Ruan, F. and Igarashi, Y. (1997a) Sphingosine-1-phosphate induces platelet activation through an extracellular action and shares a platelet surface receptor with lysophosphatidic acid. J. Biol. Chem. 272, 5291-5297.

Yatomi, Y., Igarashi, Y., Yang, L., Hisano, N., Ruomei, Q., Asanuma, N., Satoh, K., Ozaki, Y., and Kume, S. (1997b) Sphingosine-1-phosphate, a bioactive sphingolipid abundantly stored in platelets, is a normal constituent of human plasma and serum. J. Biochem. (Tokyo) 121, 969-973.

Zachary, I. and Rozengurt, E. (1992) Focal adhesion kinase ($p125^{FAK}$): a point of convergence in the action of neuropeptides, integrins, and oncogenes. Cell 71, 891-894.

Zhang, H., Buckley, N. E., Gibson, K. and Spiegel, S. (1990) Sphingosine stimulates cellular proliferation via a protein kinase C-independent pathway. J. Biol. Chem. 265, 76-81.

Zhang, H., Desai, N. N., Olivera, A., Seki, T., Brooker, G., and Spiegel, S. (1991) Sphingosine-1-phosphate, a novel lipid, involved in cellular proliferation. J. Cell Biol. 114, 155-167.

Zheng, M., Fang, H., Tsuruoka, T., Tsuji, T., Sasaki, T. and Hakomori, S. (1993) Regulatory role of GM3 ganglioside in α5β1 integrin receptor for fibronectin-mediated adhesion of FUA169 cells. J. Biol. Chem. 268, 2217-2222

Zucker, M. B. (1989) Platelet aggregation measured by the photometric method. Methods Enzymol. 169, 117-133.

Appendix

Mechanism for inhibition of haptotactic motility by sphingosine-1-phosphate

Abstract

Sphingosine-1-phosphate (Sph-1-P) has a strong inhibitory effect on chemotactic and haptotactic motility of many tumor cell lines at 0.01- 0.1 μM concentration through an unknown mechanism. We investigated the molecular mechanism by which Sph-1-P affects fibronectin-dependent haptotactic motility of serum-starved mouse melanoma B16/ F10 cells, with the following results. (i) Sph-1-P induced strong tyrosine phosphorylation of focal adhesion kinase (FAK) and paxillin, with maximal value 10 min after addition to medium. FAK phosphorylation was inhibited by *Clostridium difficile* toxin B, an inhibitor of G-protein Rho, suggesting that FAK phosphorylation is mediated by Rho. (ii) Sph-1-P enhanced GTP binding to Rho, thereby increasing GTP/ GDP ratio. Direct determination of this ratio was possible only by a new method. The binding value was maximal 5 min after addition of Sph-1-P to medium, and was much lower at 10 min, indicating that enhanced GTP-Rho binding precedes activation of FAK and paxillin, and that Rho may be the primary target of Sph-1-P. Sph-1-P added to cells did not undergo any metabolic conversion within 25 min. Therefore, the effect of Sph-1-P is not due to its metabolite. (iii) Immunofluorescence study with anti-vinculin antibody and with fluorescence-labeled phalloidin demonstrated that Sph-1-P promotes assembly of focal adhesion site with microfilaments. Cell migration is controlled by well-coordinated processes of adhesion and de-adhesion. Degree of motility is inversely correlated with number of focal adhesion sites in fibroblasts. Therefore, enhancement by Sph-1-P of Rho activation followed by enhanced

tyrosine phosphorylation of FAK and paxillin, leading to overexpression of focal adhesion sites, must be the molecular basis of inhibition of haptotactic cell motility by Sph-1-P. Sph-1-P had minimal effect on MAP kinase activity and no effect on mitogenesis in F10 cells.

Introduction

Sphingolipids and their metabolites have been implicated as modulators of signal transduction and as second messengers (Hakomori, 1990; Hakomori and Igarashi, 1995; Spiegel et al., 1996; Spiegel and Merill, 1996; Hannun, 1996; Igarashi, 1997). Sphingosine-1-phosphate (Sph-1-P), the initial product of sphingosine catabolism, induces intracellular Ca^{2+} influx in cultured smooth muscle cells through an IP3-independent pathway (Ghosh, 1990) and has a mitogenic effect on Swiss 3T3 cells by mobilizing intracellular Ca^{2+} (Zhang et al., 1991). A transient increase of Sph-1-P occurs following PDGF stimulation of Swiss 3T3 cells, indicating a possibility that Sph-1-P is a mitogenic messenger (Olivera and Spiegel, 1993). The concept of Sph-1-P as a second messenger is further supported by the fact that it mediates histamine release in human mast cells through IgE receptor FcεRI (Choi et al., 1996).

Although the mitogenic effect of Sph-1-P on various cell lines is limited, it strongly inhibits haptotactic and chemotactic motility. Sph-1-P at concentrations of 0.01-0.1 μM, added to culture medium, inhibits motility of mouse melanoma B16, Balb/c 3T3 cells, human melanoma Malme-3M, and human osteosarcoma MG-63. In contrast, it has no effect on motility of human umbilical vein endothelial cells or cow pulmonary artery endothelial cells. Other lipids tested, including Sph and N,N-dimethyl-Sph, have no motility-inhibitory effect (Sadahira et al., 1992). Sph-1-P at 0.01-0.1 μM also inhibits migration of cells through Transwell filters coated with laminin or FN. Structural requirements of Sph-1-P are strict, since dihydrosphingosine-1-P has much less effect on cell motility (Sadahira et al., 1994). The motility-inhibitory effect of Sph-1-P is presumably due to modification of lamellipodium formation and actin nucleation (Yamamura et al., 1996) (see Discussion).

Haptotactic cell motility requires dynamic interaction between matrix substrates and the "motile machinery" of cells, consisting, for example, of integrin receptors and their cytoplasmic complex in which Rho activation or inactivation determines the assembly of contractile actin-myosin filaments (see Discussion). In this paper we describe the molecular mechanism whereby Sph-1-P affects FN-dependent haptotactic cell motility, particularly its effect on Rho GTPase, FAK, and paxillin in relation to morphological formation of focal adhesion site.

Materials and Methods

Materials

Antibodies directed to FAK, Cas, H-Ras, Rho A, and phosphotyrosine (PY20) were purchased from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA). Anti-paxillin antibody was from Transduction Laboratories (Lexington, KY). Antibody to vinculin, myelin basic protein (MBP), and Protein Kinase Inhibitor Peptide (rabbit sequence) were from Sigma Chemical Co. (St. Louis, MO). Protein A/G agarose was from Santa Cruz Biotechnology, Inc. [^{32}P]orthophosphate and [γ-^{32}P]ATP were from Dupont NEN (Boston, MA). P-81 phosphocellulose paper was from Whatman Inc. (Clifton, NJ). [1-^{3}H]Sph-1-P was prepared enzymatically from [1-^{3}H]Sph with crude Sph kinase obtained from Balb/c 3T3 clone A31 cells (Mazurek et al., 1994).

Cell motility assay

Haptotactic cell motility of mouse melanoma B16/ F10 cells was assayed using Transwell compartment assembly (Costar, Cambridge, MA) with 6.5 mm diameter polycarbonate filters (8 μm pore size) (Sadahira et al., 1994). Briefly,

the lower surface of the filter was coated with human plasma FN (5 μg/ filter) (Boehringer Mannheim Biochemicals, Indianapolis, IN) and dried at 37 °C overnight. Cells were harvested after brief exposure to 0.02% EDTA, and resuspended to a concentration of $4x10^5$ cells/ ml in DMEM. DMEM (0.6 ml) was placed in the lower chamber of Transwell compartment assembly. To the upper compartment, 100 μl of cell suspension was added and incubated 3.5 h to allow spreading. After incubation for 6 h at 37 °C, cells on the lower surface of the filter were stained and counted.

Immunoprecipitation

F10 cells in a 35 mm-dish with FN coating were cultured without serum for 3.5 h, and treated with Sph-1-P. The cells were washed twice with PBS containing 1 mM Na_3VO_4, and lysed at 4 °C in 500 μl of a lysis buffer consisting of 1% Triton X, 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 2 mM NaF, 1 mM sodium orthovanadate, 1 mM PMSF, and 10 μg/ ml aprotinin. The lysates were clarified by centrifugation at 15,000 x g for 10 min and pre-cleared by incubation with Protein A/G-agarose for 1 h at 4 °C. After centrifugation (500 xg for 1 min), the supernatants were reacted with an antibody. The mixtures were placed overnight in a rotary mixer at 4 °C, added with protein A/G-Agarose, and placed again in a rotary mixer for 2 h. Agarose was washed 3 times with lysis buffer, then proteins were extracted in 2x SDS PAGE electrophoresis sample buffer heating at 95 °C for 5 min. The supernatants were subjected to SDS-PAGE.

Western blotting

Following SDS-PAGE, proteins were transferred electrophoretically to PVDF membranes (Immobilon-P, Millipore, Bedford, MA) in 25 mM Tris, 192 mM glycine, and 15% methanol at 200 mA for 2 h. The membranes were blocked in PBS containing 5% nonfat dried milk, washed in PBS, and incubated for 2 h with antibody in PBS containing 1% BSA. The membranes were washed three times in PBS containing 0.05% Tween-20, and incubated for 45 min with a secondary antibody conjugated with horseradish peroxidase. Membranes were washed in PBS 0.05% Tween-20 five times, and immunoreactive bands were visualized using chemiluminescence method with a substrate kit (Super-Signal-CL-HRP; Pierce).

Immunocytochemistry

F10 cells on FN-coated glass plates were cultured without serum for 3.5 h. The cells were treated with Sph-1-P, fixed in 2% formalin for 30 min at room temperature, permeabilized in 0.2% Triton X-100 in 50 mM Tris-HCl (pH 7.5) for 1 min, incubated with anti-vinculin or anti-phosphotyrosine antibody for 1 h at room temperature, and incubated with fluorescein isothiocyanate-conjugated anti-mouse IgG. Actin was visualized by staining with rhodamine-conjugated phalloidin.

Quantification of F-actin content

F10 cells on FN-coated plates were cultured without serum for 3.5 h. The cells were treated with Sph-1-P, fixed with 2 % formaldehyde, permeabilized with 0.1 % Triton X-100 for 5 min, and stained with rhodamine-conjugated phalloidin (Molecular Probe, Eugene, OR) for 15 min. Cells were washed in PBS, and fluorescent dye was extracted from the cells in 0.5 ml of 100 % methanol for 30 min.

Fluorescence of the methanol extract was measured by fluorometer at 550 nm excitation and 575 nm emission.

Metabolism of [1-^{3}H]Sph-1-P

Cells in a 60 mm dish were incubated with 0.1 μM [1-^{3}H]Sph-1-P (0.01 μCi) in 2.0 ml DMEM for various times, washed twice with 2.0 ml PBS, and scraped twice with 0.5 ml ice-cold methanol. One mL chloroform, 10 μl concentrated HCl, and 1 ml methanol were added to the cell suspension (final ratio of chloroform/ methanol/ concentrated HCl was 100:200:1). Lipids were extracted from the cell suspension as described previously (Bligh and Dyer, 1959). The lower chloroform phase sample was dried and resuspended in small volume of chloroform/ methanol (2:1). A portion of the lipids obtained from the lower phase was applied to silica gel 60 HPTLC plates (Merck, Darmstadt, Germany), and the plate was developed with butanol/ acetic acid/ water (3:1:1). After enhancer (Resolution TLC; E.M. Corp., Chestnut Hill, MA) treatment of the plate, autoradiography was performed with Kodak X-Omat film (Eastman Kodak, Rochester, NY) for 3 to 7 days at -80 °C. The silica gel areas which contained the radiolabeled sphingolipids were scraped off, and radioactivity was counted with a scintillation counter (Beckman).

Determination of GTP/GDP ratio in G-protein

Activity of small G-proteins was best determined by direct analysis of GTP/GDP ratio in G-protein. GTP and GDP bound to Ras were determined as described previously (Downward et al., 1990). However, the same conditions were not applicable for GTP and GDP bound to Rho. Therefore, step 3 (elution

condition) was greatly modified and step 4 (treatment with trichloroacetic acid) was added to the previous procedure (Iwabuchi et al., 1998).

Step 1: Metabolic ^{32}P labeling of cells and treatment with Sph-1-P. F10 cells in FN-coated dishes were metabolically labeled for 3.5 h with 1 mCi/ ml ^{32}P in phosphate-free DMEM (labeling medium). Cells were treated in labeling medium containing 0.1 μM Sph-1-P for 0, 3, 5, 10, or 15 min, harvested, and lysed in 0.5 ml HEPES lysis buffer (50 mM HEPES, pH 7.4, 100 mM NaCl, 20 mM $MgCl_2$, 10 mM NaF, 1 mM Na_3VO_4, 1 mM PMSF, 10 μg/ ml aprotinin, 10 μg/ ml leupeptin, 1% Triton X-100, 0.5% NP-40).

Step 2: Immunoprecipitation with anti-Rho or anti-Ras antibody. The cell lysate was centrifuged at 1300 xg for 5 min, and resultant pellet had no Rho or Ras, indicated by Western blot. The supernatant was precleared with 100 μl per ml BSA-coated charcoal for 10 min at 4°C, then centrifuged again. Supernatants were further precleared by incubation with Protein A/G agarose beads (10 μl packed) for 1 h at 4°C, centrifuged at 1300 xg for 5 min at 4°C, and incubated with 1 μg/ ml monoclonal anti-Rho A or anti-H-Ras antibody, overnight at 4°C. The incubation mixtures were added with Protein A/G agarose beads (10 μl packed) and incubated for 2 h at 4°C. The beads were washed sequentially with buffer A (50 mM HEPES, pH 7.4, 500 mM KCl, 20 mM $MgCl_2$) three times and with buffer B (50 mM HEPES, pH 7.4, 500 mM NaCl, 20 mM $MgCl_2$) twice.

Step 3: Elution. The Protein A/G agarose beads after washing with buffer B were resuspended in elution buffer (20 mM Tris/HCl, pH 7.4, 2 mM dithiothreitol, 0.2% SDS, 1 mM GTP, 1 mM GDP), heated at 65°C for 5 min, and centrifuged.

Step 4: Treatment with trichloroacetic acid and separation of GTP/ GDP. The supernatant as above was treated with 10% trichloroacetic acid and

centrifuged at 10,000 xg for 5 min at 4°C. Precipitate was discarded, and supernatant was mixed with 400 μl ether and shaken to extract trichloroacetic acid, and this procedure was repeated three times more. The upper ether layers were discarded, and aqueous lower layer was evaporated under nitrogen stream to < 20 μl (not dried completely). The concentrated solution was spotted on Ecteola cellulose plates which were then developed with 0.75 M KH_2PO_4 (pH 3.5). After autoradiography, labeled GTP and GDP spots were cut, and radioactivity was determined by liquid scintillation counting.

Determination of MAP kinase activity

B16 cells in a 35 mm dish were cultured without serum for 3.5 h and treated with Sph-1-P. The cells were washed twice with PBS containing 1 mM Na_3VO_4 and scraped off the dishes at 4 °C in 500 μl of lysis buffer consisting of 20 mM Tris-HCl (pH 8.0), 20 mM β-glycerophosphate, 2 mM EGTA, 1 mM Na_3VO_4, 2 mM DTT, 1 mM PMSF, 20 μg/ mL leupeptin and 10 μg/ ml aprotinin, and sonicated. The lysates were clarified by centrifugation at 15,000 xg for 10 min at 4 °C.

Cell lysates (15 μl) was added to a solution (5 μl) containing 1 mg/ ml MBP, 80 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM $MnCl_2$, 200 μM ATP, 8 μM Protein Kinase Inhibitor Peptide (rabbit sequence), and [γ-^{32}P]ATP (1 μCi/ sample). Following incubation for 20 min at 25 °C, the reaction was terminated by addition of a solution (3 μl) containing 0.1 N HCl and 1 mg/ ml BSA. A 10-μl aliquot was spotted onto P-81 phosphocellulose paper and washed in 0.5% phosphoric acid for 2 h. The paper was dried and ^{32}P incorporation into MBP was measured by scintillation counting.

Results

Sph-1-P inhibits FN-dependent haptotactic motility of B16 cells

Haptotactic motility of B16/ F10 cells, cultured under serum-free conditions, through Transwell filters coated with FN is shown in Figure 1. Its inhibitory effect on FN-dependent motility was maximal at 0.1 μM, and lower at higher (≥ 1.0 μM) or lower (≤ 0.01 μM) concentrations. DMS had no inhibitory effect.

Effect of Sph-1-P on tyrosine kinases, FAK, and paxillin

In order to study the effect of Sph-1-P on tyrosine phosphorylation, F10 cells cultured in serum-free medium and incubated with Sph-1-P (0.1 μM) for various durations were lysed and analyzed by Western blotting with anti-phosphotyrosine antibodies. Two bands were revealed, one with approximate Mr 125,000 and the other with approximate Mr 65,000. Both bands were enhanced by incubation with 0.1 μM Sph-1-P, with value maximal at 10 min after Sph-1-P addition and declining subsequently (Figure 2).

Because the Mr 125,000 band is close to that of FAK, serum-starved F10 cells were treated with 0.1 μM Sph-1-P and lysed, and lysates were immunoprecipitated with anti-FAK antibody and analyzed by Western blotting with anti-phosphotyrosine antibodies. Sph-1-P increased the level of tyrosine phosphorylation associated with FAK. Time course kinetics of FAK tyrosine phosphorylation show a maximal value at 10 min after Sph-1-P addition, and subsequent decline (Figure 3).

Based on the close approximation of the second band to paxillin, the lysate of serum-starved F10 cells was treated with Sph-1-P for various durations, immunoprecipitated with anti-paxillin antibody, and analyzed by Western

blotting. Tyrosine phosphorylation relative to control shows a maximal value at 10 min after Sph-1-P addition, and subsequent decline (Figure 4).

Effect of Clostridium difficile Toxin B on FAK tyrosine phosphorylation

FAK activity is controlled by Rho, a G-protein with GTPase activity mediating signaling through extracellular matrix to the actin cytoskeletal framework. Since Rho activity is specifically inhibited by *Clostridium difficile* Toxin B (see Discussion), the effect of Toxin B on Sph-1-P-dependent activation of FAK tyrosine phosphorylation was examined. In the presence of low concentration (10^{-12} M) of Toxin B, the FAK phosphorylation level was inhibited to the level of DMS-treated cells (control). Toxin B at this concentration had no effect on phosphorylation level of control cells (Figure 5) and did not cause morphological change of F10 cells. These results suggest that FAK phosphorylation is mediated by Rho.

Effect of Sph-1-P on Rho activity as determined by GTP/GDP binding ratio

G-protein activity is best demonstrated by GTP/GDP binding ratio. However, the method used for determining this ratio for Ras and other G-proteins cannot be applied to Rho, because affinity of GTP and GDP to Rho is higher than that to the other G-proteins (see Discussion). This difficulty was overcome by introduction of a new method using trichloroacetic acid (see Materials and Methods). Under this method, incubation of F10 cells with 0.1 μM Sph-1-P increased GTP/GDP ratio to 3.1 ± 0.7 (n= 3) at 5 min after Sph-1-P addition. At 10 min after addition of Sph-1-P, the ratio declined to 1.1 ± 0.2 (n= 3). The ratio was 0.9 ± 0.2 (n= 3) for control (0.1 μM DMS). A typical autoradiogram is shown in Figure 6A. Time course change of GTP/GDP ratio is shown in Figure

6B, upper panel. The ratio is maximal 5 min after introduction of Sph-1-P. Autoradiograms at different times are shown in Figure 6B, lower panel. These results indicate that activation of Rho precedes tyrosine phosphorylation of FAK and paxillin, since GTP/GDP ratio declines to ~1.0 at 10 min, when FAK and paxillin phosphorylation are maximal.

Metabolic conversion of Sph-1-P in B16/ F10 cells

There is a possibility of rapid conversion of Sph-1-P (within 10 min) to other sphingolipids or catabolites such as phosphoethanolamine or palmitaldehyde, and such metabolites, rather than Sph-1-P *per se*, may cause activity changes of Rho, FAK, and paxillin. In order to evaluate this possibility, we studied uptake and metabolic conversion of ^{3}H-labeled Sph-1-P in B16/ F10 cells (Figure 7). At 5 min after addition of Sph-1-P to medium, 4% of added Sph-1-P was incorporated into cells. There was no metabolic change within 25 min. Sph-1-P decreased and was converted to an unknown component at 45 min. Sphingomyelin was detected after 90 min. Sph-1-P was the sole radioactive component present during the first 25 min, when activation of Rho, FAK, and paxillin occurred.

Effect of Sph-1-P on MAP kinase activity

Sph-1-P (0.5- 2 μM) stimulates cell growth and proliferation through a MAP kinase pathway in Swiss 3T3 cells (see Introduction). In B16 cells, however, incubation with 0.1 μM Sph-1-P for various durations (5 min to 5 h) did not significantly activate MAP kinase. After 5-10 min incubation, activation of MAP kinase by 0.1 μM Sph-1-P was only 20% of that by vehicle (50% ethanol) (Figure 8).

A higher concentration of Sph-1-P (1 μM) produced no activation whatsoever (data not shown). DMS (0.1 μM) also had no effect.

Since activation of Ras to induce GTP loading is essential to activate the Raf kinase- MAP kinase pathway for cell proliferation, Ras activity was also examined by measuring GTP/GDP ratio. After 3 min preincubation, Ras GTP/GDP ratio was 1.9 for 0.1 μM Sph-1-P, and 1.1 for control (Figure 9), suggesting that Sph-1-P activates Ras to a minimal degree. The autoradiogram of GTP and GDP bound to Ras is shown in Figure 9.

Effect of Sph-1-P on distribution pattern of vinculin and stress fibers

Assembly of focal adhesion sites was examined as described in Materials and Methods, using fluorescence-labeled phalloidin and antibodies directed to vinculin. Sph-1-P induced increase of focal adhesion sites as indicated by anti-vinculin antibody staining, although actin filament formation remained near control level. The increase of focal adhesion sites occurred within 10 min and lasted for 5 h (Figure 10). DMS promoted no focal adhesion assembly (Figure 10). Assembly of focal adhesion sites was also examine by staining cells using anti-paxillin antibody. Sph-1-P induced increase of focal adhesion sites as indicated by anti-paxillin antibody (data not shown). Since fluorescence staining showed no significant induction of actin fiber formation (**Fig. 10**), we quantified F-actin content. Sph-1-P increased F-actin content about 2 times of control (**Fig. 11**).

Discussion

The role of sphingosine-1-phosphate (Sph-1-P) as a second messenger leading to mitogenesis or modified phenotypic expression is based on enhanced Sph-1-P synthesis following stimulation of cells by growth factors (Olivera and

Spiegel, 1993), IgE (Choi et al., 1996) or TPA (Mazurek et al., 1994). Sph-1-P may also act as a primary messenger analogous to lysophosphatidic acid (LPA), since defined receptor or receptors, shared or not shared with LPA, have been identified recently for different types of cells (Yamamura et al., 1997; Yatomi et al., 1997). An orphan receptor, endothelial differentiation gene product (EDG-1), was found recently to bind to Sph-1-P with high affinity (Lee et al., 1998). Thus, a concept that Sph-1-P elaborated and released from one cell stimulates other cells through their receptors seems plausible, in addition to the concept of Sph-1-P as a second messenger. Regardless of whether Sph-1-P acts as second or primary messenger, some types of cells (*e.g.* Swiss 3T3, mouse smooth muscle) undergo mitogenesis in response to Sph-1-P (Olivera and Spiegel, 1993) whereas many other types of cells (*e.g.* Balb/c 3T3, mouse melanoma B16, human melanoma Malme-3M, human osteosarcoma MG-63, human arterial smooth muscle) do not undergo mitogenesis; rather, their haptotactic and chemotactic motility is inhibited by Sph-1-P (Sadahira et al, 1992; Bornfeldt, 1995).

Cell motility is a complex process consisting of the following well-coordinated mechanisms. (1) A mechanism based on motility/ adhesion receptors (integrins, cadherin, tetraspan membrane glycoprotein, and CD44) and associated cytoplasmic components which connect these receptors to the cytoskeletal system. (2) Dynamic adhesion process produces cell-substratum traction. (3) Unidirectional force caused by morphological polarization results in cell body translocation. Appearance of lamellipodia showing adhesion to substratum at the forward end of a cell which is moving in a certain direction is coordinated with loss of adhesion at the back end of the cell. The exact mechanism of this coordination is unknown at this time; however, if the coordination is disrupted by overexpression of adhesion sites, motility may be

reduced or stopped (Huttenlocher et al., 1995). This concept is supported by the observation of an inverse correlation between migration and focal adhesion formation in fibroblasts (Dunlevy and Couchman, 1993). Because weak cell-substratum adhesions cannot provide traction, the cells are not able to move and do not spread well. When the cells have strong adhesion, they spread well and are immobilized. An intermediate level of attachment may produce maximal migration rate (Dimilla et al., 1991). Sph-1-P clearly promotes adhesion through initial activation of Rho and subsequent increase of focal adhesion sites, thus inhibiting F10 cell motility.

Rho regulates formation of an adhesive complex, of stress fiber formation, and of cell motility, and is active in GTP-bound form and inactive in GDP-bound form (Nobes and Hall, 1995). Recent studies indicate that Rho acts as a molecular switch to control a signal transduction pathway linking membrane receptors to the cytoskeleton (for review see Hall, 1997). Rho GTPases including rac1, cdc42, etc. regulate lamellipodium formation, which is closely associated with cell migration (Huttenlocher et al., 1995). Activated Rho inhibits hepatocyte growth factor-induced scattering in MDCK cells, presumably by increasing adhesion (Ridley et al., 1995). Rho has been implicated in regulating phosphorylation of FAK (Lacerda et al., 1996; Lacerda et al., 1997; Tsuda et al., 1997). The role of FAK is still unclear. FAK (−) cells isolated from FAK knock-out mice showed enhanced focal adhesions and reduced rate of cell migration on FN (Ilic et al., 1995). The FAK (−) cells had more focal adhesions than FAK (+) cells and an abundance of stress fibers. FAK activation was not required for focal adhesion assembly in mouse aortic smooth muscle cells (Wilson et al., 1995). However, FAK is believed to play some roles in the assembly of the focal adhesion complexes. FAK has been shown to downregulate components in focal adhesion (Hanks and Polte, 1997).

Sph-1-P was previously shown to induce enhanced FAK and paxillin activity in Swiss 3T3 fibroblasts, leading to enhanced cell proliferation. Since the process was inhibited by exoenzyme C3 transferase, which ADP-ribosylates Rho, Rho was assumed to be an upstream regulator of FAK and paxillin (Wang et al., 1997). Direct determination of Rho activity in terms of GTP binding would be ideal, but such a method has not been established. A method for determination of GTP binding to Ras is well established (Downward et al., 1990), but application of this method for determination of Rho-GTP binding is not possible because the affinity of Rho for GTP and GDP is different from its affinity for Ras. We introduced previously the use of trichloroacetic acid with successful determination of GTP/ GDP binding ratio to Rho (Iwabuchi et al., 1998). Our results indicate that Sph-1-P activates Rho and induces focal adhesion assembly in F10 cells, suggesting that the primary target of Sph-1-P is Rho.

Maximal value of GTP binding to Rho occurred at 5 min after Sph-1-P addition and declined by 10 min, whereas maximal FAK and paxillin phosphorylation occurred at 10 min after Sph-1-P addition, indicating that Sph-1-P strongly enhances GTP binding to Rho prior to enhancement of tyrosine phosphorylation of FAK and paxillin. Focal adhesion plaque formation as probed by anti-vinculin antibody occurred by 10 min after Sph-1-P addition and lasted for at least 5 h. These results indicate that Sph-1-P initially induces enhanced Rho activity, which triggers a series of signal transduction events involving FAK and paxillin, leading to increase of focal adhesion sites. FAK and paxillin are coordinately tyrosine phosphorylated at the cytoplasmic side of focal adhesion plaque (Burridge et al., 1992; for review see Hanks and Polte, 1997; Zachary and Rozengurt, 1992). Consequent enhancement of FAK and paxillin leading to increased focal adhesion sites disrupts the coordination

between front and back ends of motile cells as described above, thereby stopping migration. However, actin fiber formation, a primary reporter of Rho activation, was not significant (Figs. 10 and 11), suggesting that other pathways may be involved in focal adhesion assembly.

Metabolism experiment using [^{3}H]Sph-1-P showed that Sph-1-P may be largely responsible for these phosphorylation since Sph-1-P was mainly detected in the initial 25 min incubation (Figure 7) when the phosphorylation occurred.

Cas (p130^{cas}) is closely associated with FAK, and is co-immunoprecipitated with FAK in mouse 3T3 fibroblasts. Cas is localized in focal adhesion sites, and is tyrosine phosphorylated with kinetics similar to those of FAK during integrin-mediated cell adhesion (Hanks and Polte, 1997). However, co-immunoprecipitation experiments with B16/ F10 cells showed no association between FAK and Cas, and Sph-1-P did not induce phosphorylation of Cas in these cells (data not shown).

Sph-1-P induces MAP kinase activity and proliferation in quiescent Swiss 3T3 fibroblasts (Wu et al., 1995). However, in serum-starved melanoma cells, Sph-1-P did not stimulate MAP kinase activity significantly (Figure 8). Activation of Ras, upstream of MAP kinase, is an early event in receptor-mediated responses to many growth factors including PDGF and FCS, and is essential for cell proliferation. Ras function also has been shown to be involved in cell migration and actin polymerization and a member of the Rho family is a downstream factor (Ridley et al., 1992). Sph-1-P only slightly activated Ras (Figure 9), suggesting that Ras has no significant role in cell motility inhibition.

The functional significance of activated Rho, as claimed for mitogenesis induced by Sph-1-P in Swiss 3T3 cells (Wang et al., 1997), is unclear. The process may be characteristic of Swiss 3T3 cells, since mitogenesis and MAP kinase activation were not observed in B16 cells or Balb/c 3T3 cells. Sph-1-P

does not inhibit motility of Swiss 3T3 cells (Yamamura S., et al., unpublished data), suggesting that the signaling pathway initiated by Rho activation may be different in cells susceptible to mitogenesis induction vs. those susceptible to motility inhibition by the same signaling molecule, Sph-1-P. Further study is necessary to clarify this point.

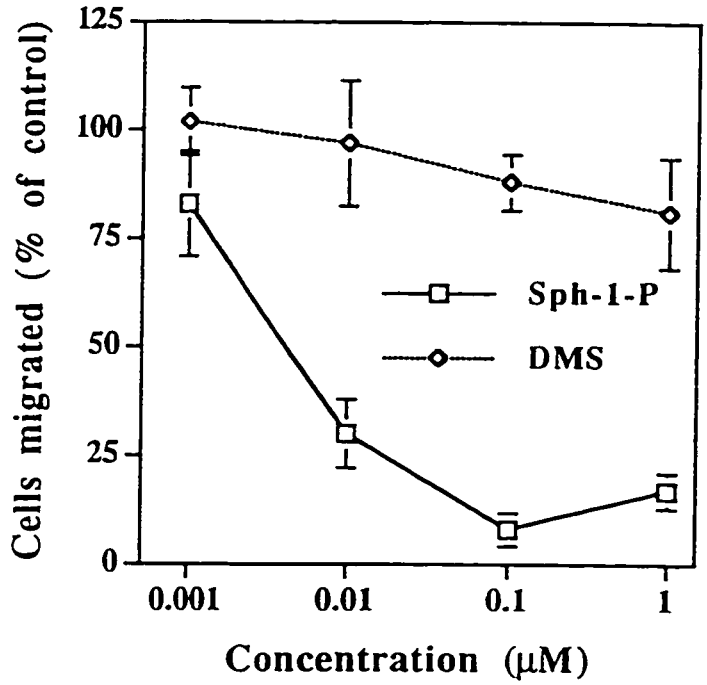


Fig. 1. Cell motility inhibition by Sph-1-P under serum-starved condition. Haptotactic cell motility of F10 cells were assayed using Transwell with FN coating. Dimethylsphingosine (DMS) was used for control. The data are presented as mean±SD (n=3).

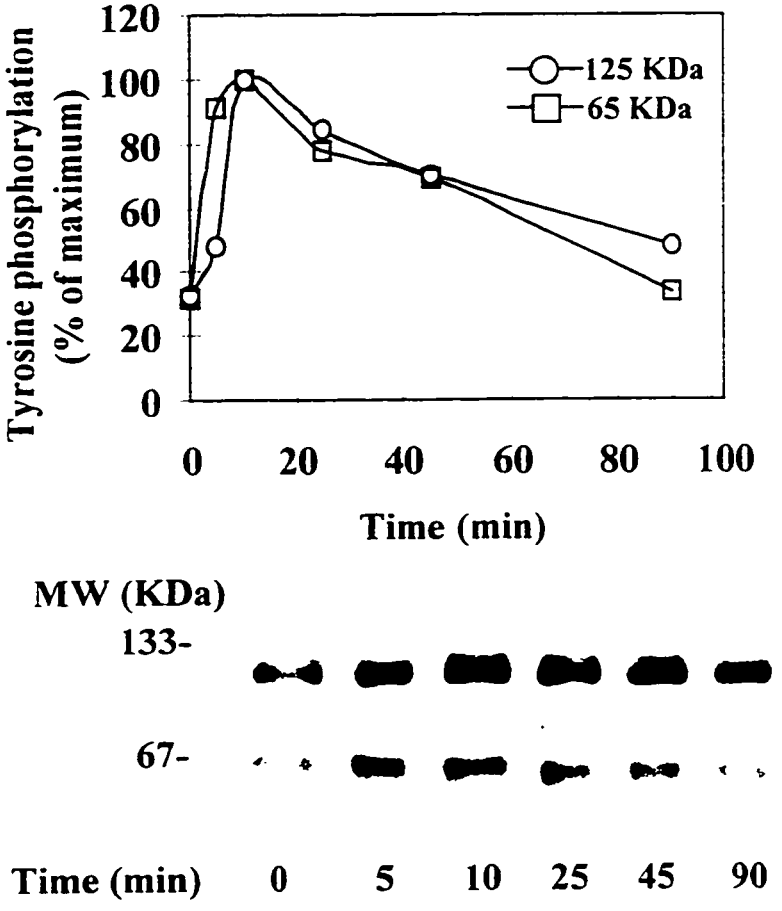


Fig. 2. Tyrosine phosphorylation of total cell lysate. F10 cells were cultured without serum for 3.5 h and treated with 0.1 μM Sph-1-P for the indicated times and lysed. The lysates were analyzed by Western blotting with anti-p-Tyr. The data are representative of three independent experiments.

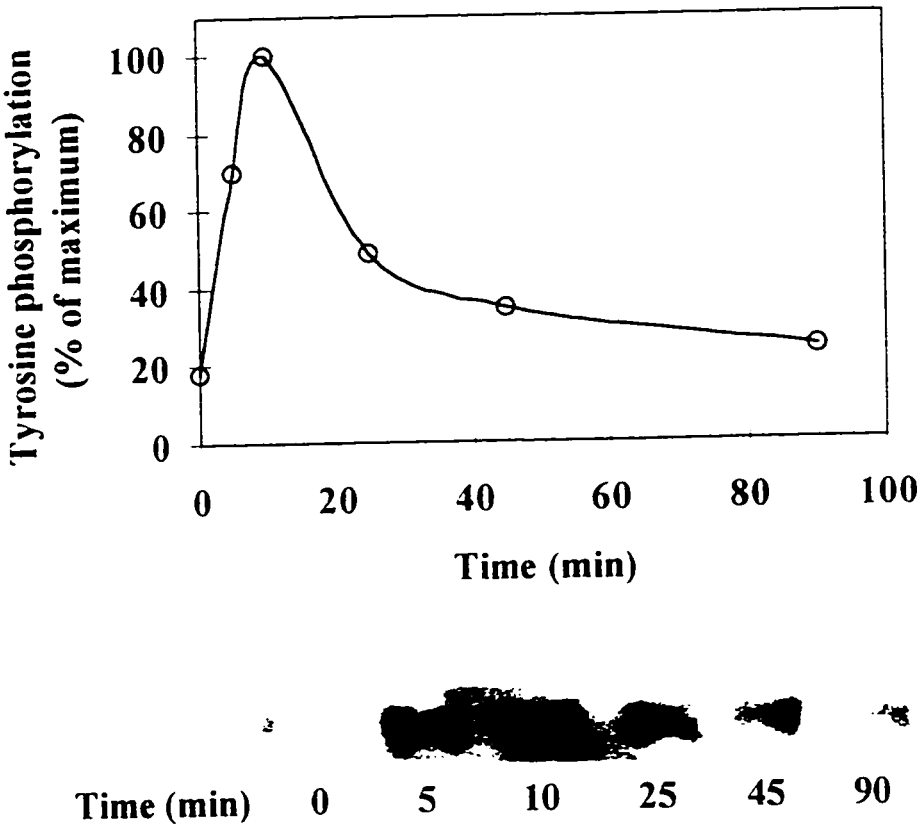


Fig. 3. Time course of Tyrosine phosphorylation of FAK by Sph-1-P. Following serum starvation for 3.5 h, F10 cells were treated with 0.1 μM Sph-1-P or DMS for indicated time and lysed, and the lysates were immunoprecipitated with anti-FAK and analyzed by Western blotting with anti-p-Tyr or anti-FAK. The data are representative of three independent experiments.

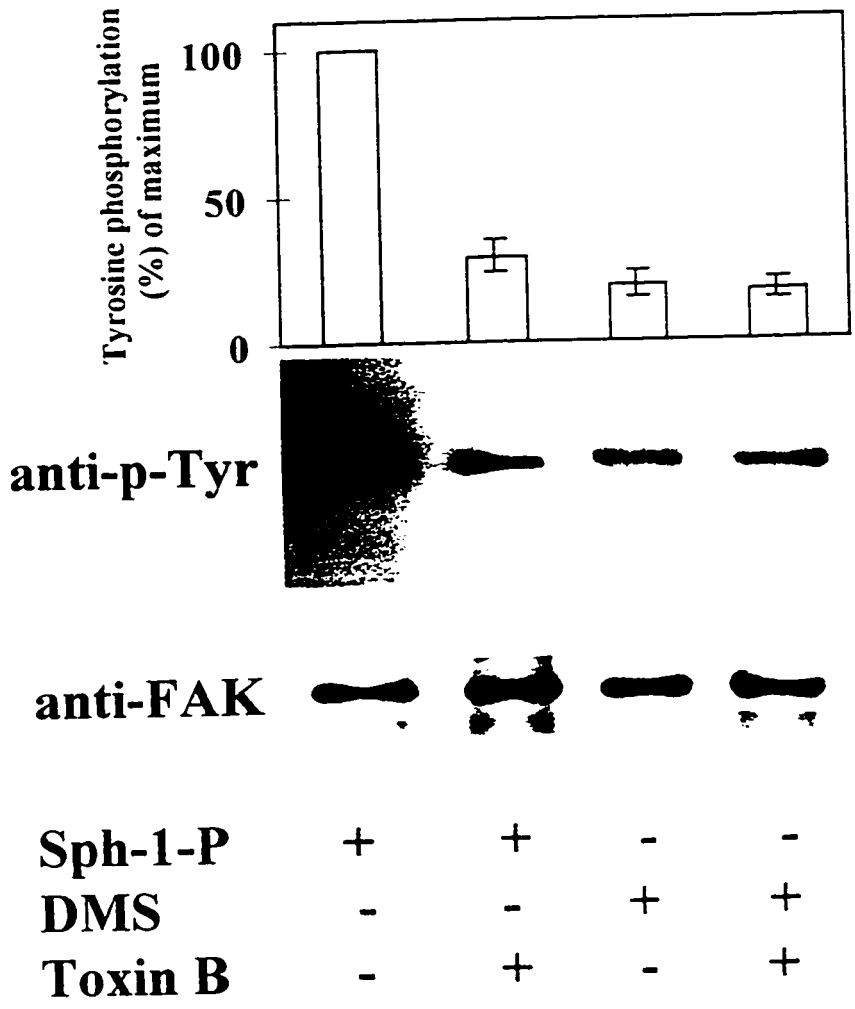


Fig. 4. Tyrosine phosphorylation of FAK by Sph-1-P and effect of C. difficile Toxin B. Following serum starvation for 3.5 h, F10 cells were incubated with or without 10^{-12} M Toxin B for 30 min. The cells were treated with 0.1 μM Sph-1-P or DMS for 10 min and lysed, and the lysates were immunoprecipitated with anti-FAK and analyzed by Western blotting with anti-p-Tyr or anti-FAK. The data are presented as mean±SD (n=3).

Fig. 5. Tyrosine phosphorylation of paxillin by Sph-1-P. A. F10 cells were cultured without serum for 3.5 h and treated with 0.1 μM Sph-1-P or DMS for 10 min and lysed, and the lysates were immunoprecipitated with anti-paxillin and analyzed by Western blotting with anti-p-Tyr or anti-paxillin. B. Time course of tyrosine phosphorylation of paxillin by Sph-1-P. F10 cells were cultured without serum for 3.5 h and treated with 0.1 μM Sph-1-P for indicated time and lysed, and the lysates were immunoprecipitated with anti-paxillin and analyzed by Western blotting with anti-p-Tyr. The data are representative of two independent experiments.

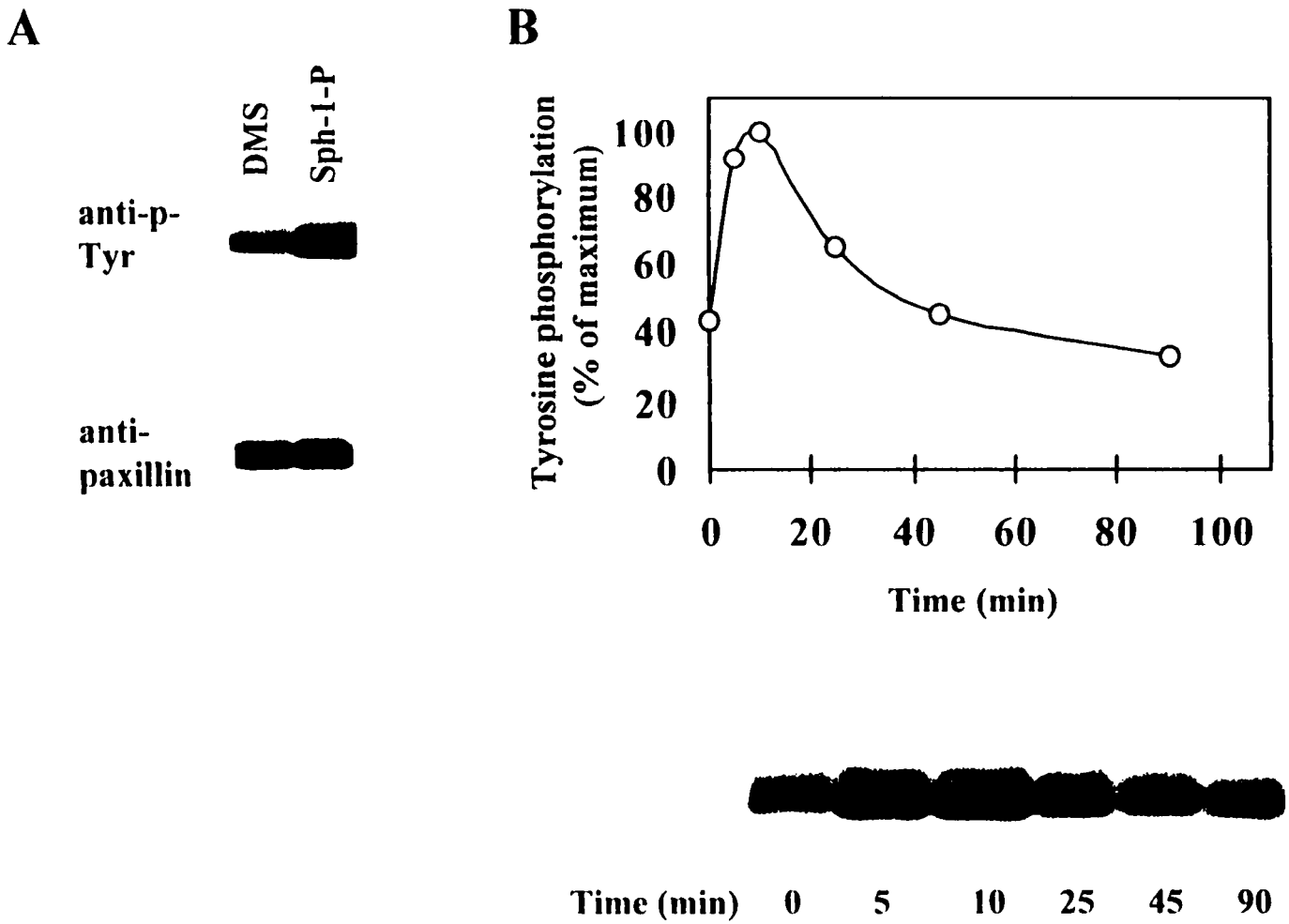
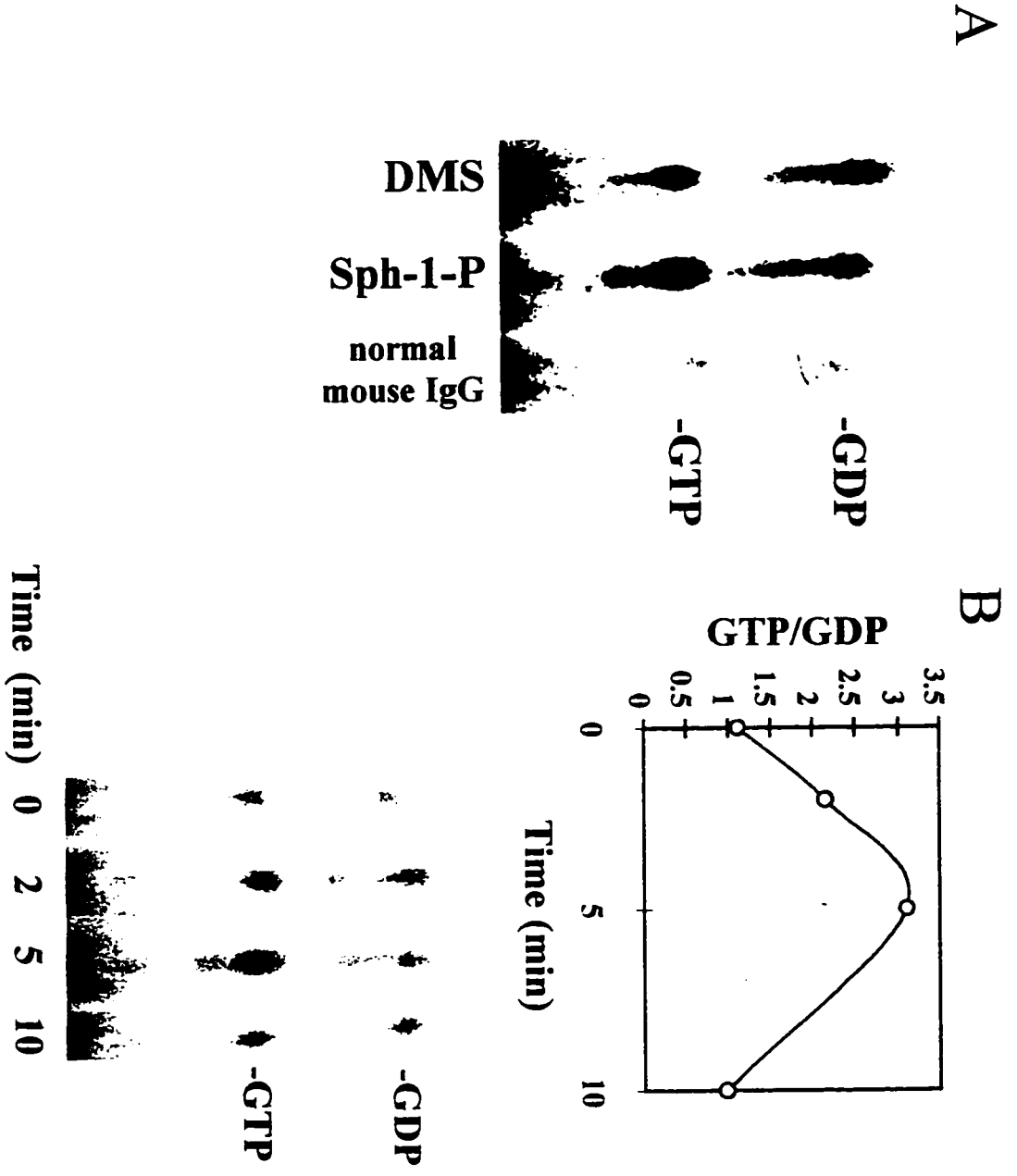

A
DMS
Sph-1-P
anti-p-Tyr
anti-paxillin
B
Tyrosine phosphorylation (% of maximum)
100
80
60
40
20
0
0
20
40
60
80
100
Time (min)
Time (min) 0 5 10 25 45 90

Fig. 6. Effect of Sph-1-P on Rho. A. F10 cells were metabolically labeled for 3.5 h with 1 mCi/ml [^{32}P] in phosphate- and serum-free DMEM, and treated with 0.1 μM Sph-1-P or DMS for 5 min. The cells were lysed, and the lysates were immunoprecipitated with anti-Rho. The GDP and GTP loaded into the Rho molecules were analyzed by thin layer chromatography. B. F10 cells were treated with 0.1 μM Sph-1-P for indicated time and time course of GTP/GDP ratio of Rho was measured. The data are representative of two independent experiments.


A
DMS
Sph-1-P
normal
mouse IgG
-GTP
-GDP
B
GTP/GDP
0
0.5
1
1.5
2
2.5
3
3.5
0
5
10
Time (min)
Time (min) 0 2 5 10
-GTP
-GDP

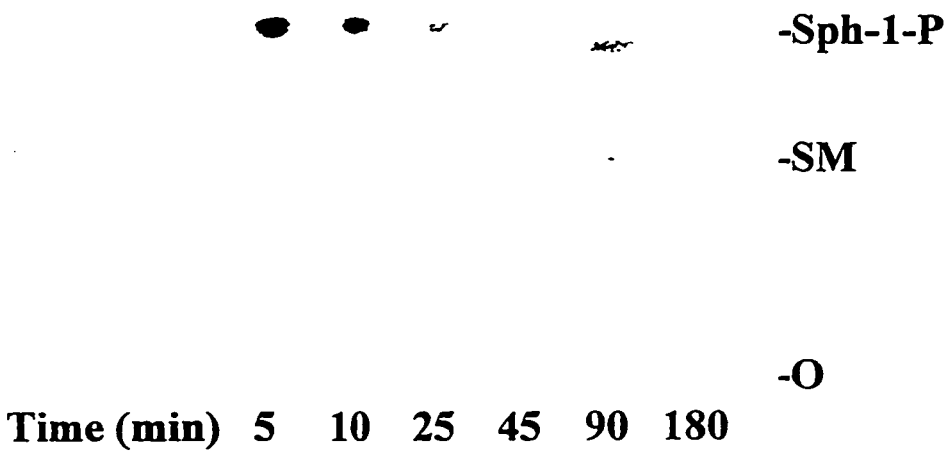


Fig. 7. Metabolism of Sph-1-P in F10 cells. [^{3}H]Sph-1-P was exogenously added to F10 cells, and incorporated [^{3}H]Sph-1-P and its metabolites were extracted and analyzed by thin layer chromatography as described in "Materials and Methods". The data are representative of two independent experiments.

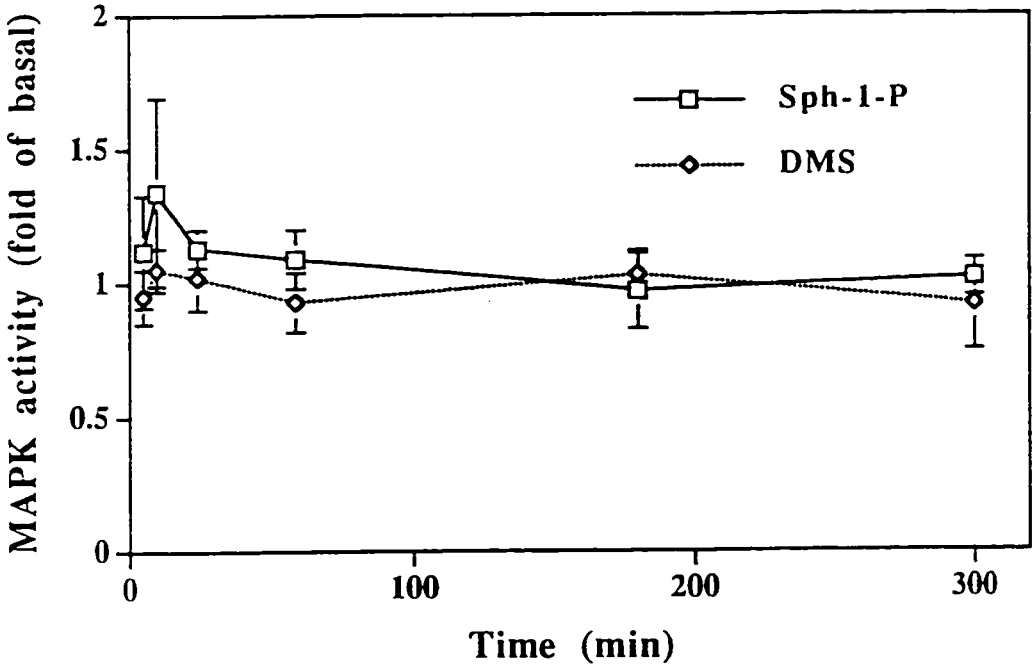


Fig. 8. Effect of Sph-1-P on MAP kinase activity. F10 cells were cultured without serum for 3.5 h and treated with 0.1 μM Sph-1-P, DMS or vehicle (50% ethanol) for at indicated times. MAP kinase activity was analyzed as described in "Materials and Methods". The data are presented as mean±SD (n=3).

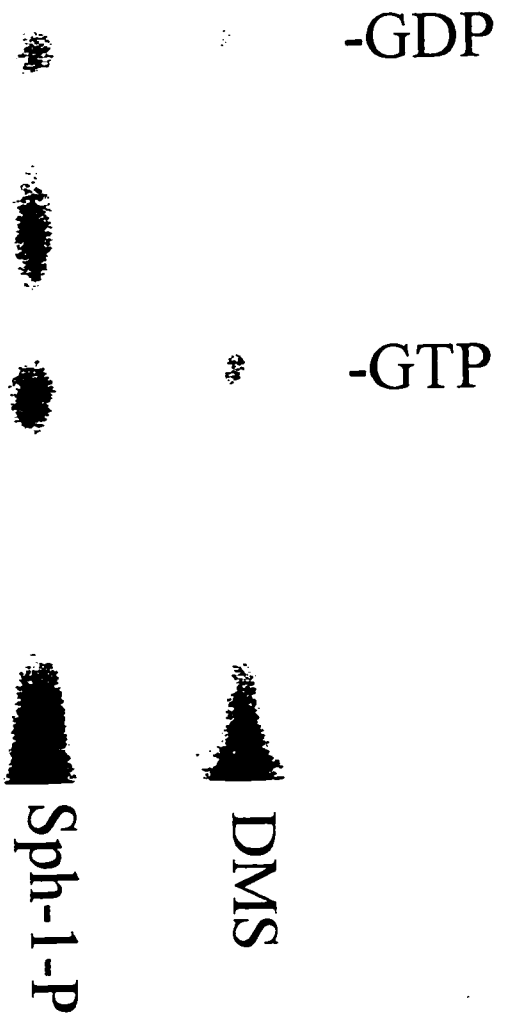
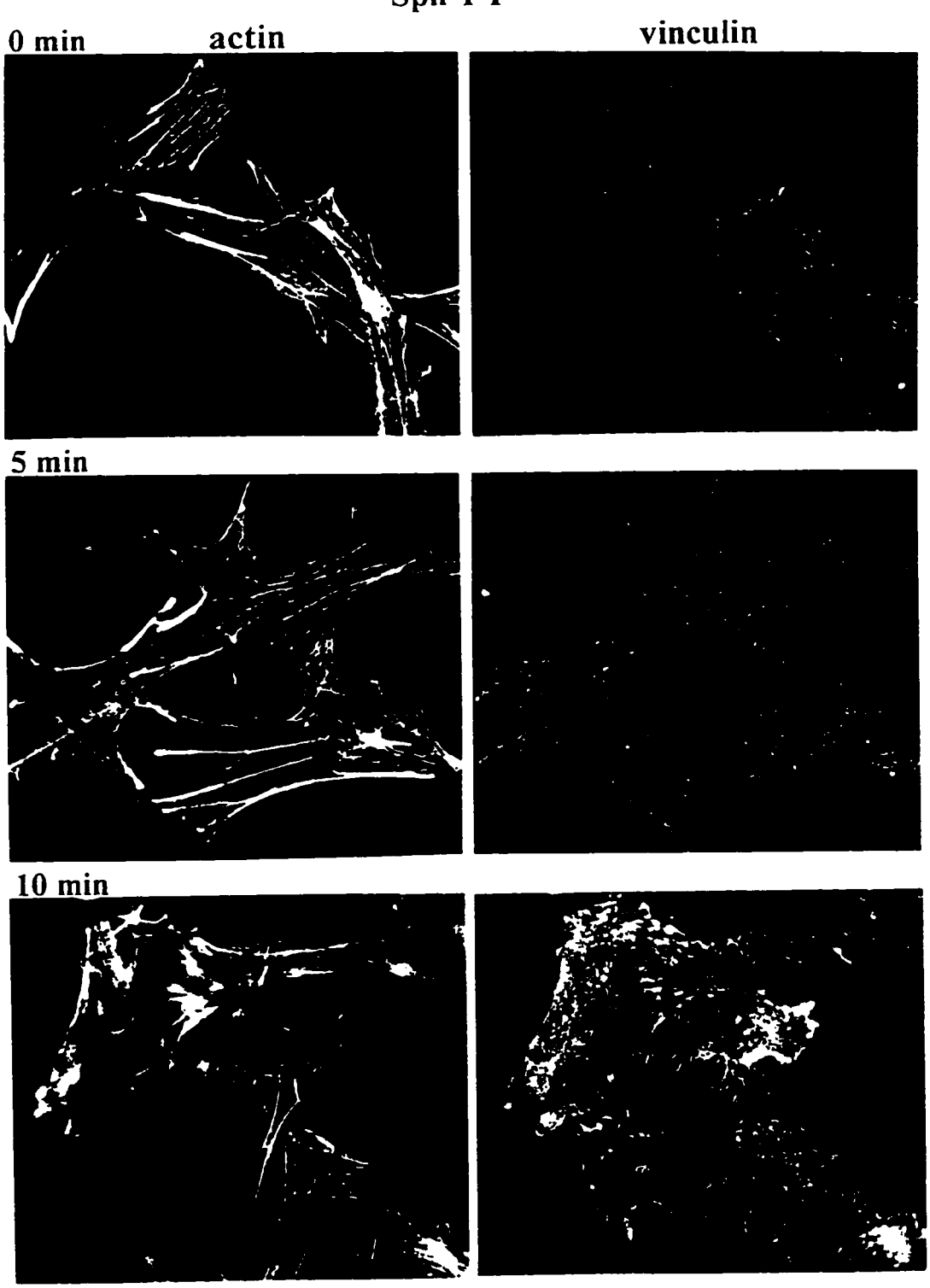
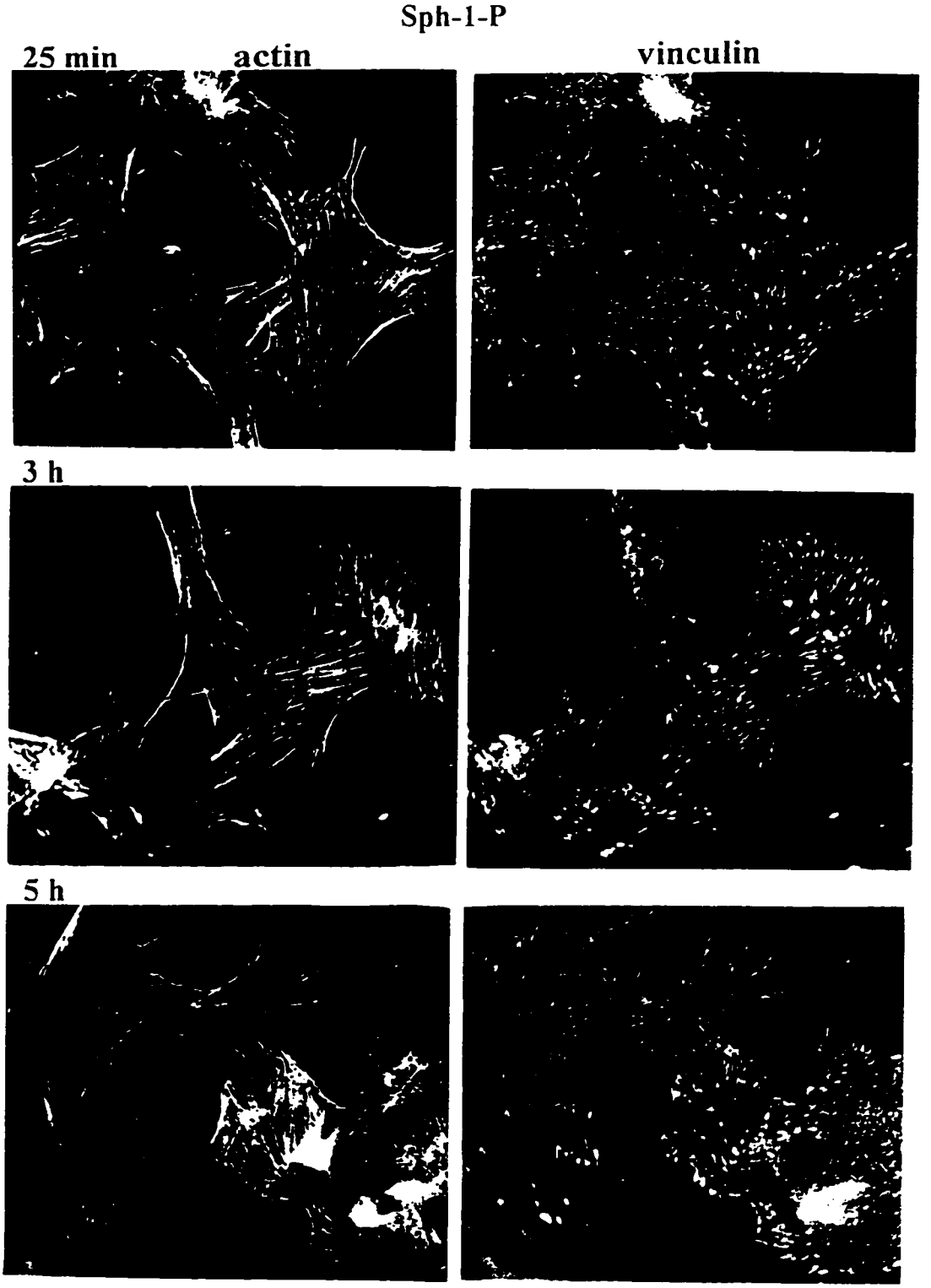


Fig. 9. Effect of Sph-1-P on Ras. F10 cells were metabolically labeled for 3.5 h with 1 mCi/ml [^{32}P] in phosphate- and serum-free DMEM, and treated with 0.1 μM Sph-1-P or DMS for 3 min. The cells were lysed, and the lysates were immunoprecipitated with anti-Ras. The GDP and GTP loaded into the Ras molecules were analyzed by thin layer chromatography. The data are representative of two independent experiments.

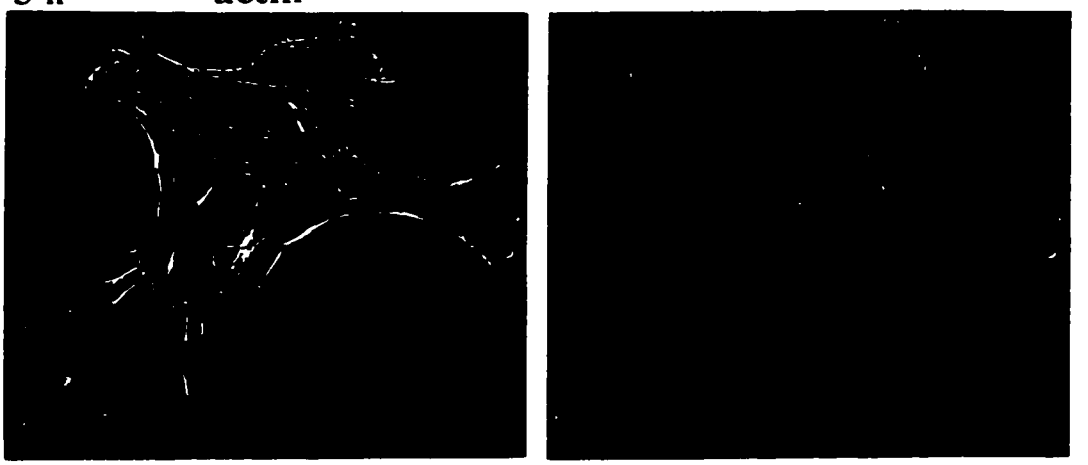
Fig. 10. Effect of Sph-1-P on focal adhesion and actin filament formation. F10 cells were cultured without serum for 3.5 h, treated with 0.1 μM Sph-1-P or DMS, fixed, permeabilized, and stained with anti-vinculin antibody or rhodamine-conjugated phalloidin.


Sph-1-P
0 min
actin
vinculin
5 min
10 min


Sph-1-P
25 min
actin
vinculin
3 h
5 h

DMS

3 h actin vinculin



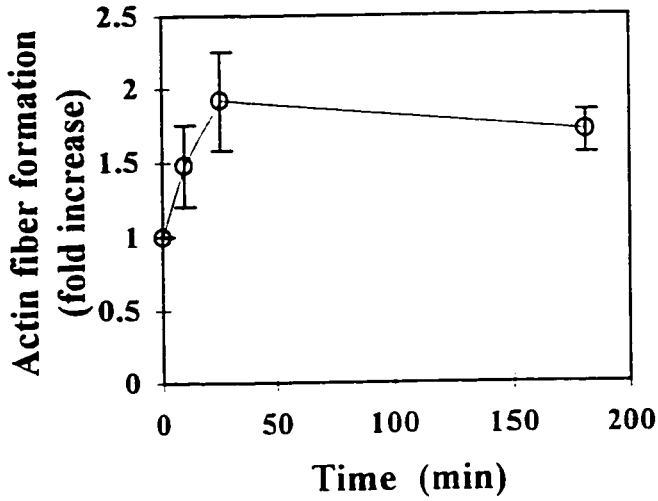


Fig. 11. Quantification of F-actin content. F10 cells on FN-coated plates were cultured without serum for 3.5 h. The cells were treated with 0.1 μM Sph-1-P for indicated time, fixed with 2 % formaldehyde, permeabilized with 0.1 % Triton X-100 for 5 min, and stained with rhodamine-conjugated phalloidin for 15 min. Cells were washed in PBS, and fluorescent dye was extracted from the cells in 0.5 ml of 100 % methanol for 30 min. Fluorescence of the methanol extract was measured by fluorometer at 550 nm excitation and 575 nm emission. The data are presented as mean+SD (n=3).

References

Bornfeldt, K. E., Graves, L. M., Raines, E. W., Igarashi, Y., Wayman, G., Yamamura, S., Yatomi, Y., Sidhu, J. S., Krebs, E. G., Hakomori, S. and Ross, R. (1995) Sphingosine-1-phosphate inhibits PDGF-induced chemotaxis of human arterial smooth muscle cells: spatial and temporal modulation of PDGF chemotactic signal transduction. J. Cell Biol. 130, 193-206.

Burridge, K., Turner, C. E., Romer L. H., (1992) Tyrosine phosphorylation of paxillin and pp125^{FAK} accompanies cell adhesion to extracellular matrix: a role in cytoskeletal assembly. J. Cell Biol. 119, 893-903.

Choi, O. H., Kim, J. and Kinet, J. (1996) Calcium mobilization via sphingosine kinase in signaling by the FcεRI receptor. Nature 380, 634-636.

DiMilla, P. A., Barbee, K., Lauffenburger, D. A. (1991) Mathematical model for the effects of adhesion and mechanics on cell migration speed. Biophys. J. 60, 15-37.

Downward, J., Graves, J. D., Warne, P. H., Rayter, S. and Cantrell, D. A. (1990) Stimulation of p21ras upon T-cell activation. Nature 346, 719-723.

Dunlevy, J. R. and Couchman, J. R. (1993) Controlled induction of focal adhesion disassembly and migration in primary fibroblasts. J. Cell Sci. 105, 489-500.

Ghosh, T. K., Bian, J. and Gill, D. L. (1990) Intracellular calcium release mediated by sphingosine derivatives generated in cells. Science 248, 1653-1656. l

Hakomori, S. (1990) Bifunctional role of glycoshingolipids. Modulators for transmembrane signaling and mediators for cellular interactions. J. Biol. Chem. 265, 18713-18716.

Hakomori, S. and Igarashi, Y. (1995) Functional role of glycosphingolipids in cell recognition and signaling. J. Biochem. 118, 1091-1103.

Hall, A. (1998) Rho GTPase and actin cytoskeleton. Science 279, 509-514.

Hanks, S. K. and Polte, T. R. (1997) Signaling through focal adhesion kinase. BioEssays, 19, 137-145.

Hannun, Y., A. (1996) Functions of ceramide in coordinating cellular responses to stress. Science 274, 1855-1859.

Huttenlocher, A., Sandborg, R. R. and Horwitz, A. F. (1995) Adhesion in cell migration. Curr. Opin.Cell. Biol. 7, 697-706.

Igarashi, Y. (1997) J. Biochem. (Tokyo) Physiology of sphingosine-1-phosphate: regulation of cell motility and platelet function. 122, 1080-1087.

Ilic, D., Furuta, Y., Kanazawa, S., Takeda, N., Sobue, K., Nakatsuji, N., Nomura, S., Fujimoto, J., Okada, M., Yamamoto, T. and Aizawa, S. (1995) Reduced

cell motility and enhanced focal adhesion contact formation in cells from FAK-deficient mice Nature 377, 539-544.

Iwabuchi, K., Yamamura, S., Prinetti, A., Handa, K. and Hakomori, S. (1998) GM3-enriched microdomain involved in cell adhesion and signal transduction through carbohydrate-carbohydrate interaction in mouse melanoma B16 cells. J. Biol. Chem. 273, 9130-9138.

Lacerda, H. M., Pullinger, G. D., Lax, A. J. and Rozengurt, E. (1997) Cytotoxic necrotizing factor 1 from Escherichia coli and dermonecrotic toxin from Bordetella bronchiseptica induce p21(rho)-dependent tyrosine phosphorylation of focal adhesion kinase and paxillin in Swiss 3T3 cells. J. Biol. Chem. 272, 9587-9596.

Lee, M-J, Brocklyn, J. R. V., Thangada, S., Liu, C. H., Hand, A. R., Menzeleev, R., Spiegel, S. and Hla, T. (1998) Sphingosine-1-phosphate as a ligand for the G protein-coupled receptor EDG-1. Science 279, 1552-1555.

Liscovitch, M. and Cantley, L. C. (1994) Lipid second messengers. Cell 77, 329-334

Mazurek, N., Megidish, T., Hakomori, S. and Igarashi, Y. (1994) Regulatory effect of phorbol esters on sphingosine kinase in Balb/c 3T3 fibroblasts (variant A 31): Demonstration of cell type-specific response. Biochem. Biophys. Res. Commun. 198, 1-9.

Nobes, C. D. and Hall, A. (1995) Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. Cell 81, 53-62.

Olivera, A., and Spiegel, S. (1993) Sphingosine-1-phosphate as second messenger in cell proliferation induced by PDGF and FCS mitogens. Nature 365, 557-560.

Ridley, A., J., Comoglio, P., M. and Hall, A. (1995) Regulation of scatter factor/hepatocyte growth factor responses by Ras, Rac and Rho in MDCK cells. Mol. Cell. Biol. 15, 1110-1122.

Ridley, A., J., Paterson, H. F., Johnson, C. L., Diekmann and Hall, A. (1992) The small GTP-binding protein rac regulates growth factor-induced membrane ruffling. Cell 70, 401-410.

Sadahira, Y., Ruan, F., Hakomori, S. and Igarashi, Y. (1992) Sphingosine-1-phosphate, a specific endogenous signaling molecule controlling cell motility and tumor cell invasiveness. Proc. Natl. Acad. Sci. USA 89, 9686-9690.

Sadahira, Y., Zheng, M., Ruan, F., Hakomori, S. and Igarashi, Y. (1994) Sphingosine-1-phosphate inhibits extracellular matrix protein-induced haptotactic motility but not adhesion of B16 mouse melanoma cells. FEBS Lett. 340, 99-103.

Spiegel, S., Foster, D. and Kolesnick, R. (1996) Signal transduction through lipid second messengers. Curr. Opin. Cell Biol. 8, 159-167.

Tsuda, T., Kusui, T. and Jensen, R. T. (1997) Neuromedin B receptor activation causes tyrosine phosphorylation of $p125^{FAK}$ by a phospholipase C independent mechanism which requires $p21^{rho}$ and integrity of the actin cytoskeleton. Biochemistry 36, 16328-16337.

Wang, F., Nobes, C., D. Hall, A. and Spiegel, S. (1997) Sphingosine-1-phosphate stimulates Rho-mediated tyrosine phosphorylation of focal adhesion kinase and paxillin in Swiss 3T3 fibroblasts. Biochem. J. 324, 481-488.

Wilson, L., Carrier, M. J. and Kellie, S. (1995) $pp125^{FAK}$ tyrosine kinase activity is not required for the assembly of F-actin stress fibers and focal adhesions in cultured mouse aortic smooth muscle cells. J. Cell Sci. 108, 2381-2391.

Wu, J., Spiegel, S., and Sturgill, T. W. (1995) Sphingosine-1-phosphate rapidly activates the mitogen-activated protein kinase pathway by a G protein-dependent mechanism. J. Biol. Chem. 270, 11484-11488.

Yamamura, S., Sadahira, Y., Ruan, F., Hakomori, S. and Igarashi, Y. (1996) Sphingosine-1-phosphate inhibits actin nucleation and pseudopodium formation to control cell motility of mouse melanoma cells. FEBS Lett. 382, 193-197.

Yamamura, S., Yatomi, Y., Ruan, F., Sweeney, E. A., Hakomori, S. and Igarashi, Y. (1997) Sphingosine-1-phosphate regulates melanoma cell motility through a receptor-coupled extracellular action and in a pertussis toxin-sensitive manner. Biochemistry 36, 10751-10759.

Yatomi, Y., Yamamura, S., Ruan, F. and Igarashi, Y. (1997) Singsong-1-phosphate induces platelet activation through an extracellular action and shares a platelet surface receptor with lysophosphatidic acid. J. Biol. Chem. 272, 5291-5297.

Zachany, I. and Rozengurt, E. (1992) Focal adhesion kinase ($p125^{FAK}$): a point of convergence in the action of neuropeptides, integrins, and oncogenes. Cell 71, 891-894.

Zhang, H., Desai, N. N., Olivera, A., Seki, T., Brooker, G., and Spiegel, S. (1991) Sphingosine-1-phosphate, a novel lipid, involved in cellular proliferation. J. Cell Biol. 114, 155-167.

VITA

Name, Family name: Yamamura
Forenames: Soichiro
Nationality: Japan
Present Address: Pacific Northwest Research Foundation
720 Broadway, Seattle, WA 98122

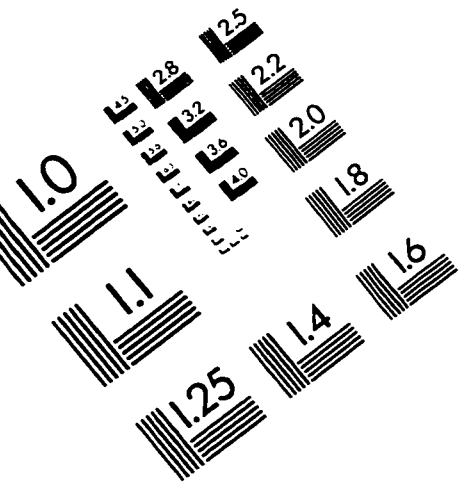
Education

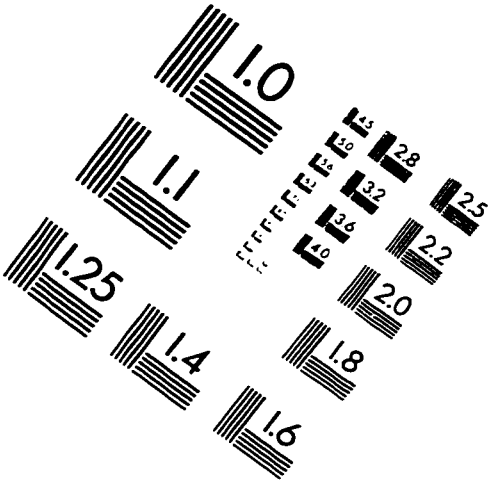
Kyoto Univ., Faculty of Engineering, Japan	B.S.	1980
Chemistry		
Kyoto Univ., Faculty of Engineering, Japan	Ph.D.	1988
Engineering		
Univ. of Washington, Dept. of Pathobiology	Ph.D.	1997
Pathobiology		

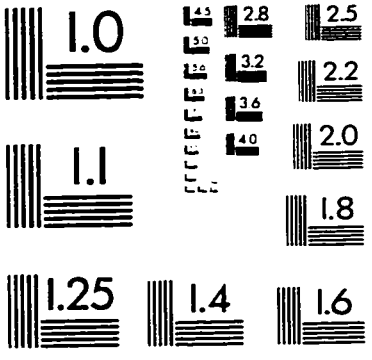
Research and Professional Experience

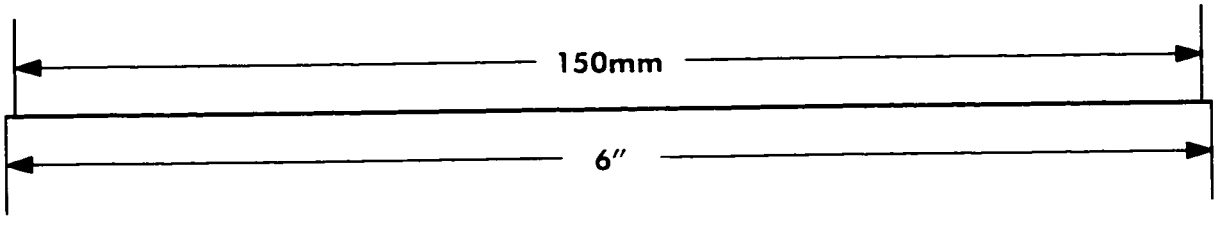
1980-92 Researcher, Osaka National Inst., MITI, Osaka, Japan,
1992-97 Research Assistant (graduate student, Univ. of Washington)

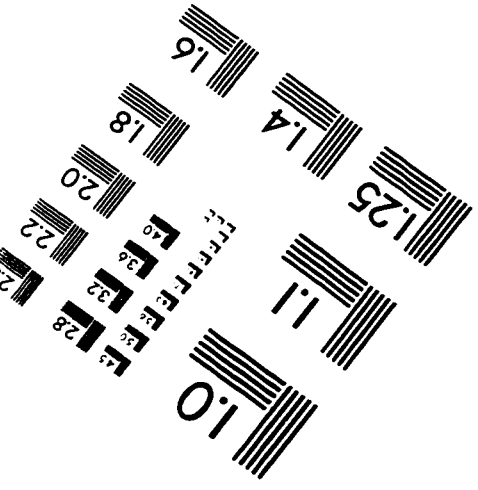
IMAGE EVALUATION
TEST TARGET (QA-3)

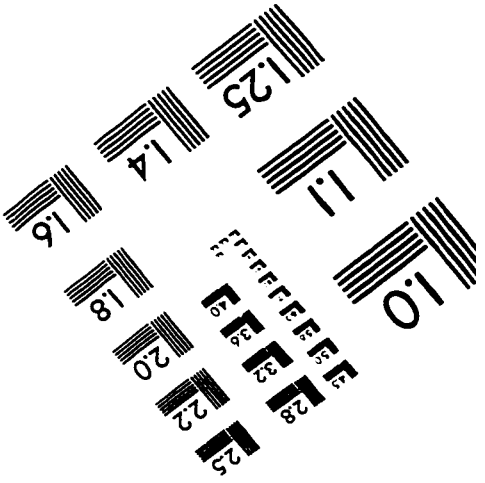












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved

99

07967

UMI

MICROFILMED 1998

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Dynamics of seasonal and interannual variability in the equatorial Pacific

by Xuri Yu

A dissertation submitted in partial fulfilment of
the requirements for the degree of

Doctor of Philosophy

University of Washington

1998

Approved by _____
Chairperson of Supervisory Committee

_____

_____

Program Authorized
to Offer Degree Oceanography

Date August 14, 1998

UMI Number: 9907967

UMI Microform 9907967
Copyright 1998, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized copying under Title 17, United States Code.

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Doctoral Dissertation

In presenting this dissertation in partial fulfilment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright law. Request for copying or reproduction of the dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, Michigan 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature _____

Date August 15, 1998

University of Washington

Abstract

Dynamics of seasonal and interannual variability in the equatorial Pacific

by Xuri Yu

Chairperson of Supervisory Committee: Affiliate Professor Michael J. McPhaden

NOAA/Pacific Marine Environmental Laboratory

The goal of this thesis is to investigate dynamics of seasonal-to-interannual variability in the equatorial Pacific upper ocean mainly using the observations from the TAO array over 1980-1997. The investigation is conducted in three relevant perspectives. First, the variability is described. At seasonal time scales, the normally westward flowing South Equatorial Current (SEC) reverses its direction with significant magnitude in boreal spring in the eastern Pacific. Seasonal variations in both the SEC and Equatorial Undercurrent (EUC), and thermocline depth at 5°N and 5°S propagate westward, but variations in thermal structure propagate eastward along the equator in the eastern and central Pacific. On interannual time scales, zonal wind anomaly is largest and in phase over the western and central Pacific. SST anomalies are significant in the region 165°E-110°W, and interannual variations in thermal structure propagate eastward, with

most of the phase shift occurring in the central Pacific.

The zonal momentum balance is then diagnosed for seasonal-to-interannual variability. At the seasonal cycle, the momentum balance is mainly between wind stress, pressure gradient and local acceleration. On interannual time scales, to zeroth order, the wind stress is balanced by the pressure gradient, indicating that the equatorial Pacific varies as a quasi-steady equilibrium. Nonlinear effects are significant at some depths on all time scales, though of secondary importance in the depth integrated momentum balance.

Finally, seasonal variations are studied in terms of equatorial waves through a simple wave dynamical model. Model results indicate that seasonal variability between 5°N and 5°S is dominated by wind-forced equatorial Kelvin waves and the first two meridional mode Rossby waves. The first two baroclinic modes dominate the solutions. The sum of the Rossby waves and Kelvin waves results in westward propagation in the equatorial zonal currents and off-equatorial thermal structure, but eastward propagation in thermal structure along the equator in the eastern and central Pacific. The simultaneous response to the zonal surface stress from the frictional dynamics also contributes to the westward propagation in the surface currents.

TABLE OF CONTENTS

List of Figures iii

List of Tables v

Chapter 1: Introduction 1

Chapter 2: Data and processing 8

2.1 Introduction 8

2.2 The TAO buoy array 8

2.3 Other data sets 15

Chapter 3: Spatial and temporal variations in the equatorial Pacific 18

3.1 Introduction 18

3.2 Examples of time series 18

3.3 Means of TAO observations 21

3.4 Seasonal variations of TAO observations and other products 25

3.5 Interannual variations of TAO observations 38

3.6 Summary 42

Chapter 4: Zonal momentum analysis of seasonal and interannual variability in the equatorial Pacific 44

4.1 Introduction 44

4.2 Diagnostic equation 45
4.3 Results 49
4.4 Neglected terms 66
4.5 Discussion and conclusions 69

Chapter 5: Wave dynamics of the seasonal cycle 72
5.1 Introduction 72
5.2 Model for estimates of Kelvin and Rossby wave variability 74
5.3 Model results 83
5.4 Discussion and conclusions 95

Chapter 6: Conclusions 100

Bibliography 104

LIST OF FIGURES

2.1 The TAO array 10

2.2 Time history of TAO sampling at locations of interest 13

3.1 Time series of the TAO data at 0°, 110°W 19

3.2 Longitudinal structure of mean fields at the equator from TAO observations 20

3.3 Latitudinal structure of mean temperature field from TAO observations 23

3.4 Monthly mean climatologies from TAO observations 24

3.5 Demeaned seasonal variations of pseudo-stress, zonal currents and temperatures along the equator from TAO observations 27

3.6 Demeaned seasonal variations in 20°C depths and surface dynamic height along the equator from TAO observations 28

3.7 Amplitudes and phases of 1 cpy harmonic of pseudo-stress, currents and etc from analysis of TAO data 29

3.8 Amplitudes and phases of 1 cpy harmonic of 20°C isotherm depth from TAO data 30

3.9 Comparison of demeaned seasonal variability between TAO and other datasets along the equator 34

3.10 Comparison of demeaned seasonal variability in 20°C isotherm depth between TAO and Levitus datasets 35

3.11 Time series of demeaned interannual variability of TAO observations. 38

3.12 Longitude-time contours of interannual variations from TAO observations 39

4.1 Vertical profiles of mean zonal pressure gradient and zonal advection 50

4.2 Seasonal fluctuations of zonally averaged zonal momentum terms . . 53

4.3 Seasonal variations in zonal wind stress, pressure gradient and zonal advection 54

4.4 Interannual fluctuations of zonally averaged zonal momentum terms 59

4.5 Interannual variations in zonal wind stress, pressure gradient and zonal advection 60

4.6 Time series of vertically integrated zonal pressure gradients and zonal transports on interannual time scales 64

5.1 Meridional structure of equatorial waves 77

5.2 Vertical profiles of the first four baroclinic modes 80

5.3 Kelvin and Rossby waves and their combination at seasonal cycle from the model 84

5.4 Mean seasonal cycle of dynamic height average for 5°N and 5°S from TAO observations and Levitus analysis 86

5.5 Reflected Rossby waves from the eastern boundary 90

5.6 Comparison of simulated seasonal variability with and without eastern boundary reflection 91

LIST OF TABLES

3.1 Estimated zonal phase speeds of the annual harmonics 31

4.1 Mean values of terms in the vertically integrated momentum balance 51

4.2 Standard deviation of terms in vertical integral of zonal momentum balance for seasonal and interannual variability 55

4.3 Cross correlation between zonal surface stress and vertically integrated zonal pressure gradient on interannual time scales 61

4.4 Cross correlation between vertically integrated zonal pressure gradient and zonal transport on interannual time scales 65

5.1 The values of baroclinic mode parameters 81

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Michael McPhaden, for his guidance, support, encouragement and faith in me. He pulled me through my difficult times and paid a great deal of attention to me and my work even during 1997/98 ENSO frenzy. Without him, this work would certainly not have been possible.

I would also like to thank my committee--Drs. William Kessler, David Battisti, Mitsuhiro Kawase and Paul Quay, whose advice and critiques are extremely invaluable through the course of my study at the University of Washignton. Discussions with Drs. Dennis Moore and Julius McCreary are instrumental in improving my understanding of equatorial wave dynamics. I also benefited from pleasant discussions with Drs. Zuojun Yu and Meghan Cronin.

I appreciate Dai McClurg, Jerry Davidson and Margie McCarty for their assistance with data processing and data presentations. My thanks also go to the TAO group of NOAA/PMEL, especially Paul Frietag, that operates the Tropical Atmosphere Ocean Array. The data collected from this important observing system make this work possible.

I also appreciate a great deal of help from Ms. Valapha Cass, our lovely secretary at NOAA/PMEL/OCRD.

I also extend my thanks to my friends and fellow students for their help, friendship and pleasure: Zhaohua Wu, Weimin Wang, Xinyou Lu, Fritz Stahr, Shuguang Song, Bob Hallberg, Dan Codiga, Jianrong Zhang, Emmanual Boss, Karl Newyear and Haili Sun.

I want to thank my wife Kan Huang and my parents for their patience,

understanding and love that inspired me in difficult times of the past six years.

This work was supported by NOAA's Office of Global Programs and Office of Oceanic and Atmospheric Research.

Chapter 1

INTRODUCTION

Ocean-atmosphere interactions in the tropical Pacific ocean have been a source of great interest in recent years because of their influence on the El Niño/Southern Oscillation (ENSO) phenomenon, a perturbation of the climate systems that affects the lives of millions of people around the globe. From an oceanographic perspective, understanding the processes that give rise to sea surface temperature (SST) anomalies on ENSO time scales is of crucial importance. These anomalies in many respects represent a year-to-year modulation of the mean seasonal cycle (e.g., Gu and Philander 1995), and are strongly determined by dynamical processes operating in the upper ocean (e.g., Seager et al. 1988; Hayes et al. 1991; Chang 1993, 1994; Chen et al. 1994; Kessler and McPhaden 1995). Hence, a better description and understanding of the dynamics of seasonal-to-interannual variability is required in the tropical Pacific for climate model development and validation. In particular, proper representation of these dynamical processes in ocean models and coupled ocean-atmosphere models is necessary for accurate simulation and forecasting of seasonal-to-interanual climate variability (e.g., Cane et al. 1986; Battisti 1988; Battisti and Hirst 1989; Suarez and Schopf 1988; Neelin 1991; Neelin et al. 1992; Mechoso et al. 1995).

Aside from the imperatives of climate prediction, oceanographers have long

been interested in the unique dynamics governing the circulation of the tropical oceans. The first modern theory of wind driven ocean circulation was formulated to explain the existence of the North Equatorial Countercurrent in the tropical Pacific (Sverdrup 1947). The discovery of the Equatorial Undercurrent (EUC) in the early 1950's (Montgomery and Stroup 1961) spawned a host of theoretical interpretations of this subsurface eastward flow (reviewed in McPhaden 1986). Among early studies, Arthur (1960) showed how Sverdrup theory was consistent with the dynamics of the EUC, and could have been used to predict its existence before it was discovered.

Sverdrup theory (1947) describes the steady state wind-driven circulation of the ocean. However, winds vary significantly across a broad range of frequencies, and the ocean adjusts to these wind changes through the excitation and propagation of planetary waves. The time it takes for these waves to cross the basin and to set up zonal pressure gradients in balance with the wind stress, compared to the time scale on which the winds vary, determines whether the ocean will be in equilibrium. At the mid-latitudes, baroclinic planetary waves propagate too slowly to adjust the ocean to equilibrium on seasonal-to-interannual time scales. Near the equator, planetary waves (namely equatorial Rossby and Kelvin waves) propagate much more rapidly such that in principle it is possible for the ocean to approach Sverdrup equilibrium to wind forcing on relatively short time scales (Philander 1979).

From Sverdrup (1947), the momentum balance for the steady circulation in a baroclinic ocean reduces to the simple expression

$$P_x = \tau_0^x$$

on the equator, where P_x is the depth integrated zonal pressure gradient and τ_0^x is the surface zonal wind stress. In this thesis, this relationship will be called 'Sverdrup momentum balance'. In the equatorial Atlantic, a basin three times narrower than the Pacific, both models and observations indicate that equatorial waves set up pressure gradients sufficiently quickly to balance seasonal and lower frequency wind stress changes (Katz 1977, 1987; Cane and Sarachik 1981; Philander and Pacanowski 1981; Katz et al. 1995). In the equatorial Pacific, the situation is less clear. In principle, the seasonal cycle should not be in equilibrium with wind forcing along the equator because the basin is too wide for waves to cross it and to set up zonal pressure gradients on time scales short compared to a season. Moreover, annual period Rossby and Kelvin waves have been observed in the tropical Pacific (e.g., Lukas and Firing 1985; Kessler and McCreary 1993), and are presumably important in generating circulation features such as the springtime reversal of the South Equatorial Current (SEC) (Philander and Chao 1991; Yu et al. 1997). On interannual time scales, equatorial waves form the cornerstone of many theories of ENSO, such as the delayed oscillator (Battisti 1988; Schopf and Suarez 1989), and these waves are evident in many empirical analyses of ENSO variability (e.g., Picaut and Delcroix 1995; Boulange and Menkes 1995). Conversely, some model studies indicate that the Sverdrup momentum balance is valid in the equatorial Pacific Ocean on interannual time scales (e.g., Schneider et al. 1995).

Up to the present, empirical diagnosis of the Sverdrup momentum balance on seasonal-to-interannual time scales in the equatorial Pacific has been limited by inadequate data. Exactly how far from steady equilibrium the equatorial Pacific

Ocean is near the equator because of equatorial wave processes is unclear, given the conflicting results of previous observational tests of the Sverdrup momentum balance (e.g., Meyers 1979a, Tsuchiya 1979; Mangum and Hayes 1984; McPhaden and Taft 1988). The purpose of this study is therefore to assess the degree to which the equatorial Pacific Ocean satisfies the Sverdrup relation on seasonal-to-interannual time scales, and the implications of this balance (or lack thereof) for current variations along the equator. To achieve this objective, we follow the methodology used in McPhaden and Taft (1988) who used $2\frac{1}{2}$ years of moored time series measurements of surface winds, temperatures and currents in the eastern equatorial Pacific (110°W-140°W) collected as part of the Equatorial Pacific Ocean Climate Studies (EPOCS) Program (Hayes et al. 1986) and TROPIC HEAT Program (Eriksen 1985). McPhaden and Taft (1988) concluded that $P_x = \tau_0^x$ applied on seasonal time scales, to within the errors of their analysis. However, these errors were relatively large, given the shortness of the records used compared to the time scale of the seasonal cycle. Their data records were also too short to analyze the dynamics of interannual variability. But for the $2\frac{1}{2}$ year record length mean, they found that the eastern Pacific equatorial Pacific was close to Sverdrup equilibrium.

In Chapter 4, we will expand on the analysis of McPhaden and Taft (1988) to incorporate many more years of moored time series data, and to incorporate a much wider range of longitudes (110°W-165°E) in the analysis. We will rely primarily on data from the Tropical Atmosphere Ocean (TAO) Array, which consists of about 70 moorings spanning the equatorial Pacific basin between about 8°S and 8°N (McPhaden 1995). The array measures surface winds, ocean temperatures and ocean currents. Many of the sites have been instrumented for a

sufficiently long period (more than 6 years) to compute reliable monthly mean climatologies for this period of time and to estimate interannual variability. The primary conclusions from the analysis are: (1) on seasonal time scales, the balance is mainly between wind stress, pressure gradient and local acceleration and the difference between the stress and pressure gradient provides accelerations and decelerations for zonal currents; (2) on interannual time scales, local acceleration is much smaller than either zonal wind stress or zonal pressure gradient, so to zeroth order, the wind stress is balanced by the pressure gradient, indicating that the equatorial Pacific varies as a succession of Sverdrup momentum balances from one year to the next.

We further study the dynamics of seasonal variability in the equatorial Pacific from a different perspective in Chapter 5. Specifically we analyze dominant equatorial wave modes in the seasonal variability using equatorial wave theory since linear dynamics is a good approximation and equatorial waves are pronounced in seasonal variability as we will conclude in Chapter 4. Previous studies have shown that the dynamics of seasonal time scale variability in the upper ocean are governed by the local response to wind forcing, and by the excitation and propagation of Rossby and Kelvin waves (Meyers 1979a, b; Kindle 1979; Busalacchi and O'Brien 1980; Kessler 1990; Kessler and McCreary 1993; Minobe and Takeuchi 1995; Yang et al. 1997). Certain aspects of the seasonal cycle, however, are still not well understood. Among these are the fact that winds, SST, and zonal currents along the equator propagate westward, while 20°C isotherm depth along the equator propagates eastward in the central and eastern Pacific (e.g., Meyers 1979a, b; Horel 1981; Lukas and Firing 1985; McPhaden and Taft 1988; Kessler 1990). Based on XBT analysis, Meyers (1979a, b) first

suggested the eastward propagation in 20°C along the equator was due to equatorial Kelvin wave dynamics, whereas off-equatorial dynamics were governed by local Ekman pumping and westward propagating Rossby waves. Recent modeling studies support this basic hypothesis (Minobe and Takeuchi 1995; Yang et al. 1997), although there are ambiguities among these recent studies on how precisely annual period equatorial Kelvin and Rossby waves combine to produce the full range of observed variability near the equator.

In addition, an important phenomenon that has not been fully understood is the springtime direction reversal of the South Equatorial Current (SEC) in the eastern equatorial Pacific. This reversal of the normally westward flowing SEC was first observed by Puls (1895) in ship drift data, and later confirmed from moored time series data (Halpern 1987; McPhaden and Taft 1988). Several hypotheses have been proposed for why the SEC runs counter to the easterly trade winds in the eastern Pacific; these hypotheses typically involve various combinations of local wind forcing, equatorial Kelvin wave dynamics, and/or seasonal modulation of high frequency instability waves (Philander et al. 1987; McPhaden and Taft 1988; Philander and Chao 1991). The relationship of the springtime reversal to other aspects of the seasonal cycle, e.g., the prominent annual Rossby wave, has not been fully clarified however. We will propose a hypothesis for this relationship in Chapter 5.

There are several advantages of using TAO data for this study. One is that high frequency variability which might otherwise be aliased in infrequent shipboard measurements can be averaged out, yielding more sharply defined features of the mean seasonal cycle. Another is that contemporaneous data for winds, SST, upper ocean temperatures are available, all with similar

temporal/spatial resolution and covering large regions not routinely sampled by standard shipping routes. Finally, unique to the TAO data is the ability to compute climatologies of upper ocean velocity at several mooring sites along the equator (q.v. McPhaden and McCarty 1992; McCarty and McPhaden 1993).

This thesis is outlined as follows. Chapter 2 describes the data sets and processing procedures used in this study. Chapter 3 presents means and seasonal-to-interannual variations computed from TAO data, and comparison of TAO mean seasonal variations with climatologies based on ship winds, upper ocean thermal data, drifting buoys, and blended in situ/satellite SST analyses. These comparisons are performed as a check on the representativeness of the TAO climatologies, which are based on relatively short records. Chapter 4 presents the analysis of the zonal momentum balance for the long term mean and seasonal-to-interannual variability. Diagnosis of the mean seasonal cycle in terms of wind forced Kelvin and Rossby waves will be presented in Chapter 5, followed by the important conclusions of the thesis in Chapter 6.

Chapter 2

DATA AND PROCESSING

2.1. Introduction

We rely most heavily in this study on data from the TAO buoy array (McPhaden 1993). We will also use several other datasets to compare with the seasonal cycle computed from TAO data to determine the representativeness of TAO seasonal analyses relative to those based on longer term records. These other datasets are the Florida State University (FSU) winds (Legler 1991), Reynolds and Smith SST (Reynolds and Smith 1995), drifter surface currents (Reverdin et al. 1994) and Levitus temperatures (Levitus et al. 1994). TAO observations are included in some of these other datasets, but they represent a relatively small percentage of the overall data in these datasets. Therefore, our comparisons will not be overly biased by this overlap. There are other surface wind, temperature and current products, which we could compare the TAO analyses to, but the above climatologies are representative and sufficient for our purposes. In this chapter, we will describe the TAO dataset in detail in Section 2.2 and then briefly introduce the other datasets in Section 2.3.

2.2. The TAO buoy array

2.2.1. TAO measurements and derived data

The TAO array was implemented as part of TOGA program for improved

description, understanding and prediction of seasonal-to-interannual climate variability. The array presently consists of nearly 70 deep ocean moorings with 2°-3° meridional spacing and 10°-15° longitudinal spacing across the equatorial Pacific, spanning from 8°S to 8°N (McPhaden 1998) (Fig. 2.1). Most of these buoys are Autonomous Temperature Line Acquisition System (ATLAS) thermistor chain moorings (Hayes et al. 1991), which measure temperatures at the surface and 10 subsurface depths down to 500 m, as well as surface winds, relative humidity and air temperatures. In addition, Mechanical Current Meters (MCM) and Acoustic Doppler Current Profilers (ADCP) are installed on the buoys at five locations along the equator (Fig. 2.1), collecting time series of currents in the upper 250-300 m (McPhaden 1993). The MCMs started operation at 140°W and 110°W in early 1980's and at 165°E in mid-1980's. ADCPs were installed at all the five locations indicated in Fig. 2.1 for operation in early 1990's except at 170°W, where ADCPs were first deployed in 1988 (Weisberg and Hayes 1995). Data are sampled at 15-minute to 2-hour intervals depending on instrument, then processed to daily averages. Details of sampling characteristic instrumentation and measurement errors can be found in McPhaden and McCarty (1992), McCarty and McPhaden (1993), Mangum et al (1994) and Plimpton et al (1997a, b).

In addition to random sampling errors, a potential source of bias error enters into our calculation of near surface currents at 0°, 170°W. Measurements at this location were made primarily with upward looking ADCPs mounted on subsurface floats at depths between 250-300 m. Side lobe reflections from the surface contaminate the upper 10% (25-30 m) of the profile (Plimpton et al. 1997b), so velocities at 10-20 m depth were computed by linear extrapolation of the vertical gradient from deeper levels, as in Weisberg and Hayes (1995). We

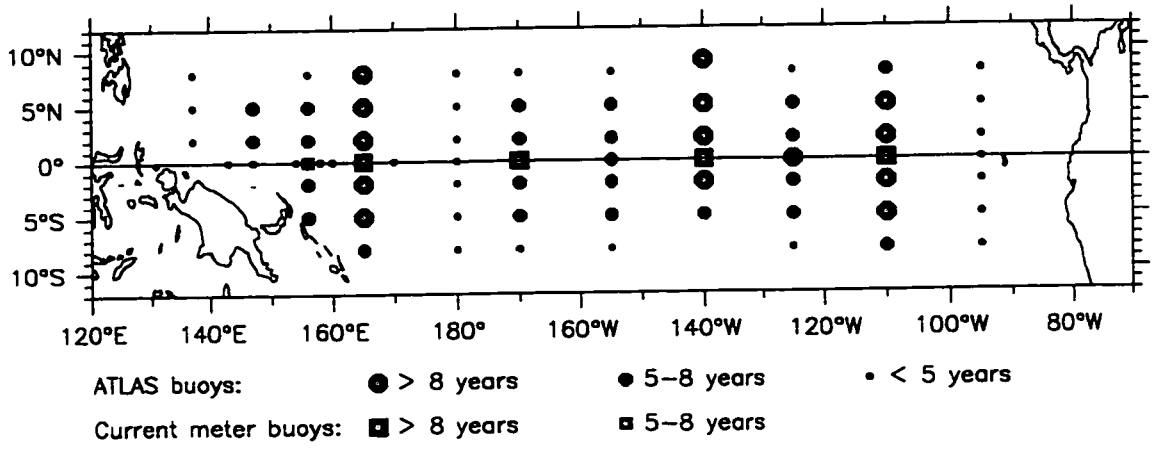


Fig. 2.1. The TAO array as of May 1997. Circles are ATLAS thermistor chain buoys, while squares indicate current meter moorings. The size of the symbols indicates length of time a buoy has been in the water at each location.

tested the accuracy of this procedure using data at 0°,140°W where an extrapolated time series could be compared with a 10 m time series from a mechanical CM. The CM and extrapolated ADCP time series of monthly mean zonal velocity had a cross correlation coefficient of 0.96, and the standard deviation of the differences was about 6 cm s^{-1}. However, the record length mean of the extrapolated time series was too strong to the west by 11 cm s^{-1}. Whereas vertical shears on which this extrapolation is based are stronger at 140°W than 170°W, the test nonetheless points out a potential for our shallow 10-20 m velocity estimates at 0°, 170°W to be biased westward by several cm s^{-1}.

Time series of daily-averaged wind components are used to compute surface wind stress, which is defined as

$$\vec{\tau} = \rho_a C_d |\vec{U}| \vec{U},$$

where $\vec{\tau}$ is vector surface wind stress, ρ_a is the air density, C_d is the drag coefficient and $\vec{U}$ is the vector wind speed. The winds at 10 m above the sea surface are estimated from TAO observations at 4 m using the algorithm described in daSilva et al (1994), assuming a neutrally stable surface boundary layer. The air density and drag coefficient are set 1.225 kg m^{-3} and 1.2×10^{-3}, respectively. Uncertainty in C_d may lead to errors of 20% in wind stress (Zeng et al. 1997).

Dynamic height is calculated relative to a reference level of 500 dbar from daily-averaged temperatures from the TAO and a mean temperature-salinity relationship based on the Levitus (1994) and Levitus et al. (1994) temperature and salinity climatologies. The root-mean-square error for the calculation is about 3-4 dyn. cm in the western Pacific and 2-3 dyn. cm in the eastern Pacific, mainly due

to the use of the mean temperature-salinity relationship, neglected variability below 500 dbar and relatively coarse vertical resolution of the TAO temperature observations (Busalacchi et al. 1994).

2.2.2. *Compiling TAO mean climatologies*

We compute mean climatologies of surface pseudo-stress, temperature and dynamic height along the equator, 5°N and 5°S, and current along the equator. For this purpose, we use the records from the locations where the buoys have operated for more than 5 years by the end of 1996, restricting ourselves to the period 1988 - 1996 which encompassed most of the growth in the TAO array. Longitudinally, we limit ourselves to the region 110°W-165°E. Although there are sites along 156°E that were occupied during 1991-96, the records at this longitude are gappy and the seasonal cycle is too weak to be reliably extracted with confidence from 5 year long records. Timelines for each mooring site are shown in Fig. 2.2.

For those sites retained in our analysis, we compute the mean seasonal cycle in the following manner. First, the daily averaged time series were averaged to monthly means. Second, monthly means of each individual month of the year were averaged over the time period of January 1988 - December 1996. Lastly, the resulting monthly climatologies were smoothed by a 1-2-1 filter to remove the intraseasonal variability (Kessler and McPhaden 1995; Kessler et al. 1995).

Ninety percent confidence intervals for each individual month of the climatology were estimated based on the Student t-distribution, assuming that each year is independent. The standard deviation for each month required in this calculation is estimated from year-to-year variations around the sample mean for a given month, so the estimated errors are mainly due to interannual variability

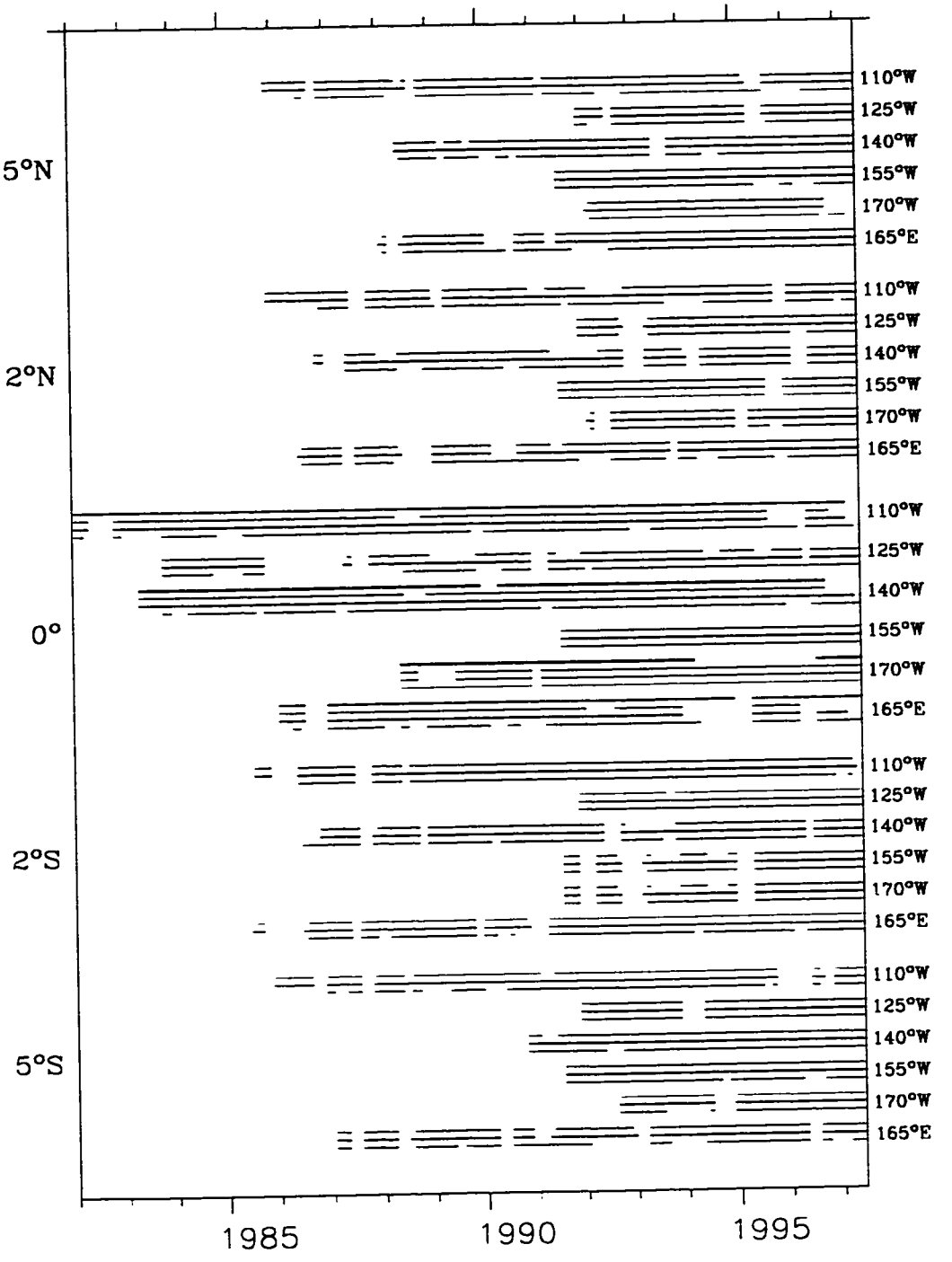


Fig. 2.2. Time history of sampling at locations of our interest. At y = 0°, each buoy has four lines: surface winds (lower thin); SST (medium thick); dynamic height (upper thin); and currents between 80 - 100 m (thick). At y = ±2° and ±5°, each buoy has three lines for surface winds, SST and dynamic height.

around the mean climatology.

To analyze the characteristics of seasonal variations, the monthly mean climatologies are fit to a five-parameter harmonic regression model for annual means, and 1 and 2 cycle per year (cpy) harmonics. This five-parameter harmonic regression is also applied to the entire time series. We find that harmonics from these two approaches have very small differences. Error variances are computed from residuals around these regressions, assuming 7 degrees of freedom (12 months less than 5 parameter fit). Standard errors for means, amplitudes and phases of the harmonics are derived using standard formulae (Draper and Smith 1981).

2.2.3. Compiling TAO interannual variability

Interannual variations of surface wind stress, ocean temperature, dynamic height and currents along the equator were estimated only from time series of these variabiles with length of 6 years or longer by May of 1997 for this analysis. Timelines for each mooring site are shown in Fig. 2.2. Overall data return for winds, temperatures and currents is about 85%.

To reduce missing data for surface winds and ocean temperatures at the equator, averages between 2°S and 2°N were used to fill gaps, which should not introduce errors because of high meridional coherence within 2° of the equator (Harrison and Luther 1990; Kessler et al. 1996). A different method was used to fill gaps in zonal velocity time series since there are no velocity measurements off the equator from the TAO observing system. We start with MCM time series at 165°E, 140°W and 110°W since they are longer than those from the ADCPs. Measurements of currents from the two different types of buoys are highly

correlated (Plimpton et al. 1997b), so records from the ADCPs are used to fill gaps in the time series from the MCMs at these three longitudes when available. At 170°W, the ADCP time series only are used because there were no MCM data at this location.

After filling data gaps using these methods, remaining gaps three months wide or less were filled by linear interpolation in time. This method works reasonably well for seasonal and ENSO time scales, which are the primary focus of our study. Data at 125°W before 1988 and the velocity data at 170 °W after 1994 are not used in this study because of the large gaps in these time series.

Interannual variations around the mean seasonal cycle were smoothed by a thirteen-month running-mean filter and three-month triangle filter to remove noise with frequencies equal or higher than 12 months after mean climatologies are subtracted from the time series.

2.3. Other data sets

2.3.1. The Florida State University wind product

The FSU wind product is a monthly analysis of ship and buoy winds on a 2° by 2° grid over the tropical Pacific (Legler 1991). Its mean climatology is computed in the same manner as from TAO winds, but for two different lengths of records. One is averaged over 36 years, from January 1961 to December 1996. Another is averaged over 1988-96, i.e., the same period as the TAO climatology. In addition to making comparisons between TAO and FSU pseudo-stress in Chapter 3, we will use the FSU product to force a Kelvin wave characteristic model in Chapter 5.

We note that FSU analysis procedures do not take into account the different anemometer heights of the various sources of wind information from ships and buoys (D. Legler, personal communication 1997). Thus, we will make no adjustments to either climatology when making our comparisons. If the FSU climatology were representative of winds at 10 m height, the computed pseudostresses based on TAO buoy winds at 4 m height might be 15% weaker than the FSU winds under conditions of neutral stability. This difference is no larger than uncertainties in drag coefficients used to compute stresses, and would not fundamentally alter our conclusions about the degree of consistency between the two climatologies.

2.3.2. Levitus et al (1994) temperature climatology

This dataset is an objective analysis of all historical temperature data between 1900 and 1993. Data sources include the National Oceanographic Data Center (NODC) archives, the NODC Oceanographic Data Archaeology and Rescue (NODAR) project and the Intergovernmental Oceanographic Commission (IOC) Global Oceanographic Data Archaeology and Rescue (GODAR) project (Levitus et al 1994). The data consist primarily of vertical profiles from expendible bathythermographs (XBT), conductivity-temperature-depth (CTD), mechanical bathythermograph (MBT) casts and digital bathythermograph (DBT) casts. The irregularly sampled data are mapped to standard depth levels through Lagrangian interpolation and then binned into 1° latitude by 1° longitude squares with optimum interpolation described in Gandin (1963). To produce climatological means, all the historical data in a given one-degree square within a given month are averaged as a representative of the monthly mean for the

square. Various techniques are used to quality control the data. Errors in analyzed fields are usually less than 1°C in the Pacific Ocean.

2.3.3. Reynolds and Smith SST

This dataset is a monthly global sea surface temperature climatology constructed from two intermediate climatologies on a 1° latitude by 1° longitude grid (Reynolds and Smith 1995). The first is for a 30-year (1950-79) period between roughly 40°S and 60°N based on in situ (ship and buoy) SST data, supplemented by 4 years (1982-85) of satellite SST retrievals. The second is based on monthly analyses using in situ SST data, satellite SST retrievals, and sea-ice coverage data over a 12-year period (1982-93). Uncertainties in SST's from this climatology are typically about 0.6°C.

2.3.4. Reverdin surface currents

This gridded data set is constructed as described by Reverdin et al. (1994) from drifter displacement data during the period January 1987 to April 1992. The records from more than 1000 surface drifting buoys were merged with currents from TAO current-meter moorings at 5 sites (4 along the equator at 165°E, 170°W, 140°W and 110°W and one at 7°N, 140°W). These data were optimally fitted into regular grids of 1° latitude by 5° longitude. The drifters were initially drogued at 15 meters. Sampling errors due to high-frequency variations are of the order of 7 cm s^{-1} at most except near the boundaries; systematic error can exceed 5 cm s^{-1} near 20°N and 20°S.

Chapter 3

SPATIAL AND TEMPORAL VARIATIONS IN THE EQUATORIAL PACIFIC

3.1. Introduction

In this chapter, we present the primary description of the mean and seasonal-to-interannual variability of thermal structure and circulation in the equatorial Pacific compiled from the TAO observations. Examples of TAO time series are presented first in Section 3.2. Mean fields from the TAO array are then presented in Section 3.3, and the seasonal and interannual variations around these means are described in Section 3.4 and Section 3.5, respectively. In addition, in Section 3.4, comparisons with other data sets are made for seasonal variations in order to examine the consistency of the TAO climatologies relative to those from the other data sets. Section 3.6 summarizes the results.

3.2. Examples of time series

Fig. 3.1 shows, as an example, time series of surface stress, temperatures and zonal currents at 0°, 110°W, which are the longest of the moored time series. At this location, both zonal and meridional stress have strong seasonal variations, but interannual variations are relatively weak. Sea surface temperature (SST) displays large variations on both seasonal and interannual time scales, while thermocline temperatures have more pronounced interannual signals, indicated by

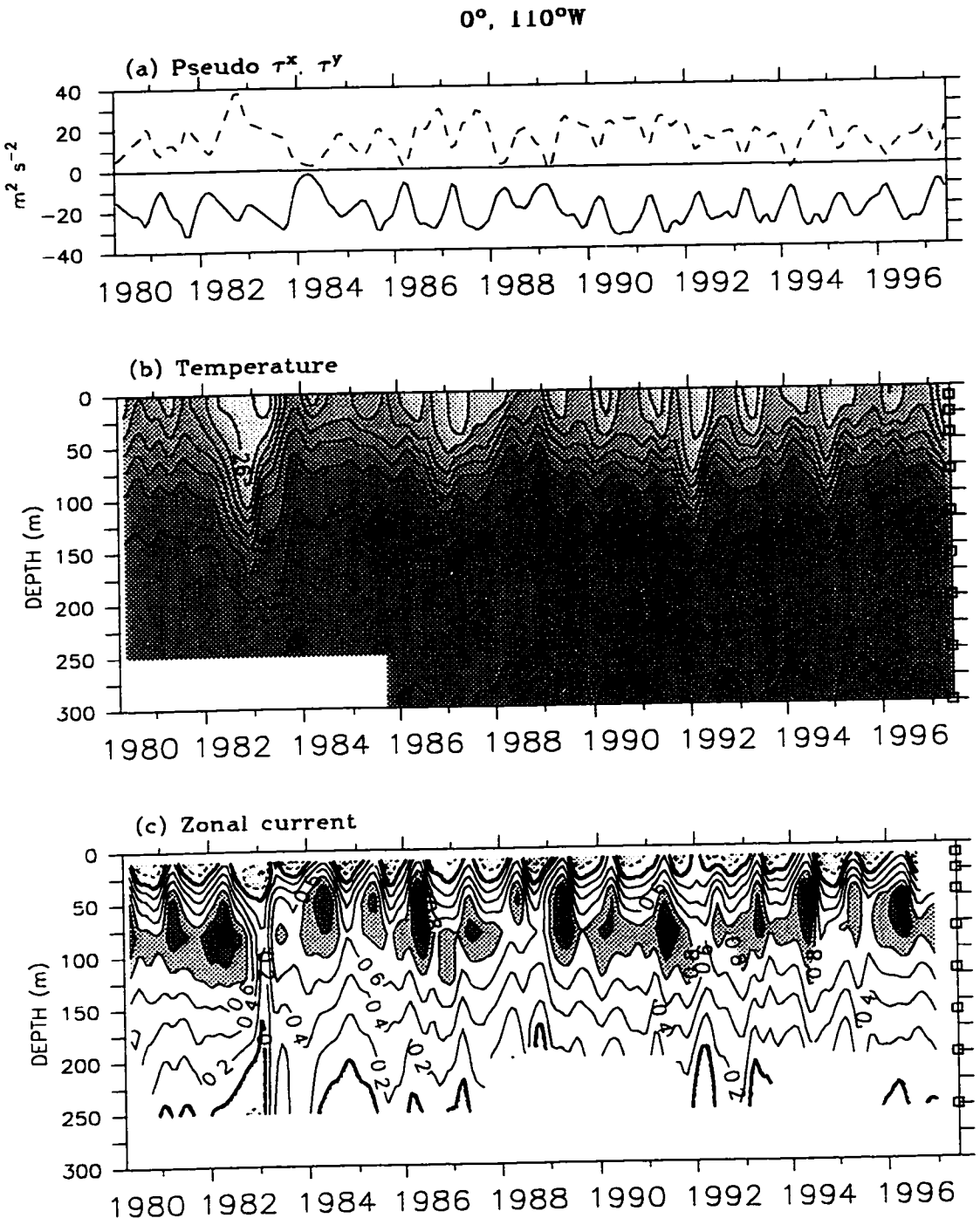


Fig. 3.1. The TAO data at 0°, 110°W. (a) Time series of zonal (solid) and meridional (dashed) surface pseudo-stress ($m^2\ s^{-2}$); (b) depth-time contours of temperatures (interval: 2°C); (c) depth-time contours of zonal currents (interval: 0.2 m s^{-1}). Westward currents are lightly shaded and the eastward EUC is heavily shaded. Small squares are depths of measurement.

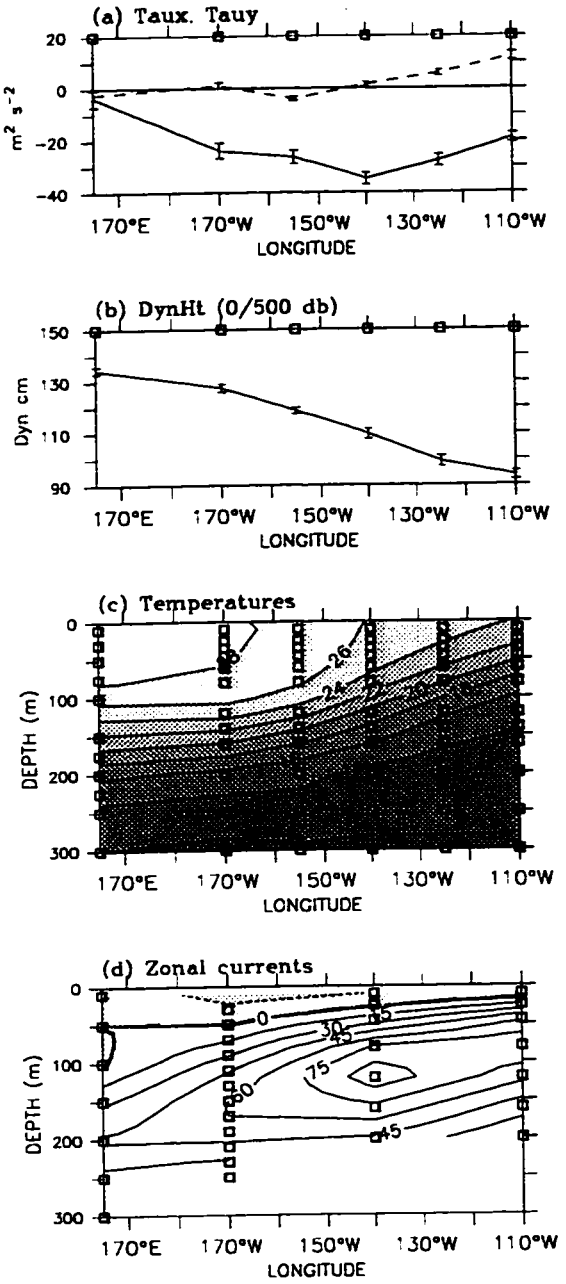


Fig. 3.2. Longitudinal structure of mean fields at the equator from TAO observations. (a) surface zonal (solid) and meridional (dashed) pseudo-stress ($m^2 s^{-2}$). Vertical bars are 90% confidence intervals. (b) surface dynamic height relative to 500 db (dyn. cm). (c) temperature (°C). The 90% confidence limits are ±0.3, ±0.7 and ±0.2 °C above, at and below the thermocline, respectively. (d) zonal currents (cm s^{-1}). The 90% confidence limits are ±7, ±9 and ±4 cm s^{-1} above, at and below the EUC core, respectively. Small squares indicate locations of measurements. ADCP resolution at 0°, 170°W is 10 m; only every other 10 m measurement depth is shown. Westward flow is shaded.

deepening of thermocline in warm El Niño years (i.e., 1982-83, 1986-87, 1991-92, 1994-95). Weak cold La Niña events occurred in 1984-85, 1995-1996, and a strong La Niña event occurred in 1988-89.

Zonal currents show strong variations in both surface and subsurface layers at this location (Fig. 3.1c). One of the most prominent features is the reversal of the mean westward South Equatorial Current (SEC), in boreal spring and early summer of almost every year. Flow at these times is against the easterly trade winds in the eastern Pacific. The dynamics of this "springtime reversal" (Halpern 1987) have been a subject of several recent studies (Philander et al. 1987; McPhaden and Taft 1988; Philander and Chao 1991; Yu et al. 1997), and its possible mechanism is further discussed in this thesis.

3.3. Means of TAO observations

Fig. 3.2 shows the longitudinal profiles of mean zonal and meridional winds and surface dynamic height (0/500 db), and the longitude-depth contours of temperature and zonal current means along the equator. Sea surface temperature (SST) decreases to the east, from 29°C at 165°E to 23°C at 110°W. Coldest SSTs are found in the eastern Pacific where the thermocline is shallow and equatorial upwelling is efficient at cooling the surface. The thermocline slopes upward towards the east in response to zonal wind stress forcing (Fig. 3.2a). Using the 20°C isotherm as an indicator, thermocline depth decreases from about 160 m at 165°E to about 50 m at 110°W (Fig. 3.2c). Surface dynamic height rises to the west by about 40 dyn cm, likewise in response to zonal surface wind stress forcing.

The core of the eastward Equatorial Undercurrent (EUC) slopes upward to

the east in the thermocline Fig. (3.2d) with maximum flow around 140°W. The undercurrent is driven by the eastward zonal pressure gradient, which in turn is set up by the surface easterlies (McPhaden 1981; McCreary 1981). The mean surface South Equatorial Current along the equator is toward the west, in the direction of the surface wind stress. Maximum westward flow in the SEC occurs at 0°, 170°W. This maximum is probably real but may be accentuated by a westward bias in mean currents of several cm s^{-1} as described the previous chapter. The equatorial speed of the SEC is much weaker than that of the Undercurrent, even though the SEC is directly forced by the stress. Philander and Pacanowski (1980) argued that the westward momentum generated by the surface stress is reduced by vertical advection of eastward momentum due to upwelling, resulting in the weaker surface current than one would expect from linear theory (McPhaden 1981; McCreary 1981). Diagnostic analyses of the mean zonal momentum balance are consistent with this hypothesis (Bryden and Brady 1985; McPhaden and Taft 1988).

The meridional-depth structures of temperature at 165°E, 140°W and 110°W are shown in Fig. 3.3. The most prominent feature at 140°W and 110°W is shoaling of isotherms toward the equator in the upper 50-100 m, a signature of cold water brought up to the surface layer by trade-wind-forced equatorial upwelling. At 165°E, the effects of upwelling are not evident in temperature because the easterly surface stress is very weak and the thermocline is relatively deep. Another important feature, shown in all three locations, is vertical stretching of the thermocline near the equator. This feature is associated with the geostrophic shear of the westward surface current and eastward undercurrent current in the thermocline (Lukas and Firing 1983).

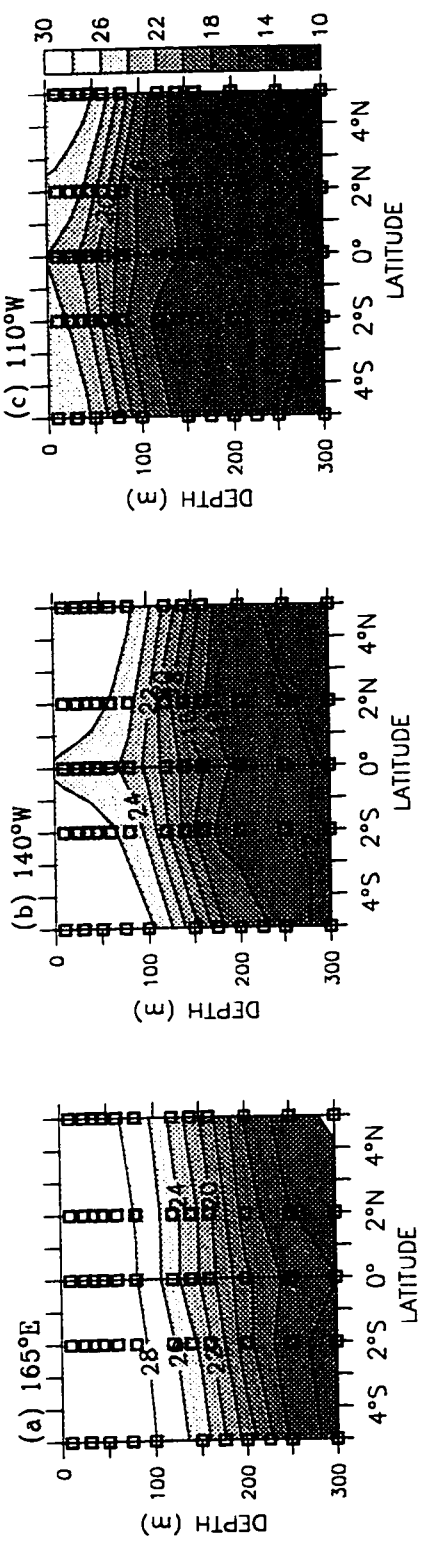


Fig. 3.3. Latitudinal structure of mean temperature field from TAO observations along 110°W, 140°W and 165°E. The 90% confidence limits are the same as in Fig. 2. Small squares indicate locations of measurements.

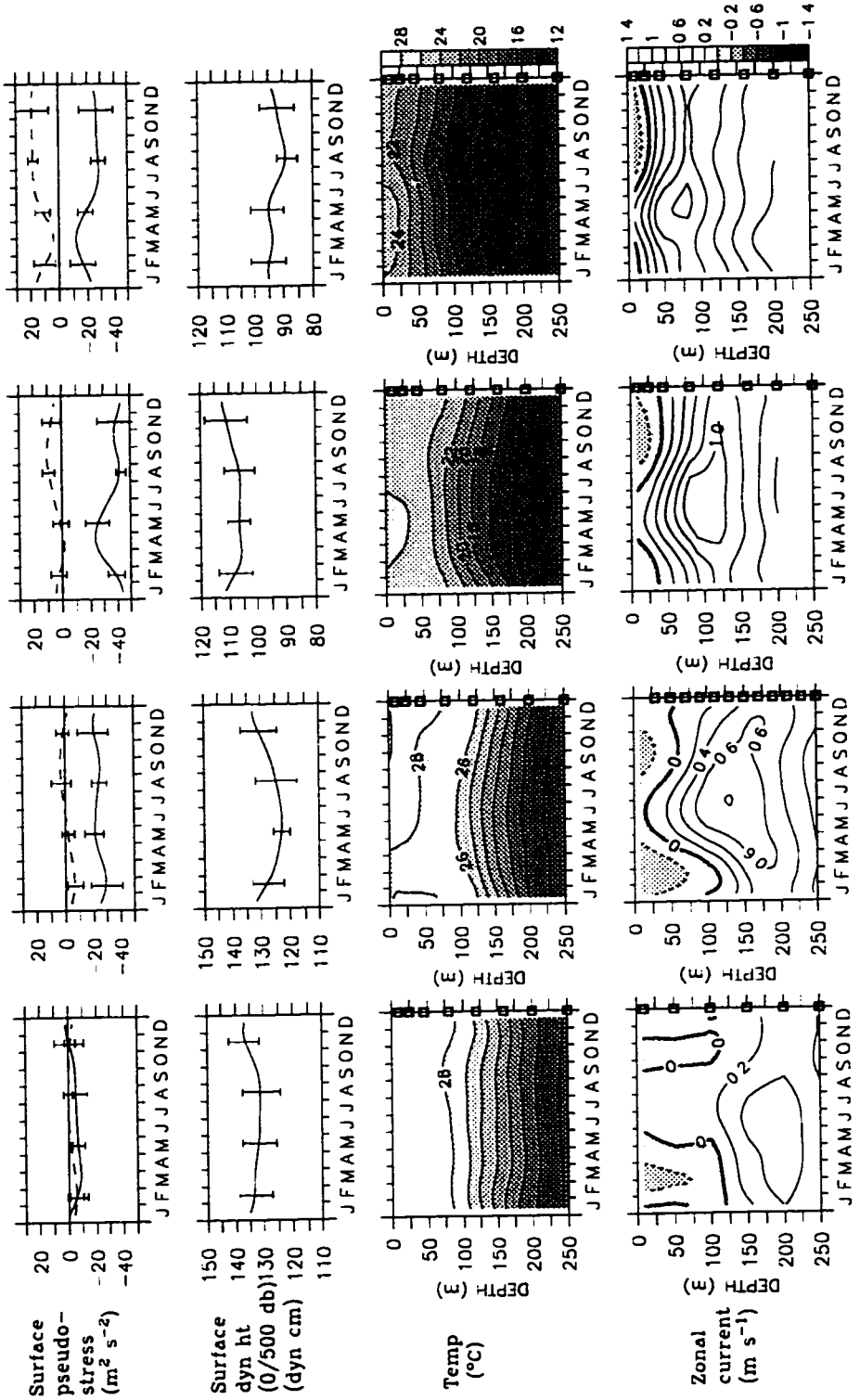


Fig. 3.4. Monthly mean climatologies of surface winds ($m^2\ s^{-2}$), surface dynamic height relative to 500 db (dyn. cm), temperature (°C) and zonal currents ($cm\ s^{-1}$) at (a) 0°, 165°E, (b) 0°, 170°W, (c) 0°, 140°W, and (d) 0°, 110°W. The 90% confidence intervals are indicated every third month for winds and dynamic heights. Shading indicates westward flow. Small squares indicate locations of measurements.

3.4. Seasonal variations of TAO observations and other products

3.4.1 TAO observations

Fig. 3.4 shows time series of the monthly mean climatologies of surface winds and surface dynamic height (0/500 db), and the depth-time structure of the climatologies of temperatures and zonal currents at 165°E, 170°W, 140°W and 110°W at the equator. The seasonal variation in both zonal and meridional winds is much larger in the eastern Pacific than in the western Pacific (Fig. 3.4, upper panels). In the eastern Pacific, the winds are from the southeast quadrant with minimum speeds in boreal spring and maximum speeds in boreal fall or winter. The variation of meridional winds is out of phase with the zonal component in the eastern Pacific, i.e., weak easterlies are associated with weak southerlies and vice versa. SST in the eastern Pacific ocean undergoes a prominent seasonal variation; its range is 2-3°C, with highest temperatures in boreal spring and lowest in boreal fall (Fig. 3.4c & d). In the western Pacific, where the surface winds have very weak seasonal variability and the thermocline is deep, the seasonal variation in SST is much weaker.

The upper thermocline undergoes its largest seasonal vertical migration in the central and eastern Pacific; west of the dateline these migrations are of much smaller amplitude (Fig. 3.4). At 170°W, the thermocline begins to shoal in January and is at its shallowest in May-June (Fig. 3.4b, lower middle panel). At 140°W and 110°W, the thermocline starts to shoal in late January, but it does not reach its shallowest depth until June-July at 140°W and August at 110°W. The seasonal range in surface dynamic height is typically less than 15 dyn cm, with minima (maxima) surface heights corresponding roughly to periods of shallowest (deepest) thermocline depth.

The zonal surface current has significant seasonal variations at all four locations (Fig. 3.4, lower panels) with a largest magnitude in the eastern Pacific. In boreal spring, the current flows eastward at speeds of 20-30 cm s^{-1} at 140°W and 110°W while the surface zonal winds, although weaker, are still easterly. This "springtime reversal" of the SEC reaches its maximum in April at 110°W and in May at 140°W. Further west at 170°W, 10 m currents nearly reverse in June (and may actually reverse when taking account of the probable westward bias in our estimated 10 m mean flow at this site), while at 165°E zonal velocity reverses and flows against the surface winds in June-August. At both 170°W and 165°E there is also a prominent semi-annual component to the 10 m zonal velocity, which at 165°E results in a second period of eastward flow in November-January.

Zonal currents from the surface to the EUC core vary seasonally in phase (Fig. 3.4). In boreal spring, the EUC reaches its maximum in the eastern Pacific of about 120 cm s^{-1} at 140°W and about 105 cm s^{-1} at 110°W. Also the EUC core undergoes a seasonal vertical migration, reaching its shallowest depths in boreal spring (~80 m at 140°W and ~60 m at 110°W), and its deepest depths in early boreal winter. At 170°W, the scenario is similar but the variations lag those further to the east. The seasonal variation of the EUC is relatively weak at 165°E, compared to further to the east, though still apparent with a maximum in boreal spring at a depth of around 200 m.

Seasonal variations in winds, temperature, dynamic height and zonal current are highlighted by subtracting the long-term means from monthly mean climatologies (Figs. 3.5 and 3.6). Along the equator, zonal winds progress westward, as do SST and zonal currents in both the surface and subsurface layers.

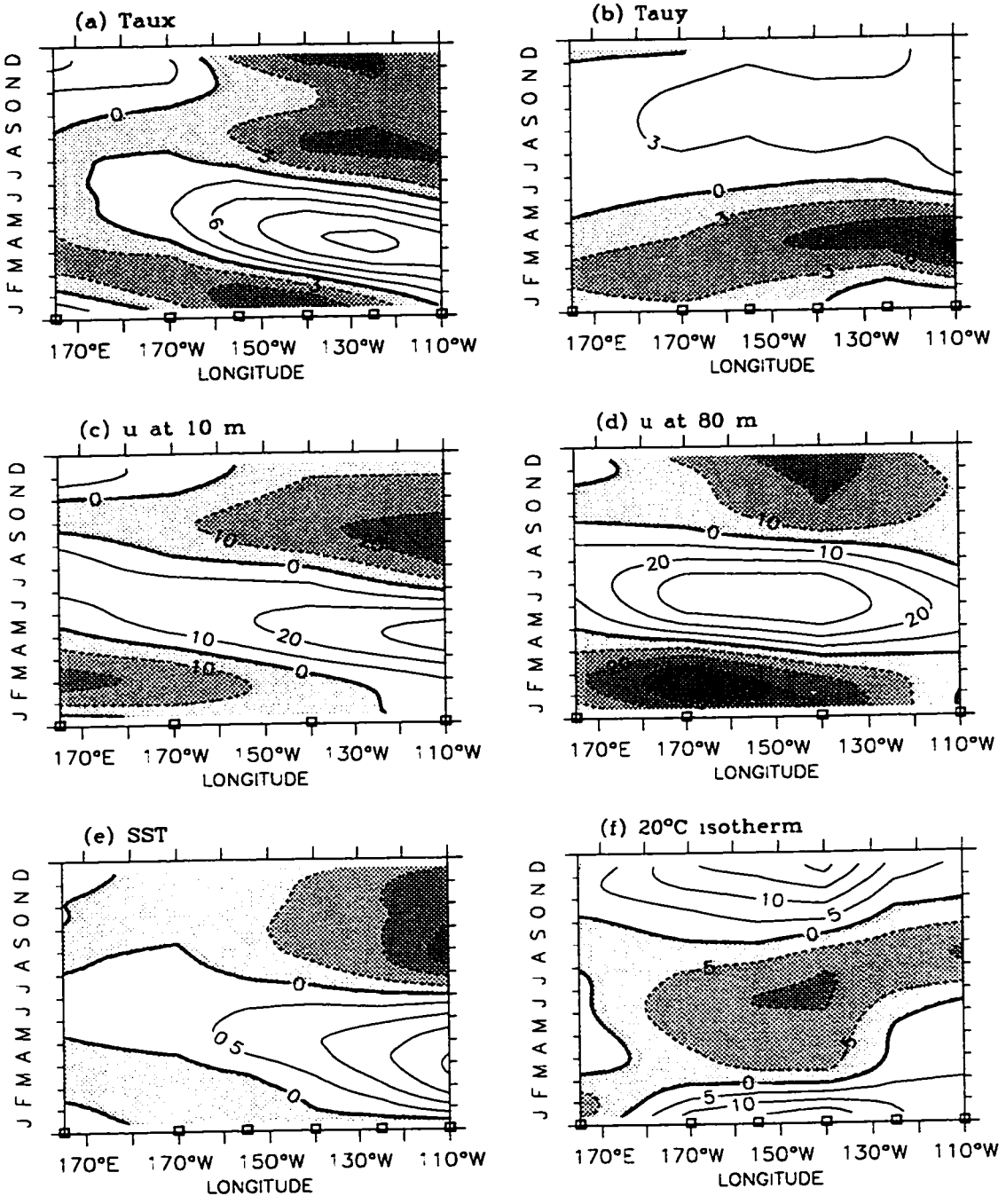


Fig. 3.5. Demeaned seasonal variations along the equator from TAO array observations. (a) Zonal pseudo-stress (m^2 s^{-2}); (b) meridional pseudo-stress (m^2 s^{-2}); (c) zonal current at 10 m (cm s^{-1}); (d) zonal current at 80 m (cm s^{-1}); (e) SST (°C); and (f) depth of 20°C isotherm (m). The squares on the lower axis are measurement longitudes. Negative anomalies are shaded.

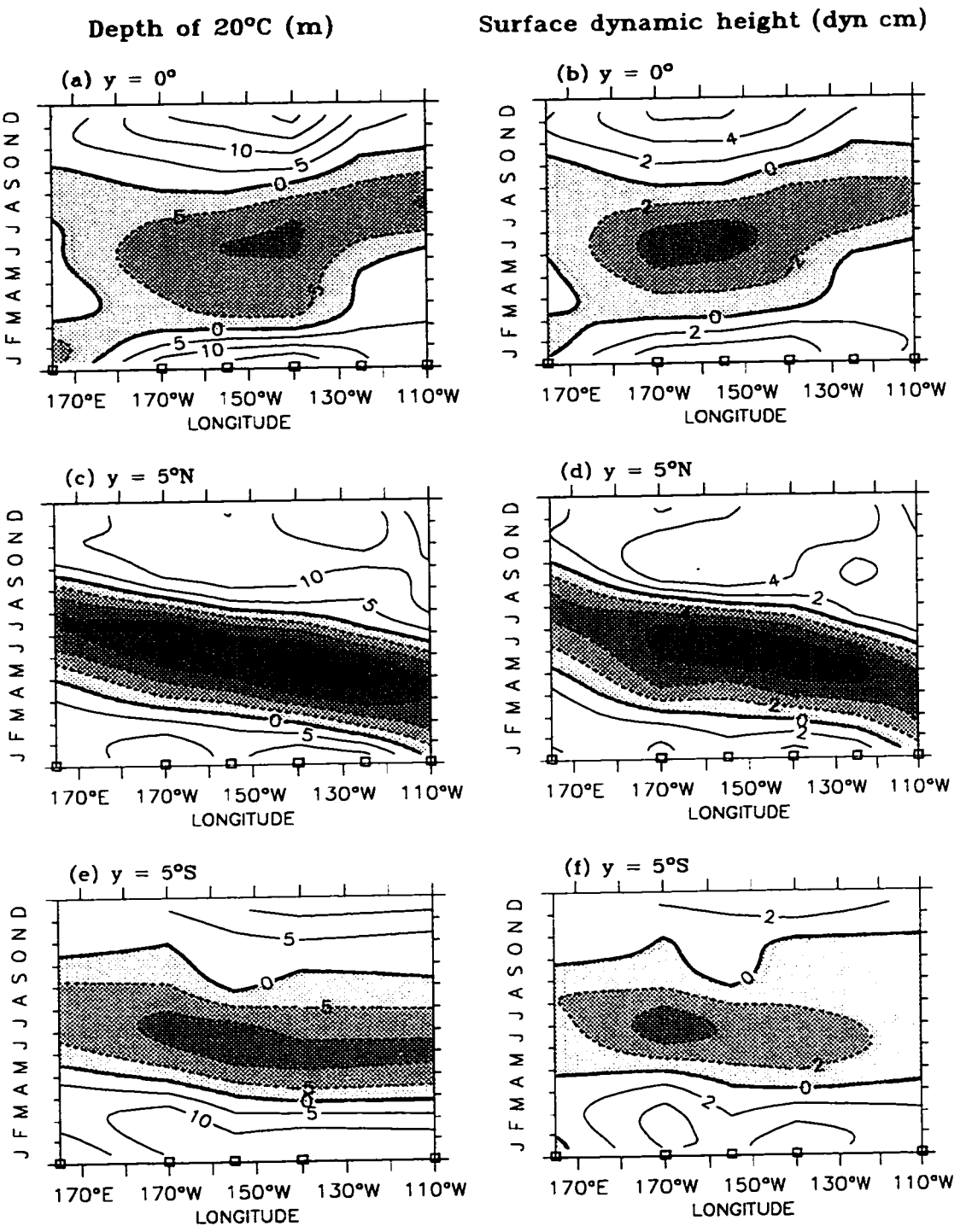


Fig. 3.6. Demeaned seasonal variations along the equator from TAO observations. Left panels: depth of 20°C isotherm (m). Right panels: surface dynamic height relative to 500 db (dyn. cm). The squares on the lower axis are measurement longitudes. Negative anomalies are shaded.

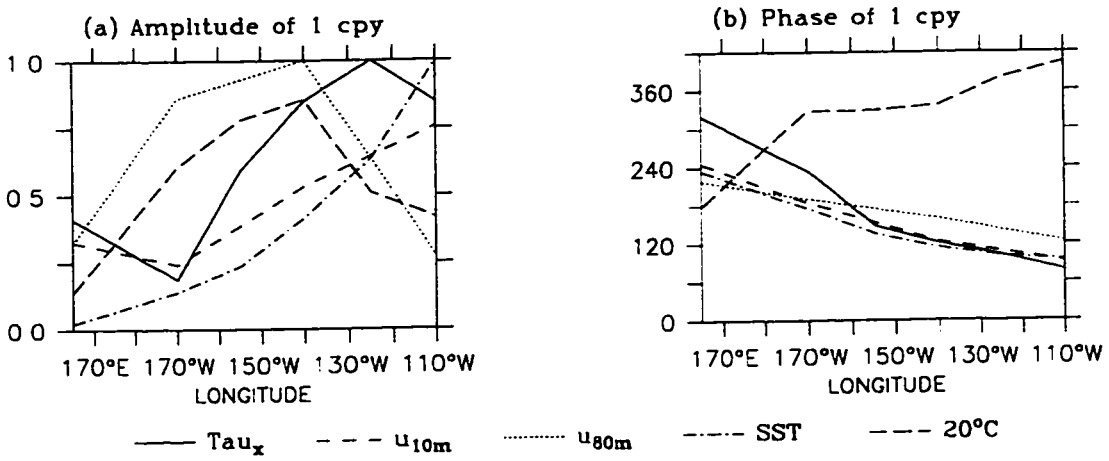


Fig. 3.7. Amplitudes and phases of 1 cpy harmonic based on least squares regression analysis of TAO data. Solid line: pseudo-stress τ^x; short dashed line: u_{10m}; dotted line: u_{80m}; dotted-dashed line: SST; long dashed line: 20°C isotherm. Amplitudes are each normalized by their maximum. τ^x_{max} = 9.6 $m^2 s^{-2}$; u_{max} = 0.37 m s^{-1}; SST_{max} = 2.0 °C; $D^{20°C}_{max}$ = 19.4 m.

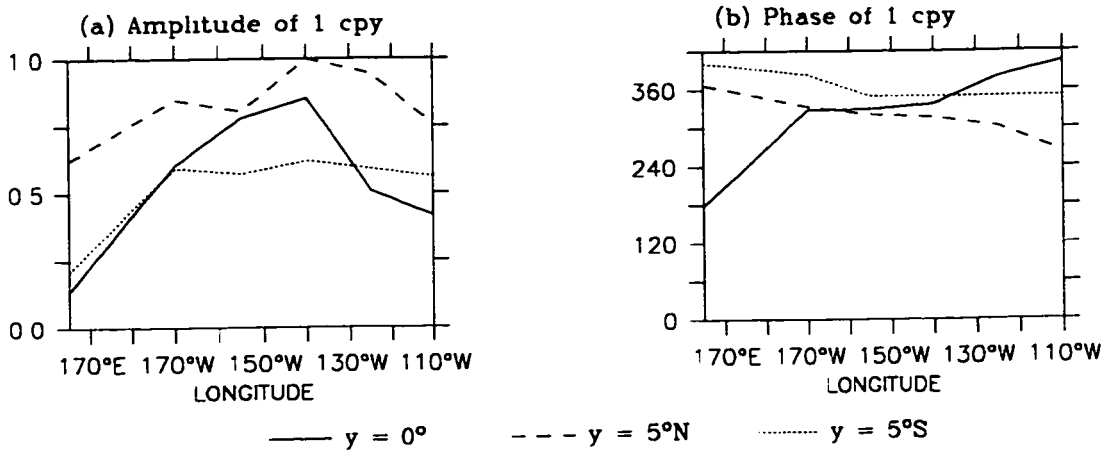


Fig. 3.8. Amplitudes and phases of 1 cpy harmonic based on least squares regression analysis of 20°C isotherm depth from TAO data. Solid line: y = 0°; dashed line: y = 5°N; dotted line: y = 5°S. Amplitudes are normalized $D^{20°C}_{max}$ = 19.4 m.

Table 3.1. Estimated zonal phase speeds (m s^{-1}) of the annual harmonics for mean climatologies. The speeds are derived from an amplitude weighted mean least square regression of phases along the equator. One standard error is indicated for each estimate.

A. TAO climatologies

τ^x		-0.43 ± 0.13
u_{10m}		-0.73 ± 0.27
u_{80m}		-1.13 ± 0.45
SST		-0.87 ± 0.23
20°C depth	y=0°	0.65 ± 0.21*
	y=5°N	-1.11 ± 0.37
	y=5°S	-0.90 ± 0.55
Surface dynamic height	y=0°	0.49 ± 0.18*
	y=5°N	-0.92 ± 0.29
	y=5°S	-0.85 ± 0.67

B. Other climatologies

FSU τ^x		-0.38 ± 0.06
Reverdin u		-0.59 ± 0.08
Reynolds SST		-0.68 ± 0.09
Levitus 20°C depth	y=0°	0.67 ± 0.21*
	y=5°N	-0.97 ± 0.17
	y=5°S	-0.83 ± 0.16
Levitus Surface dynamic height	y=0°	0.47 ± 0.15*
	y=5°N	-1.00 ± 0.20
	y=5°S	-0.76 ± 0.38

(* Fit only to longitudes 155°W-110°W)

In contrast, vertical displacements of the 20°C isotherm and dynamic height have a different character east and west of about 170°W. East of this longitude, a dominant annual period fluctuation propagates eastward, whereas west of 170°W there is a suggestion of westward propagation and a more pronounced semi-annual cycle.

The time-longitude structure of the seasonal variability in 20°C isotherm depth and surface dynamic height along 5°N shows very clear westward propagation across the entire basin (Fig. 3.6c & d). At 5°S, westward propagation is also apparent in 20°C and dynamic height primarily west of 140°W.

Meridional wind stress tends to propagate eastward in the western and central Pacific. However, meridional wind stress is relatively inefficient at exciting dynamical responses in the equatorial ocean on seasonal time scales (Gent et al 1983), so hereafter we will focus discussion primarily on zonal wind stress.

To further analyze the seasonal cycle, monthly mean climatologies are fit to a five-parameter (annual mean, 1 cpy and 2 cpy) harmonic regression model. Then the phases and amplitudes of the harmonics are used to estimate their zonal phase speeds using an amplitude weighted linear least squares fit. We note that the 1 cpy harmonic is generally larger than the 2 cpy harmonic and the amplitude of the 2 cpy harmonic exceeds one standard error less often than the amplitude of the 1 cpy harmonic (not shown). Therefore, in subsequent discussion, we will emphasize 1 cpy variations, noting where appropriate the importance of 2 cpy variations in the seasonal cycle.

The annual harmonic typically has its maximum amplitude in the eastern and central Pacific, and phase increases to the west in all the variables except the

20°C isotherm depth and dynamic height along the equator (Fig. 3.7b and Fig. 3.8b; dynmaic height is not shown in these figures because it has very similar phase features as the depth of 20°C isotherm). The annual harmonic of zonal pseudo-stress has its maximum amplitude in the eastern Pacific (Fig. 3.7a), with a zonal phase speed of -0.43±0.13 m s^{-1} (Table 3.1). The annual harmonic of SST has significant amplitude only east of the dateline; its phase increases to the west with a phase speed of -0.87±0.23 m s^{-1}. Annual harmonics of surface zonal current and 80 m zonal current have significant amplitude across the entire region 165°E-110°W (Fig. 3.7a) with phase speeds -0.73±0.27 m s^{-1} and -1.13±0.45 m s^{-1}, respectively.

Annual period fluctuations in 20°C depth and dynamic height along the equator are largest in the central Pacific. Conversely, amplitudes at 165°E are only marginally significant and the phases there are highly uncertain. For that reason, phase speeds for the annual harmonic in 20°C depth and dynamic height cannot be reliably estimated in the western Pacific. For the eastern and central Pacific where 20°C isotherm depths and dynamic height amplitudes are highly significant, phase speeds along the equator are eastward at 0.5-0.7 ms^{-1} (Table 3.1).

The annual harmonic of 20°C isotherm depth and dynamic height is larger at 5°N than at either 0° or 5°S. This is partly due to the stronger seasonal wind stress and wind stress curl forcing north of the equator (see more discussion in Section 5.2). Phase speeds are westward at about 1 m s^{-1} on average at both 5°N and 5°S (Table 3.1). The relatively large uncertainty in phase speeds at 5°S results from the lack of zonal progression in phase east of 140°W. Nevertheless, at both 5°N and 5°S, these phase speeds are significantly different from the eastward phase speed in 20°C and dynamic height along the equator of about 0.5-0.7 m s^{-1}.

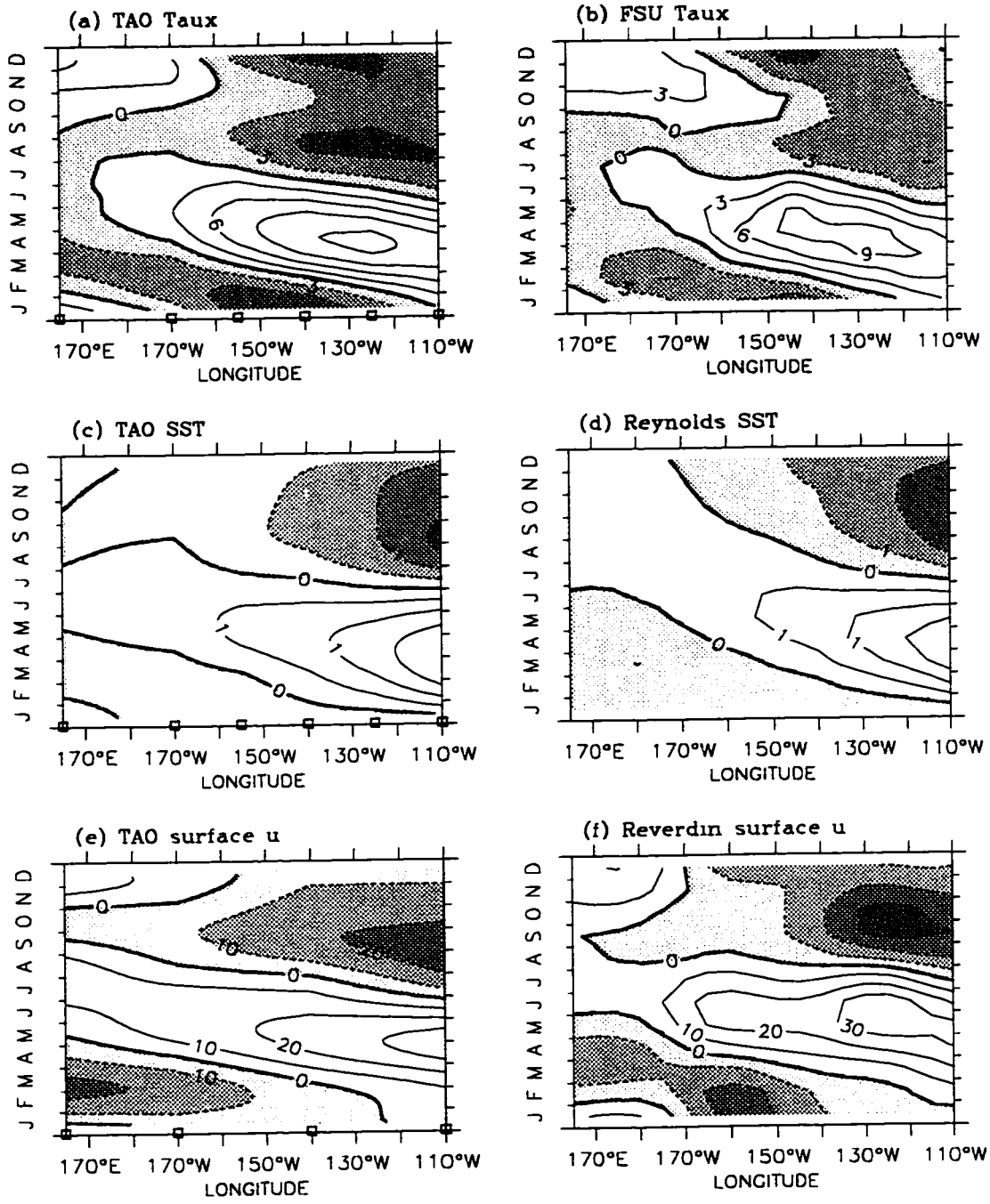


Fig. 3.9. Comparison of demeaned seasonal variability between TAO and other datasets along the equator. Squares on the lower axis are measurement longitudes for TAO. The negative is shaded. (a), (b): zonal pseudo-stress τ^x from TAO and FSU ($m^2\ s^{-2}$). (c), (d): SST from TAO and Reynolds and Smith. (°C). (e), (f): surface zonal current from TAO and Reverdin et al (1994) ($cm\ s^{-1}$).

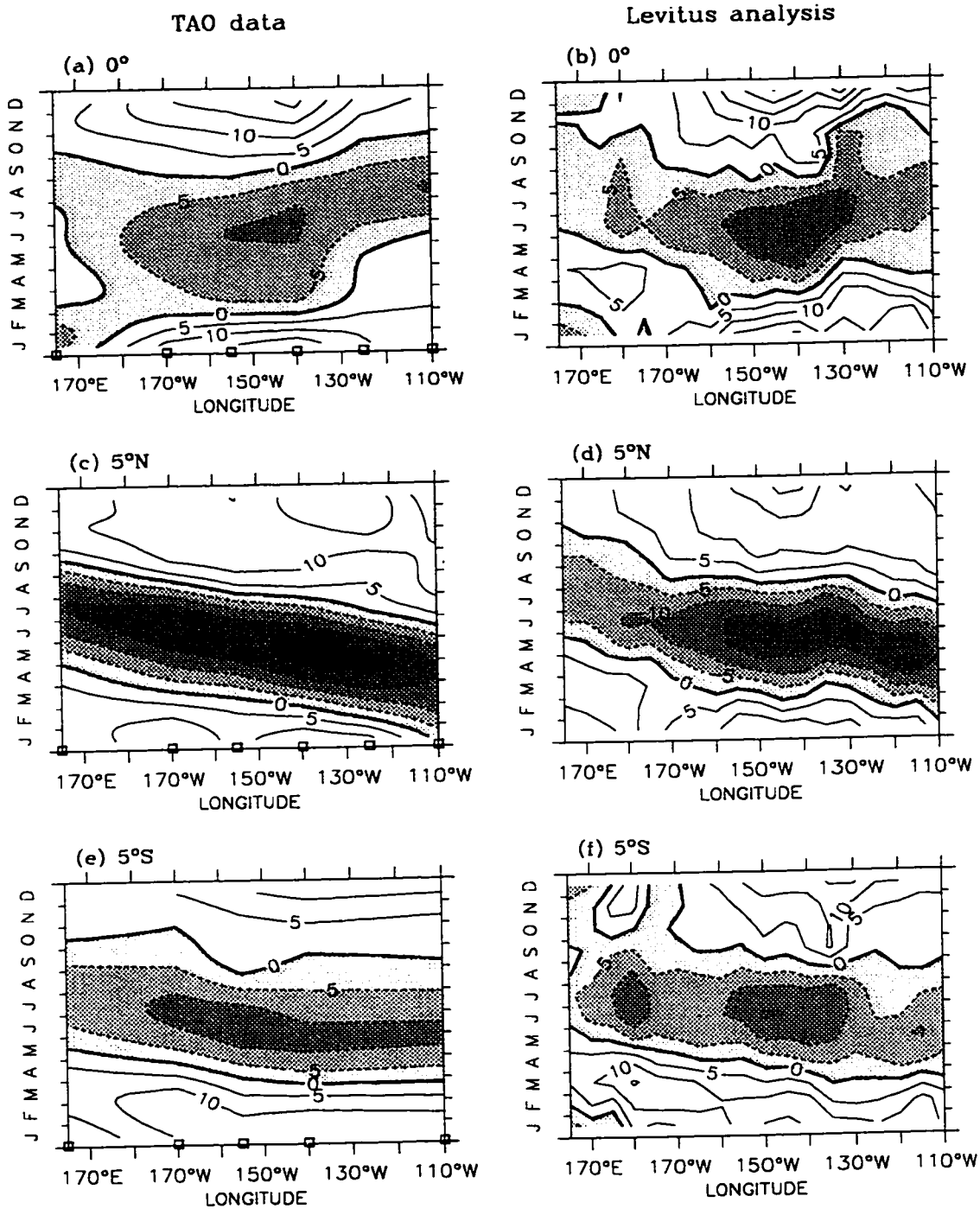


Fig. 3.10. Comparison of demeaned seasonal variability in 20°C isotherm depth between TAO and Levitus datasets. Squares on the lower axis are measurement for TAO. Contour intervals are 5 m. Negative anomalies (shading) means shoaling. Left panels: TAO data. Right panels: Levitus data.

3.4.2. Comparison with other climatologies

The zonal surface pseudo-stress along the equator from the TAO array is consistent with FSU pseudo-stress (Fig. 3.9a & b). In both, variations progress to the west with a similar phase speed (Table 3.1), but with relatively larger amplitudes in the TAO data. The amplitude difference could be due to the fact that FSU mean climatology is averaged over 1961-1996 while the TAO mean climatology is averaged over only 1988-1996. Indeed, the amplitude of the seasonal variability is more consistent with that in the TAO climatology in the eastern Pacific if the FSU climatology is averaged over the same period as the TAO data. Lukas and Firing (1985) also analyzed the surface winds collected during the Hawaii-to-Tahiti Shuttle experiment from ship, buoys, satellite and island weather stations. The annual harmonic amplitude is about 30% weaker in their analysis than in our analysis of TAO observations, but the phase speed of -0.41 m s^{-1} they obtained is essentially identical to our estimate of -0.43 m s^{-1}, considering the uncertainties in the calculations.

Variations in SST from the TAO array are similar to those of Reynolds SST along the equator, capturing the major large zonal-scale features such as westward phase progression and magnitude of seasonal cycle (Fig. 3.9c & d). The two analyses agree particularly very well in the central and eastern Pacific, where seasonal variability is largest. The Reynolds SST shows a slower westward phase progression than TAO west of 160°W, but in this region seasonal variations in both products are weak.

Fig. 3.9 also displays longitude-time contours of zonal surface current along the equator from the TAO buoy array and Reverdin et al near-surface velocity analysis. Both analyses show similar large scale features, including

westward propagation of the annual cycle and a 2 cpy variation in the western Pacific. The zonal phase speed of the annual cycle in the Reverdin et al zonal currents along the equator is -0.59 m s^{-1}, comparable to the phase speed from our analysis of TAO buoy data (Table 3.1). There are no subsurface current climatologies against which to compare those from the TAO array.

Fig. 3.10 compares the TAO climatology for the 20°C isotherm depth with Levitus temperature analysis, both on and off the equator. The best agreement is along 5°N, both in amplitude and phase. At 5°S, TAO and Levitus amplitudes are similar, but the Levitus analysis shows a more consistent westward phase increase across the basin, particularly east of 140°W. Least-squares determined phase speeds for TAO and Levitus analyses are nearly identical at 5°S, but the greater consistency of the Levitus phase progression leads to more reliable phase speed estimates at this latitude (Table 3.1). It may be that TAO record lengths of typically 5-6 years are insufficient to extract subtleties of phase characteristics in the southeastern equatorial Pacific where annual period signals are relatively weak.

Along the equator, both TAO and Levitus data sets exhibit eastward phase propagation at about 0.5-0.7 m s^{-1} for the annual harmonic in the central and eastern Pacific (Fig. 3.10a & b; Table 3.1). Likewise both suggest westward propagation, and a relatively strong semi-annual cycle, in the western Pacific. In the Levitus data, the dividing longitude between eastward and westward phase propagation along the equator (around 155°W) is more strongly demarcated than in the TAO data, probably due to the better zonal resolution in the Levitus analysis. Other detailed differences are evident. In addition to errors of various types, these differences are mainly due to the differences in data sources and data

distributions for TAO and Levitus climatologies as shown in Sections 2.2 and 2.3.

3.5. Interannual variations of TAO observations

Interannual fluctuations, illustrated at 0° 110°W and 0° 165°E (Fig. 3.11), are dominated by ENSO time scale variations. The fluctuations at 0° 140°W are not included in Fig. 3.11 because they are very similar to those at 0° 110°W. Zonal wind stress variations on interannual time scales are larger in the western Pacific than in the eastern Pacific. Conversely, meridional wind stress variations are stronger in the eastern Pacific because of meridional shifts in the position of the Intertropical Convergence Zone (ITCZ). Interannual variations in zonal currents are comparable to seasonal variations (Figs. 3.4 & 3.5), but unlike seasonal variations, interannual variations in zonal currents do not have strong vertical coherence and are not trapped above the EUC core. Interannual variations in temperature have high coherence in and above the thermocline at 110°W with maximum amplitudes in the thermocline between 50-100 m. At 165°E, maximum temperature amplitudes are also found in the thermocline at depths of 100-200 m, but these are less coherent with surface temperature variabtions as compared to further east. The SST anomaly is large at 110°W because the mean thermocline is very shallow there and SST variations are dominated by anomalous vertical displacements of the thermocline (Battisti and Sarachik 1995).

Interannual variations along the equator for winds, SST, thermocline depth, dynamic height and zonal transport per unit width in the EUC (depths between 80-200 m) are shown in Fig. 3.12. Only data from 1988 to 1996 are presented due to limited extent of the TAO array before 1988 (Fig. 2.2). The 18°C instead of 20°C isotherm is shown because the 20°C isotherm breaks the surface during the 1988-

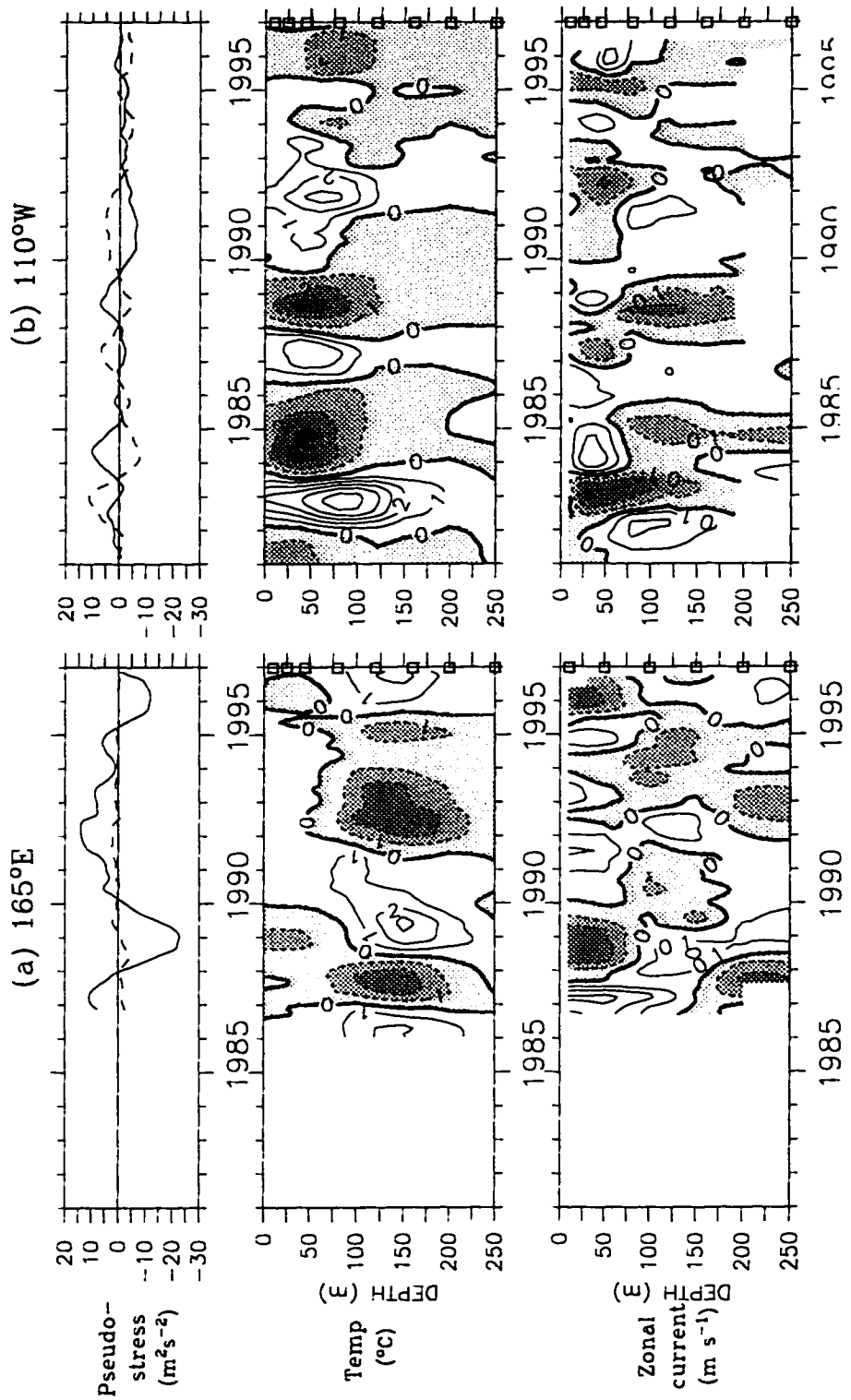


Fig. 3.11. Demeaned interannul variations from the TAO. Upper panels: time series of zonal (solid) and meridional (dashed) wind stress; middle panels: depth-time contours of temperatures (interval: 1°C); and lower panels: depth-time contours of zonal currents (interval: 0.1 m s^{-1}). In middle and lower panels, shading indicates negative fluctuations. Small squares indicate depths of measurements.

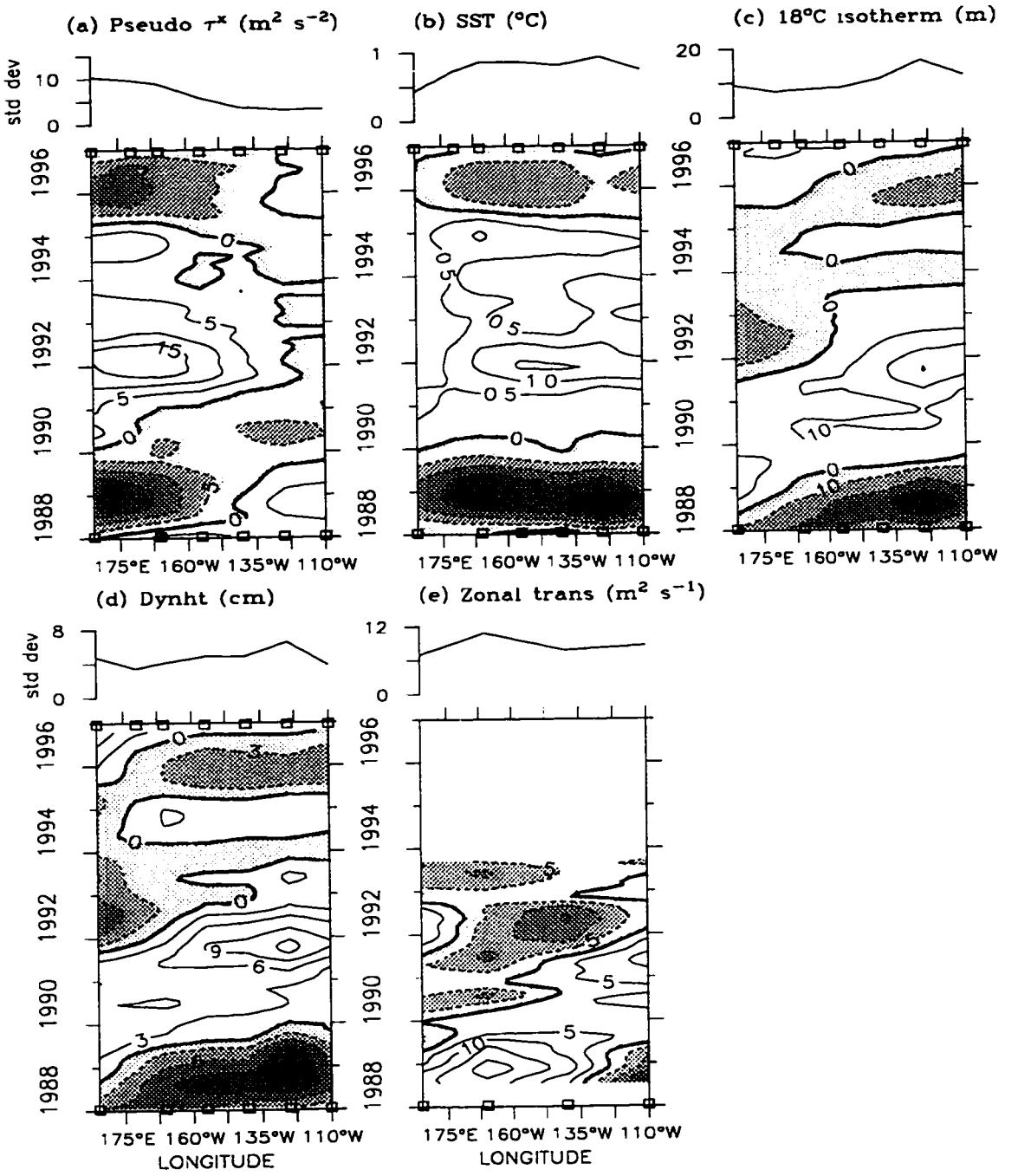


Fig. 3.12. Time series of demeaned interannual variability of TAO observations along the equator. Negative anomalies are shaded. (a) Zonal stress (interval: 5 m^2 s^{-2}); (b) SST (interval: 0.5°C); (c) depth of 18°C isotherm (interval: 10 m); (d) surface dynamic height (interval: 3 dyn. cm); and (e) zonal transport per unit width over depths of 80-200 m (interval: 5 m^2 s^{-1}). In (a)-(d), the squares on the lower (upper) axis are measurement locations before (after) September 1993 and in (e), the squares on the lower axis indicate the locations for the entire period of time. The curve above each panel is standard deviation of the variation in that panel.

89 La Niña. Zonal transport per unit width is shown only through September 1993 (Fig. 3.12e) because velocity data at 170°W are available only up to March 1994 and edge effects due to smoothing the data further shorten the time series.

Zonal surface stress is roughly in phase in the western and central Pacific, where it has maximum amplitudes. Variations in zonal stress in the eastern Pacific are weak and roughly 180° out of phase with those west further. SST variations have amplitudes typically of 0.5-1.5°C and are highly correlated with depth anomaly of 18°C isotherm in the eastern Pacific (Fig. 3.12b & c), where the mean thermocline depth is shallow.

Thermocline depth and dynamic height have very similar phase patterns with maximum amplitudes in the eastern Pacific (Fig. 3.12c & d). Both 18°C depth and dynamic height exhibit eastward phase propagation, but most of the phase shift occurs in the central Pacific. This phase propagation is similar to that shown in analytic study of Cane and Sarachik (1981) and in the model simulations of Battisti and Hirst (1989) and Schneider et al (1995). In contrast, neither SST nor surface wind stress displays similarly consistent eastward propagation from 1986 to 1996 (Fig. 3.12a & b).

Although longitude-time contours of zonal transport are constructed from only four locations along the equator, the transport variations exhibit a well defined pattern of interannual variability (Fig. 3.12e). Extrema generally occur in the central Pacific between 140°W-170°W. During the cold ENSO event of 1988-89, when surface easterly winds were anomalously strong in the central and western Pacific, subsurface zonal transport in the central Pacific was anomalously eastward (i.e., the EUC was unusually strong). During the warm event of 1991-92, when the easterly winds relaxed in the central and western Pacific, zonal transport

was anomalously westward between 140°W-170°W (i.e., the EUC was unusually weak).

3.6. Summary

TAO data are used to document mean fields, seasonal cycle and interannual variability in the equatorial Pacific. The mean fields show familiar features in thermal structure and ocean circulation, also found in some historic data compilations (e.g., Legler 1991; Levitus et al. 1994; Reynolds and Smiths 1995). For instance, westward trade winds in the equatorial Pacific drive the westward flowing SEC, which is weakened by nonlinearity. Also due to the winds, warm water piles up in the western Pacific and cold water is brought up in the eastern Pacific, so that thermocline slopes upward to the east. The EUC core also shoals to the east.

The seasonal cycle from the TAO data is well defined, consistent with other climatologies, all of which are compiled from much longer records except the Reverdin surface currents. Along the equator, the normally westward surface zonal current (i.e., SEC) reverses its direction with a significant magnitude in the eastern and central Pacific in boreal spring and early summer. This seasonal variation in the zonal surface current propagates westward, as do seasonal variations in the EUC and zonal surface winds at the equator. At 5°N and 5°S, the seasonal variations in the 20°C isotherm and dynamic height also propagate westward. Conversely at the equator in the eastern and central Pacific, the variations in 20°C isotherm depth and dynamic height propagate eastward.

On interannual time scales, the zonal wind anomaly is largest in the western and central Pacific and also zonally coherent at the equator, but about

180° out of phase with the anomaly in the eastern Pacific. SST anomaly is significant in the entire region 165°E-110°. Interannual variations in thermocline depth and dynamic height tend to propagate eastward with most of the phase shift occurring in the central Pacific. The SST anomaly is highly correlated with anomalous changes of thermocline depth in the eastern Pacific.

Chapter 4

ZONAL MOMENTUM ANALYSIS OF SEASONAL AND INTERANNUAL VARIABILITY IN THE EQUATORIAL PACIFIC

4.1. Introduction

In Chapter 3,we presented seasonal and interannual variability in both thermal structure and ocean circulation in the equatorial Pacific. Seasonal variations in the 20°C isotherm depth and dynamic height at ±5° propagate westward, which was interpreted as long Rossby waves by linear dynamics in some previous studies (e.g., Meyers 1979b; Kessler 1990). Phase features of seasonal variations in thermal structure at the equator were also explained in terms of linear equatorial waves (Meyers 1979a; Kessler and McCreary 1993; Minobe and Takeuchi 1995). Seasonal variations of zonal currents in our analysis show westward propagation along the equator, which can also be related to linear equatorial waves. On the other hand, interannual variability in our analysis in Chapter 3 does not show clear phase propagation, that can be directly related to equatorial waves. While a few empirical studies stated that equatorial waves were evident (e.g., Picaut and Delcroix 1995; Boulanger and Menkes 1995), some model results indicated that the Sverdrup momentum balance, i.e., $P_x = \tau_0^x$, is approximately valid in the equatorial Pacific on interannual time scales (e.g., Schneider et al. 1995).

One empirical approach to investigate whether the equatorial Pacific ocean is in equilibrium with the surface forcing or not is to analyze the momentum balance using observational data. McPhaden and Taft (1988) diagnosed the zonal momentum balance for seasonal variability in the eastern Pacific using time series of about three years from the Equatorial Pacific Ocean Climate Studies (EPOCS) Program (Hayes et al. 1986) and TROPIC HEAT Program (Eriksen 1985). They concluded that for the depth integrated momentum balance both local acceleration and nonlinearity were not significant, but uncertainty in their analysis was large because of short time series. In this chapter, we expand on their analyses to incorporate much longer time series of moored data from the TAO array. Derivation of a diagnostic equation basically follows their methodology. We derive the diagnostic zonal momentum equation in Section 4.2, then present the results of the diagnosis in Section 4.3. The neglected zonal momentum terms are briefly discussed in Section 4.4. Section 4.5 gives discussion and conclusions.

4.2. Diagnostic equation

The zonal momentum equation in the stratified ocean can be written as

$$\rho(u_t + uu_x + vu_y + wu_z - fv) + p_x = \rho(Au_z)_z + \rho\nabla\bullet(K\nabla u) \qquad (4.1)$$

where u, v and w are zonal, meridional and vertical velocity components, p is pressure, ρ is density, f is the local vertical component of earth rotation rate, A and K are vertical and horizontal eddy viscosity, respectively. The sign ∇ is a horizontal gradient operator.

The Coriolis term in (4.1) can be neglected since $f = 0$ on the equator.

Estimating the vertical friction term is difficult due to vertically varying eddy viscosity and the need to compute second derivatives of zonal velocity. Second derivatives can be very noisy, particularly when estimated from data at only 7-8 depths in the upper 250 m as in most of the TAO velocity data. To avoid this problem, we focus on the vertically integrated zonal momentum equation, although we examine the vertical structures of those terms whose vertical profiles can be reasonably well estimated. Integrating (4.1) from the sea surface to some depth $-H$, we have

$$\rho(\overline{u_t} + \overline{uu_x} + \overline{vu_y} + \overline{wu_z}) + \overline{p_x} = \tau_0^{(x)} - \tau_{-H}^{(x)} + \rho\nabla\bullet(\overline{K\nabla u}) \tag{4.2}$$

where τ_0^x and τ_{-H}^x are zonal surface wind stress and zonal shear stress at $-H$, respectively, and the overbar over a term denotes vertical integral of the term in (4.1), e.g.,

$$\overline{q} = \int_{-H}^{0} q\, dz,$$

where q can be any oceanic variable in (4.1). The term τ_{-H}^x is estimated as

$$\tau_{-H}^x = \rho A u_z/_{-H} \tag{4.3}$$

Based on data availability, H was set at 250 m in the western Pacific and 200 m in the central and eastern Pacific. When momentum terms are averaged zonally within a zonal region, the depth of 200 m was used if any longitudes of this region do not have data available at 250 m. There are no velocity data at the sea surface,

so the velocity data at 10 m are linearly extrapolated to the surface in order to calculate depth integrals over 200 m or 250 m consistent with the depth integrated pressure gradient. The differences between velocity integrals to 0 m and 10 m are small (only a few percent) in any case.

The data from the TAO array are discrete in time, so centered finite differences are used to estimate local accelerations. The data are also discrete with coarse resolution in longitude, so to relate estimates of zonal gradients to other terms in (4.2), we zonally average the momentum balance to get,

$$\rho[\langle \overline{u}_t \rangle + \frac{(\overline{u^2_{x_e}} - \overline{u^2_{x_w}})}{(2\Delta x)} + \langle \overline{vu}_y \rangle + \langle \overline{wu}_z \rangle] + \frac{(\overline{p}_{x_e} - \overline{p}_{x_w})}{(\Delta x)}$$
$$= \langle \tau_0^{(x)} \rangle - \langle \tau_{-H}^{(x)} \rangle + \rho \langle \nabla \bullet (\overline{\boldsymbol{K} \nabla u}) \rangle \qquad (4.4)$$

where x_e and x_w are eastern and western longitudes, respectively, and $\Delta x = x_e - x_w$. The angle brackets denote the average over these two longitudes, e.g.,

$$\langle q_t \rangle = \frac{1}{\Delta x} \int_{x_w}^{x_e} q_t \, dx,$$

where q is any oceanic variable in (4.3). Eq. (4.4) will be used to estimate zonal momentum terms by using all available time series of TAO observations over intervals of 165°E-170°W 170°W-140°W and 140°W-110°W. Before computing these terms, the data are processed to 5-day averages to filter out variability that is due to short time scale, spatially incoherent fluctuations like inertia-gravity waves and small scale locally wind forced variations. Then the mean, seasonal and interannual variations in momentum terms are obtained through procedures

described in Section 2.2.

In (4.4), zonal averaged gradients, $(\overline{u_{x_e}^2}-\overline{u_{x_w}^2})/(2\Delta x)$ and $(\overline{p_{x_e}}-\overline{p_{x_w}})/(\Delta x)$, are exact, but averaging other terms zonally (i.e., $\rho<\overline{u_t}>$, $<\tau_0^x>$ and $<\tau_{-H}^x>$) may cause errors because of coarse zonal resolution. For example, velocity data at only two end points are available for estimating $\rho<\overline{u_t}>$ at adjacent current meter mooring locations. On seasonal time scales, u and τ_0^x propagate westward with a phase speed of about 0.8 m s^{-1} in the central and eastern Pacific as shown in Chapter 3. Using the formulae for a pure progressive sine wave in Bryden (1977), zonal averages of $\rho<\overline{u_t}>$, $<\tau_0^x>$ and $<\tau_{-H}^x>$ over 30° intervals would be lower than their true values by 8%. This error is not large enough to affect our analysis. We will show later that on interannual time scales, u_t is much smaller than surface stress across the entire region over 165°E-110°W, so the errors in $\rho<\overline{u_t}>$ will not affect the main conclusions of this study. Zonal coherence for interannual variations in winds is in general sufficiently high that zonal averages based on TAO mooring data should be reasonably representative of true zonal averages (Fig. 3.12; also see McPhaden and Taft 1988; Kessler et al. 1996).

Our data do not allow estimation of the nonlinear term $\rho<\overline{vu}_y>$ or $\rho<\overline{wu}_z>$. Nor can the horizontal diffusion $\rho<\nabla\bullet(\overline{\boldsymbol{K}\nabla u})>$ be reliably estimated. But this diffusion term has zonal scales much smaller than the scales of the three zonal averaging intervals, so it is negligible when averaged zonally in these intervals. Larger zonal scale turbulence due to tropical instability waves is included in meridional advection, $\rho v u_y$. In the next section we will concentrate on those terms that we can directly estimate, and return to the issue of neglected nonlinearity in Section 4.4.

4.3. Results

4.3.1. Time means

Fig. 4.1 shows vertical profiles of mean $<p_x>$ and $\rho<uu_x>$ in intervals 165°E-170°W, 170°W-140°W and 140°W-110°W. Mean local acceleration $\rho<u_t>$ is not statistically different from zero and is not shown. Consistent with a weak mean surface stress in the western Pacific, the mean zonal pressure gradient is weaker in the western Pacific than in the central and eastern Pacific. In all these three intervals, $<p_x>$ decreases with depth and is larger than zonal advection $\rho<uu_x>$ in the upper 100 m. Below 100 m, $\rho<uu_x>$ at some locations may be as large as $<p_x>$ (e.g. 110°-140°W and 170°W-165°E).

Table 4.1 lists mean $<\tau_0^x>$, $<\tau_{-H}^x>$ and vertical integrals of $<p_x>$ and $\rho<uu_x>$ in these three zonal intervals. These results indicate that surface wind forcing is mainly balanced by zonal pressure gradient. Specifically, to within the errors of the analysis, vertically integrated pressure gradient $<\overline{p_x}>$ and $<\tau_0^x>$ are not distinguishable with 90% confidence. The strongest mean zonal wind stresses and pressure gradients are between 110°W and 170°W, with both diminishing between 170°W and 165°E. Zonal advection is about 20% at most of either $<p_x>$ or $<\tau_0^x>$.

Assuming A = 1.2 $cm^2\ s^{-1}$ in (4.3) as derived in Munk (1966), zonal shear stress at 250 m in the western Pacific and at 200 m in the central and eastern Pacific is at least about 40 times smaller than the other terms (Table 4.1). So, though A is subject to considerable uncertainty (Peters et al. 1988), this term is negligible compared to the other terms.

Therefore, we would conclude from this analysis that on a long term average, the equatorial Pacific ocean is basically in Sverdrup momentum balance

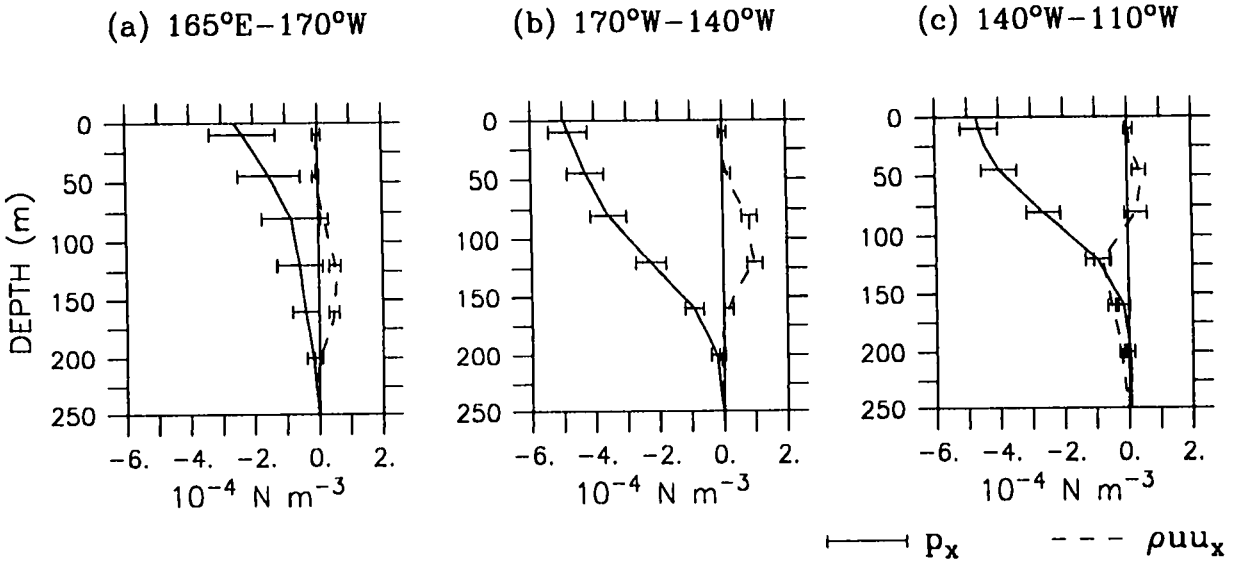


Fig. 4.1. Vertical profiles of mean zonal pressure gradient (solid) and zonal advection (dashed) (*10^{-4} N m^{-3}). Horizontal bars indicate 90% confidence intervals.

Table 4.1. Mean values of terms in the vertically integrated momentum balance. The integral is over 0-250 m for 165°E -170°W and over 0-200 m for 170 W-140 W and 140°W-110°W. Ninety percent confidence interval is indicated for each estimate. Units are $*10^{-2}$ N m^{-2}.

	$<\tau_0^x>$	$<\overline{p}_x>$	$\rho<\overline{uu}_z>$	$<\tau_{-H}^x>$
165°E -170°W	-2.36 ± 0.76	-2.11 ± 1.13	0.43 ± 0.13	0.05 ± 0.01
170°W-140°W	-5.08 ± 0.63	-5.54 ± 0.50	0.73 ± 0.14	0.04 ± 0.01
140°W-110°W	-5.11 ± 0.51	-4.18 ± 0.49	-0.37 ± 0.15	0.06 ± 0.01

with the zonal wind stress. The inexact balance between the mean stress and pressure gradient in Table 4.1 is due to relatively weak depth integrated nonlinear terms 100 m. Below 100 m, $\rho\langle uu_x\rangle$ at some locations may be as large as $\langle p_x\rangle$ (e.g. 110°-140°W and 170°W-165°E) in (4.4) which we have not been able to estimate, plus any bias errors in our analysis.

4.3.2 Seasonal fluctuations

Fig. 4.2 shows the depth-time contours of seasonal variations in zonal pressure gradient $\langle p_x\rangle$, local acceleration $\rho\langle u_t\rangle$ and zonal advection $\rho\langle uu_x\rangle$ in the intervals 165°E-170°W, 170°W-140°W and 140°W-110°W. In these three zonal intervals, $\langle p_x\rangle$ decreases in amplitude with depth and is approximately in phase in the upper 150 m. Local acceleration, which is also vertically coherent most the time of year, is comparable in amplitude to pressure gradient throughout the upper 150 m in all three zonal intervals. Like zonal velocity itself, the acceleration has the maximum below the surface but above the EUC core which is located at depths of about 100 m in the eastern Pacific and about 180 m in the western Pacific as shown in Fig. 3.2. Significant local accelerations indicate that on seasonal time scales, ocean variations are not in equilibrium with surface forcing. Zonal advection is generally smaller in amplitude than $\langle p_x\rangle$ and $\rho\langle u_t\rangle$ and has a less uniform vertical structure. It is comparable to either of the latter terms only around the depth of the EUC core.

The residuals, $\langle p_x\rangle + \rho\langle u_t\rangle + \rho\langle uu_x\rangle$, are shown in the bottom panels of Fig. 4.2. These residuals include vertical stress gradients and other nonlinear terms, $-(\langle vu_y\rangle + \langle wu_z\rangle)$ as shown in Eq. (4.4), in addition to errors due to sampling problems and estimation of the momentum terms. The residuals in the

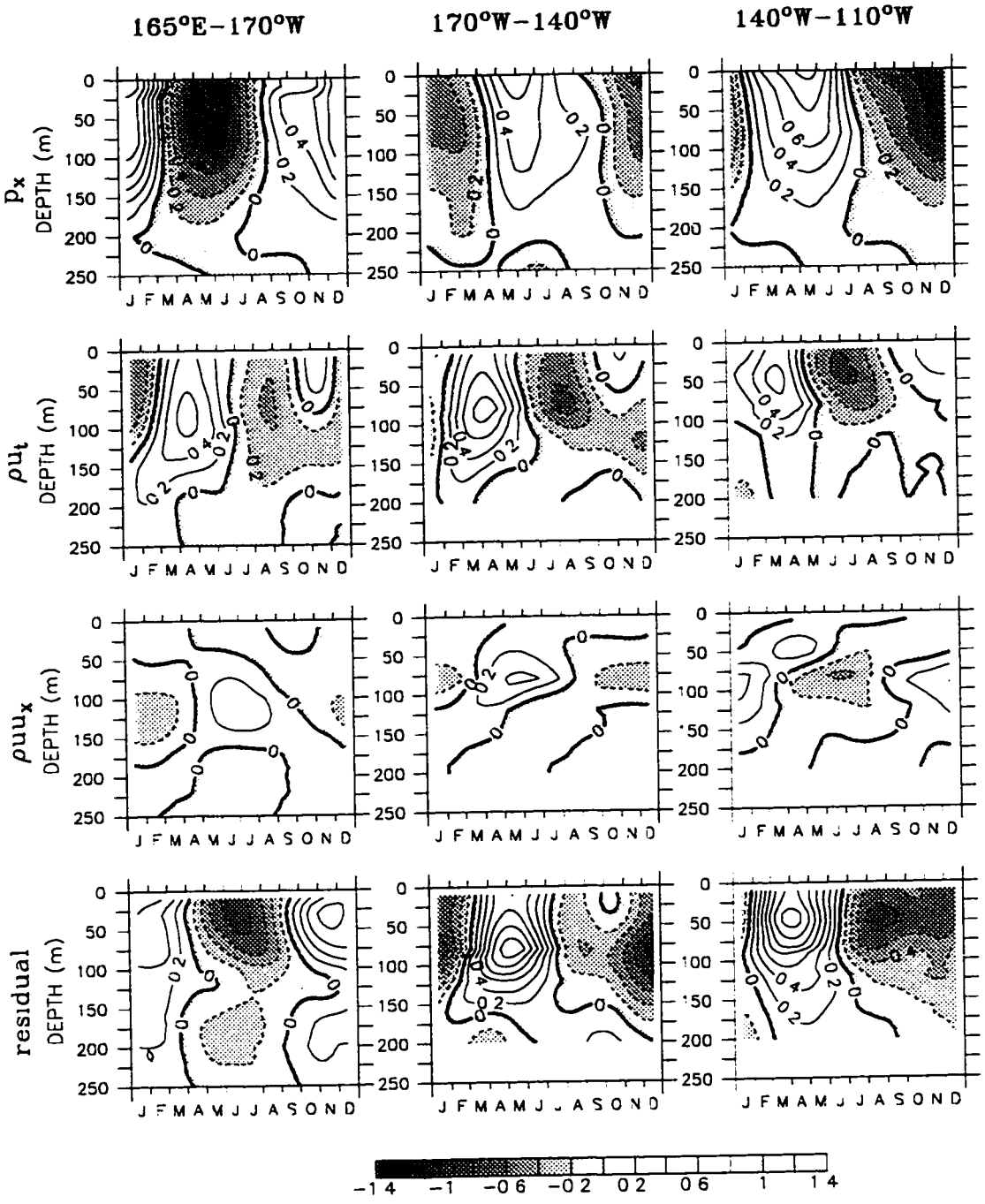


Fig. 4.2. Depth-time contours of seasonal fluctuations in zonally averaged zonal momentum terms. upper panels: zonal pressure gradient, $\langle p_x \rangle$; upper middle panels: local acceleration, $\rho\langle u_t \rangle$; lower middle panels: zonal advection, $\rho\langle uu_x \rangle$; lower panels: residual = $\langle p_x \rangle + \rho\langle u_t \rangle + \rho\langle uu_x \rangle$. Intervals are $0.2*10^{-4}$ N m^{-3}.

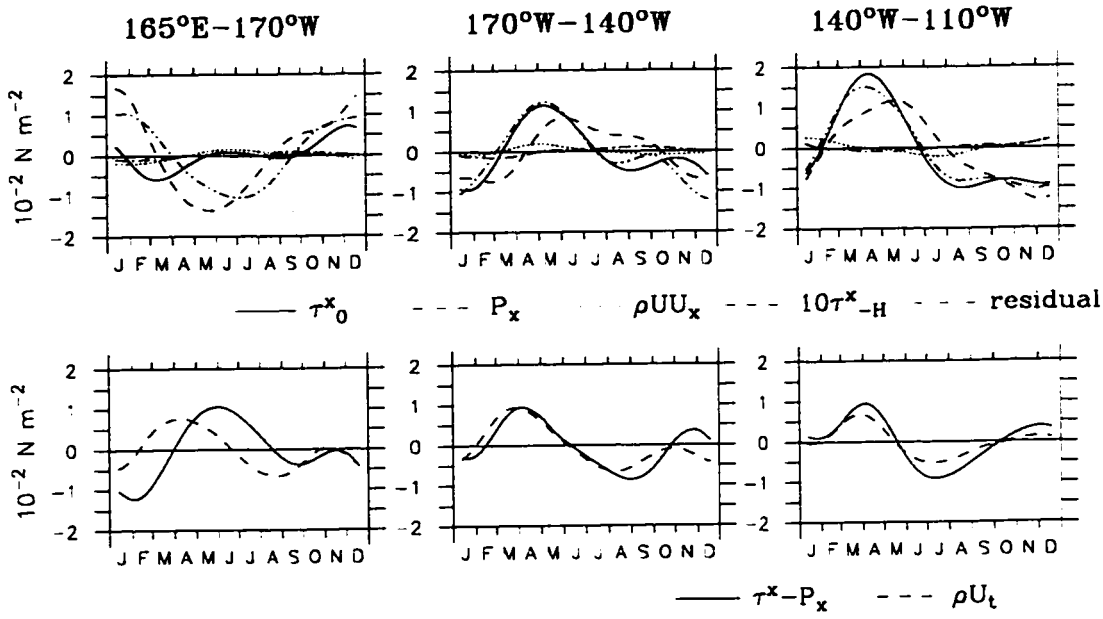


Fig. 4.3. Upper panels: seasonal variations in zonal wind stress $<\tau_0^x>$ (solid), depth-integrated pressure gradient $<\overline{p}_x>$ (Short dashed) and zonal advection $\rho<\overline{uu}_x>$ (dotted), vertical shear stress, $<\tau_{-H}^x>$ (dot-dashed) and depth-integrated residua defined in Fig. 4.2 (double dot-dashed). The H for vertical integral and τ_{-H}^x is set 250 m in 165°E-170°W and 200 m in 170°W-140°W and 140°W-110°W, respectively. Lower panels: τ_0^x - $<\overline{p}_x>$ (solid) and $\rho<\overline{u}_t>$ (dashed). Units are $*10^{-2}$ N m^{-2}.

Table 4.2. Standard deviation of terms in vertically integrated zonal momentum balance for seasonal and interannual variability in three longitudinal intervals. The integral is over 0-250 m for 165°E -170°W and 0-200 m for 170°W-140°W and 140°W-110°W. Units are $*10^{-2}$ N m^{-2}.

a. Seasonal

	$<\tau_0^x>$	$<\overline{p_x}>$	$\rho<\overline{u_t}>$	$\rho<\overline{uu_x}>$	$<\tau_{-H}^x>$
165°E-170°W	0.34	0.95	0.46	0.11	<0.01
170°W-140°W	0.62	0.55	0.49	0.09	0.01
140°W-110°W	1.01	0.82	0.35	0.14	0.01

b. Interannual

	$<\tau_0^x>$	$<\overline{p_x}>$	$\rho<\overline{u_t}>$	$\rho<\overline{uu_x}>$	$<\tau_{-H}^x>$
165°E-170°W	1.75	2.02	0.07	0.20	<0.01
170°W-140°W	1.02	1.18	0.08	0.12	<0.01
140°W-110°W	0.68	0.64	0.09	0.17	<0.01

intervals 170°W-140°W and 140°W-110°W, especially in the upper 100 m, each are strongly correlated with the local surface wind stress in three intervals (Fig. 4.3). Even semi-annual variations correspond between the residuals in the surface layer and the surface stress. So, upon vertical integration, the time series of the residuals match the surface stress very well, as shown in Fig. 4.3, suggesting that the depth-integrated nonlinear terms $\rho\langle\overline{vu_y}\rangle$ and $\rho\langle\overline{wu_z}\rangle$ be small compared with the surface wind stress. But the fact that the residuals have maxima below the sea surface may suggest that these two terms be significant at some depths. In the interval 165°E-170°W, however, the residual is not strongly correlated with the surface stress there and its vertical integral in the upper ocean does not match the stress.

Estimates of depth integrated momentum terms in (4.4) are also shown in Fig. 4.3. In the central and eastern Pacific, the vertical integral of zonal pressure gradient $\langle\overline{p_x}\rangle$ and the zonal surface stress $\langle\tau_0^x\rangle$ have similar amplitudes, but $\langle\tau_0^x\rangle$ tends to lead $\langle\overline{p_x}\rangle$ by about two months. In the western Pacific, seasonal fluctuations in $\langle p_x\rangle$ are stronger than those further to the east whereas local surface wind stress variations are relatively weak. Vertical integrals of $\rho\langle uu_x\rangle$ are small compared with vertical integrals of $\langle p_x\rangle$ or $\rho\langle u_t\rangle$. In the interval 140°W-110°W, where $\rho\langle\overline{uu_x}\rangle$ is largest, it is only about 17% of the amplitude of $\langle\overline{p_x}\rangle$ (Table 4.2).

Fig. 4.3 also shows estimates of seasonal variations in zonal shear stress, τ_{-H}^x, at 250 m in 165°E-170°W, and 200 m in 170°W-140°W and 140°W-110°W (upper panels), assuming $A = 1.2\ cm^2\ s^{-1}$. This term generally is at least 50 times smaller than the surface stress on seasonal time scales (Table 4.2), so it is negligible on these time scales and will be neglected in subsequent discussion of

seasonal variability. The relative importance of local accelerations for the mean seasonal cycle is emphasized by taking differences between zonal surface stress and zonal pressure gradient in the depth integrated momentum equation (4.4),

$$\Delta F = \tau_0^x - < \overline{p_x} >$$

and comparing them with $\rho<\overline{u_t}>$ (Fig. 4.3, lower panels). In the intervals 110°W-140°W and 140°W-170°W, ΔF is in phase with $\rho<\overline{u_t}>$, and nearly of the same amplitude. These results suggest that the imbalance between zonal pressure gradient and wind stress between 110°W and 170°W goes mainly into accelerating zonal currents. Therefore, linear dynamics may well describe the seasonal cycle to zeroth order, with nonlinearity playing a secondary role in seasonally evolving zonal currents in the eastern and central Pacific.

In the interval between 165°E-170°W, however, the ΔF does not match $\rho<\overline{u_t}>$ well in either amplitude or phase. The latter has a smaller magnitude than ΔF and leads ΔF in general by a couple of months. Accounting for zonal advection improves the balance only slightly. It is possible that neglected nonlinear terms are more significant in the western Pacific than in the eastern Pacific. On the other hand, analysis errors may be larger in the western Pacific. Averaging zonally as in (4.4) may introduce larger errors into estimates of surface stress and local acceleration in the western Pacific than in the central and eastern Pacific. Zonal stress and velocity are available only at the two extremes of the 165°E-170°W interval, and at these two longitudes, the stress tends to be out of phase for the seasonal cycle (Figs. 3.4 and 3.5). Also, discrepancy in zonal pressure gradient may be larger in the western Pacific than in the central or eastern Pacific due to

two factors. First, the mean T-S relationship we use to estimate dynamic height may, due to relatively large fluctuations in salinity, cause more discrepancy in the seasonally varying pressure gradients in the western Pacific than in the eastern Pacific (Busalacchi et al. 1994; Cronin and McPhaden 1997). Second, the reference level of 500 m may not be deep enough in the western Pacific, where active upper ocean is deeper than farther east, to capture all the variability in dynamic height in the vertical integral of zonal momentum balance. However, there are no adequate observations to quantitatively estimate how significantly zonal averaging, mean T-S relation and shallow reference level each affect the accuracy of our zonal momentum analysis for the seasonal variability.

In summary, between 110°W and 170°W, zonal surface stress and pressure gradient are the two main forces in the depth integrated zonal momentum balance for seasonal variations. Their difference drives the prominent accelerations and decelerations of seasonally varying zonal currents. The fact that these linear terms balance each other between 110°W and 170°W suggests that seasonal variability in the equatorial circulation can be approximated by linear dynamics, with nonlinear terms being of secondary importance. The seasonal balance is less clear between 165°E and 170°W, perhaps due to errors in the analysis and/or the terms that we are not able to estimate.

4.3.3. Interannual fluctuations

Fig. 4.4 shows the depth-time contours of interannual variations in zonal pressure gradient $\langle p_x \rangle$, local acceleration $\rho\langle u_t \rangle$ and zonal advection $\rho\langle uu_x \rangle$ over 165°E-170°W and 140°W-110°W. Fig. 4.5 shows vertical integrals of these terms

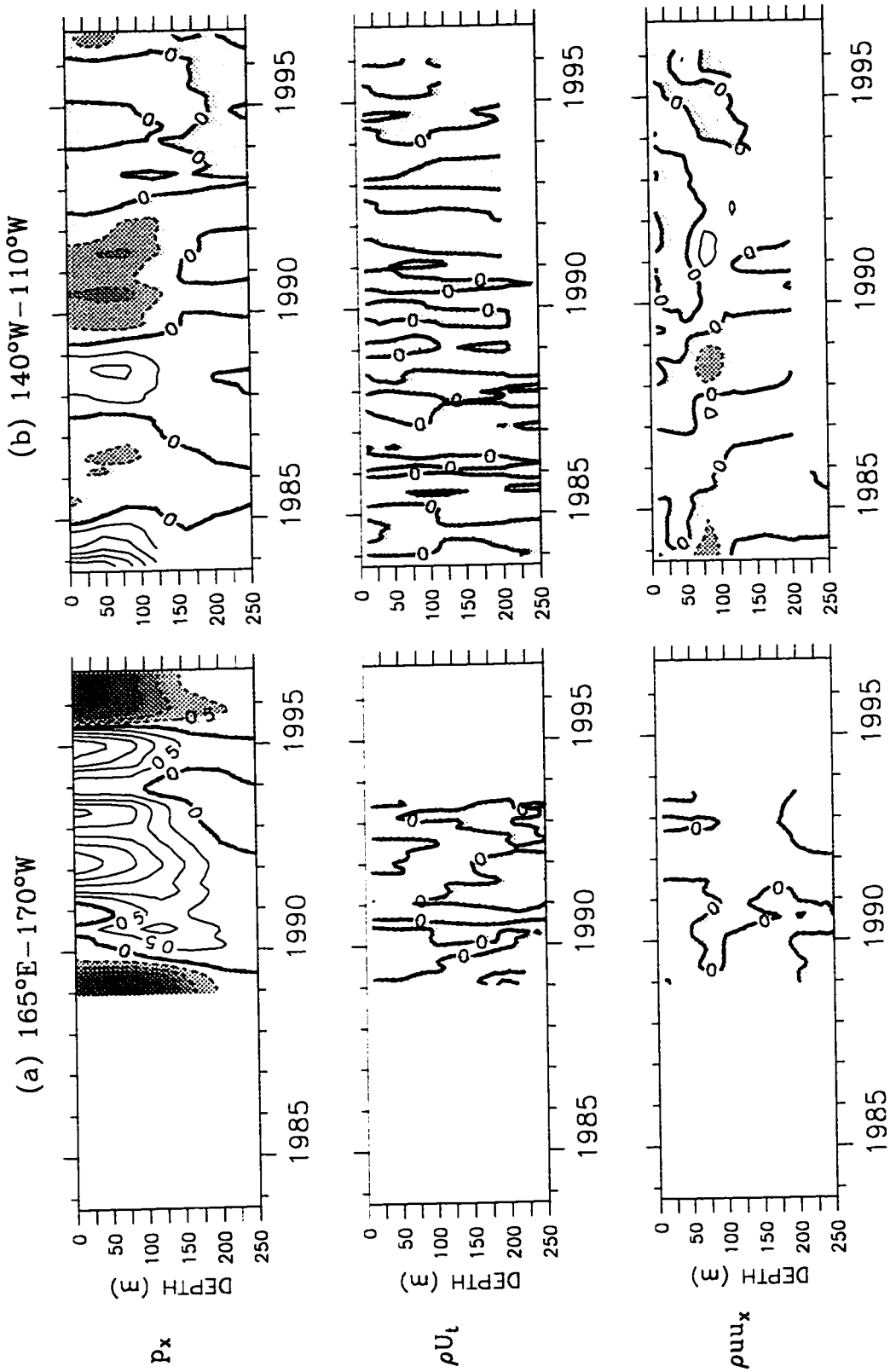


Fig. 4.4. Depth-time contours of interannual fluctuations in zonally averaged zonal momentum terms. upper panels: $<p_x>$; middle panels: $\rho<u_t>$; lower panels: $\rho<uu_x>$. Intervals are $0.5*10^{-4}$ N m^{-3}.

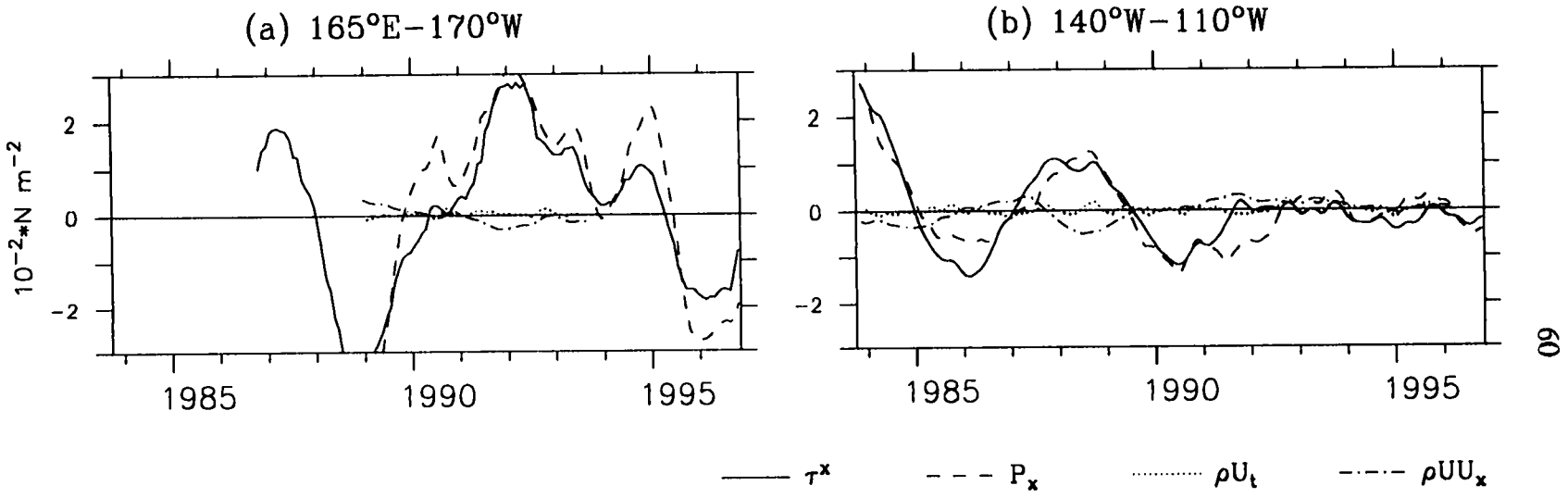


Fig. 4.5. Interannual fluctuations of zonal momentum terms in (4.4): zonal wind stress (solid), zonal pressure gradient (dashed), local acceleration (dotted) and zonal advection (dot-dashed). The depth integral is over the upper 250 m in 165°E-170°W and 200 m in 140°W-110°W, respectively. Units are $*10^{-2}$ N m^{-2} .

Table 4.3. Cross correlation between zonal wind stress and depth integrated pressure gradient in three zonal intervals on interannual time scales. The vertical integral is over 0-250 m for 165°E-170°W and 0-200 m for 170°W-140°W and 140°W-110 W. The calculation is over the indicated times of data availability. The 90% significance levels for the correlations are estimated, using the method described in Davis (1976).

Interval	165°E-170°W	170°W-140°W	140°W-110°W
Data availability	12/88-10/96	12/88-10/96	11/83-10/96
Cross correlation	0.93	0.88	0.80
90% significance	0.80	0.78	0.66

and surface wind stress $<\tau_0^x>$. We have not shown variations in $<\tau_{.H}^x>$ in Fig. 4.5 for the sake of clarity; this term is much smaller than the dominant terms in the momentum balance on interannual time scales as we found earlier for the seasonal cycle (Table 4.2). Also the interval over 170°W-140°W is not included in Figs. 4.4 and 4.5 because the major points can be illustrated with just eastern and western Pacific time series. As for the seasonal cycle, the mean T-S relationship used for estimates of dynamic height may produce larger errors in the western Pacific than in the eastern Pacific. However, interannual variations in zonal wind stress and induced pressure gradient are much larger than for seasonal variations in the western Pacific (Figs. 4.3 and 4.5), so the signal-to-noise ratio is much higher for interannual variability. Therefore, the effect of errors resulting from use of a mean T-S relationship for dynamic height calculations may not be as severe for interannual variations as for seasonal variations in the western Pacific.

The zonal pressure gradient is highly coherent in the upper 100 m in the intervals 165°E-170°W and 140°W-110°W (Fig. 4.4). It is also in phase with $<\tau_0^x>$, unlike for the seasonal cycle. The zonal wind stress and vertical integral of zonal pressure gradient have similar amplitudes (Fig. 4.5; Table 4.2) and are significantly correlated at the 90% confidence level (Table 4.3). Local accelerations are much smaller than the zonal pressure gradient in the upper 150 m. Moreover, vertically integrated accelerations are at least one order of magnitude smaller than either vertically integrated pressure gradient or zonal wind stress (Fig. 4.5; Table 4.2). Therefore, in contrast to variations on seasonal time scales, the ocean is in a state of quasi-equilibrium with the surface forcing on interannual time scales.

In the interval of 140°W-110°W, interannual variations of zonal advection

are in general significantly smaller than those in pressure gradient except around EUC core depths of about 100 m (Fig. 4.4). Zonal advection is also not as vertically coherent as $<p_x>$, so upon vertical integration, its amplitude is further reduced relative to the amplitude of either $<\overline{p}_x>$ or $<\tau_0^x>$ (Fig. 4.5). As shown in Table 4.2, the former is only about 25% of the latter. West of 140°W, the amplitude of $\rho<\overline{uu}_x>$ is even smaller compared to the pressure gradient and wind stress. These results indicate that there is an approximate balance between surface wind stress and vertically integrated pressure gradient across the entire region of 165°E-110°W and that the upper ocean responds to surface wind forcing as a succession of quasi-equilibrium states in Sverdrup momentum balance on interannual time scales.

Dynamic theories, both linear and nonlinear, predict that if ocean responds to wind forcing as a succession of quasi-equilibrium states, the strength of the zonal currents, particularly the EUC, will be in phase with variations in wind forcing and zonal pressure gradient (Charney and Spiegel 1971; McPhaden 1981; McCreary 1981). The linear theories further predict that maximum zonal currents in the thermocline are found east of the maximum zonal wind stress and pressure gradient because of damped Kelvin wave radiation (McPhaden 1981; McCreary 1981). To examine this hypothesis that zonal currents in the thermocline should vary in phase with the zonal pressure gradient, we present time series of vertically integrated zonal pressure gradient and zonal transport per unit width in the thermocline (80-200 m) at the end longitudes of the three zonal averaging intervals (Fig. 4.6). Consistent with linear theory, cross correlations between zonal pressure gradients and zonal transports at the east end longitudes are significantly nonzero with maxima at zero or near zero lag (Table 4.4). These correlations are also

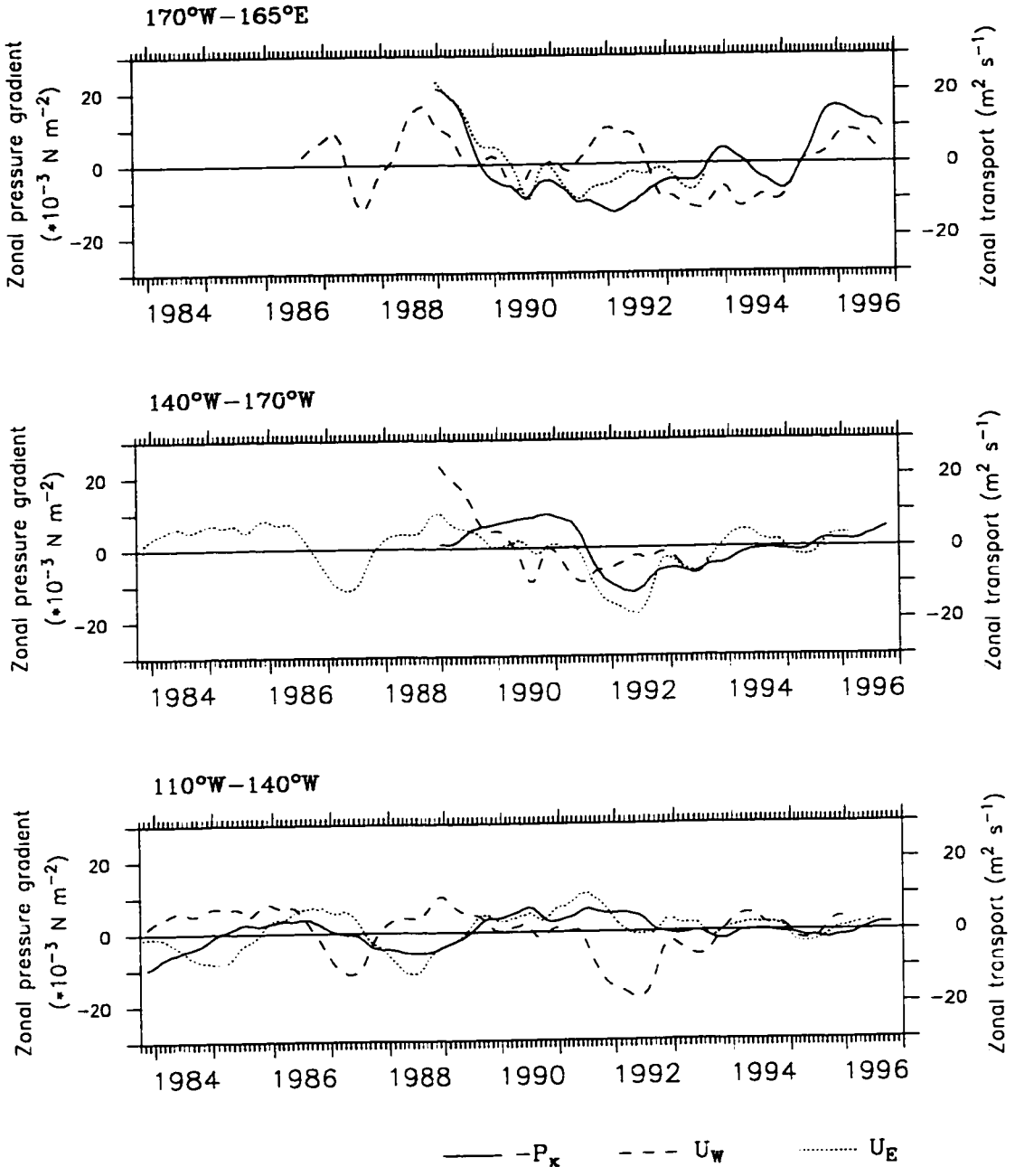


Fig. 4.6. Time series of vertically integrated zonal pressure gradient (solid) and zonal transport per unit width (dashed: west endpoint; dotted: east endpoint) on interannual time scales. Pressure gradients and transports are vertically integrated over depths of 80 m to 200 m. Units for the pressure gradient and transport are $*10^{-3}$ N m^{-2} and m^2 s^{-1}, respectively.

Table 4.4. Cross-correlation coefficients at zero lag between vertically integrated zonal pressure gradient and zonal transport on interannual time scales. Pressure gradients and transports are vertically integrated over depths of 80 m and 200 m. Subscripts W, E and 90 mean the west endpoint, the east endpoint, and 90% significance, respectively. Values above the 90% significance are highlighted.

	r_W	r_E	r_{90}
110°W-140°W	-0.21	**0.72**	0.64
140°W-170°W	0.77	**0.85**	0.81
170°W-165°E	0.83	**0.96**	0.96

consistently higher than those between zonal pressure gradient and zonal transport at the west end longitudes, the latter of which are generally not significantly different from zero with 90% confidence.

We repeated this correlation analysis to examine the relationship between zonal wind stress and zonal transport variations in the thermocline on interannual time scales. As expected, results were similar to those cited above for zonal pressure gradient and zonal transports, because of the near balance between $<\overline{p}_x>$ and $<\tau_0^x>$ along the equator. This covariability between anomalous wind stress and zonal transport at EUC depths in the thermocline is also evident in Figure 3.12. Zonal transport in the central Pacific varies in phase with the wind stress in the central and western Pacific, while the transport anomaly in the eastern Pacific is correlated with the wind stress forcing slightly to its west. Also consistent with linear theories, maximum variability in zonal transport anomalies occurs east of the maximum variability in wind forcing, as shown in standard deviations of interannual variations in the zonal wind stress and zonal transport in Fig. 3.12.

4.4. Neglected terms

We are not able to estimate nonlinear terms other than uu_x due to lack of data. However, as shown in previous studies (Philander and Pacanowski 1980; Bryden and Brady 1985; McPhaden and Taft 1988; Qiao and Weisberg 1997), the nonlinear terms uu_x and wu_z in the mean tend to cancel each other since the EUC flows upwards to the east along the isopycnals. After taking this tendency for cancellation into account, the net effect of nonlinearity is to internally redistribute zonal momentum in the upper ocean. The redistribution reduces the strength of the SEC along the equator and produces a narrower, vertically stretched EUC with

larger eastward mass transports relative to that expected from linear dynamics (McPhaden 1981; McCreary 1981). Since the nonlinear terms have smaller vertical scales than the pressure gradient term, they are reduced even further relative to the pressure gradient upon vertical integration.

We have also not been able to estimate lateral turbulent diffusion directly from our data. Near the equator, lateral diffusivity is due mainly to tropical instability waves which have their largest amplitudes between about 110°W and 140°W (Halpern et al. 1988). A major source of energy for these waves is the unstable meridional shear between the Equatorial Undercurrent and the South Equatorial Current (Hansen and Paul 1984; Luther and Johnson 1990). Bryden and Brady (1989) estimated the magnitude of the mean lateral eddy stress to be 1.6×10^{-2} N m^{-2} averaged between 110°W and 152°W, based on moored time series measurements at those two longitudes. The dominant contribution to the average stress came from the meridional flux of zonal momentum at 110°W. Estimates of mean eddy stresses derived from other published reports (e.g., Johnson and Luther 1994; Qiao and Weisberg 1997) are consistent with Bryden and Brady's results in both magnitude, and in the decrease in amplitude of the stresses between the eastern and central Pacific. Compared to our mean zonal pressure gradient and wind stress estimates between 110°W and 170°W (Table 4.2), Lateral eddy stresses computed by Bryden and Brady (1989) are 2.5 to 3 times smaller. Hence, lateral diffusion is likely to be significant in the mean zonal momentum balance of the equatorial Pacific (particulary between 110°W and 140°W), though of secondary importance relative to the mean zonal wind stress and pressure gradient.

Meridional and vertical advection and tropical instability waves are all

seasonally and interannually modulated (Philander et al. 1985). However, there are no reliable estimates of seasonal or interannual variations in the advection and lateral eddy stresses associated with these waves. Like zonal advection uu_x, these terms may not be negligible at particular depths and dismissed easily, as suggested in Section 4.3.2. For example, using monthly averaged mean climatology of surface ocean currents (Reverdin et al. 1994), scale analyses show that ρvu_y and ρwu_z have the order of 10^{-4} N m^{-3} in the surface layer, comparable to zonal pressure gradient there (Fig. 4.2). Tropical instability waves may enhance the amplitude of ρvu_y, but it is likely to remain the same order. However, these nonlinear terms may not significantly affect the vertically integrated zonal momentum balance due to two reasons. First, these terms tend to cancel each other. In the surface layer, ρvu_y is always positive due to tropical instability waves. It has the minimum (maximum) magnitude in boreal spring (fall or winter) when the instability waves are least (most) active (Harrison, personal communication), so its seasonal variation around the long-term mean contributes eastward momentum (i.e., negative ρvu_y) to local acceleration in boreal spring and westward momentum in fall or winter. The term ρwu_z, on the other hand, is usually negative in the surface layer due to Ekman transport induced upwelling. It has the minimum magnitude in boreal spring when the upwelling weakens and maximum in fall or winter, so its seasonal cycle has westward momentum in boreal spring, but eastward momentum contributing to local acceleration in fall or winter. Therefore, ρvu_y and ρwu_z tend to cancel each other in the surface layer, where ρuu_x is negligible (Fig. 4.2). Likewise, we can conclude that these three terms tend to cancel each other in deeper layers on seasonal time scales. There are not enough data to scale analyze interannual fluctuations of these momentum

terms, but the instability waves tend to weaken (strengthen) in a warm El Niño (cold La Niña) year while the upwelling weakens (strengthens). Therefore, as for seasonal time scales, meridional and vertical advection have a tendency of cancelling each other while zonal advection is small in the surface layer on interannual time scales.

The second reason is that nonlinear terms and lateral diffusion in general have smaller spatial scales than pressure gradient and wind forcing. So averaging vertically and zonally would reduce their significance relative to the latter terms.

Cancellation among these terms and due to spatial scales may not be complete. However, we would surmise that sum of the terms is likely to play a secondary role on seasonal to interannual time scales, given the generally good agreement we find between explicitly estimated terms in the zonal momentum balance equation.

4.5. Discussion and conclusions

Measurements from the TAO array have been used to analyze the zonal momentum balance in the mean, and on seasonal and interannual time scales in the equatorial Pacific. In the mean, this analysis verifies that the Sverdrup momentum balance is a very good approximation for the depth integrated mean circulation along the equator. This balance indicates that the mean easterly winds set up the eastward zonal pressure gradients, which in turn drives the EUC (Philander and Pacanowski 1980; McPhaden 1981; McCreary 1981; Wacongne 1989). Nonlinearity redistrubutes zonal momentum internally, but has vertical scales shorter than those of the zonal pressure gradient. Thus, in the depth integrated momentum balance, nonlinear terms are reduced in magnitude relative

to the dominant pressure gradient and wind stress terms. Lateral diffusion due to tropical instability waves is at its largest in the interval 110°W-140°W, but of secondary importance in the zonal momentum balance.

On seasonal time scales, the zonal momentum balance is mainly between surface stress, pressure gradient and local acceleration. The difference between surface stress and pressure gradient leads to accelerations and decelerations of the zonal currents. These accelerations indicate that wave processes are important and that on seasonal time scales the equatorial Pacific is not in equilibrium with wind forcing. Details of wave generation and propagation in the equatorial Pacific will be discussed in Chapter 5 (also see Minobe and Takeuchi 1995).

On interannual time scales, in contrast, local acceleration is very small relative to either zonal surface wind stress or pressure gradient. The surface stress is principally balanced by the pressure gradient, suggesting that, to zeroth order, the ocean response to wind forcing is characterized by a succession of quasi-equilibrium states. The fact that the Sverdrup momentum balance holds at zeroth order does not mean, however, that equatorial waves are unimportant on interannual time scales. In other words, the stress and ocean response are not in a state of exact equilibrium, but a state of quasi-equilibrium where the imbalance gives rise to the interannual variability. Although these waves are not immediately obvious from our analysis because local accelerations are small, waves are the only mechanism by which the ocean can adjust to large scale wind forcing. The adjustment is relatively fast compared to the time scale of ENSO, which dominates interannual variability in the tropical Pacific, so that explicit wave propagation is not evident.

In addition, associated with the Sverdrup momentum balance at zeroth

order is an unsatisfied western boundary condition. This condition allows for time varying, nearly zonally independent fluctuations in pressure (i.e., heat content) which are the result of very low frequency wave processes pumping heat into and out of the equatorial band on ENSO time scales (Wyrtki 1985; Cane and Zebiak 1987; Springer et al. 1990). Schneider et al. (1995) showed that these wave processes are not in exact equilibrium with surface wind stress forcing, and that the details of anomalous thermocline topography along the equator result from sum of these waves plus east-west tilting in response to zonal wind stress variations. The Sverdrup momentum balance explains that part of the solution that relates wind stress variations to zonal tilts of the thermocline, but does not explain variations in zonal mean depth of the thermocline across the basin on interannual time scales (Cane 1992).

Chapter 5

WAVE DYNAMICS OF THE SEASONAL CYCLE

5.1 Introduction

As shown in Chapter 3, seasonal variability in the 20°C isotherm depth and surface dynamic height at both 5°N and 5°S propagate westward with a phase speed of about 1 m s^{-1}. Equatorial zonal currents also propagate to the west with a similar phase speed. In contrast, variability in 20°C isotherm depth and dynamic height along the equator propagates eastward with phase speeds of about 0.5 - 0.7 m s^{-1} in the eastern and central Pacific. All these phase features suggest that linear dynamics may be a good approximation for the seasonal variations. In Chapter 4, through analysis of zonal momentum balance, we conclude that nonliearity may be significant at some depths, but not significant in the vertically integrated zonal momentum balance. Therefore, the nonlinearity may distort the spatial structures of gravest mode waves, which have much larger spatical scales than the nonlinearity, but the linear wave dynamics is capable of capturing main phase features of these waves.

In some previous studies (Meyers 1979b; Kessler 1990), westward propagation in thermal structure on seasonal time scales at ±5° was interpreted as first baroclinic mode mid-latitude (i.e. quasi-geostrophic) Rossby waves. The quasi-geostrophic approximation has the advantage of simplifying the diagnostic equations for upper ocean thermal structure, in that the vorticity balance can be

localized and evaluated at a single latitude. However, at low latitudes, scale analysis indicates that quasi-geostrophy is not strictly valid, as the Coriolis force becomes vanishingly small near the equator. Linear theory predicts that fluctuations on time scales longer than about one day become trapped in an equatorial wave guide, and that within this wave guide variability is coherent across a range of latitudes. Consistent with this theory, recent studies of seasonal variability equatorial Pacific (e.g., Kessler and McCreary 1993; Minobe and Takeuchi 1995) have been successful in the application of equatorial wave principles to the interpretation of model results and ocean observations.

Our observations of coherent westward propagating velocity and temperature structures between 5°N and 5°S likewise argue for an interpretation in terms of equatorial wave dynamics. Specifically, it is hypothesized that much of the observed westward propagation in ocean variables between 5°N and 5°S is due to long, nondispersive equatorial Rossby waves. Other equatorial wave modes must be responsible, however, for eastward propagation of thermocline depth observed along the equator. The only logical candidate at low frequency is the equatorial Kelvin wave. We can exclude short Rossby waves and mixed Rossby gravity waves from consideration because their phase speeds are westward and their zonal wavelengths are on the order of hundreds, rather than thousands, of kilometers.

Observed eastward phase speeds in 20°C depth and dynamic height along the equator do not correspond to low baroclinic mode free Kelvin wave speeds based on linear theory, as will be elaborated on below. It is found that these phase speeds result in part from a mix of Kelvin and long, nondispersive Rossby wave modes. In addition, wind forcing is prevalent along the entire zonal extent of the

basin. Under these circumstances, variability at a particular location will result from both local and remote forcing.

In this chapter, a simple model is presented to simulate wind forced Kelvin and Rossby wave variability based on the equatorial wave characteristic rays. The model was first developed by Gill and Clarke (1974), using only winds and mean ocean density structure as input. Recently, Minobe and Takeuchi (1995) used this type of model to simulate seasonal wave variability in the equatorial Pacific. Like them, we will conclude that the two gravest baroclinic mode Kelvin waves and the first meridional mode Rossby waves of the first two baroclinic modes dominate at the seasonal cycle within 5°of the equator. In addition, using this formalism we quantitatively demonstrate why zonal currents propagate westward, but thermal structure tends to propagate eastward along the equator in the eastern and central Pacific.

We derive the simple dynamical model in Section 5.2. Then model results and comparison with the observation are presented in Section 5.3, followed by discussion and conclusions in Section 5.4.

5.2. Model for estimates of Kelvin and Rossby wave variability

Following McCreary (1981), the momentum and continuity equations for a vertical baroclinic mode are

$$(\frac{\partial u_n}{\partial t} + \frac{A}{c_n^2}u_n) - \beta y v_n + \frac{1}{\rho}\frac{\partial p_n}{\partial x} = \frac{\tau^x}{\rho \int_{-H}^{0} \psi_n^2 dz} \tag{5.1a}$$

$$\beta y u_n + \frac{1}{\rho}\frac{\partial p_n}{\partial y} = \frac{\tau^y}{\rho \int_{-H}^{0} \psi_n^2 dz} \tag{5.1b}$$

$$(\frac{\partial}{\partial t} + \frac{A}{c_n^2})\frac{p_n}{\rho c_n^2} + \frac{\partial u_n}{\partial x} + \frac{\partial v_n}{\partial y} = 0 \tag{5.1c}$$

where u and v are zonal and meridional velocities, p is kinematic pressure (actual pressure divided by density), c_n is the eigenvalue for the *n-th* vertical mode, A is a constant related to the vertical friction and diffusion, H is the full depth of the ocean, (τ^x, τ^y) is the surface wind stress and ψ_n is the *n-th* vertical structure function, determined by solution of a Sturm-Liouville eigenvalue problem. We have made long wave approximation by neglecting $(\partial_t^2 + A/c_n^2)\, v_n$ in (5.1b). This approximation filters out all equatorial waves except low frequency Kelvin waves, which propagate nondispersively eastward, and long Rossby waves, which propagate nondispersively westward.

Wind forced equatorial waves can be written as

$$u_{mn} = S_{mn}(x, z, t)\, \phi_0(y), \tag{5.2a}$$

$$p_{mn} = c_n S_{mn}(x, z, t)\, \phi_0(y); \tag{5.2b}$$

for Kelvin waves (with $m = -1$), and

$$u_{mn} = S_{mn}(x, z, t)\left(\sqrt{\frac{m}{m+1}}\phi_{m+1}(y) - \phi_{m-1}(y)\right), \tag{5.3a}$$

$$p_{mn} = c_n S_{mn}(x, z, t)\left(\sqrt{\frac{m}{m+1}}\phi_{m+1}(y) + \phi_{m-1}(y)\right), \tag{5.3b}$$

for Rossby waves (with $m \geq 1$). The index m is a meridional mode number and ϕ_m is a Hermite function of order m. The ϕ_m implicitly depend on vertical mode number n through scaling of the latitudinal coordinate, y, by the equatorial Rossby radius of deformation

$$\lambda_n = (c_n/\beta)^{1/2}. \tag{5.4}$$

The meridional structures of u and p for the Kelvin wave and first two meridional mode Rossby waves, determined by ϕ_m, are shown in Fig. 5.1.

The $S_{mn}(x, z, t)$ for a wave mode (m, n) is an integral over wind projection onto this mode, expressed as

$$S_{mn}(x, z, t) = S_{mn}[X, z, t+\frac{(X-x)}{c_n}] + \alpha_n(z)\int_X^x b_{mn}[\xi, t+\frac{(\xi-x)}{c_n}]\exp[-r_n(x-\xi)]\,d\xi \tag{5.5}$$

where X is the western (eastern) boundary for Kelvin (Rossby) waves, r_n is a

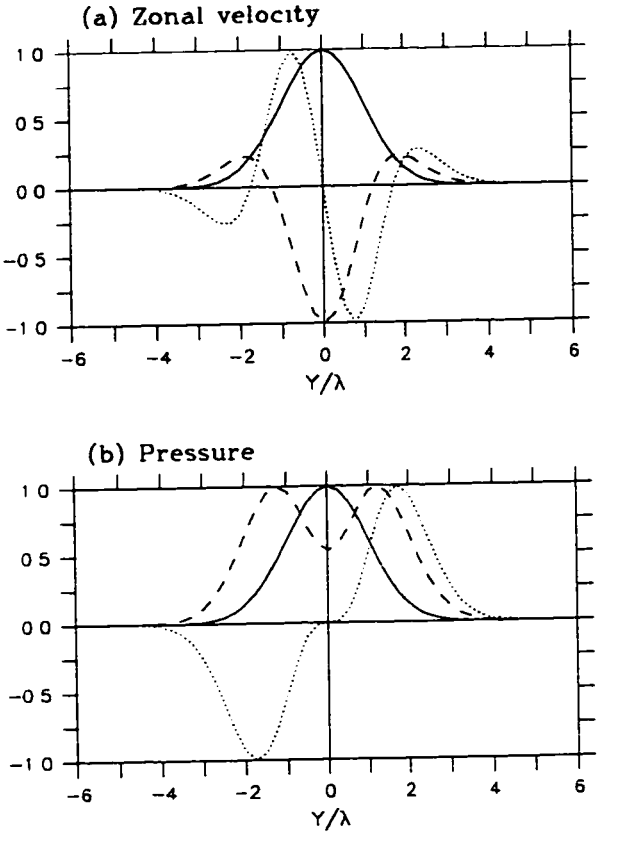


Fig. 5.1. Meridional structure of equatorial waves. Amplitudes are normalized and y-axis is non-dimensionalized by the Rossby radius of deformation given by equation (5.4). (a) Zonal velocity; (b) pressure. Solid: Kelvin wave; dashed: first meridional mode Rossby wave; dotted: second meridional mode Rossby wave.

damping coefficient, $\alpha_n(z)$ is an amplitude of coefficient, b_{mn} is the projection of the winds onto the wave mode. At low frequency limit, the b_{mn} can be expressed as (Gill and Clarke 1974)

$$b_{-1n} = \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} \frac{\Phi_0(y)\,\tau^x(x,y,t)}{\lambda_n} dy \qquad (5.6a)$$

for Kelvin waves, and

$$b_{mn} = \int_{-\infty}^{\infty} \left[\frac{\Phi_{m+1}(y)}{\sqrt{m+1}} - \frac{\Phi_{m-1}(y)}{\sqrt{m}}\right] \frac{\tau^x(x,y,t)}{\lambda_n}\, dy \qquad (5.6b)$$

for Rossby waves. The total ocean response to the wind forcing is sum of all these waves, i.e., $\sum_{n=1}^{\infty}(\sum_{m=1}^{\infty} S_{mn} + S_{-1n})$. The first term in (5.5) is the wave amplitude at the western (eastern) boundary for a Kelvin (Rossby) wave.

$S_{mn}(x, z, t)$ can be any dependent variable as specified by the coefficient $\alpha_n(z)$. For the pressure and zonal velocity, $\alpha_n(z)$ is (Kessler and McPhaden 1995), respectively,

$$\alpha_n^{\rho}(z) = \frac{\psi_n(0)\psi_n(z)}{\rho H} ,$$

$$\alpha_n^{u}(z) = \frac{\psi_n(0)\psi_n(z)}{\rho c_n H} ,$$

where ρ is water density.

In this study, the $\psi_n(z)$ and c_n are estimated from the Hawaii-Tahiti Shuttle experiment, which took place between 150°W-158°W (Wyrtki and Kilonsky 1984). The $\psi_n(z)$ were obtained with a surface mixed layer of 75 m depth. The purpose of this mixed layer is to simulate the frictional response to the surface stress as described in Zebiak and Cane (1987). The values of the first four modes are shown in Fig. 5.2 and Table 5.1, with H assumed to be 4000 m. The seasonal cycle of winds from TAO observations is available only from 165°E to 110°W, so we use linearly interpolated daily values of FSU winds to force the model. With the time step (Δt) of one day, the zonal resolution (Δx) is set at 2° to satisfy the numerical stability criterion $\Delta t < \Delta x/c_n$.

The integral in (5.6) is taken only over 13°S - 13°N due to wind data availability. However, these limits of integration are sufficient because the *e*-folding latitude of $\phi_m(y)$ is significantly less than 13° for a few gravest meridional modes.

For forced Kelvin waves, the eastward integral along the wave characteristic in (5.5) starts at 150°E, where the land effect is not significant. For forced Rossby waves, the westward integral starts at 80°W. These integration along wave characteristic paths will involve negative time (i.e., $t+\frac{(\xi-x)}{c_n} < 0$). When this occurs, twelve months will be added to the time since seasonal cycle is periodic. This approach is similar to initializing an ocean numerial model with mean climatology winds for a several years.

The drag coefficient and air density we use to convert FSU pseudostress to stress are 1.2×10^{-3} and 1.2 kg m^{-3}, respectively. The damping coefficient is a variable function of vertical mode n (Gent 1985) with specific values given by

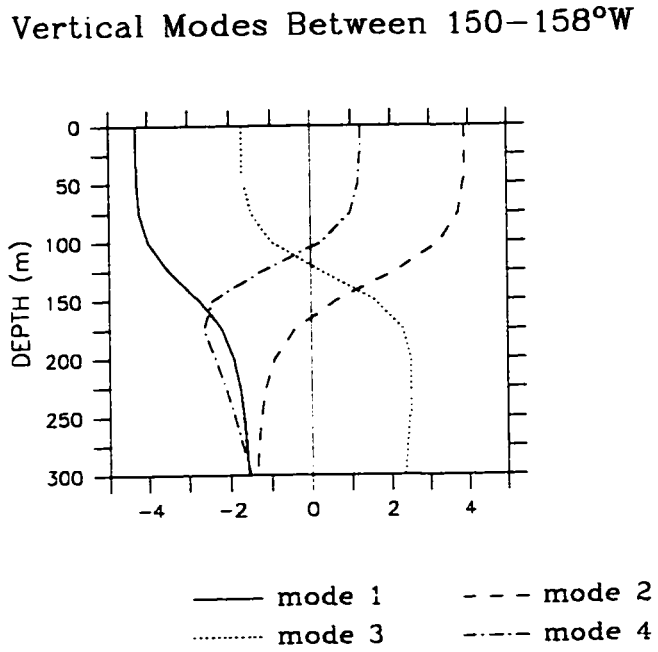


Fig. 5.2. Vertical profiles of the first four baroclinic modes for horizontal velocity components and pressure based on data from the Hawaii-Tahiti Shuttle experiment 150-158°W. Solid: first mode; dashed: second mode: dotted: third mode; dashed-dotted: fourth mode.

Table 5.1. The values of baroclinic mode parameters for the first four modes based on data from Hawaii-Tahiti Shuttle experiment 150°W - 158°W.

	Mode 1	Mode 2	Mode 3	Mode 4
Equivalent Depth H_n (cm)	76.0	30.9	11.5	5.9
Wave speed $c_n=(gH_n)^{1/2}$ (m s^{-1})	2.73	1.74	1.06	0.76
Rossby Radius $R=(c_n/\beta)^{1/2}$ (km)	346	276	216	183

$$r_n = A/c_n^2 = (c_1/c_n)\ (6\ months)^{-1}. \tag{5.7}$$

This equation yields values of $r_1 = (6\ \text{months})^{-1}$ and $r_2 = (4\ \text{months})^{-1}$, similar to those recommended by Picaut et al. (1993).

In this study, only gravest mode waves are included in the model solution in order to clearly demonstrate the basic wave dynamics in the seasonal cycle. Specifically, four gravest baroclinic modes and the first meridional mode Rossby wave for each baroclinic mode are considered in this study. How higher mode waves may improve the model simulation relative to the observations is discussed in Section 5.3.4.

In numerical model studies (Zebiak and Cane 1987; Battisti 1988; Neelin 1991; Chen et al. 1995), the surface mixed layer of about 50 m is assumed to be governed by wind-driven frictional dynamics, so the surface wind stress exerts on the layer as a body force and the layer responds simultaneously as a slab. This simple dynamics may produce a solution with zonal surface currents of reasonable amplitude compared with the observation (Chen et al. 1995) with a extremely large damping coeffient r_n. However, the in-phase relation between the simulated zonal currents from this dynamics and the surface wind stress is not observed (Fig. 3.4). In addition, Fig. 4.2 shows that zonal pressure gradient and local acceleration are comparable in magnitude to vertical stress gradient even in the surface mixed layer. Therefore, wave processes are important in adjusting the ocean response to the surface winds in this layer. By including an imbeded surface mixed layer, our simple wave model is able to produce both the remote ocean response through

equatorial waves and local ocean response.

5.3. Model results

5.3.1. Wind-forced waves

We begin by considering just forced wave solutions in the interior ocean, since previous studies (e.g., Kessler and McCreary 1993; Minobe and Takeuchi 1995) have suggested that boundary reflections are of secondary importance for the mean seasonal cycle in the equatorial Pacific. Boundary reflections will be considered explicitly in Section 5.3.2. We will find that they are weak in the longitude range sampled by the TAO array.

Fig. 5.3 shows sum of first four baroclinic mode Kelvin waves for zonal surface current and sea level along the equator from the model (top panels). Sea level has been computed by dividing surface pressure by $\rho_o g$. Both the current and sea level have consistent eastward propagation. The first and second vertical modes explain 38% and 53% of the variance of the simulated zonal velocity at the surface, respectively, while the third and fourth modes together explain only 9% of the variance of the simulated currents at the surface. For sea level, the first two baroclinic modes explain 60% and 38% of the variance of the simulation, respectively.

The sum of the four gravest vertical modes for first meridional mode Rossby waves shows consistent westward propagation along the equator (Fig. 5.3, middle panels). The first mode explains 76% of total variance in the sea level; the second explains 19% of the total variance. Similarly, the first two vertical mode Rossby waves explain 63% and 29% of the variance in zonal velocity, respectively. Therefore, for both Kelvin and Rossby waves, the two gravest

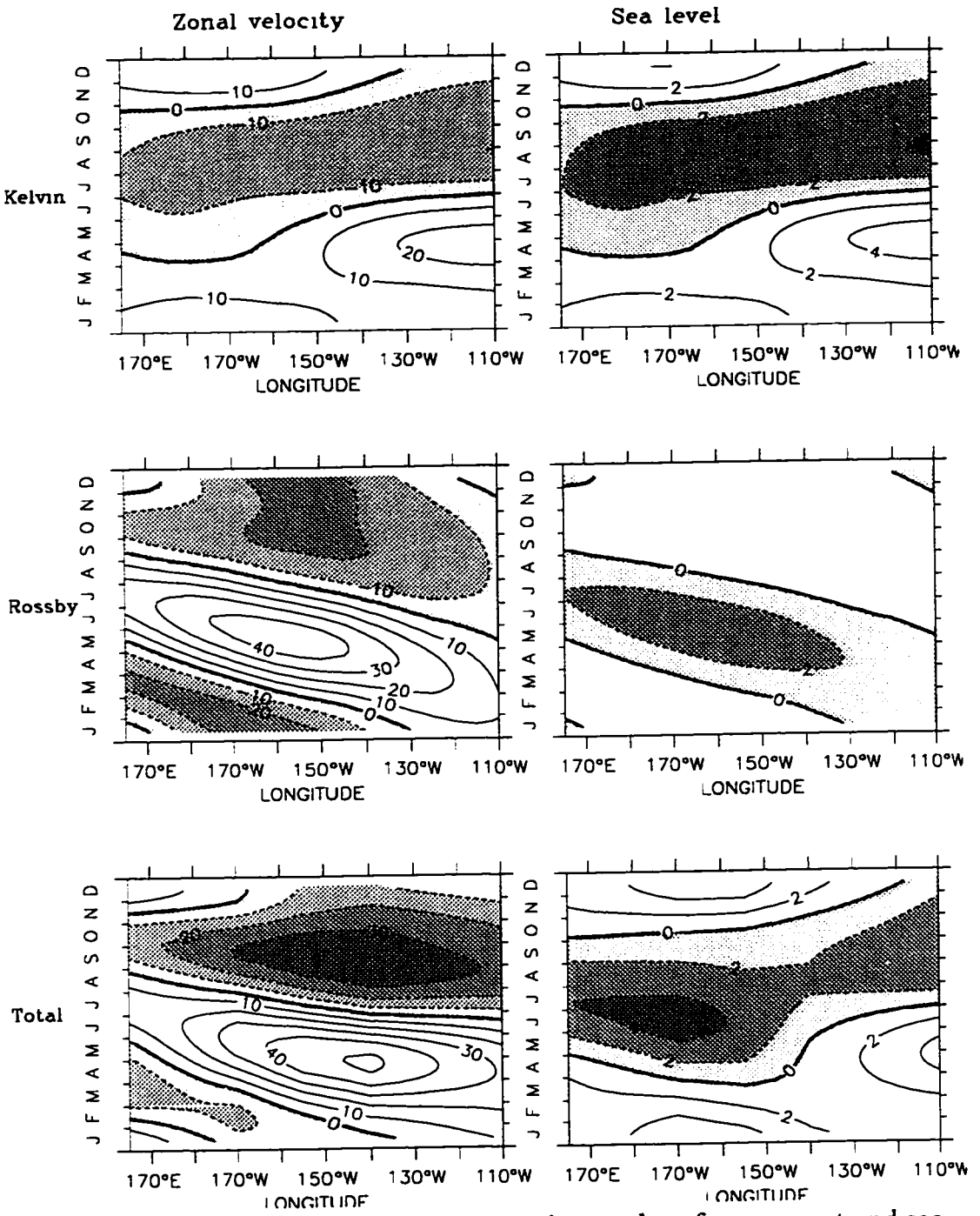


Fig. 5.3. Model Kelvin and Rossby waves in zonal surface current and sea level and their combination at the seasonal cycle. Upper panels: sum of four gravest baroclinic mode Kelvin waves; middle panels: sum of the four gravest baroclinic mode, first meridional mode Rossby waves; lower panels: combination of Kelvin and Rossby waves. Contour intervals for zonal current and sea level are 10 cm s^{-1} and 2 cm, respectively.

vertical modes are dominant in seasonal variability along the equator. For zonal velocity, Rossby waves together have larger amplitude than Kelvin waves, particularly in the central Pacific. For sea level, on the other hand, amplitude of the Kelvin waves is usually larger in the seasonal cycle.

The combination of Kelvin and Rossby waves from the model is also shown in Fig. 5.3 (bottom panels). Despite the simplicity of the model, zonal velocity and sea level are in general comparable to those in the observations (cf., Fig. 3.4c, Fig. 3.5b). Along the equator, the model exhibits westward propagation of zonal currents with a seasonal cycle similar to that observed. Sea level propagates eastward east of about 160°W and westward west of that longitude. For comparison with the phase speeds in Table 5.1, model surface zonal currents propagate westward at -0.66±0.13 m s^{-1} and model sea level propagates eastward at 0.52±0.12 m s^{-1} between 155°W and 110°W. However, there are differences between the model and the observations. For example, the modeled currents are too strong with extrema centered near 140°W rather than 110°W as in the observations. Also, the 1 cpy amplitude in sea level due to the Rossby waves, is too large in the western Pacific. These differences are expected due to the simplicity of the model. Two main dynamical processes that are not considered in the model are surface mixing and nonliearity. The surface mixing is represented by higher vertical mode waves in McCreary (1981). So neglecting higher verical modes may cause errors in surface ocean circulation. As shown in Chapter 4, nonliearity is significant around the depths of the EUC core for seasonal and interannual variability. Neglecting the nonlinearity in the model may cause discrepancy from the observation, especially around the EUC depths.

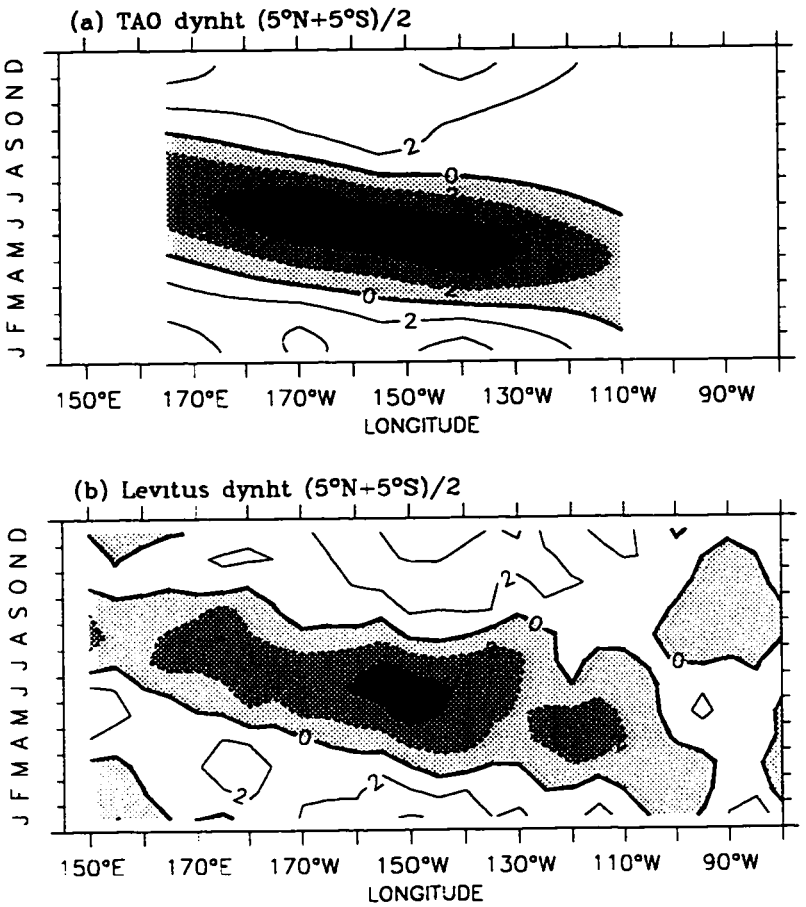


Fig. 5.4. Mean seasonal cycle of dynamic height average for 5°N and 5°S. (a) TAO observations; (b) Levitus analysis. Negative anomalies are shaded. Contour interval is 2 dyn. cm.

Observations at ±5° can also be compared with the wave model. At ±5° Kelvin wave variability is small (Fig. 5.1), so that variability at these latitudes should be dominated by Rossby waves. The average of dynamic height at 5°N and 5°S, namely

$$dyn(5^{o}) = 0.5\,[dyn(5^{o}N) + dyn(5^{o}S)], \tag{5.8}$$

is mainly due to meridionally symmetric long Rossby waves and should have similar phase features to Rossby wave variability at the equator. This is indeed the case, comparing the simulated Rossby waves in Fig. 5.3 with dyn(5°) derived from both the TAO data and Levitus analysis (Fig. 5.4).

From these results, we can explain the strikingly different phase in zonal currents, dynamic height and thermocline depth along the equator for the seasonal cycle. Rossby and Kelvin wave contributions to the pressure field are roughly comparable in magnitude, with eastward propagating Kelvin waves having greater influence in the east and westward propagating Rossby waves having greater influence in the west. However, for zonal velocity west of 120°W, Rossby wave amplitudes exceed Kelvin wave amplitudes, so that westward propagation in zonal velocity is evident over a much broader range of longitudes.

Why do Rossby waves explain more variance in zonal currents than Kelvin waves? This is related to the differences in the relationship between zonal velocity and pressure for these two waves (Fig. 5.1). Specifically, for a Kelvin wave, the ratio of pressure to zonal velocity at the equator is (Moore and Philander 1977),

$$\frac{p_{-1n}(0)}{u_{-1n}(0)} = c_n \quad .$$

But for a mode 1 Rossby wave, the ratio is

$$\frac{p_{1n}(0)}{u_{1n}(0)} = \frac{c_n}{3} \quad .$$

So we have

$$\frac{u_{1n}(0)}{u_{-1n}(0)} = 3\,\frac{p_{1n}(0)}{p_{-1n}(0)} \quad .$$

Therefore, for a situation where Rossby and Kelvin wave pressures may be comparable, Rossby waves will explain much more variance in zonal currents than in pressure field.

5.3.2. Boundary reflections

In the proceeding section, we made the assumption that the reflection at the eastern and western boundaries was not significant. We now examine that assumption more closely to further support our conclusion that the seasonal variability consists predominantly of wind forced equatorial waves.

a. Western boundary

At the western boundary, the amplitudes of reflected Kelvin waves can be estimated from the amplitude of incoming Rossby waves based on equatorial wave theory. As suggested by Cane and Sarachik (1981), the western boundary condition in the long wave limit can be approximated as

$$\int_{-\infty}^{+\infty} (u_i + u_{-1n})/_{x=Xw}\, dy = 0, \tag{5.9}$$

where u_i are incident Rossby waves and u_{-1n} are the reflected Kelvin waves. If the incident variability is dominated by first baroclinic mode, first meridional mode long Rossby waves, which is the case for the seasonal cycle, then (5.9) becomes,

$$\int_{-\infty}^{+\infty} (u_{11} + u_{-11})/_{x=Xw}\, dy = 0.$$

From (5.3b), we have

$$p_{-11}(X_W, y, t) = \frac{1}{2}\, p_{11}\,(X_W, y_0, t)\, \frac{\phi_0(y)}{\sqrt{\frac{1}{2}\phi_2(y_0) + \phi_0(y_0)}},$$

where y_0 is some latitude. For $y_0 = 5°$, we have

$$p_{-11}(X_w, y, t) \approx 0.76\, p_{11}(X_w, 5°, t)\, \phi_0(y). \tag{5.10}$$

Fig. 5.4b shows averages of seasonal variations in surface dynamic height at 5°N and 5°S from the Levitus climatology. This analysis extends further west than the TAO data with adequate records for compiling a reliable mean climatology in the

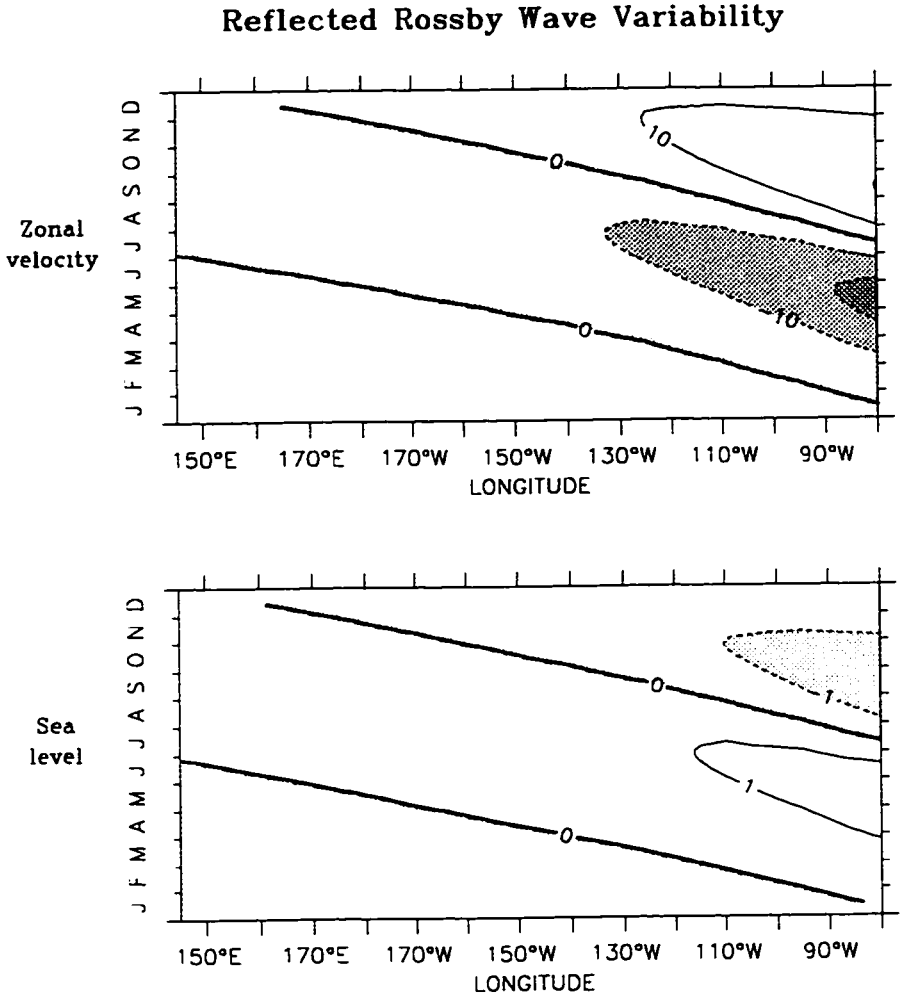


Fig. 5.5. Reflected Rossby waves from the eastern boundary. Upper panel: zonal surface velocity; lower panel: sea level. Contour intervals for zonal velocity and sea level are 10 cm s^{-1} and 1 cm, respectively.

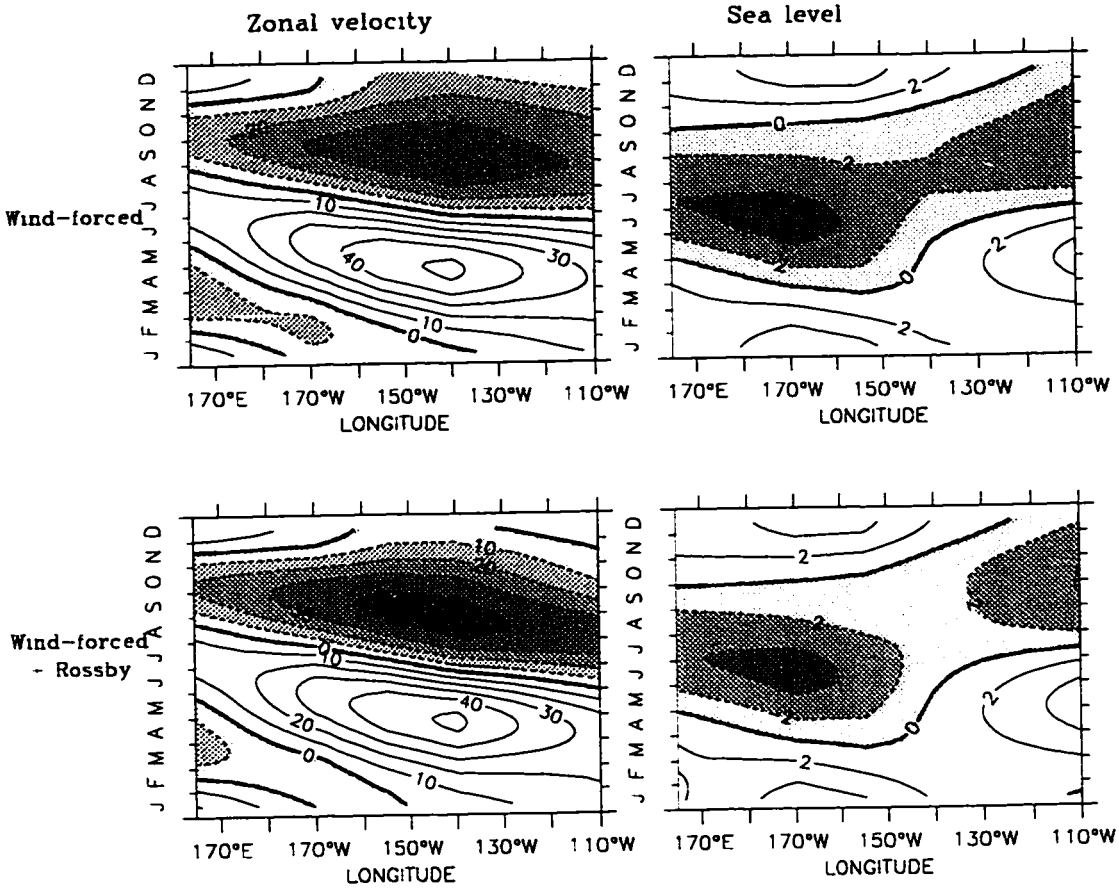


Fig. 5.6. Comparison of simulated seasonal variability in zonal surface current and sea level with and without eastern boundary reflection. Upper panels: sum of seasonal Kelvin and Rossby waves directly forced by surface wind stress; lower panels: the combination of variability in the upper panels and reflected Rossby waves in Fig. 5.5. Contour intervals for zonal current and sea level are 10 cm s^{-1} and 2 cm, respectively.

far western Pacific. The averaged seasonal variations at 5°N and 5°S have amplitudes of about 2.5 dyn cm at their westernmost extent (150°E). Assuming that these amplitudes extend all the way to the western boundary, the amplitude of the reflected Kelvin wave would be about 1.3 dyn cm as determined from (5.10), which is about 25% of the amplitude of the observed seasonal variability at the equator (Fig. 3.6b). This estimate is based on the assumption that the boundary is straight. Since the real boundary is much more complex, reflection will be less efficient and the resulting Kelvin waves might be even smaller. In addition, these waves will be damped while propagating to the east. We, therefore, conclude that reflection at the western boundary is not significant relative to seasonal wind forced variations in central and eastern Pacific (Fig. 3.5).

b. Eastern boundary

At the eastern boundary, incoming seasonal variability consists mainly of the first two baroclinic mode Kelvin waves. In the idealized case of a straight meridional boundary, these waves reflect into an infinite sum of symmetric mode long Rossby waves (Moore 1968). Using the formalism in McCreary (1985), we computed the amplitudes of the first 10 reflected waves for each of the four vertical modes and plotted their sum along the equator. In this sum, the first meridional mode Rossby wave accounts for most of the reflected wave variance near the equator. Most of this variance is confined to east of 120°W (Fig. 5.5).

Adding reflected waves to the interior forced wave solutions shows that in the longitude range 110°W-165°E, where we have been able to compute mean seasonal cycles from the TAO data, eastern boundary reflections make very little impact on the overall structure of the zonal currents and sea level (Fig. 5.6). These

results are consistent with those of previous authors (e.g. Kessler and McCreary 1993l; Minobe and Takeuchi 1995) who found that eastern boundary reflections contribute little to surface layer variability in the interior of the basin because of vertical propagation of kinetic energy.

5.3.3. Sensitivity to damping

The details of our long equatorial wave simulation depend to a certain extent on choice of damping coefficients. In this study we set the coefficients using (5.7) with a 6 month damping time scale for the first baroclinic mode and a 4 month damping time scale for the second baroclinic mode. To test sensitivity of our simulations to different coefficients, we reran the model with $r_n = (c_1/c_n)^2$ (24 months)$^{-1}$ as in Gent (1985) and Minobe and Takeuchi (1995). The model results are very similar to those with the shorter time scale damping given by (5.7) except that amplitudes are about 20% larger. We favor the stronger damping because the model results match TAO observations better, and because stronger damping is consistent with Picaut et al's (1993) recommendation for modeling equatorial long waves.

5.3.4. Higher baroclinic and meridional mode waves

As shown in Section 5.3.1, the third and fourth baroclinic mode waves explain only about 10% of the variance of the simulated surface zonal velocity and sea level for both equatorial Kelvin and Rossby waves. Higher baroclinic modes are very likely even less significant compared with the third and fourth modes. However, these high mode waves together may improve the model solution as pointed out in McCreary (1980). He showed that high modes together affect local

ocean response in the surface mixed layer to surface wind stress significantly, which may be underestimated in our simple wave model due to the neglect of higher vertical modes.

Two even numbered gravest mode Rossby waves (i.e., the second and fourth meridional modes that are anti-symmetric around the equator in zonal velocity and pressure) are simulated using the simple model, but are not presented in this thesis because they do not contribute significantly to variability on the equator, which is our primary focus here. These anti-symmetric meridional mode Rossby waves are excited by seasonal wind stress and wind stress curl forcing that are stronger north of the equator than south of the equator (Gent 1985; McPhaden et al. 1988). They enhance the seasonal variations in thermocline depth and dynamic height at 5°N and yet reduce the variations at 5°S by about 9% in amplitude, but the differences between these two latitudes due to these modes are not as large as those observed (Fig. 3.5 and Fig. 3.7). Higher meridional mode Rossby waves may contribute even less than these two modes. For example, the third meridional mode Rossby waves summed over all four vertical modes are only about 7% the amplitude of first meridional mode waves of the four gravest baroclinic modes at the equator in the model simulation.

Kessler and McCreary (1993) suggested that the significant asymmetry in the oceanic background state in the east-central Pacific may be another main reason for the asymmetry of the seasonal variability. The mean thermocline is tighter and shallower north of the equator than south of the equator (Fig. 3.3). The mean ocean circulation is also asymmetric about the equator, with the SEC extending across the equator to 2°-3°N, North Equatorial Countercurrent covering 3°-10°N and North Equatorial Current further north. However, it has not yet been

estimated how significantly the background asymmetry would affect the asymmetry of seasonal variations about the equator.

5.4. Discussion and Conclusions

Diagnostic calculations suggest that, at 5°N and 5°S, seasonal variations are dominated by the gravest two baroclinic mode, first meridional mode long Rossby waves. At the equator, in addition to these Rossby waves, the first two baroclinic mode Kelvin waves are excited by zonal wind forcing. The sum of the Rossby waves and Kelvin waves results in westward propagation in the zonal currents but eastward propagation in thermocline depth and dynamic height along the equator in the eastern and central Pacific. Minobe and Takeuchi (1995) also found that the first two vertical mode wind-forced Kelvin and Rossby waves dominated mean seasonal variability in the equatorial Pacific, and that eastern and western boundary reflections were relatively unimportant. However, they did not address the different direction of phase propagation along the equator for zonal currents, 20°C isotherm depths and dynamic height.

In the surface mixed layer, fritional dynamics described in Zebiak and Cane (1987) contributes to the westward propagation as well because the ocean response from this dynamics is in phase with the surface wind stress, which itself propagates westward. Sorting out the relative importance of the wave dynamics and frictional dynamics in this layer is the subject of future research.

While the diagnostic model we used is very simple and easy to track dominant wave modes for dynamical explanation of seasonal variability, it has significant limitations due to neglect of several dynamical processes. Among them, wave-mean flow interactions, nonlinearity in seasonal variability and horizontal

variations in the mean background state are three most prominent processes. These processes are possible to distort spatial structures of equatorial waves. For example, with existence of the mean SEC and EUC, maximum amplitude of the first baroclinic Kelvin wave would be around the EUC core depth instead of the surface layer as the linear wave dynamics predicts (McPhaden and Ripa 1990). A ocean model including these processes may erase the discrepancy of our simulation from the observations.

Yang et al. (1997) proposed a different hypothesis than ours to explain why the seasonal variation of thermocline depth propagates eastward along the equator. When forced by the 1 and 2 cpy harmonics of surface winds, their linear model did not produce the right sense of eastward propagation in upper layer thickness. So between 150°E-120°W they added an extra abruptly changing zonal wind patch, which excited an eastward propagating Kelvin wave. However, our results indicate that this extra ad hoc forcing is not necessary to generate eastward propagation in upper layer thickness (or thermocline depth in our case). An essential difference between our study and theirs is how the winds are projected onto the ocean. In Yang et al, a Gaussian meridional structure was assumed for zonal wind stress, with a 3° radius of deformation. Compared to the actual winds projected onto the Kelvin wave as given by b_{-1n} from (5.6a), their winds have seasonal variations about 30% too weak on average between 140°E-160°W, and about 10% too strong on average between 160°W-100°W. Both the underestimate in the western Pacific and overestimate in the eastern Pacific would suppress Kelvin wave variability and exaggerate Rossby wave variability. The zonal wind patch they added largely overlapped with the region where their assumed meridional structure for the winds leads to underestimates in annual period

forcing. Our results, in which the actual meridional structures are used for the wind forcing, do not require such adjustments.

As discussed in Chapter 3, the westward SEC reverses its direction in the eastern Pacific in late boreal spring and early summer although surface winds are always easterly. This phenomenon has puzzled scientists for more than a century since Puls (1895) first discovered it. Based on this study and previous studies of the mean circulation in the equatorial Pacific, we suggest that this springtime reversal is the result of a weak westward mean current at the surface, on which seasonal wave variability is superimposed. The weak mean westward surface flow results from upward advection of eastward momentum from the EUC core towards the surface, and poleward advection of westward momentum put into the ocean by the mean wind stress (e.g., Philander and Pacanowski 1980; Bryden and Brady 1985; McPhaden and Taft 1988). The nonlinearity that gives rise to this mean circulation pattern is relatively weak, however, and can be thought of qualitatively as a first order correction to the mean circulation as we discuss in Chapter 4 (also see McPhaden 1981). Superimposed on this mean circulation are seasonal variations in linear, wind-forced Kelvin waves and Rossby waves. In combination these waves produce a tendency for significant eastward flow in late boreal spring and early summer when the trade winds relax, a tendency that is strong enough to overcome the westward mean flow near the surface. The zonal current reversal propagates westward as a result of the current variations associated with the annual period Rossby wave. The Rossby wave weakens west of about 160°W as it propagates out of the dominant forcing region (Fig. 3.4a) and as it leaks energy into the deep ocean (Lukas and Firing 1985; Kessler and McCreary 1993). Hence, at 170°W where westward mean flow in the SEC is strongest, the springtime

reversal is weak or nonexistent. Further to the west at 165°E where the mean SEC is weak and semi-annual variations are pronounced, eastward surface currents are observed to run against the wind in June-August. This hypothesis may be verified by studies with more complex ocean models which are capable of simulating both the mean circulation and seasonal variability in the circulation reasonably.

Yu et al. (1997) also argued that nonlinearity in the mean circulation and springtime weakening of the southeasterly trades play an important role in occurrence of the springtime reversal in the SEC in the eastern Pacific. However, our interpretation differs in some important respects. Yu et al. concluded that the springtime weakening of the local southeasterly trades in the eastern Pacific is the primary cause of the reversal, while Kelvin waves excited in the western and central Pacific reduce the strength of the reversal. They did not make a link between the springtime reversal and the annual period Rossby wave, however, nor did they offer an explanation for the westward propagation of the springtime reversal.

Gu et al (1997) have recently pointed out that seasonal time scale variations are not stationary in the eastern equatorial Pacific. Using TAO data, they showed that thermocline depth variations at periods near one year were more pronounced in the 1990's than in the 1980's. Our definition of the seasonal cycle, which relies most heavily on data from the 1990's, may therefore contain some biases due to decadal time scale variations like those discussed in Gu et al. These biases, if present, are not a serious concern in general given the comparability of our seasonal analyses and those based on longer term data sets. However, the implications of this nonstationary for understanding ENSO is a topic worth further consideration. Also, Gu et al suggested, but did not demonstrate, that changes in

seasonal variations between the 1980's and the 1990's were due to changes in wind forcing and equatorial ocean wave responses. If our interpretation of the mean seasonal cycle in terms of wind forced equatorial Kelvin and Rossby waves is correct, the diagnostic framework presented in this paper may provide a means to quantitatively evaluate the role of these processes in modulating of the seasonal cycle in the equatorial Pacific.

In summary, our analysis offers some new insights into seasonal variations in the equatorial Pacific Ocean. But neglect of such processes as wave-mean flow interactions and zonal variations in the background state from our simple diagnostic model is likely to result in discrepancies in the modeled seasonal variations from the observed ones. Also, although we have invoked nonlinearity of the mean state as part of our explanation for the springtime reversal of the SEC, we cannot explicitly compute important nonlinear terms in the zonal momentum balance. Finally, we have concentrated our interpretation primarily on hemispherically symmetric aspects of the seasonal cycle, although there are significant asymmetries, particularly in the eastern Pacific, related to hemispheric asymmetries in zonal wind stress, wind stress curl forcing and the mean oceanic background state. Thus, there is a need for additional work, particularly with numerical models, to examine in more detail the hypotheses we have proposed in this study.

Chapter 6

CONCLUSIONS

In this thesis, we studied the dynamics of seasonal-to-interannual variability in the equatorial Pacific using the data from the Tropical Atmosphere Ocean (TAO) buoy array. We begin with presenting well defined mean climatology and interannual variability, focusing on the region of 165°E-110°W, in the equatorial Pacific compiled from the observations of the TAO over 1980-1997. Analyzing these compiled data, we find pronounced zonal phase features in seasonal variability, which may be interpreted in terms of linear equatorial wave dynamics. On the other hand, we do not find much phase information in interannual variability which can be clearly related to equatorial waves. To investigate whether the equatorial Pacific ocean is in equilibrium with the forcing, we then analyze zonal momentum balance on seasonal and interannual time scales through a diagnostic momentum equation using the TAO data. The analysis confirms that linear dynamics is a good approximation for both seasonal and interannual variability in the equatorial Pacific with nonlinearity being significant only at some depths. At the seasonal cycle, equatorial wave signals are prominent and the ocean is not in equilibrium with the forcing, while interannual variability is in quasi-equilibrium. Last, through a simple dynamical model, we further analyze seasonal variations in terms of equatorial waves and concluded that the variations consist primarily of the two gravest baroclinic mode equatorial Kelvin

waves and the two gravest baroclinic mode, first meridional mode long Rossby waves. Detail results of each chapter of the thesis are outlined further below.

In Chapter 3, we present mean fields, seasonal cycle and interannual variability derived from the TAO data in the equatorial Pacific. The mean fields show familiar features in thermal structure and ocean circulation such as the westward flowing South Equatorial Current (SEC), eastward flowing Equatorial Undercurrent (EUC), the warm pool in the western Pacific, the cool tongue in the eastern Pacific and thermocline slope along the equator, which are also found in some historic data. At the seasonal cycle, the following features are most important: (1) the normally westward SEC reverses its direction with significant magnitude in late boreal spring in the eastern and central Pacific; (2) seasonal variations in both the SEC and EUC, and 20°C isotherm depth at 5°N and 5°S propagate westward with a phase speed of about 1 m s^{-1}; but (3) along the equator, variations in 20°C isotherm depth and dynamic height propagate eastward at 0.5-0.7 m s^{-1} in the eastern and central Pacific. These results are confirmed by consistency check with other data sets. The propagation features suggest that the seasonal variations may be interpreted as equatorial waves of linear dynamics. On interannual time scales, we find that zonal wind anomaly is largest and in phase in the western and central Pacific. SST anomaly is significant in entire region of 165°E-110°W, and interannual variations in thermocline depth and dynamic height tend to propagate eastward with most of the phase shift in the central Pacific. The SST anomaly is highly correlated with anomalous thermocline depth in the eastern Pacific.

In Chapter 4, measurements from the TAO array are used to analyze the zonal momentum balance on the mean, seasonal and interannual time scales in the

equatorial Pacific based on a diagnosis of the zonal momentum equation. This analysis verifies some of hypotheses many model studies proposed on the mean circulation in the equatorial Pacific, such as the Sverdrup momentum balance between the ocean response and surface wind forcing, with zonal advection being of secondary significance in the depth integrated zonal momentum balance. On seasonal time scales, the zonal momentum balance is mainly between surface stress, pressure gradient and local acceleration. The difference between the surface stress and pressure gradient provides net forcing for accelerations and decelerations of zonal currents. These accelerations indicate that wave processes are pronounced and that the equatorial Pacific is not in equilibrium with wind forcing on seasonal time scales. On interannual time scales, in contrast, local acceleration is very small relative to either surface stress or pressure gradient. The surface wind stress is principally balanced by the pressure gradient, suggesting that, to zeroth order, the oceanic response to wind forcing is characterized by a succession of quasi-equilibrium states. Consistent with theories of the Equatorial Undercurrent, zonal transports in the thermocline vary simultaneously with the strength of the zonal pressure gradient and the easterly trade winds. In both seasonal and interannual variability, zonal advection is significant only around the EUC depths, its significance tends to reduce in the vertically integrated zonal momentum balance.

As shown in Chapter 3, the seasonal variations have prominent features of linear equatorial waves. Furthermore, analysis of zonal momentum balance in Chapter 4 demonstrates that linear wave dynamics are appropriate to describe main seasonal variability in both currents and thermal structures with possible discrepancies due to neglect of such nonlinear processes as wave-mean interaction

and nonlinearity in seasonal variability. Therefore, in Chapter 5, these seasonal variations are interpreted by means of a simple dynamical model based on linear equatorial wave theory. Model results, which agree well with the TAO observations, indicate that seasonal variability between 5°N and 5°S is dominated by wind-forced equatorial Kelvin waves and first meridional mode Rossby waves. The first two baroclinic modes dominate the solutions. The sum of the Rossby waves and Kelvin waves results in westward propagation in the equatorial zonal currents and off-equatorial thermal structure and yet eastward propagation in thermocline depth and dynamic height along the equator in the eastern and central Pacific. The simultaneous response to the zonal surface stress from the frictional dynamics also contributes to the westward propagation of the zonal velocity in the surface layer.

BIBLIOGRAPHY

Arthur, R.S., 1959: A review of the calculation of ocean currents at the equator. Deep-Sea Res., 6, 287-297.

Anderson, D.L.T., and J.P. McCreary, 1985: Slowly propagation disturbances in a coupled ocean-atmosphere model. J. Atmos. Sci., 42, 615-629.

Balmaseda, M.A., M.K. Davey, and D.L.T. Anderson, 1995: Decadal and seasonal dependence of ENSO prediction skill. J. Climate, 8, 2705-2715.

Battisti, D. S., 1988: Dynamics and thermodynamics of a warming event in a coupled atmosphere-ocean model, J. Atmos. Sci., 45, 2889–2919.

Battisti, D.S. and E. Sarachik, 1995: Understanding and Predicting ENSO. Reviews of Geophysics, 33, 1367-76

Battisti, D.S., and Hirst, A.C., 1989: Interannual variability in the tropical atmosphere-ocean system: Influence of the basic state and ocean geometry. J. Atmos. Sci., 46, 1687-1711.

Boulanger, J.-P., and C. Menkes, 1995: Propagation and reflection of long equatorial waves in the Pacific ocean during the 1992–1993 El Niño. J. Geophys. Res., 100, 25,041–24,059.

Bryden, H.L., 1977: Geostrophic comparisons of morred measurements of current and temperature during the Mid-Ocean Dynamics Experiment. Deep-Sea Res., 24, 677-681.

Bryden, H.L., and E.C. Brady, 1985: Diagnostic model of the three-dimensional circulation in the upper equatorial Pacific Ocean. J. Phys. Oceanogr., 15,

1255-1273.

Bryden, H.L., and E.C. Brady, 1989: Eddy momentum and heat fluxes and their effects on the circulation of the equatorial Pacific ocean. J. Marine Res., 47, 55-79.

Busalacchi, A.J., M.J. McPhaden, and J. Picaut, 1994: Variability in equatorial Pacific sea surface topography during the verification phase of the TOPEX/POSEIDON mission, J. Geophys. Res., 99, 24,725-24,738.

Busalacchi, A.J., and J.J. O'Brien, 1980: The seasonal variability in the tropical Pacific. J. Phys. Oceanogr., 10, 1929-1952.

Cane. M.A., 1983: Oceanographic events during El Niño. Science, 222, 1189-1193.

Cane. M.A., 1992: Comments on "The fast-wave limit and interannual oscillations". J. Atmos. Sci., 49, 1947-1949.

Cane, M.A., M. Munnich and S.E. Zebiak, 1990: A study of self-excited oscillations of the tropical ocean-atmosphere system. Part I: Linear analysis. J. Atmos. Sci., 47, 1562-1577.

Cane, M.A., and Sarachik, E.S., 1981: The response of a linear baroclinic equatorial ocean to periodic forcing. J. Mar. Re., 39, 651-693.

Cane, M. A., S. C. Dolan, and S. E. Zebiak, 1986: Experimental forecasts of the 1982/83 El Niño. Nature, 321, 827–832.

Cane, M.A., and S.E. Zebiak, 1987: Prediction of El Niño events using a physical model. Atmospheric and Oceanic Variability, J. Cattel, Ed., Roy. Meteor. Soc., 153-182.

Chang, P., 1993: Seasonal cycle of sea surface temperature and mixed layer heat budget in the tropical Pacific Ocean. Geophys. Res. Lett., 20, 2079-2082.

Chang, P., 1994: A study of the seasonal cycle of sea surface temperature in the tropical Pacific Ocean using reduced gravity models, J. Geophys. Res., 99, 7725-7741.

Chao, Y., and S.G.H. Philander, 1993: On the structure of the southern oscillation. J. Clim., 6, 450-469.

Charney, J.G. and S.L. Spiegel, 1971: The structure of wind-driven homogeneous layer neat the equator. Deep Sea Res. 6, 303-310.

Chen, D., L.M. Rothstein, and A.J. Busalacchi, 1994: A hybrid vertical mixing scheme and its application to tropical ocean models. J. Phys. Oceanogr., 24, 2156-2179.

Cronin, M., and M.J. McPhaden, 1997: The upper ocean heat balance in the western equatorial Pacific warm pool during September-December 1992, J. Geophys. Res., 102, 8533-8553.

daSilva, A.M., C.C. Young and S. Levitus, 1994: Atlas of surface marine data. Volume 1: Algorithms and procedures. NOAA Atlas NESDIS 5, Washington, D.C.

Davis, R.E., 1976: Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. J. Phys. Oceanogr., 6, 249-266.

Draper, N.R. and H. Smith, 1981: Applied regression analysis. Wiley, New York, 709p.

Eriksen, C.E., 1985: The Tropic Heat Program: An overview. Eos, Trans. Amer. Geophys. Union, 66, 50-54.

Frankignoul, D., F. Bonjean and G. Reverdin, 1996: Interannual variability of surface currents in the tropical Pacific during 1987-1993, J. Geophys. Res., 101, 3629-3647.

Gandin, L.S., 1963: Objective analysis of meteorological fields. Gidrometeorol Izdat, Leningrad (translation by Israel program for Scientific Translations, Jerusalem, 1966, 242 pp.)

Gent, P.R., 1985: The annual cycle in the central equatorial Pacific ocean. J. Marine Res., 43, 743-759.

Gent, P.R., K. O'Neill, and M.A. Cane, 1983: A model of the semiannual oscillation in the equatorial Indian Ocean. J. Phys. Oceanogr., 13, 2148-2160.

Gill, A.E., and A.J. Clarke, 1974: Wind-induced upwelling, coastal currents and sea-level changes. Deep-Sea Res. 21, 325-345.

Gu, D., S.G.H. Philander, and M.J. McPhaden, 1997: The seasonal cycle and its modulation in the eastern tropical Pacific Ocean. J. Phys. Oceanogr., in press.

Halpern, D., 1987: Observations of annual and El Niño thermal and flow variations along the equator at 0° 110°W and 0° 95°W during 1980-1985. J. Geophys. Res., 92, 8197-8212.

Halpern, D., R.A. Knox and D.S. Luther, 1988: Observations of 20-day period meridional current oscillations in the upper ocean along the Pacific equator. J. Phys. Oceanogr., 18, 1514-1534.

Harrison, D. E. and Luther, D. S., 1990: Surface winds from tropical Pacific islands: Climatological statistics, J. Clim., *3*, 251–271.

Hayes, S. P., L. J. Mangum, J. Picaut, A. Sumi, and K. Takeuchi, TOGA TAO, 1991: A moored array for real-time measurements in the tropical Pacific Ocean, Bull. Am. Meteorol. Soc., 72, 339–347.

Hayes, S.P., P. Chang, and M.J., McPhaden, 1991: Variability of the sea surface

temperature in the eastern equatorial Pacific during 1986-88. J. Geophys. Res., 96, 10,553-10,566.

Hirst, A.C., 1988: Slow instabilities in tropical ocean basin-global atmosphere models. J. Atmos. Sci., 43, 606-630.

Horel, J.D., 1981: On the annual cycle in the tropical Pacific atmosphere and ocean, Mon. Weather Rev., 110, 1863-1878.

Katz, E. J., 1981: Dynamic topography of the sea surface in the equatorial Atlantic. J. of Marine Res., 39, 53-63.

Katz, E.J., 1987: Seasonal response of the sea surface to the wind in the Equatorial Atlantic. J. Geophys. Res., 92(C2), 1885-1893

Katz, E. J., J. A. Carton, and A. Chakraborty, 1995: Dynamics of the equatorial Atlantic from altimetry. J. Geophys. Res., 100, 25,061–25,068.

Kessler, W.S., 1990: Observations of long Rossby waves in the northern tropical Pacific. J. Geophys. Res., 95, 5183-5217.

Kessler, W.S., and J.P. McCreary, 1993: The annual wind-driven Rossby wave in the subthermocline equatorial Pacific. J. Phys. Oceanogr., 23, 1192-1207.

Kessler and McPhaden, 1995: Oceanic equatorial waves and the 1991-93 El Niño. J. of Climate, 8, 1758-1774.

Kessler, W.S., M.J. McPhaden and K.M. Weickmann, 1995: Forcing of Intraseasonal Kelvin Waves in the Equatorial Pacific. J. Geophys. Res., 100, 10 613-10 631.

Kessler, W. S., M. C. Spillane, M. J. McPhaden, and D. E. Harrison, 1996: Scales of variability in the equatorial Pacific inferred from the Tropical Atmosphere-Ocean (TAO) array, *J.* Climate, 2999-3024.

Kindle, J.C., 1979: Equatorial Pacific ocean variability - Seasonal and El Niño

time scales. Dissertation, Florida State University.

Latif, M., and N.E. Graham, 1992: How much predictive skill is contained in the thermal structure of an oceanic GCM? J. Phys. Oceanogr., 22, 951-962.

Latif, M., T.P. Barnett, M.A. Cane, M. Fluegel, N.E. Graham, H. von Storch, J.-S. Xu, and S.E. Zebiak, 1994: A review of ENSO prediction studies. Clim. Dyn., 9, 167-179.

Legler, K.M., 1991: Producing surface wind products for oceanographers. Proc. IGOSS/IOC Ocean Products Workshop, Tokyo, Japan, Intergovernmental Oceanic Commission, 59-76.

Levitus, S., and T.P. Boyer, World Ocean Atlas 1994, 4 volumes, Washington, D.C., NOAA/NESDIS/NODC, 1994.

Levitus, S., T. P. Boyer, and J. Antonov, World Ocean Atlas 1994, Volume 5: Interannual Variability of Upper Ocean Thermal Structure, pp., NOAA Atlas NESDIS 5, Washington, D.C.

Li, T., and S.G.H. Philander, 1996: On the annual cycle of the eastern equatorial Pacific. J. Climate, 9, 2986-2998.

Lukas, R., and E. Firing, 1983: The geostrophic balance of the Pacific Equatorial Undercurrent. Deep-Sea Res. 31, 61-66.

Lukas, R., and E. Firing, 1985: The annual Rossby wave in the central equatorial Pacific Ocean. J. Phys. Oceanogr., 11, 55-67.

Mangum, L.J., H.P. Freitag, and M.J. McPhaden, 1994: TOGA-TAO array sampling schemes and sensor evaluations. Proc. Oceans '94 OSATES, Vol. II, Brest, France, 402-406.

Mangum. L.J., and S.P. Hayes, 1984: The vertical structure of the zonal pressure gradient in the eastern equatorial Pacific. J. Geophys. Res., 89, 10,441-449.

McCarty, M.E. and M.J. McPhaden, 1993: Mean seasonal cycles and Interannual variations at 0°, 165°E during 1986-1992. NOAA Tech. Memo. ERL-PMEL-98, 64p.

McCreary, J., 1981: A linear stratified ocean model of the equatorial undercurrent. Philos. Trans. R. Soc., London, 298, 603-645.

McCreary, J., 1984: Equatorial beams. J. Mar. Res., 42, 395-430.

McCreary, J., 1985: Modelling equatorial oceanic circulation. Annu. Rev. Fluid Mech., 17. 359-409.

McCreary, J., and Lukas, R., 1986: The response of the equatorial ocean to a moving wind field. J. Geophys. Res., 91, 11691 - 11705.

McPhaden, M.J., 1981: Continuously stratified models of the steady state equatorial ocean. J. Phys. Oceanogr., 11, 337-354.

McPhaden, M.J., 1993: TOGA-TAO and the 1991-1993 El Niño-Southern oscillation event. Oceanography, 6, 36-44.

McPhaden, M.J., 1995: The Tropical Atmosphere-Ocean array is completed. Bull. Amer. Meteor. Soc., 76, 739-741.

McPhaden, M. J., A. J. Busalacchi, and J. Picaut, 1988: Observations and wind forced model simulations of the mean seasonal cycle in tropical Pacific sea surface topography, J. Geophys. Res., 93, 8131–8146.

McPhaden, M.J., and collaborators, 1998: The Tropical Ocean Global Atmosphere (TOGA) observing system: a decade of progress. J. Geophys. Res., in press.

McPhaden, M.J., H.P. Freitag, S.P. Hayes, B.A. Taft, Z. Chen and K. Wyrtki, 1988: The response of the equatorial Pacific to a westerly wind burst in May 1986. J. Geophys. Res., 93, 10 589-10 603.

McPhaden, M.J. and B.A. Taft, 1988: Dynamics of seasonal and intraseasonal variability in the eastern equatorial Pacific. J. Phys. Oceanogr., 18, 55-67.

McPhaden, M.J. and M.E. McCarty, 1992: Mean seasonal cycles and interannual variations at 0°, 110°W and 0°, 140°W during 1980-1991. NOAA Tech. Memo. ERL-PMEL-95, 118p.

McPhaden, M.J., J.A. Proehl, and L.M. Rothstein, 1987: On the structure of low frequency equatorial waves. J. Phys. Oceanogr., 17, 1555-1559.

Mechoso, C.R. and collaborators, 1995: The seasonal cycle over the tropical Pacific in coupled ocean-atmosphere general circulation models. Mon. Wea. Rev., 123, 2825-2838.

Meyers, G., 1979a: Annual variation in the slope of the 14°C isotherm along the equator in the Pacific Ocean. J. Phys. Oceanogr., 9, 885-891.

Meyers, G., 1979b: On the annual Rossby wave in the tropical North Pacific Ocean. J. Phys. Oceanogr., 9, 663-674.

Meyers, G., and J.-R. Donguy, 1980: An XBT network with merchant ships. Trop. Ocean-Atmos. Newslett., 2, 6-7.

Minobe, S. and K. Takeuchi, 1995: Annual period equatorial waves in the Pacific Ocean. J. Geophys. Res., 100(C9), 18,379-392.

Montgomery, R.B., and E.D. Stroup, 1962: Equatorial waters and currents at 150°W in July-August 1952,. Johns Hopkins Oceanographic Studies, 2, 68pp.

Moore, D.W., 1968: Planetary-gravity waves in an equatorial ocean. Ph.D. thesis, Harvard University, Cambridge, Massachusetts.

Moore, D.W. and Philander, S.G.H., 1977: Modelling of the tropical oceanic circulation. In 'The Sea' (E. D. Goldberg, I.N. McCave, J.J. O'Brien, and

J.H. Steele, eds.), Vol. 6, pp. 319-361, Wiley, New York.

Munk, W.H., 1966: Abyssal recipes, Deep-sea Res., 13, 707-730.

Neelin, J. D., 1991: The slow sea surface temperature mode and the fast-wave limit: Analytic theory for tropical interannual oscillations and experiments in a hybrid coupled model, J. Atmos. Sci., 48, 584–606.

Neelin, J.D., and collaborators, 1992: Tropical air-sea interaction in general circulation models. Climate Dyn., 7, 73-104.

Peters, J., M.C. Gregg and J.M. Toole, On the parameterization of equatorial turbulence, J. Geophys. Res., 93, 1199-1218, 1988.

Philander, S.G.H., 1979: Variability of the tropical ocean, Dynamics of Atmospheres and Oceans, 3, 191-208.

Philander, S. G. H., and collaborators, 1985: Long waves in the equatorial Pacific Ocean, Eos Trans. AGU, 66, 154.

Philander, S.G.H. and Y. Chao, 1991: On the contrast between the seasonal cycles of the equatorial Atlantic and Pacific Oceans. J. Phys. Oceanogr., 21, 926-931.

Philander, S.G.H., W.J. Hurlin, and A.D. Seigel, 1987: Simulation of the seasonal cycle of the tropical Pacific Ocean. J. Phys. Oceanogr., 17, 1986-2002.

Philander, S.G.H. and R. Pacanowski, 1980: The generation of equatorial currents. J. Geophys. Res., 85(C2), 1123-1136.

Philander, S.G.H., and R.C. Pacanowski, 1981: Response of equatorial oceans to periodic forcing. J. Geophys. Res., 86, 1903-1916.

Picaut, J, C. Menkes, J.P. Boulanger, and Y. Penhoat, 1993: Dissipation in a Pacific equtorial long wave model, TOGA Notes, 11-15.

Picaut, J., and T. Delcroix, 1995: Equatorial wave sequence associated with warm

pool displacements during the 1986-1989 El Niño-La Niña. J. Geophys. Res., 100(C9), 18,393-408.

Plimpton, P.E., H.P. Freitag, and M.J. McPhaden, 1997a: ADCP velocity errors from pelagic fish schooling around equatorial moorings. J. Atmos. Ocean. Tech., in press.

Plimpton, P.E., H.P. Freitag, M.J. McPhaden, and R.H. Weisberg, 1997b: Using echo intensity to correct moored ADCP data for fish-bias errors at 0°, 170°W. NOAA Tech. Memo. ERL PMEL-xxx, 163 pp. (In press).

Puls, C., 1895: Oberflachentemperaturen und Stromungsverhaltmisse des Aquatorialgurtels des Stillen Ozeans. Dtsch. Arch. Seewarte, 18, 1-38.

Qiao, L., and R.H. Weisberg, 1997: The zonal momentum balance of the Equatorial Undercurrent in the central Pacific. J. Phys. Oceanogr., 27, 1094-1119.

Rasmusson, E.M. and T. Carpenter 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. Mon. Weather Rev. 110, 354-384.

Reverdin, G., C. Frankignoul, E. Kestenare and M.J. McPhaden, 1994: Seasonal variability in the surface currents of the equatorial Pacific. J. Geophys. Res., 99, 20 323-20 343.

Reynolds, R. W. and T. M. Smith, 1995: A high resolution global sea surface temperature climatology. J. Climate, 8, 1571-1583.

Schneider, E.K., B. Huang and J. Shukla, 1995: Ocean wave dynamics and El Niño, J. of Climate, 8, 2415-2439.

Seager, R., S.E. Zebiak, and M.A. Cane, 1988: A model of the tropical sea surface temperature climatology. J. Geophys. Res., 93, 1265-1280.

Springer. S.R., M.J. McPhaden, and A.J. Busalacchi, 1990: Oceanic heat content variability in the tropical Pacific during the 1982-1983 El Niño, J. Geophys. Res., 95, 22,089-22,101.

Suarez, M.J., and P.S. Schopf, 1988: A delayed action oscillator for ENSO, J. Atmos. Sci., 45, 3283-3287.

Sverdrup, S.U., 1947: Wind-driven currents in a baroclinic ocean with application to the equatorial currents of the eastern Pacific. Proc. Natl. Acad. Sci., 33, 318-326.

Tsuchiya, M., 1979: Seasonal variation of the equatorial zone geopotential gradient in the eastern Pacific Ocean. J. Marine Res., 37(3), 399-407.

Wacongne, S, 1989: Dynamical regimes of a fully nonlinear stratified model of the Atlantic Equatorial Undercurrent. J. Geophys. Res., 94, 4801-4815.

Webster, P.J., and S. Yang, 1992: Monsoon and ENSO: Selectively interactive systems. Q.J.R. Meteorol. Soc., 118, 877-926.

Weisberg, R.H., and S.P. Hayes, 1995: Upper ocean variability on the equator in the west-central Pacific at 170°W. J. Geophys. Res., 100, 20,485-20,498.

Weisberg, R.H., and S.P. Hayes, 1995: Upper ocean variability on the equator in the west-central Pacific at 170°W. J. Geophys. Res., 100, 20,485-20,498.

Wyrtki, K., and B. Kilonsky, 1984: Mean water and current structure during the Hawaii to Tahiti shuttle experiment. J. Phys. Oceanogr. 14, 242-254.

Wyrtki., K., 1985: Water displacements in the Pacific and the genesis of El Niño cycles, J. Geophys. Res., 90, 7129-7132.

Xie, S.-P., 1995: Interaction between the annual and interannual variations in the equatorial Pacific. J. Phys. Oceanogr. 25. 1930-1941.

Xie, S., A. Kubokawa, and K, Hanawa, 1989: Oscillations with two feedback

processes in a coupled ocean-atmosphere model, J. Climate, 2, 946-964.

Yang, Y.J., T.Y. Tang, and R.H. Weisberg, 1997: Basin-wide zonal wind stress and ocean thermal variations in the equatorial Pacific Ocean. J. Geophys. Res., 102, 911-927.

Yin, F.L. and E.S. Sarachik, 1993: Dynamics and heat balance of steady equatorial undercurrents. J. Phys. Oceanogr., 23, 1647-1669.

Yu, Z., P.S. Schopf, and J.P. McCreary, 1997: On the annual cycle of upper ocean circulation in the eastern equatorial Pacific. J. Phys. Oceanogr., 27, 309-324.

Zebiak, S.E., 1993: Air-sea interaction in the equatorial Atlantic region, J. of Climate, 6, 1567-1586.

Zebiak, S.E., and M.A. Cane, 1987: A model El Niño-Southern Oscillation.Mon. Weather Rev., 115, 2262-2278.

Zeng, X., M. Zhao, and R.E. Dickinson, 1997: Intercomparison of Bulk aerodynamic algorithms for the computation of sea surface fluxes using the TOGA COARE and TAO data. J. Climate, submitted.

CURRICULUM VITAE

XURI YU

EDUCATION

Ph.D. Oceanography, University of Washington, Seattle WA, August 1998 (Expected)

M.S. Oceanography, Ocean University of Qingdao, China, July 1989

B.S. Oceanography and Applied mathematics, Ocean University of Qingdao, China, July 1986

PROFESSIONAL EXPERIENCE

Graduate Research Assistant, University of Washington, February 1992 - present. Investigated dynamics of the tropical Pacific ocean through data analysis and model simulation. Wrote scientific papers. Taught graduate students as a teaching assistant.

Instructor and research associate, Ocean University of Qingdao, China July 1989 - July 1991. Taught courses in oceanography and mathematics. Conducted research on storm surges through numerical models.

PROFESSIONAL SOCIETIES

American Geophysical Union, 1995-present

The Oceanography Society, 1996-present

PUBLICATIONS

Yu, X. and M.J. McPhaden, 1998: Dynamical analysis of seasonal and interannual variability in the Equatorial Pacific. J. Phys. Oceanogr. (accepted).

Yu, X. and M.J. McPhaden, 1997: Seasonal variability in the equatorial Pacific. J. Phys. Oceanogr. (in press).

Yu, X. and J.Y. Wang, 1991: Study of sea-bottom friction coefficient by application on three-dimensional model. Marine Science (China), 6, 57-60.

Yu, X. and J.Y. Wang, 1991: Study of ocean bottom friction: applying to storm surges and astronomical tides. J. of Ocean Univ. of Qingdao, 21, 10-20.

CONFERENCE PRESENTATIONS

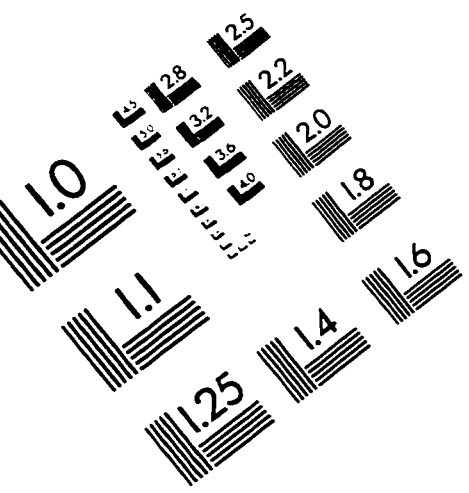
Ocean Science Meeting, San Diego, February 1998, "Seasonal variability in the equatorial Pacific."

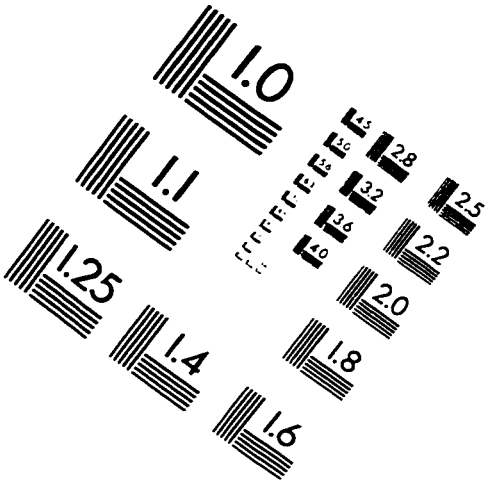
El Niño International Scientific Conference, Melbourne, Australia, April 1996, "Slow eastward propagation in equatorial Pacific During ENSO"

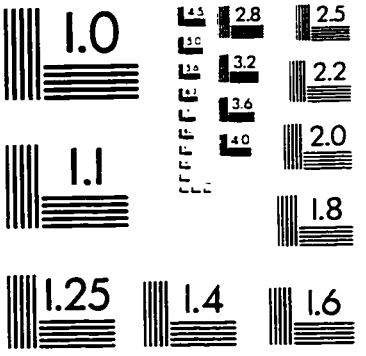
Pacific Basin Meeting, Honolulu, July 1995, "Dynamics of seasonal-to-interannual variability in the equatorial Pacific Ocean."

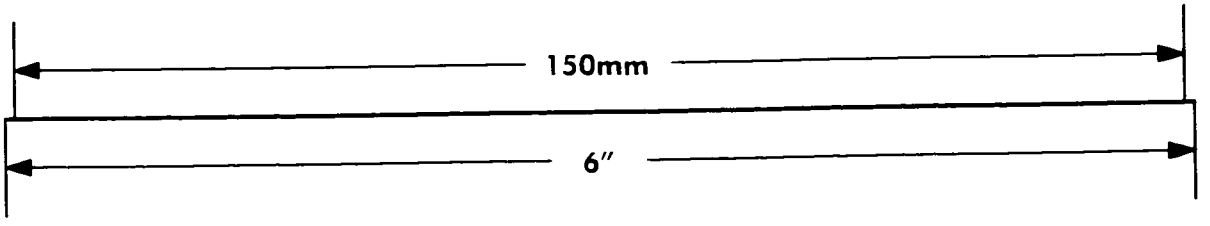
Storm Surge and Tides International Conference, Beijing, China, August 1990, "Study on sea bottom friction coefficient by using a numerical model."

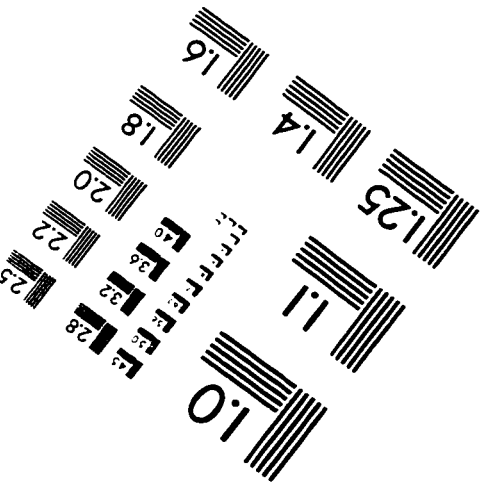
IMAGE EVALUATION
TEST TARGET (QA-3)

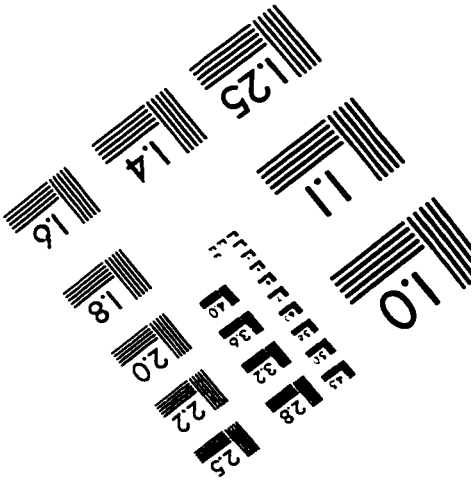












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved

99

07968

UMI

MICROFILMED 1998

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Detection and Imaging of Targets in the Presence of Clutter Based on Angular Correlation Function

by

Guifu Zhang

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

University of Washington

1998

Approved by _____ July 7, 1998
(Chairperson of Supervisory Committee)

_____

_____

Program Authorized
to Offer Degree Electrical Engineering

Date July 30, 1998

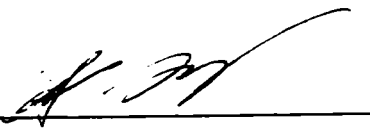
UMI Number: 9907968

UMI Microform 9907968
Copyright 1998, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized
copying under Title 17, United States Code.

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date 7-24-98

University of Washington

Abstract

Detection and Imaging of Targets in the Presence of Clutter Based on Angular Correlation Function

by Guifu Zhang

Chairperson of Supervisory Committee

Professor Leung Tsang

Department of Electrical Engineering

The detection of targets in clutter environments is a problem of practical importance in both military and civil uses. It is quite often that targets are embedded in clutter environment. Examples are: mines and piples buried under ground, tanks hidden in trees, submarines under water, boats on ocean surface, tumors inside bodies and so forth. Among all of the examples, the surrounding media (random media or random rough surfaces) can also scatter and radiate waves, causing clutter. The clutter can be comparable to or larger than the target signal, which can obscure the targets and make the detection of targets difficult.

In the past, radar cross section (RCS) had been studied for target detection. It is difficult to distinguish target scattering from clutter with the RCS measurements when the background clutter is strong. Recently, advanced techniques, such as wavelet transform, polarimetric measurements, image processing, and matched filtering method have been used. In this thesis, we study a method based on the calculation of angular correlation function (ACF) of two received wave fields corresponding two incident waves.

ACF of random media scattering was first studied. This includes the derivation of general expression, volume scattering, rough surface scattering, and averaging techniques.It has been shown that the memory effect is a result of the statistically tranlational invariance of random scattering. Random rough surface scattering exhibits a memory line, while volume scattering exhibits two memory dots corresponding to self-correlation and reciprocity path. For the purpose of target detection, the memory effect is suggested to be avoided to suppress the clutter.

Numerical studies of target detection consist of two steps: (1) calculate the scattered wave fields, and (2) process the simulated data. To obtain wave fields scattered by a target embedded in clutter, we formulated problems exactly by using surface and volume integral equations. The equations are then solved numercally with the use of the method of moment and fast methods, which give exact solution of Maxwell's equations. Coherent addition approximation has also been used to generate data for SAR and ACF processing. It has been found that ACF is more effective in suppressing the clutter due to the random media and rough surface scattering than RCS, and ACF imaging gives larger signal-clutter ratios and finer resolution than the conventional SAR imaging. That is because the memory effect of random media scattering has been avoided and the frequency dependence of target scattering is overcome in ACF imaging.

TABLE OF CONTENTS

List of Figures iv

Chapter 1: Introduction 1

1.1 Historical Review 1

1.2 Angular Correlation Function Versus Scattering Cross Section 3

1.3 An Overview of This Dissertation 4

Chapter 2: Angular Correlation Function of Random Media Scattering 8

2.1 Introduction 8

2.2 General Derivation of Memory Effect for Random Media Scattering . 9

2.3 Angular Correlation Function of Wave Scattering from Randomly Distributed Scatterers 13

2.3.1 Foldy-Lax Multiple Scattering Equation 13

2.3.2 Analytical Method 14

2.3.3 Numerical Simulations 19

2.3.4 Results 21

2.4 Angular Correlation Function for Rough Surface Scattering 23

2.4.1 Analytical Expression 24

2.4.2 Results 26

2.5 Averaging Techniques 26

2.5.1 Definitions of Average Techniques 27

2.5.2 Numerical results 28
2.6 Conclusions and Discussions 29

Chapter 3: The Detection of a Buried Object Embedded in Random Discrete Scatterers under a One-Dimensional Random Rough Surface 37

3.1 Introduction 37
3.2 Formulation 38
3.3 Numerical Results and Analysis 40

Chapter 4: Angular Correlation Function and Scattering Coefficient of Electromagnetic Waves Scattered by A Buried Object Under A Two-Dimensional Rough Surface 50

4.1 Introduction 50
4.2 Formulation 52
4.2.1 Incident Wave 52
4.2.2 Surface Integral Equations 53
4.2.3 Matrix Equation and Its Solution 55
4.2.4 Statistics of Scattered Fields 56
4.3 Results and Discussions 58
4.4 Conclusions and Discussions 60

Chapter 5: Application of Angular Correlation Function of Clutter Scattering and Correlation Imaging in Target Detection 69

5.1 Introduction 69
5.2 Formulation 71
5.3 Monte Carlo Simulations of SAR data 74
5.4 Circular SAR Imaging 74

5.5 Linear SAR Imaging 77
5.6 Conclusions and Discussions 80

Chapter 6: Summary and Future Research **90**

6.1 Summary 90
6.2 Future Research 91
6.2.1 High-order Moments 91
6.2.2 Multiple scattering effect on imaging 91
6.2.3 Bistatic imaging at a single frequency 92
6.2.4 Applications of non-Gaussian random field 93

Bibliography **99**

LIST OF FIGURES

1.1 Configuration of angular correlation function for random medium scattering. 7

2.1 Configuration of angular correlatio function for wave scattering from randomly distributed scatterers and random rough surfaces 30

2.2 3-D plot of analytical results of ACF for $f_v = 10\%$ at the reference angles $(\theta_{i1}, \theta_{s1}) = (20°, -40°)$. Results are shown as a function of the $sin(\theta_{i2})$ and $sin(\theta_{s2})$. 31

2.3 Analytical solutions of ACF for the volume density $f_v = 1, 5, 10,$ and 30%. The reference angles are at $(\theta_{i1}, \theta_{s1}) = (20^o, -40^o)$. 32

2.4 Comparison of analytical, numerical and experimental results for the volume density $f_v = 1, 5, 10,$ and 30%. The reference angles are at $(\theta_{i1}, \theta_{s1}) = (20°, -40°)$. 33

2.5 ACF along the memory line for TE and TM cases. $\epsilon_r = (16.67 + i1.15)$ for 30% moisture, $\epsilon_r = (3.71 + i0.13)$ for 5% moisture, $h = 0.05\lambda$, $l_o = 0.1\lambda$, and $L = 20\lambda$. (a): $\theta_{i1} = 20^o$, $\theta_{s1} = 40^o$. (b): $\theta_{i1} = 20^o$, $\theta_{s1} = -20^o$. 34

2.6 Comparison between SPM and numerical simulations for TE cese. $\epsilon_r = (16.67 + i1.15)$ for 30% moisture, $\epsilon_r = (3.71 + i0.13)$ for 5% moisture. $h = 0.05\lambda$, $l_o = 0.1\lambda$, and $L = 20\lambda$. (a): $\theta_{i1} = 20^o$, $\theta_{s1} = 40^o$. (b): $\theta_{i1} = 20^o$, $\theta_{s1} = -20^o$. 35

2.7 3-D plots of ACF magnitude by different averaging methods. Reference angles are ($\theta_{i1} = 20^o$, $\theta_{s1} = -20^o$). Dielectric constant of region 1 is $\epsilon_r = 3.7 + i0.13$. $h = 0.35\lambda_o$, $l = 1.0\lambda_o$, $L = 40\lambda_o$, g = L/4. (a): One realization. (b): Realization averaging over 100 rough surfaces. (c): Frequency averaging over a frequency band of $0.5f_o$ to $1.5f_o$. (d): Angular averaging over an angular range of $\theta - 10^o$ to $\theta + 10^o$. . . . 36

3.1 Configuration of angular correlation of wave scattering from a rough surface with a buried object embedded in small scatterers. The two incident angles are θ_{i1} and θ_{i2} with respective scattered angles θ_{s1} and θ_{s2}. 44

3.2 ACF for wave scattering by a rough surface and 300 scatterers based on frequency averaging. (a): 3-D plot of the ACF. (b): ACF magnitude along and perpendicular to the memory line. 45

3.3 3-D plots of ACF for an object embedded in scatterers under a surface. (a): Under a flat surface. (b): Under a rough surface ($h = 0.1\lambda_o, l = 0.5\lambda_o$). 46

3.4 Comparisons between ratios (in dB) of the ACF ("*") with object that without object and ratios of intensity ("o"). (a) rough surface without scatterers ($h = 0.1\lambda_o$, $l = 0.5\lambda_o$) (b) flat surface with 300 scatterers. 47

3.5 Results of Figure 3.4 after angular averaging. 48

3.6 Comparisons between ratios (in dB) of the ACF with object that without object and ratios of intensity for the case in which both the rough surface and the random discrete scatterers are present. The left column is the results by frequency averaging only ("*": ACF, "o": Intensity). The right column is the results obtained by taking angular averaging over the frequency averaging results of the left column (—: ACF, —: Intensity). 49

4.1 Configuration of EM wave scattering by a 3-D object buried under a 2-D rough surface. 61

4.2 Discretization of a sphere into 80 triangle patches. 62

4.3 Comparison of MoM and Mie scattering for the radar cross section of a PEC sphere. 63

4.4 Comparison between the solution of the matrix equation for the object on left-hand side and that on right-hand side. 64

4.5 ACF and RCS of EM wave scattering by a PEC sphere buried under a 2-D rough surface for hh polarization component. Parameters are: $a = 0.3\lambda$, $d = 0.6\lambda$;$L_x = L_y = 8.0\lambda$, $h = 0.02\lambda$, $l_x = l_y = 0.5\lambda$, $r_d = 3.5\lambda$, $\theta_{i1} = 20°$, $\theta_{s1} = -40°$, and $\theta_{i2} = 20°$. 65

4.6 RCS of EM wave scattering by a PEC sphere buried under a 2-D rough surface for co-pol. and cross-pol. components. Parameters are: $a = 0.3\lambda$, $d = 0.6\lambda$;$L_x = L_y = 8.0\lambda$, $h = 0.02\lambda$, $l_x = l_y = 0.5\lambda$, $r_d = 3.5\lambda$, $\theta_{i1} = 20°$, $\theta_{s1} = -40°$, and $\theta_{i2} = 20°$. 66

4.7 ACF of EM wave scattering by a PEC sphere buried under a 2-D rough surface for co-pol. and cross-pol. components. Parameters are: $a = 0.3\lambda$, $d = 0.6\lambda$;$L_x = L_y = 8.0\lambda$, $h = 0.02\lambda$, $l_x = l_y = 0.5\lambda$, $r_d = 3.5\lambda$, $\theta_{i1} = 20°$, $\theta_{s1} = -40°$, and $\theta_{i2} = 20°$. 67

4.8 PACF of EM wave scattering by a PEC sphere buried under a 2-D rough surface between co-pol. and cross-pol. components. Parameters are: $a = 0.3\lambda$, $d = 0.6\lambda$:$L_x = L_y = 8.0\lambda$, $h = 0.02\lambda$, $l_x = l_y = 0.5\lambda$, $r_d = 3.5\lambda$, $\theta_{i1} = 20°$, $\theta_{s1} = -40°$, and $\theta_{i2} = 20°$. (a): hhvh, (b) vvhv . 68

5.1 Configurations of SAR imaging for targets embedded in small scatterers. (a) Circular SAR, (b) linear SAR 82

5.2 Simulated image of targets embedded in clutter by field imaging for circular SAR 83

5.3 Simulated image of targets embedded in clutter by ACF imaging for circular SAR. 84

5.4 Simulated image of targets embedded in clutter by ACF focusing for circular SAR. 85

5.5 Simulated image of targets embedded in clutter by field imaging for linear SAR. 86

5.6 Simulated image of targets embedded in clutter by FACF imaging for linear SAR. 87

5.7 Simulated image of targets embedded in clutter by FCF imaging for linear SAR. 88

5.8 Simulated image of targets embedded in clutter by ACF imaging for linear SAR. 89

6.1 Multiple scattering effect on imaging for volume scattering 94

6.2 Bistatic imaging of three spheres at different frequency. (a): Three spheres with radius of $a_1 = 0.175\lambda_o$, $a_2 = 0.235\lambda_o$ and $a_3 = 0.35\lambda_o$. (b): Imaging at frequency of $f_1 = f_o$, $k_o a_1 = 1.17$. (c): Imaging at frequency of $f_2 = 0.75 f_o$, $k_2 a_2 = 1.17$. (d): Imaging at frequency of $f_3 = 0.5 f_o$, $k_3 a_3 = 1.17$. 95

6.3 Bistatic imaging for a target cylinder embedded in small cylinders. Parameters are $a_t = 5cm$, $a_c = 1.67cm$, and $L_x = L_z = 200cm$. 96

6.4 Bistatic imaging for a target cylinder embedded in small cylinders under a rough surface. Parameters are $a_t = 10cm$, $a_c = 0.5cm$, $N = 300$, and $L_x = 800cm$. 97

6.5 Comparison of bistatic scattering coefficients of the random rough surfaces with Gaussian (solid line), Rayleigh (dash line), and exponential (dotdash line) distributions. Results are obtained by averaging over 100 realizations. Parameters are $h = 0.5\lambda$, $l = 1.0\lambda$, $\theta_i = 30^o$, and $\epsilon_r = 3.7 + 0.13i$. 98

Acknowledgement

I would like to express my apprciations to Professor Leung Tsang for his guidance and encouragement during my study at University of Washington. He funded me and served as the chairperson of my Ph. D. supervisory committee. I am indebted to my mother for her sacrifices to further my eduation. I am grateful to my wife, Min Chen, for her strong support and my daughter, Hanzi Zhang, for her happiness to balance my school life.

I appreciate the contributions of my Ph. D. supervisory committee members, Professor Chi Chan, Akira Ishimaru, Yasuo Kuga, George M. Pigott and Jothiram Vivekanandan. I'd like to thank ERSL program coordinator, Mrs. Noel Henry, for her kind helps. Great appreciations are also expressed to ERSL graduate students: Kyung Pak, Chongxiao Chen, Li Li, Charlie Le, Steve Chan, Charles E. Mandt, Todd Elson, Ji-Hae Yea, Jun Liu, Libing Huang, Qin Li, Chien-min Lin, Chi-te Chan, Juhn Rockway, John Thomas, Song-Wu and Junhao Cha for their supports, friendship, and helps.

Thanks are also given to the Laboratory of Applications and Computations in Electromagnetics and Optics. The research in this dissertation was supported by Office of Naval Research.

Chapter 1

INTRODUCTION

1.1 Historical Review

The detection and imaging of targets embedded in clutter environments are of practical importance in both military and civil uses [29]. Targets can be under or above a random rough surface such as ground surface and ocean surface, or hidden in random media like vegetation such as a tank under tree. Targets can also be found inside inhomogenous medium, like tumors inside human body. The random media (rough surface, random discrete scatterers and inhomogeneities) also scatter, absorb, and radiate waves, which can be comparable to or larger than that from the targets [9][42]. In this case, the clutter can obscure the targets, making the detection and imaging of targets in the presence of clutter difficult.

The received signal from targets surrounded by clutter is usually random. Statistical characteristics of the received signal are studied for target detection. The mean of radar cross section (RCS), correlation function, and probability density function (PDF) attract most interests in detection problems. The received signal is processed with an ambiguity function or with a matched filter [4]. Recently, various radar systems and data processing methods have been used to detect targets embedded in clutter. Numerical simulations have been used to study the scattering properties [43]. Wavelet transform techniques have been used to separate targets from clutter [49]. Polarimetric measurements have been made for the detection of targets buried in a natural snowpack to enhance the visibility [48]. SAR imaging has attracted great

interest [1][33].

Image processing is to obtain detailed information of a target region from wide-band and large angular range measurements. The conventional SAR imaging method can be called field imaging, in which the target function is obtained from the measured fields by inverse Fourier transform or by focusing. The focusing method is also called correlation imaging since the field is correlated with a reference signal. The field focusing method can be improved by introducing a filter function, such as a matched filter, if the scattering property of the target is known. However, if the target scattering function is not known, the matched filtering method can not be applied.

Since Feng et al. [6] discovered the "memory effect" of random scattering of waves by random media in 1988, the study of angular correlation function (ACF) has attracted considerable attention [19][18][10][13]. The ACF is the correlation function of two scattered fields in directions $\hat{k}_{s1}$ and $\hat{k}_{s2}$ corresponding to two incident waves in directions $\hat{k}_{i1}$ and $\hat{k}_{i2}$, respectively. It is defined as

$$\begin{aligned} \Gamma(\hat{k}_{s1}, \hat{k}_{i1}; \hat{k}_{s2}, \hat{k}_{i2}) &= < F(\hat{k}_{s1}, \hat{k}_{i1}) F^*(\hat{k}_{s2}, \hat{k}_{i2}) > \\ &= < F(\theta_{s1}, \phi_{s1}; \theta_{i1}, \phi_{i1}) F^*(\theta_{s2}, \phi_{s2}; \theta_{i2}, \phi_{i2}) > \end{aligned} \tag{1.1}$$

$F(\hat{k}_s, \hat{k}_i)$ is the scattering amplitude from incident direction $\hat{k}_i$ into scattered direction $\hat{k}_s$. A strong correlation, called the angular memory effect, is only exhibited on the memory line for rough surface scattering and memory dots for volume scattering. The memory effect of random media scattering is a result of statistical translational invariance of random media. It has been confirmed by experiments and numerical simulations [10][43][13]. The ACF for angular pairs far away from the memory effect is very small in magnitude. Therefore, ACF can be used in the detection of targets embedded in clutter.

Correlation functions are obtained by taking averages over a product of two signals. Thus, a key step to calculating the ACF is how to take averages. For random media scattering, the average is usually taken over many realizations of random media or

rough surfaces [42][9]. For the purpose of target detection, however, the realization average may not be applicale since a target is usually associated with one realization of random medium and random rough surface. We have used frequency averaging and angular averaging to obtain the ACF. Numerical results have shown that detection of targets by ACF with frequency and angular averaging are alternative ways to obtaining averaged results.

1.2 Angular Correlation Function Versus Scattering Cross Section

To simply illustrate the detection methods for a target in clutter environment, we express the total scattering amplitude based on the coherent addition approximation[53]. For a medium consisting of a target and randomly distributed scatterers, the total scattering amplitude can be expressed in terms of the scattering amplitude of the target $F_t(\hat{k}_s, \hat{k}_i)$ and the scattering amplitudes of the random scatterers $f_n(\hat{k}_s, \hat{k}_i)$

$$F(\hat{k}_s, \hat{k}_i) = e^{i(\bar{k}_i - \bar{k}_s)\cdot\bar{r}_t} F_t(\hat{k}_s, \hat{k}_i) + \sum_{n=1}^{N} e^{i(\bar{k}_i - \bar{k}_s)\cdot\bar{r}_n} f_n(\hat{k}_s, \hat{k}_i) \tag{1.2}$$

where the positions are $\bar{r}_t$ for the target and $\bar{r}_n$ (n=1,2, $\cdots$, N) for random scatterers. To have a strong clutter effect, we assume that $|F_t| >> |f_n|$ and $\sum_{n=1}^{N} |f_n|^2 > |F_t|^2$. That is scattering by the target is much larger than a single "random" particle, but the summation of scattering intensity from the random scatterers is larger than that from the target.

Conventionally, the scattering cross section (RCS) is studied for target detection, which is

$$\sigma(\hat{k}_s, \hat{k}_i) \;=\; |F_t(\hat{k}_s, \hat{k}_i)|^2 + \sum_{n=1}^{N} |f_n(\hat{k}_s, \hat{k}_i)|^2 \tag{1.3}$$

Therefore, the clutter contribution is a positive number added to the signal in RCS measurement. The ensemble average is usually calculated over many realizations in experiments and numerical simulations, which may not be applicable in target detection.

However, the angular correlation function (ACF) is obtained by substituting (1.2) into (1.1), as given by

$$\Gamma(\hat{k}_{s1},\hat{k}_{i1};\hat{k}_{s2},\hat{k}_{i2}) = e^{i\Delta\bar{k}\cdot\bar{r}_t}F_t(\hat{k}_{s1},\hat{k}_{i1})F_t^*(\hat{k}_{s2},\hat{k}_{i2}) + \sum_{n=1}^{N} < e^{i\Delta\bar{k}\cdot\bar{r}_n} > f_n(\hat{k}_{s1},\hat{k}_{i1})f_n^*(\hat{k}_{s2},\hat{k}_{i2}) \quad (1.4)$$

In Eq. (1.4), the first term is the contribution from the target and the 2nd term is that from the clutter scattering. There is a phase term $e^{i\Delta\bar{k}\cdot\bar{r}_n}$, which is a random function since scatterers are randomly distributed. After taking an average, the phase term becomes zero except for $\Delta\bar{k} = \bar{0}$ for which we can obtain the "memory condition". The ACF becomes significantly large for the angular pairs that satisfy the memory condition and vanishes for $\Delta\bar{k} \neq \bar{0}$. Although Eq. (1.4) is shown for single scattering, the memory effect is exact and is true for multiple scattering [43] (see next chapter). Let $\bar{k}_{i1} = \bar{k}_{i2}$ and $\bar{k}_{s1} = \bar{k}_{s2}$ in Eq. (1.4), the ACF reduces to RCS as given by (1.3). Therefore, the ACF contains more information and is better than RCS for target detection. The average can be taken over frequencies, angles or their combinations for one realization of the medium.

1.3 An Overview of This Dissertation

As we see above, conventional detection and imaging methods do not work well for the detection of targets embedded in clutter environments. In this thesis, we study the ACF method for target detection based on theoretical analyses and numerical simulations.

In chapter 2, we study the ACF and its memory effects for random media scattering. First, we give the general derivation of the angualr correlation function based on the statitically translational invariance of Green's function. Next, we study the ACF and the memory effect for rough surfaces scattering and volume scattering based on analytical methods and numerical simulations. Then, averaging techniques are dis-

cussed and illustrated for the target detection. In this case, realization averaging may not be applicable, frequency averaging and angular averaging are needed.

In chapter 3, we study the ACF detection method for 2-D scattering problem in which a 2-D target is embedded in discrete scatterers under a 1-D rough surface. The problem is formulated by using surface integral equations for the rough surface and the surface of the target, volume integral equations for the random discrete scatterers. The equations are then solved numercally with the use of the method of moment. Frequency averaging is used to obtain the ACF and RCS. The results with and without the target are compared. It is found that ACF gives a ratio 10 dB larger than RCS.

In chapter 4, we study the ACF method for 3-D scattering problem in which a 3-D target is buried under a 2-D rough surface. The surface integral equations of wave fields are used for the rough surface and the surface of the buried object. The surface fields are then solved by the method of moments. The rough surface is discretized by a canonical grid, and the target sphere is discretized by using triangle patches. The scattered wave field from the object is represented by the rough surface field so that the matrix equation can be solved efficiently using the sparse-matrix canonical-grid (SMCG) method. Numerical simulations are illustrated for a perfectly conducting sphere buried under a 2-D rough surface. Both scattering coefficient and angular correlation function (ACF) are calculated based on azimuthal angular averaging. It is found that ACF is more effective in suppressing the clutter due to the rough surface scattering, and the cross-polarization components of ACF can be better than co-polarization components for target detection.

In chapter 5, we will show how to combine the traditional SAR imaging with the new ACF processing by conducting numerical simulations of 3-D scattering problems. The angular correlation imaging method is discussed and illustrated by using Monte Carlo simulations. One special case of correlation imaging is angular correlation imaging (ACF imaging). Angular correlation imaging consists of calculating ACF

with focusing to yield the angular correlation images and then averaging is taken over different images to obtain the image. Numerical simulations of ACF focusing and ACF imaging based single averaging and double averaging are shown.

In last chapter, we summary the research done for this dissertation and propose research topics for futher studies.

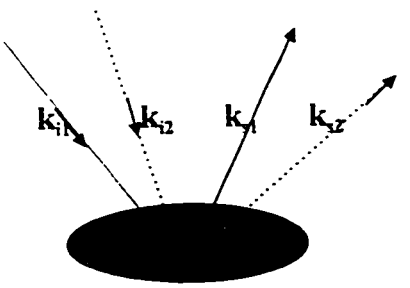


Figure 1.1: Configuration of angular correlation function for random medium scattering.

Chapter 2

ANGULAR CORRELATION FUNCTION OF RANDOM MEDIA SCATTERING

2.1 Introduction

The wave scattering by rough surfaces and random media such as turbulence and randomly distributed discrete scatterers has been studied extensively [34] [9] and [42]. Theoretical analysis, experimental measurements, and numerical methods have all been used for the problems. The statistical characteristics of the scattered wave fields have been studied.

But most of the works in the past had been mainly concentrated on the study of the scattering coefficient which represents the scattered wave intensity. It is true that correlation function and higher-order moments have also been studied. In the previous studies, however, the correlation function was calculated for two scattered wave fields with one wave incidence.

Since Feng first found the "memory effect" in 1988 [6], the study of angular correlation function attracts great attention. The angular correlation function is the correlation function of two scattered fields in directions θ_{s2} and θ_{s1} corresponding to two incident waves in directions θ_{i2} and θ_{i1}, respectively (Figure 2.1). Recent studies revealed that the angular correlation function of scattered fields by random media and rough surfaces can exhibit a strong correlation called angular memory effect [6] [19] [18] [10] [13] [43]. The memory line obeys the angular relation of $sin\theta_{s2} - sin\theta_{s1} = sin\theta_{i2} - sin\theta_{i1}$. It is analogous to a phase matching condition for reflection by a flat boundary.

For rough surface scattering, it has been shown that the values of the ACF are small except along the memory line of the incident and scattered directions. Random rough surfaces can be statistically translational invariant. Thus, it exhibits a phase matching condition when ensemble averages of the scattered fields are taken. This has been proved by using the statistically translational invariance of Green's function [43] and verified by theoretical models and experiments [10] [13]. For a single random rough surface, the memory effect of the ACF has also been shown by using frequency averaging [55].

For volume scattering, the phase matching condition needs to be satisfied in both horizontal and vertical directions, i.e. $sin\theta_{i1} - sin\theta_{s1} = sin\theta_{i2} - sin\theta_{s2}$ and $cos\theta_{i1} + cos\theta_{s1} = cos\theta_{i2} + cos\theta_{s2}$. The memory effect for volume scattering is different from that of a rough surface and the memory line is narrower and shorter [54] [14].

In this chapter, we study the ACF of wave scattering by rough surfaces and random media. First, we will give the general derivation of angular correlation function. Then, we study the ACF of wave scattering from randomly distributed scatterers by using analytical and numerical methods. For rough surface scattering, we use small perturbation method and numerical simulations to calculate the ACF. Last, three kinds of averaging techniques are illustrated and discussed.

2.2 General Derivation of Memory Effect for Random Media Scattering

An important property of angular correlation function of random media scattering is memory effect. The memory effect is a result of the generalized phase matching condition due to statistically translational invariance of clutter (volume scattering and rough surface scattering). In this section, we derive rigorously the general expression of angular correlation function and the phase matching condition for volume scattering based on statistically translational invariance of Green's function.

Let $G(\overline{r}, \overline{r}')$ be the exact Green's function for the boundary value problem that

includes volume scattering effects. Assuming statistically translational invariance, then the first and the second moment of the Green's function have

$$< G(\bar{r} + \Delta\bar{r}, \bar{r}' + \Delta\bar{r}) >=< G(\bar{r}, \bar{r}') > \tag{2.1}$$

and

$$< G(\bar{r}_1 + \Delta\bar{r}, \bar{r}'_1 + \Delta\bar{r})G^*(\bar{r}_2 + \Delta\bar{r}, \bar{r}'_2 + \Delta\bar{r}) >=< G(\bar{r}_1, \bar{r}'_1)G^*(\bar{r}'_2, \bar{r}'_2) > \tag{2.2}$$

where <> denotes ensemble average. Note that the translational invariance holds when averages are taken. It does not hold for each realization. This is the condition of statistically translational invariance.

Let the exact scattered field be ψ_s due to J_i. Then

$$\psi_s(\bar{r}) = \int d\bar{r}' G(\bar{r}, \bar{r}') J_i(\bar{r}') \tag{2.3}$$

Since the incident wave is a plane wave with $e^{\imath\bar{k}_i\cdot\bar{r}}$ dependence, we have

$$J_i(\bar{r}' + \Delta\bar{r}) = e^{\imath\bar{k}_i\cdot\Delta\bar{r}} J_i(\bar{r}') \tag{2.4}$$

The first moment, the coherent wave field, is

$$< \psi_s(\bar{r}) >= \int d\bar{r}' < G(\bar{r}, \bar{r}') > J_i(\bar{r}') \tag{2.5}$$

The incident field direction is $\bar{k}_i$. Displacement of $\bar{r}$ by $\Delta\bar{r}$ gives

$$\begin{aligned} < \psi_s(\bar{r} + \Delta\bar{r}) > &= \int d\bar{r}' < G(\bar{r} + \Delta\bar{r}, \bar{r}') > J_i(\bar{r}') \\ &= \int d\bar{r}' < G(\bar{r} + \Delta\bar{r}, \bar{r}' + \Delta\bar{r}) > J_i(\bar{r}' + \Delta\bar{r}) \\ &= \int d\bar{r}' < G(\bar{r}, \bar{r}') > J_i(\bar{r}' + \Delta\bar{r}) \\ &= e^{\imath\bar{k}_i\cdot\Delta\bar{r}} \int d\bar{r}' < G(\bar{r}, \bar{r}') > J_i(\bar{r}') \end{aligned} \tag{2.6}$$

The last equality is a result of the fact that the antenna J_i creates incident wave k_i. Also

$$< \psi_s(\bar{r} + \Delta\bar{r}) >= e^{\imath\bar{k}_s\cdot\Delta\bar{r}} < \psi_s(\bar{r}) > \tag{2.7}$$

Thus, substituting (2.7) into (2.6) and using (2.3) gives

$$k_{ix} = k_{sx}, \quad k_{iy} = k_{sy}, \quad k_{iz} = k_{sz} \tag{2.8}$$

i.e.

$$\begin{aligned} \sin\theta_i \cos\phi_i &= \sin\theta_s \cos\phi_s \\ \sin\theta_i \sin\phi_i &= \sin\theta_s \sin\phi_s \\ -\cos\theta_i &= \cos\theta_s \end{aligned} \tag{2.9}$$

Eq. (2.9) is the general form of phase matching condition for coherent wave. Within the scattering plane ($\phi_i = \phi_s = 0$ or π), we have

$$\begin{aligned} \sin\theta_i &= \sin\theta_s \\ -\cos\theta_i &= \cos\theta_s \end{aligned} \tag{2.10}$$

This is the Snell's law for wave propagation and reflection.

The second moment, correlation function, is the average of the product of two scattered fields, as given by

$$< \psi_s(\bar{r}_1)\psi_s{}^*(\bar{r}_2) >= \int d\bar{r}'_1 \int d\bar{r}'_2 < G(\bar{r}_1, \bar{r}'_1)G^*(\bar{r}_2, \bar{r}'_2) > J(\bar{r}'_1)J^*(\bar{r}'_2) \tag{2.11}$$

The two incident field directions are $\bar{k}_{i1}$ and $\bar{k}_{i2}$. Displacement of $\bar{r}$ and $\bar{r}'$ by $\Delta\bar{r}$ gives

$$\begin{aligned} < \psi_s(\bar{r}_1 &+ \Delta\bar{r})\psi_s{}^*(\bar{r}_2 + \Delta\bar{r}) > \\ &= \int d\bar{r}'_1 \int d\bar{r}'_2 < G(\bar{r}_1 + \Delta\bar{r}, \bar{r}'_1)G^*(\bar{r}_2 + \Delta\bar{r}, \bar{r}'_2) > J_{i1}(\bar{r}'_2)J^*_{i2}(\bar{r}'_2) \\ &= \int d\bar{r}'_1 \int d\bar{r}'_2 < G(\bar{r}_1 + \Delta\bar{r}, \bar{r}'_1 + \Delta\bar{r})G^*(\bar{r}_2 + \Delta\bar{r}, \bar{r}'_2 + \Delta\bar{r}) > \\ &\quad J_{i1}(\bar{r}'_1 + \Delta\bar{r})J^*_{i2}(\bar{r}'_2 + \Delta\bar{r}) \\ &= \int d\bar{r}'_1 \int d\bar{r}'_2 < G(\bar{r}_1, \bar{r}'_1)G^*(\bar{r}_2, \bar{r}'_2) > J_{i1}(\bar{r}'_1 + \Delta\bar{r})J^*_{i2}(\bar{r}'_2 + \Delta\bar{r}) \\ &= e^{i\bar{k}_{i1}\cdot\Delta\bar{r}}e^{-i\bar{k}_{i2}\cdot\Delta\bar{r}} \int d\bar{r}'_1 \int d\bar{r}'_2 < G(\bar{r}_1, \bar{r}'_1)G^*(\bar{r}_2, \bar{r}'_2) > J_{i1}(\bar{r}'_1)J^*_{i2}(\bar{r}'_2) \end{aligned} \tag{2.12}$$

The last equality is a result of the fact that the antenna J_{i1} creates incident wave k_{i1} and the antenna J_{i2} creates incident wave k_{i2}. Also

$$< \psi_s(\bar{r} + \Delta\bar{r})\psi_s{}^*(\bar{r}' + \Delta\bar{r}) >= e^{i(\bar{k}_{s1} - \bar{k}_{s2})\cdot\Delta\bar{r}} < \psi_s(\bar{r})\psi_s{}^*(\bar{r}') > \tag{2.13}$$

Thus, substituting (2.13) into (2.12) gives

$$\begin{aligned} k_{ix1} - k_{ix2} &= k_{sx1} - k_{sx2} \\ k_{iy1} - k_{iy2} &= k_{sy1} - k_{sy2} \\ k_{iz1} - k_{iz2} &= k_{sz1} - k_{sz2} \end{aligned} \tag{2.14}$$

i.e.

$$\begin{aligned} \sin\theta_{i1}\cos\phi_{i1} - \sin\theta_{i2}\cos\phi_{i2} &= \sin\theta_{s1}\cos\phi_{s1} - \sin\theta_{s2}\cos\phi_{s2} \\ \sin\theta_{i1}\sin\phi_{i1} - \sin\theta_{i2}\sin\phi_{i2} &= \sin\theta_{s1}\sin\phi_{s1} - \sin\theta_{s2}\sin\phi_{s2} \\ -\cos\theta_{i1} + \cos\theta_{i2} &= \cos\theta_{s1} - \cos\theta_{s2} \end{aligned} \tag{2.15}$$

Eq. (2.15) is the general form of phase matching condition of the 2nd-order angular correlation function. Within the scattering plane, it reduces to

$$\begin{aligned} \sin\theta_{i1} - \sin\theta_{i2} &= \sin\theta_{s1} - \sin\theta_{s2} \\ -\cos\theta_{i1} + \cos\theta_{i2} &= \cos\theta_{s1} - \cos\theta_{s2} \end{aligned} \tag{2.16}$$

In the same way above, we can derive the statistical phase matching condition of MXNth-order angular correlation function for random media scattering as follows

$$\begin{aligned} \sum_{m=1}^{M} k_{ixm} - \sum_{n=1}^{N} k_{ixn} &= \sum_{m=1}^{M} k_{sxm} - \sum_{n=1}^{N} k_{sxn} \\ \sum_{m=1}^{M} k_{iym} - \sum_{n=1}^{N} k_{iyn} &= \sum_{m=1}^{M} k_{sym} - \sum_{n=1}^{N} k_{syn} \\ \sum_{m=1}^{M} k_{izm} - \sum_{n=1}^{N} k_{izn} &= \sum_{m=1}^{M} k_{szm} - \sum_{n=1}^{N} k_{szn} \end{aligned} \tag{2.17}$$

Thus, the statistical phase matching condition for random media scattering has been proved. For the scatterers randomly distributed in a finite region, the statistically translational invaniance may not be satisfied. Next, we study the second-order angular correlation function of wave scattering by random scatterers distributed in a finite region and by random rough surfaces based on anlytical methods and numerical simulations.

2.3 *Angular Correlation Function of Wave Scattering from Randomly Distributed Scatterers*

Wave scattering from randomly distributed scatterers has been studied analytically, numerically, and experimentally [9], [42], and [12]. Independent scattering approximation and its generalized form – radiative tranfer theory have been used in the study of wave propagation and scattering in random media. Recently, wave approach has been used to obtain rigorous results for wave scattering from dense medium [17] vegetation [53]. The results obtained by using wave approach have provided satisfied explanation of new phonomena such as back-scattering enhancement and memory effect [41] [40] [10]. In this section, we will derive analytical solutions for the angular correlation function using the second-order approximation and conduct numerical simulations [14] [50].

2.3.1 Foldy-Lax Multiple Scattering Equation

We consider a case in which an electromagnetic (EM) field is incident on a volume composed of many infinitely long cylinders. The wave field satisfies the Foldy-Lax self-consistent equation, in which all the multiple scattering effects are included [15], [9], [42]. It is written as

$$\psi_{ex}^{(j)}(\bar{r}) = \psi_{inc}^{(j)}(\bar{r}) + \sum_{\ell=1,\ell\neq j}^{N} \psi_{s}^{(\ell)}(\bar{r}) \tag{2.18}$$

where the incident wave field is $\psi_{inc}^{(j)}(\bar{r})$, $\psi_{ex}^{(j)}(\bar{r})$ is the exciting field to the jth scatterer and $\psi_s^{(\ell)}(\bar{r})$ is the scattered field from the ℓth scatterer to the jth scatterer.

After Eq. (2.18) is solved, the total scattered field is the contributions from all of the scatterers as given by

$$\psi_s(\bar{r}) = \frac{e^{\imath kr}}{\sqrt{r}} F(\hat{k}_s, \hat{k}_i) \tag{2.19}$$

with a total scattering amplitude of

$$F(\hat{k}_s, \hat{k}_i) = \sum_{j=1}^{N} e^{-\imath \bar{k}_s \cdot \bar{r}_j} f_j(\hat{k}_s, \hat{k}_i) \psi_{ex}(\bar{r}_j) \tag{2.20}$$

where $f_j(\hat{k}_s, \hat{k}_i)$ is a scattering amplitude of jth cylinder for the incident wave vector $\hat{k}_i$ and scattered wave vector $\hat{k}_s$.

Next, we will solve Eq. (2.18) using (a) an analytical method and (b) Monte Carlo simulations. The scattered fields are obtained and the ACF of randomly distributed cylinders is calculated.

2.3.2 Analytical Method

In the analytical method, the total scattering amplitude is calculated iteratively up to the second-order. We assume that the incident wave is a plane wave. The exciting field can be expressed as

$$\psi_{ex}(\bar{r}_j) = e^{\imath \bar{k}_i \cdot \bar{r}_j} + \sum_{\ell=1, \ell \neq j}^{N} \frac{e^{ik|\bar{r}_j - \bar{r}_l|}}{\sqrt{|\bar{r}_j - \bar{r}_l|}} f_l(\hat{k}_s, \hat{k}_i) e^{\imath \bar{k}_i \cdot \bar{r}_l} \tag{2.21}$$

Substituting (2.21) into (2.20) gives the total scattering amplitude as

$$F(\hat{k}_s, \hat{k}_i) \cong \sum_{j=1}^{N} e^{i(\bar{k}_i - \bar{k}_s) \cdot \bar{r}_j} f_j(\hat{k}_s, \hat{k}_i) + \sum_{j=1}^{N} \sum_{\ell=1, \ell \neq j}^{N} \frac{e^{ik|\bar{r}_j - \bar{r}_l|}}{\sqrt{|\bar{r}_j - \bar{r}_l|}} f_j(\hat{k}_s, \hat{k}_{ji}) f_l(\hat{k}_{ls}, \hat{k}_i) e^{\imath \bar{k}_i \cdot \bar{r}_l} e^{-\imath \bar{k}_s \cdot \bar{r}_j} \tag{2.22}$$

The mutual correlation function is then obtained by taking an average of the product of (2.22) and its complex conjugate at two pairs of arguments. For the simplification of calculation, the cross terms are neglected since they are usually small due to random

phase. By taking into account the attenuation of wave propagation between two cylinders, we can express the mutual correlation function as

$$\begin{aligned}\Gamma(\hat{k}_{s1},\hat{k}_{i1};\hat{k}_{s2},\hat{k}_{i2}) \cong & < \sum_{j=1}^{N} f_j(\hat{k}_{s1},\hat{k}_{i1}) f_l^*(\hat{k}_{s2},\hat{k}_{i2}) e^{i(\bar{k}_{i1}-\bar{k}_{s1})\cdot\bar{r}_j} e^{-i(\bar{k}_{i2}-\bar{k}_{s2})\cdot\bar{r}_j} > \quad (2.23)\\ & + < \sum_{j=1}^{N}\sum_{\ell=1,\ell\neq j}^{N} f_j(\hat{k}_{s1},\hat{k}_{jl}) f_l(\hat{k}_{jl},\hat{k}_{i1}) f_j^*(\hat{k}_{s2},\hat{k}_{jl}) f_l^*(\hat{k}_{jl},\hat{k}_{i2})\\ & \frac{e^{-c|\bar{r}_j-\bar{r}_l|}}{\sqrt{|\bar{r}_j-\bar{r}_l|}} e^{i\bar{k}_{i1}\cdot\bar{r}_l} e^{-i\bar{k}_{s1}\cdot\bar{r}_l} e^{-i\bar{k}_{i2}\cdot\bar{r}_j} e^{i\bar{k}_{s2}\cdot\bar{r}_l} >\\ & + < \sum_{j=1}^{N}\sum_{\ell=1,\ell\neq j}^{N} f_j(\hat{k}_{s1},\hat{k}_{jl}) f_l(\hat{k}_{jl},\hat{k}_{i1}) f_j^*(\hat{k}_{s2},-\hat{k}_{jl}) f_l^*(-\hat{k}_{jl},\hat{k}_{i2})\\ & \frac{e^{-c|\bar{r}_j-\bar{r}_l|}}{\sqrt{|\bar{r}_j-\bar{r}_l|}} e^{i\bar{k}_{i1}\cdot\bar{r}_l} e^{-i\bar{k}_{s1}\cdot\bar{r}_l} e^{-i\bar{k}_{i2}\cdot\bar{r}_j} e^{i\bar{k}_{s2}\cdot\bar{r}_l} >\end{aligned}$$

where the transport attenuation coefficient c is obtained by the diffusion approximation [9] and given by

$$c = a + s(1-\overline{\mu}) \qquad (2.24)$$

with the absorption coefficient a, scattering coefficient s, and the mean cosine of the bistatic scattering cross section $\overline{\mu}$.

The use of the transport attenuation coefficient c of Eq. (2.24) is based on the comparisons of the analytical results with the numerical results and experimental results. While working on the problem, we tried the other two approaches: second - order scattering without attenuation and that with normal attenuation. The result without attenuation over-estimated the magnitude of the correlation function obtained by the exact method, while that with normal attenuation under-estimated the exact method. However, the result obtained by the second-order scattering with transport attenuation coefficient agrees very well with that of numerical and experimental results. It can be explained as follows: Attenuation of the coherent wave in random media is due to absorption and scattering by the scatterers. However, the part of the scattered wave close to the forward-scattering direction contains high

degrees of correlation and will not significantly contribute to the reduction (attenuation) of the mutual correlation function. To take into account this effect, we borrowed an idea from the diffusion approximation. We need to point out that the transport attenuation coefficient is chosen after conducting an extensive comparison between the second-order method and exact numerical calculations.

Eq. (2.23) is the general expression of the mutual correlation function up to second-order scattering. The first term is the correlation function of the first-order scattered wave. The second and third terms are due to the second-order scattering. They are called the ladder and cyclic terms, respectively. The ensemble average in Eq. (2.23) can be carried out with some approximations. First, we assume that the product of the scattering amplitudes and the exponential terms which contain scatterer positions are independent, and the ensemble average can be performed separately. In the rigorous formulation, the averaging of the scatterer positions and the scattering amplitude should be treated as a whole, but it is a difficult and complicated problem. To simplify the calculation, we de-couple the averaging into two parts and obtain an approximation. This is based on the fact the Green's function and scattering amplitude are not strongly correlated and the scatterer positions are independent of the scattering amplitude. The ensemble average of the scattering amplitudes is given by

$$\gamma_1 =< f(\hat{k}_{s1}, \hat{k}_{i1}) f^*(\hat{k}_{s2}, \hat{k}_{i2}) > \tag{2.25}$$

$$\gamma_{2L} =< f(\hat{k}_{s1}, \hat{k}_{jl}) f(\hat{k}_{jl}, \hat{k}_{i1}) f^*(\hat{k}_{s2}, \hat{k}_{jl}) f^*(\hat{k}_{jl}, \hat{k}_{i2}) > \tag{2.26}$$

$$\gamma_{2C} =< f(\hat{k}_{s1}, \hat{k}_{jl}) f(\hat{k}_{jl}, \hat{k}_{i1}) f^*(\hat{k}_{s2}, -\hat{k}_{jl}) f^*(-\hat{k}_{jl}, \hat{k}_{i2}) > \tag{2.27}$$

Secondly, the probability density function of the scatterer position can be written in terms of the volume and pair distribution function for the first- and second-order distributions as

$$p(\bar{r}) = \frac{1}{V} \tag{2.28}$$

$$p(\bar{r}_j, \bar{r}_l) = \frac{g(\bar{r}_j, \bar{r}_l)}{V^2} \tag{2.29}$$

Using the above approximations and Eq. (2.25)-(2.29), the normalized angular correlation function is given by

$$\begin{aligned}\Gamma(\hat{k}_{s1},\hat{k}_{i1};\hat{k}_{s2},\hat{k}_{i2}) &= \gamma_1\frac{n}{L_x}\int e^{i(\bar{k}_{i1}-\bar{k}_{s1})\cdot\bar{r}_j}\,e^{-i(\bar{k}_{i2}-\bar{k}_{s2})\cdot\bar{r}_j}\,d\bar{r}_j \\ &+ \gamma_{2L}\frac{n^2}{L_x}\int\int\frac{e^{-c|\bar{r}_j-\bar{r}_l|}}{\sqrt{|\bar{r}_j-\bar{r}_l|}}\,e^{i(\bar{k}_{i1}-\bar{k}_{i2})\cdot\bar{r}_l}e^{-i(\bar{k}_{s1}-\bar{k}_{s2})\cdot\bar{r}_j}\,g(\bar{r}_j,\bar{r}_l)d\bar{r}_j d\bar{r}_l \\ &+ \gamma_{2C}\frac{n^2}{L_x}\int\int\frac{e^{-c|\bar{r}_j-\bar{r}_l|}}{\sqrt{|\bar{r}_j-\bar{r}_l|}}\,e^{i(\bar{k}_{i1}+\bar{k}_{s2})\cdot\bar{r}_l}e^{-i(\bar{k}_{s1}+\bar{k}_{i2})\cdot\bar{r}_j}\,g(\bar{r}_j,\bar{r}_l)d\bar{r}_j d\bar{r}_l\end{aligned} \tag{2.30}$$

where the number density is n. and L_x is the horizontal size of the volume.

To simplify the integrals in Eq. (2.30), we let

$$\begin{aligned}\bar{r}_d &= \bar{r}_l-\bar{r}_j, \quad \bar{r}_c=\frac{1}{2}(\bar{r}_l+\bar{r}_j)\\ \Delta\overline{K}_d &= (\bar{k}_{i1}-\bar{k}_{s1})-(\bar{k}_{i2}-\bar{k}_{s2})\\ \overline{K}_L &= \frac{1}{2}(\bar{k}_{i1}-\bar{k}_{i2}+\bar{k}_{s1}-\bar{k}_{s2})\\ \overline{K}_C &= \frac{1}{2}(\bar{k}_{i1}+\bar{k}_{s2}+\bar{k}_{s1}+\bar{k}_{i2})\end{aligned}$$

With the above transformations. Eq. (2.30) becomes

$$\begin{aligned}\Gamma(\hat{k}_{s1},\hat{k}_{i1};\hat{k}_{s2},\hat{k}_{i2}) &= \gamma_1\frac{n}{L_x}\int e^{i\Delta\overline{K}_d\cdot\bar{r}_j}\,d\bar{r}_j \\ &+ \gamma_{2L}\frac{n^2}{L_x}\int e^{i\Delta\overline{K}_d\cdot\bar{r}_c}\,d\bar{r}_c\int\frac{e^{-cr_d}}{r_d}\,e^{i\overline{K}_L\cdot\bar{r}_d}\,g(r_d)d\bar{r}_d \\ &+ \gamma_{2C}\frac{n^2}{L_x}\int e^{i\Delta\overline{K}_d\cdot\bar{r}_c}\,d\bar{r}_c\int\frac{e^{-cr_d}}{r_d}\,e^{i\overline{K}_C\cdot\bar{r}_d}\,g(r_d)d\bar{r}_d\end{aligned} \tag{2.31}$$

For the scatterers distributed in a rectangular region with sides of L_x and L_z, the integral of the center coordinate $\bar{r}_c$ is given by

$$\begin{aligned}I_c &= \int_{-L_x/2}^{L_x/2}e^{i\Delta K_{dx}x}\,dx\int_{-L_z}^{0}e^{i\Delta K_{dz}z+cz/\alpha}\,dz \\ &= L_xL_z\,sinc\,(\Delta K_{dx}L_x/2)\left(1-e^{i\Delta K_{dz}L_z-cL_z/\alpha}\right)/\left(cL_z/\alpha+i\Delta K_{dz}L_z\right)\end{aligned} \tag{2.32}$$

where $\frac{1}{\alpha}=\left(\frac{1}{cos\theta_{i1}}+\frac{1}{cos\theta_{s1}}+\frac{1}{cos\theta_{i2}}+\frac{1}{cos\theta_{s2}}\right)/2$. From Eq. (2.32), it is clear that the correlation function is non-zero if $\Delta K_{dx}=0$. Since ΔK_{dx} is a function of angles and

frequencies, there exists certain relationships which give $\Delta K_{dx} = 0$. If we fix the frequency, Eq (2.31) becomes ACF and we have a relationship given by

$$sin\theta_{i1} - sin\theta_{s1} = sin\theta_{i2} - sin\theta_{s2} \tag{2.33}$$

For a fixed pair of $(\theta_{i1}, \theta_{s1})$, $sin\theta_{i2}$ and $sin\theta_{s2}$ are linearly related and on a line called "memory line" which is the result of statistically translational invariance in the horizontal direction.

It is noticed that the wave number in the vertical direction also affects the ACF of volume scattering. When the attenuation is negligibly small, the L_z dependence of Eq. (32) becomes another *sinc* function, which results in

$$cos\theta_{i1} + cos\theta_{s1} = cos\theta_{i2} + cos\theta_{s2} \tag{2.34}$$

Eqs. (2.33) and (2.34) give two points in the θ_{i2}-θ_{s2} plane as their solution, which we will call "memory dots". Away from these two dots, the ACF is very small. This is the result of a statistically translational invariance condition satisfied in both horizontal and vertical directions. When the attenuation becomes substantial, however, the scattered fields from different z-positions will undergo different amount of attenuation at the surface (observation point). In this case the tranlational invariance in vertical direction given in Eq.(2.34) will not be satisfied. Similar to the rough surface case, the memory line becomes broad as the volume concentration of the medium increases.

The integral over the difference coordinate $\bar{r}_d$ can be simplified as

$$\begin{aligned} I_d &= \int \frac{e^{-c r_d}}{r_d} e^{i\bar{K}\cdot\bar{r}_d} g(r_d) d\bar{r}_d \\ &= 2\pi \int e^{-c r_d} J_o(K r_d) g(r_d) dr_d \end{aligned} \tag{2.35}$$

Since the Bessel function and the exponential function vanish as the argument becomes very large, the main contribution in Eq. (35) comes from the region of

small Kr_d. When the attenuation is small, r_d can be large, which requires K to be small. Therefore, there are two points with large correlation on the memory line, which correspond to $\theta_{i2} = \theta_{i1}$ and $\theta_{s2} = \theta_{s1}$ for the ladder term, and $\theta_{s2} = -\theta_{i1}$ and $\theta_{i2} = -\theta_{s1}$ for the cyclic term. When the attenuation is large, however, r_d needs to be small, which allows K to be large. So, we can see a broad line for the second-order scattering. In this case, the second-order scattering is dominated by the pairs of scatterers closed to each other.

Substituting (32) and (35) into (31) gives

$$\begin{aligned}\Gamma(\hat{k}_{s1}, \hat{k}_{i1}; \hat{k}_{s2}, \hat{k}_{i2}) &= [\gamma_1 n + \gamma_{2L} n^2 2\pi \int e^{-c r_d} J_o(K_L r_d) g(\bar{r}_d) dr_d \\ &+ \gamma_{2C} n^2 2\pi \int e^{-c r_d} J_o(K_C r_d) g(r_d) dr_d] \\ &\cdot (1 - e^{i\Delta K_{dz} L_z - c L_z/\alpha})/(c/\alpha + i\Delta K_{dz}) \\ &\quad sinc(\Delta K_{dx} L_x/2)\end{aligned} \tag{2.36}$$

2.3.3 Numerical Simulations

In Monte Carlo simulations, we solve the Foldy-Lax equation exactly for wave scattering from the randomly distributed cylinders for each realization. Then, the angular correlation function is obtained by taking averages over all the realizations.

First, we expand the incident plane wave $\psi_{inc}(\bar{r}) = e^{i\bar{k}_i \cdot \bar{r}}$, as an expansion of cylindrical wave function as

$$\psi_{inc}^{(j)}(\bar{r}) = e^{i\bar{k}_i \cdot \bar{r}_j} \sum_{n=-\infty}^{\infty} a_n^{(j)} J_n(k_0|\bar{r} - \bar{r}_j|) e^{in\phi_{rr_j}} \tag{2.37}$$

where $a_n^{(j)} = (-1)^n e^{-in\phi_j - in\pi/2}$.

Then, we write the exciting field and the internal field for cylinder j as the series expansion of the cylindrical wave functions as follows.

Exciting field

$$\psi_{ex}^{(j)}(\bar{r}) = \sum_{n=-\infty}^{\infty} b_n^{(j)} J_n(k_0|\bar{r} - \bar{r}_j|) e^{in\phi_{rr_j}} \tag{2.38}$$

Internal field

$$\psi_{int}^{(j)}(\bar{r}) = \sum_{n=-\infty}^{\infty} c_n^{(j)} J_n(k_1|\bar{r} - \bar{r}_j|)e^{in\phi_{rr_j}} \tag{2.39}$$

Substituting(2.39) and the Green's function into the integral expression of Huygens' principle. we obtain the scattered field in terms of internal coefficient

$$\psi_s^{(\ell)}(\bar{r}) = \sum_{n=-\infty}^{\infty} A_n^{(\ell)} H_n^{(1)}(k_0|\bar{r} - \bar{r}_\ell|)e^{in\phi_{rr_j}} c_n^{(\ell)} \tag{2.40}$$

with $A_n^{(\ell)} = \frac{i\pi a_\ell}{2}[J_n(k_1 a_\ell)J_n'(k_0 a_\ell)k_0 - J_n(k_0 a_\ell)J_n'(k_1 a_\ell)k_1]$.

By employing the extinction theorem, and using the expansion of Green's functions. we have the relation between the exciting coefficient and the internal coefficient

$$b_n^{(\ell)} = B_n^{(\ell)} c_n^{(\ell)} \tag{2.41}$$

with $B_n^{(\ell)} = \frac{i\pi a_\ell}{2}[J_n(k_1 a_\ell)H_n^{(1)\prime}(k_0 a_\ell)k_0 - H_n^{(1)}(k_0 a_\ell)J_n'(k_1 a_\ell)k_1)]$

Substituting (2.37),(2.38), and (2.40) into (2.18), we get

$$\begin{aligned}\sum_{n=-\infty}^{\infty} b_n^{(j)} J_n(k_0|\bar{r} - \bar{r}_j|)e^{in\phi_{rr_j}} &= e^{i\bar{k}_i\cdot\bar{r}_j} \sum_{n=-\infty}^{\infty} a_n^{(j)} J_n(k_0|\bar{r} - \bar{r}_j|)\, e^{in\phi_{rr_j}} \\ &+ \sum_{\ell=1,\ell\neq j}^{N} \sum_{n'=-\infty}^{\infty} A_{n'}^{(\ell)} H_{n'}^{(1)}(k_0|\bar{r} - \bar{r}_\ell|)e^{in'\phi_{rr_\ell}} c_{n'}^{(\ell)}\end{aligned} \tag{2.42}$$

Using equation (2.41) and the translation addition theorem.

$$H_{n'}^{(1)}(k_0|\bar{r} - \bar{r}_\ell|)e^{in'\phi_{rr_\ell}} = \sum_{n=-\infty}^{\infty} H_{n-n'}^{(1)}(k_0|\bar{r} - \bar{r}_j|)\, J_n(k_0|\bar{r} - \bar{r}_j|)e^{in\phi_{rr_j} - i(n-n')\phi_{r_\ell r_j}} \tag{2.43}$$

in equation (2.42), we obtain

$$B_n^{(j)} c_n^{(j)} = e^{i\bar{k}_i\cdot\bar{r}_j} a_n^{(j)} + \sum_{\ell=1,\ell\neq j}^{N} \sum_{n'=-\infty}^{\infty} A_{n'}^{(\ell)} H_{n-n'}^{(1)}(k_0|\bar{r} - \bar{r}_j|)\, e^{-i(n-n')\phi_{r_\ell r_j}} c_{n'}^{(\ell)} \tag{2.44}$$

It can be written in a matrix form $[Z][c] = [a]$. After $c_n^{(\ell)}$ is solved, the total scattered field can be obtained and the total scattering amplitude is given by

$$F(\theta_s, \theta_i) = \sqrt{\frac{2}{\pi k}} e^{-i\pi/4} \sum_{\ell=1}^{N} e^{-i\bar{k}_s\cdot\bar{r}_\ell} \sum_{n=-\infty}^{\infty} (-i)^n A_n^{(\ell)} e^{in\phi_{rr_\ell}} c_n^{(\ell)} \tag{2.45}$$

The ACF can then be calculated as follows:

$$\Gamma(\theta_{s1},\theta_{i1};\theta_{s2},\theta_{i2}) = \sum_{n_r=1}^{N_r} F_{n_r}(\theta_{s1},\theta_{i1})F^*_{n_r}(\theta_{s2},\theta_{i2})/L_x/N_r \qquad (2.46)$$

where the realization index is n_r, and N_r is the total number of realizations used for ensemble averaging.

2.3.4 Results

The ACF is calculated for horizontally polarized EM wave scattering by cylinders with a diameter of 3 mm at a frequency of 100 GHz. The relative dielectric constant of the cylinders is $\epsilon_r = 4.5 + i0.2$. The albedo of the cylinder scattering is 0.6127. The scattering configuration of the problem is shown in Figure 2.1. Figure 2.2 shows the analytical ACF given in Eq. (2.36) as a function of the variable angles θ_{i2} and θ_{s2}. A high level of correlation is shown along the memory line, and the two peaks correspond to the auto-correlation and reciprocity points on the memory line. Similar to the rough surface case, the memory line is very narrow and the ACF decreases rapidly away from the memory line.

In Figures 2.3 and 2.4, the ACF along the memory line is shown for the reference angles of ($\theta_{i1} = 20^o$ and $\theta_{s1} = -40^o$) and for the fractional volume densities of 1, 5, 10, and 30%. In Figure 2.3, the contribution from the first-order, second-order ladder, and second-order cyclic terms are plotted separately. As expected, the first-order is dominant for a low density case. The second-order ladder and cyclic terms give a rise at the auto-correlation and reciprocity points, respectively. For high density cases such as Figure 2.3c and d, the contribution from the second-order terms becomes significant. In addition, the first-order term shows only a broad response, and two peaks visible for a low density case (Figure 2.3a) effectively disappear. The attenuation coefficient of the reduced coherent wave is usually given by the sum of the absorption and scattering coefficients. In order to obtain good agreement between the second-order solution and numerical simulations, we found that the transport

attenuation coefficient must be modified to include the effects of diffuse intensity as shown in Eq. (2.24). This is because the part of the scattered wave close to the forward-scattering direction does not really reduce the correlation since the phase information is preserved. The mean cosine $\overline{\mu}$ of a 3 mm cylinder is approximately 0.546. This approach combines the benefits of second-order scattering approximation and the diffusion approximation. The coherent interaction is still preserved in the second-order approximation and the over-estimation of the second-order scattering effect is corrected by using the modified transport attenuation coefficient.

Figure 2.4 shows the comparison with numerical simulations and millimeter wave (MMW) experiments. Because of limited computer resources, the numerical simulations were conducted for 1, 5, 10 and 30% cases only. The positions of 100 cylinders are randomly generated in a rectangular region of L_x X L_z with $L_x = 9cm$. The result is averaged over 50 realizations. Unlike the numerical simulations, the analytical results do not have the middle bump. This difference may be due to the lack of higher-order terms in the analytical solution. MMW experiments were conducted with test samples which were fabricated using polystyrene molds and glass rods with a diameter of 3 mm and a variance of ± 0.4 mm [50]. The dimensions of the sample holders are 40 cm long and 25 cm high for 1, 5, 10 and 30% samples, respectively. Using independent scattering assumption, the optical thickness (τ) of the samples was found to be 1.4367, 7.1835, and 14.367 for 1, 5, 10 and 30% at 100 GHz, respectively. A MMW bistatic scatterometer based on the HP8510C MMW vector network analyzer was used for the experiments. A detailed description of the MMW system can be found in [30].

We have obtained the analytical solution for the ACF based on the modified second-order approximation. The approach is based on the coherent summation of the scattered waves which preserves the interference effects such as backscattering enhancement. Two main assumptions have been used in the derivation.

The first assumption is that we include the transport attenuation coefficient in

the second-order scattering rather than the normal attenuation coefficient for coherent wave. The adoption of the transport attenuation coefficient was done after conducting extensive comparisons between the proposed analytical solution and the numerical and experimental results. The conventional second-order scattering approximation without attenuation over-estimates the correlation function. while that with normal attenuation under-estimates the correlation function. The modification of the attenuation coefficient includes the contribution due to incoherent wave. By comparison with the exact numerical simulations. we found that the correction factor is given by $(1-\overline{\mu})$ which also appears in the diffusion equation.

Another assumption is that the averaging of the scatterer positions and the scattering amplitude can be performed separately. In the exact formulation, the averaging of the scatterer positions and the scattering amplitude should be treated as a whole. To simplify the calculation. we de-coupled the averaging into two parts. This is based on the fact that the Green's function and scattering amplitude are not strongly correlated and the scatterer positions are independent to the scattering amplitude. The present approach gives good agreements with numerical and experimental results.

2.4 Angular Correlation Function for Rough Surface Scattering

Wave scattering by random rough surfaces has been studied extensively in terms of scattering coefficient. Recently, the angular correlation function of the scattered wave field by rough surfaces has attracted great attention and has been studied based on analytical, numerical and experimental methods [18], [19], [22], [10], [13], [43]. The analytical method used for the angular correlation function of rough surface scattering was based on the Kirchhorff approximation which is a good approach when the surface has a large radius of curvature compared with the incident wavelength. In thsi section, we study the angular correlation function for the rough surfaces with small roughness using small perturbation method [9], [42]. The results are also compared with that

obtained by numerical simulations.

2.4.1 *Analytical Expression*

Consider a plane wave incident on an one dimensional rough surface

$$\psi_i = e^{\imath k_{ix} x - \imath k_{iz} z} \tag{2.47}$$

where $k_{ix} = k_o sin\theta_i$, and $k_{iz} = k_o cos\theta_i$. Let k_o and k_1 be the wave numbers of the upper and lower region, respectively.

By using the first order small perturbation method, the scattered field is

$$\psi_{s\alpha}^{(1)}(\overline{r}) = \frac{e^{\imath k_o r}}{\sqrt{r}} \psi_{s\alpha}^{(1)N}(\hat{k_s}, \hat{k_i}) \tag{2.48}$$

where

$$\psi_{s\alpha}^{(1)N}(\overline{r}) = \frac{1}{4}\sqrt{\frac{2}{\pi k_o}} e^{-\imath\frac{\pi}{4}} \int dx' \int dk_x e^{i(k_x - k_{sx})x'} P_\alpha(k_x, k_{ix}, k_{sx}) F(k_x - k_{ix}) \tag{2.49}$$

In (2.49), $F(k_x)$ is the Fourier transform of the random height of the surface, and $\alpha = h, v$ denotes the polarizations. Also nd dielectric

$$P_h(k_x, k_{ix}, k_{sx}) = \frac{(\epsilon_r - 1)k_o^2}{(k_{iz} + k_{tz})(k_{oz} + k_{1z})} 2ik_{iz}(k_{oz} + k_{sz}) \tag{2.50}$$

$$P_v(k_x, k_{ix}, k_{sx}) = \frac{(\epsilon_r - 1)(k_{tz}k_{1z} - \epsilon_r k_{ix}k_x)}{(\epsilon_r k_{iz} + k_{tz})(\epsilon_r k_{oz} + k_{1z})} 2ik_{iz}(k_{oz} + k_{sz}) \tag{2.51}$$

where $k_{oz} = \sqrt{k_o^2 - k_x}$, $k_{sz} = \sqrt{k_o^2 - k_{sx}}$, $k_{1z} = \sqrt{k_1^2 - k_x}$ and $k_{tz} = \sqrt{k_1^2 - k_{ix}}$.

Substituting Eq.(2.49) and its conjugate into Eq.(1.1), we obtain the mutual correlation function of wave scattering from rough surface

$$\begin{aligned} \Gamma_\alpha^{(S)}(1,2) &= < \psi_{s\alpha}^{(1)N}(\hat{k}_{i1}, \hat{k}_{s1}\psi_{s\alpha}^{(1)N}(\hat{k}_{i2}, \hat{k}_{s2} >_r / L \\ &= \frac{1}{8\pi L}(k_{o1}k_{o2})^{-1/2} \int\int dx'dx'' \int\int dk_x' dk_x'' e^{i(k_x' - k_{s1x})x' - i(k_x'' - k_{s2x})x''} \\ &\quad P_\alpha(k_x', k_{i1x}, k_{s1x}) P_\alpha^*(k_x'', k_{i2x}, k_{s2x}) < F(k_x' - k_{i1x}) F^*(k_x'' - k_{i2x}) >_r \end{aligned} \tag{2.52}$$

For statistically homogenous random rough surfaces, the random spectrum has the property

$$< F(k'_x - k_{i1x})F^*(k''_x - k_{i2x}) >_r = W(k'_x - k_{i1x})\delta[(k'_x - k_{i1x}) - (k''_x - k_{i2x})] \quad (2.53)$$

where W is the spectrum of the rough surface, and $<>_r$ denotes realization average. The condition of Eq.(2.53) is also known as statistically translational invariance. Making transformations

$$\begin{aligned} x_c &= (x' + x'')/2, & x_d &= x' - x'', \\ k_{icx} &= (k_{i1x} + k_{i2x})/2, & k_{idx} &= k_{i1x} - k_{i2x}, \\ k_{scx} &= (k_{s1x} + k_{s2x})/2, & k_{sdx} &= k_{s1x} - k_{s2x} \end{aligned}$$

in (2.53), we perform the integrals and obtain

$$\begin{aligned} \Gamma_\alpha^{(S)}(1,2) &= \frac{1}{4}(k_{o1}k_{o2})^{-1/2} e^{ik_{o1}r_1 - ik_{o2}r_2} sinc[(k_{idx} - k_{sdx})L/2] \\ &\quad P_\alpha(k_{scx} + \frac{1}{2}k_{idx}, k_{i1x}, k_{s1x}) P_\alpha^*(k_{scx} - \frac{1}{2}k_{idx}, k_{i2x}, k_{s2x}) W(k_{scx} - k_{icx}) \end{aligned} \quad (2.54)$$

where L is the length of the rough surface that correspond to the extent of the rough surface under the antenna beam width. The angular correlation function and the frequency correlation function are special cases of equation (2.54).

The angular correlation function is obtained by setting $k_{o1} = k_{o2}$ in (2.54). It can be seen that angular correlation can only exists when $|k_{idx} - k_{sdx}|L$ is small. In the limit of infinite L, the $\sin x/x$ function becomes a delta function with nonzero values at

$$\sin\theta_{i1} - \sin\theta_{i2} = \sin\theta_{s1} - \sin\theta_{s2} \quad (2.55)$$

For fixed θ_{i1} and θ_{s1}, $\sin\theta_{i2}$ and $\sin\theta_{s2}$ are linearly related by Eq. (2.55). This linear relation is called memory line. The angular width of the memory line is the order of $\frac{\lambda}{L}$.

The frequency correlation function can be obtained from (2.54) by setting $\theta_{i1} = \theta_{i2}$ and $\theta_{s1} = \theta_{s2}$. Due to the *sinc* function in Eq.(2.54), the frequency correlation is only

appreciable when $|k_{idx} - k_{sdx}|L$ is small. This shows the width in frequency domain of

$$\delta f = \frac{1}{\sin\theta_i - \sin\theta_s}\frac{\lambda_o}{L}f_o \tag{2.56}$$

To get independent samples, the difference of the two frequencies must be large enough so that $|f_1 - f_2| > |\delta f|$ satisfied.

2.4.2 Results

The angular correlations defined by (2.54) are calculated for the fixed reference angles of θ_{i1} and θ_{s1}. We plot the angular correlation along the memory line for $h = 0.05\lambda_o$, $l = 0.1\lambda_o$, and $L = 10\lambda_o$ and show the results in Figure 2.5 and 2.6.

Figure 2.5 shows the comparisons of angular correlation between horizontally and vertically polarized cases for dielectric surfaces. The dielectric constant is obtained for the soil by using the empirical model [47] for the moisture content of 5% and 30%. They are (3.17.0.5) and (16.4.3.7), respectively. The reference angles $(\theta_{i1}, \theta_{s1})$ are $(20^o, 40^o)$ for figure 2.5a, and $(20^o, -20^o)$ for figure 2.5b. The correlation for TM case is not necessary bigger than that for TE case, but dependent on the references angles and the soil moisture.

Figure 2.6 shows the comparisons of the results obtained by the small perturbation method and that by numerical simulations. They agree well. Results show that along the memory line, the angular correlation function of wave scattering from slightly rough surfaces has broad angular dependence instead of the two peaks that are exhibited for very rough surface and random media.

2.5 Averaging Techniques

In the studies of waves in random media, it is customary to use realization averaging to calculate the scattering cross section. However, for target detection in which a target is always surrounded by one realization of random medium or under a single

rough surface, the realization averaging may not be applicable and other means of taking a coherent average must be investigated. In this section, we will define the ACF of the scattered wave based on realization, frequency, and angular averaging methods. It is possible to obtain the ACF for wave scattering from one realization of medium if the data are available over a wide frequency bandwidth or angular range.

2.5.1 *Definitions of Average Techniques*

Realization averaging: The ensemble average is obtained by taking averaging over different samples (rough surfaces) with the same statistics.

$$\Gamma_r(\theta_{s1},\theta_{i1};\theta_{s2},\theta_{i2}) = \frac{1}{N_r}\sum_{n=1}^{N_r} \psi_s^N(\theta_{s1},\theta_{i1},n)\psi_s^{N*}(\theta_{s2},\theta_{i2},n)/\sqrt{P_1P_2} \tag{2.57}$$

where n denotes the realization index, and N_r is the number of realizations. P_1 and P_2 are total power flux of the two incident waves, respectively. In general, many independent samples (realizations) must be generated to obtain stable result.

Frequency averaging: Frequency averaging takes an ensemble average over a frequency bandwidth Δf centered at f_o.

$$\Gamma_f(\theta_{s1},\theta_{i1};\theta_{s2},\theta_{i2}) = \frac{1}{N_f}\sum_{n=1}^{N_f} \psi_s^N(\theta_{s1},\theta_{i1};f_n)\psi_s^{N*}(\theta_{s2},\theta_{i2};f_n)/\sqrt{P_1P_2} \tag{2.58}$$

where N_r is the number of frequencies over the frequency range $f_o - \Delta f < f_n < f_o + \Delta f$, and n is the frequency index for f_n.

Angular averaging: The angular averaging is defined by small changes of incident and scattering angles around the fixed angles.

$$\Gamma_a(\theta_{s1},\theta_{i1};\theta_{s2},\theta_{i2}) = \frac{1}{N_a}\sum_{n=1}^{N_a} \psi_s^N(\theta_{s1}+\delta_n,\theta_{i1}+\delta_n)\psi_s^{N*}(\theta_{s2}+\delta_n,\theta_{i2}+\delta_n)/\sqrt{P_1P_2} \tag{2.59}$$

where N_a is the number of the angles, and δ_n is the small angular difference for index n.

Averaging with weighting: Since the scattered waves are usually frequency and angle dependent, it is better to use a weighting function in the averaging. The advantage of using weighting function is to make each sample is equally weighted so

that the maximum number of independent samples can be obatained. The detailed illustration and numerical results can be found in the reference [51].

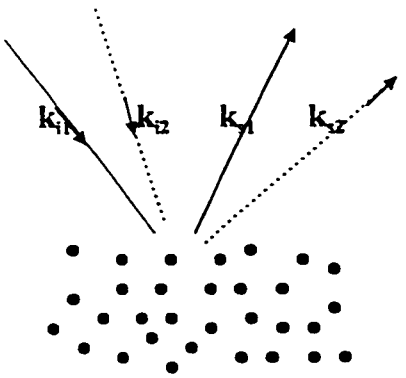
2.5.2 Numerical results

Figure 2.7 shows the ACF magnitude with different ensemble averaging methods. The random rough surface is generated by using the spectrum method [35] with a Gaussian height distribution and Gaussian correlation function. The test surface was generated with a random number and results for different surfaces with the same statistics are quite similar. In numerical simulations, a tapered plane wave is used as an incident wave to eliminate the effect of the rough surface. The tapering parameter g is chosen to be $L/4$ in the simulations. Figure 2.7a is the result for one rough surface without any averaging. Figure 2.7b is that of realization averaging taken over 100 rough surfaces. It is clear that the existence of the memory line becomes apparent if a sufficient number of independent samples are included. Figure 2.7c shows the ACF magnitude of a single rough surface based on the frequency averaging method with 50 equally spaced samples over the frequency bandwidth of $0.5f_o$ to $1.5f_o$. The large ridges in Figure 2.7c correspond to the ψ_2 measured in the forward, including specular, direction. Although the ACF magnitude by frequency averaging is noisier than that of realization averaging, a distinct memory line is clearly visible. To suppress fluctuation in Figure 2.7c, more independent samples must be included in the averaging process. It was estimated that only about 10 independent samples can be obtained with the bandwidth of $0.5f_o$ to $1.5f_o$ using Eq.(2.58). For the frequency averaging to be effective, a much wider bandwidth, which is available in ultra-wideband radars, may be required. Figure 2.7d shows the ACF magnitude by the angular averaging method given in Eq.(2.59). The results are smoother than those without averaging (Figure 2.7a), but the memory line is not as clearly visible as that of frequency averaging. Using SPM, we estimated that only 3 independent samples are available within a 20^o angular range for the value of L that we have chosen.

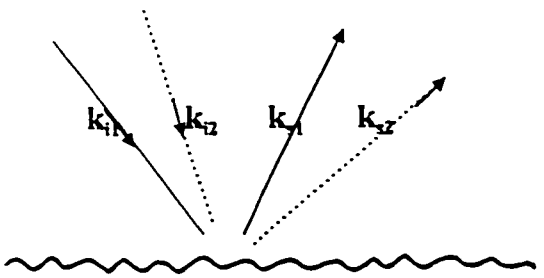
2.6 *Conclusions and Discussions*

In this chapter. we have studied the angular correlation function of scattered wave fileds from random media and random rough surfaces. The general expressions of ACF are first derived based on the statistically translational invariance of Green's function. It is shown that the memory effect of ACF is a result of phase-matching condition. The ACF of wave scattering from randomly distributed scatterers and that from random rough surfaces are then studied using analytical methods and numercal simulation. We have shown that the memroy effect exhibits a broad line for rough surface scattering and memory dots for for volume scattering, which corresponds to self-correlation and reciprocity path. Away from the memory effect. the ACF is very small (theoretically zero), which can be used for suppressing clutter. We also studied different averaging techniques: realization averaging, frequency averaging, and angular averaging, for the calculation of ACF. It is shown that frequency averaging also show memory effect. angular averaging smooths out the clutter. and weighting function can be helpful to obtain enough independent samples.

We have shown that the ACF of wave scattering from different medium exhibits different memory effect depending on the statistical homogenity. ACF can be used to ditinguish different kinds of clutter (randam media or rough surfaces). Since wave scattering from a deterministic target does not have the statistically translational invariance, ACF for target scattering is different from that of clutter. Therefore. ACF can be used for the detection of targets embedded in clutter environment.



(a)



(b)

Figure 2.1: Configuration of angular correlatio function for wave scattering from randomly distributed scatterers and random rough surfaces

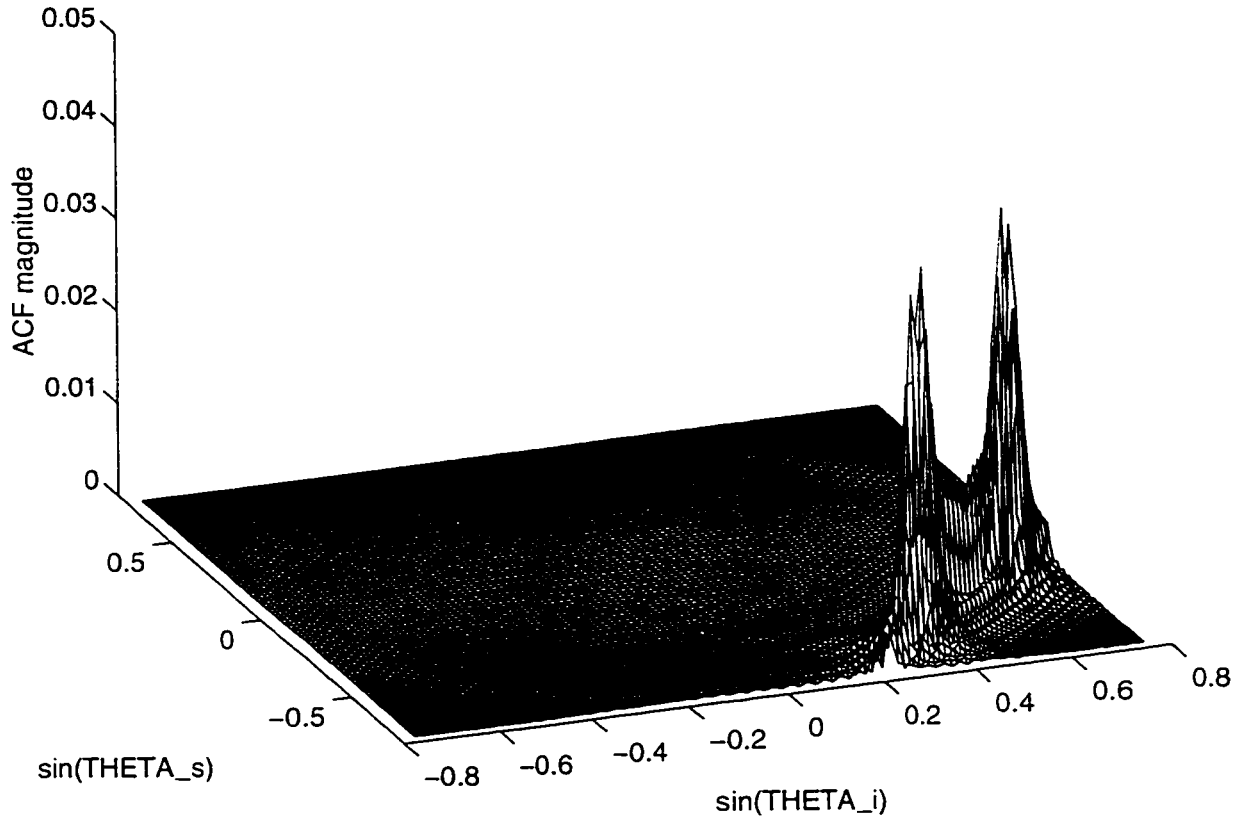


Figure 2.2: 3-D plot of analytical results of ACF for $f_v = 10\%$ at the reference angles $(\theta_{i1}, \theta_{s1}) = (20°, -40°)$. Results are shown as a function of the $sin(\theta_{i2})$ and $sin(\theta_{s2})$.

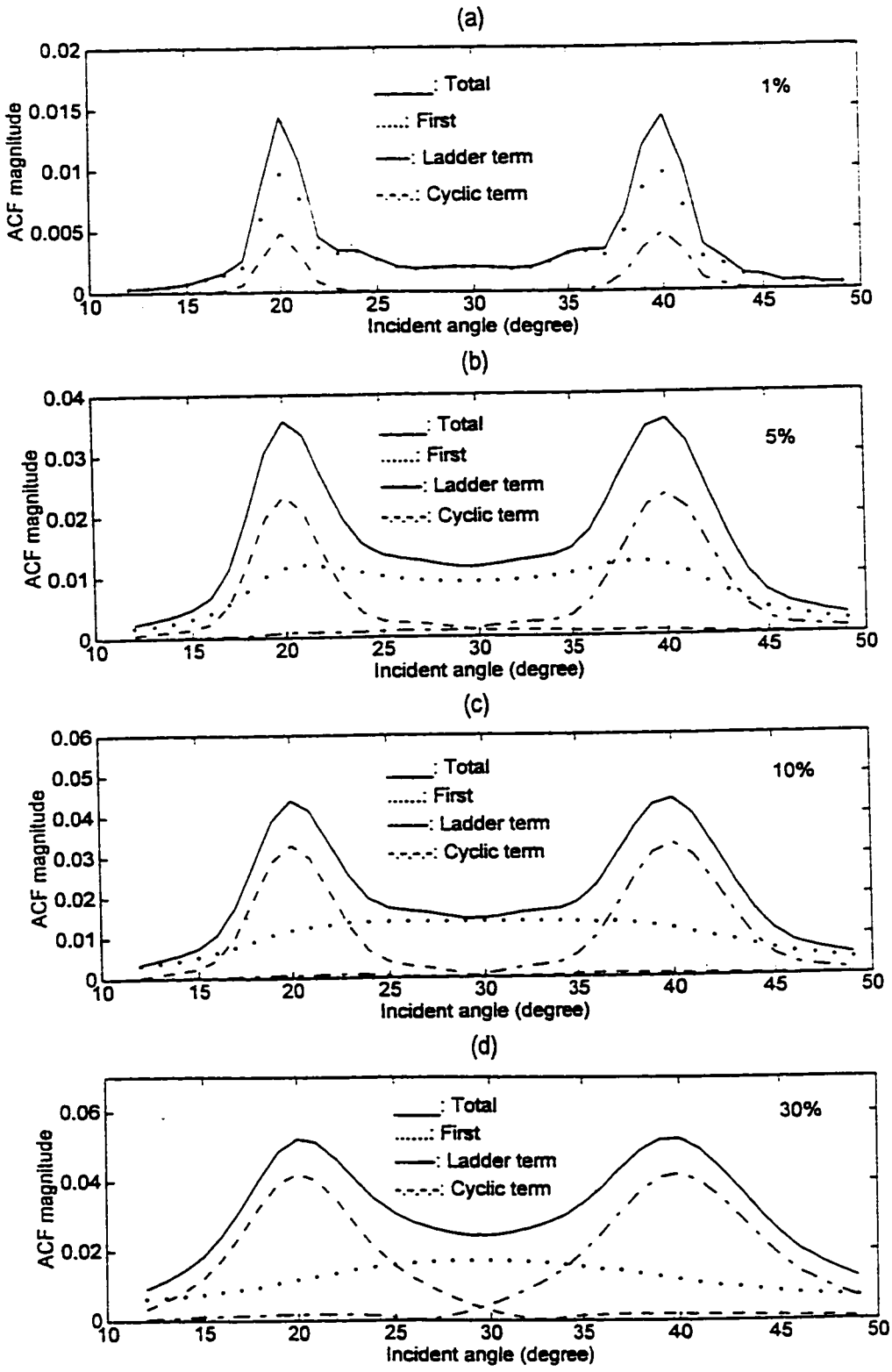


Figure 2.3: Analytical solutions of ACF for the volume density $f_v = 1, 5, 10$, and 30%. The reference angles are at $(\theta_{i1}, \theta_{s1}) = (20^o, -40^o)$.

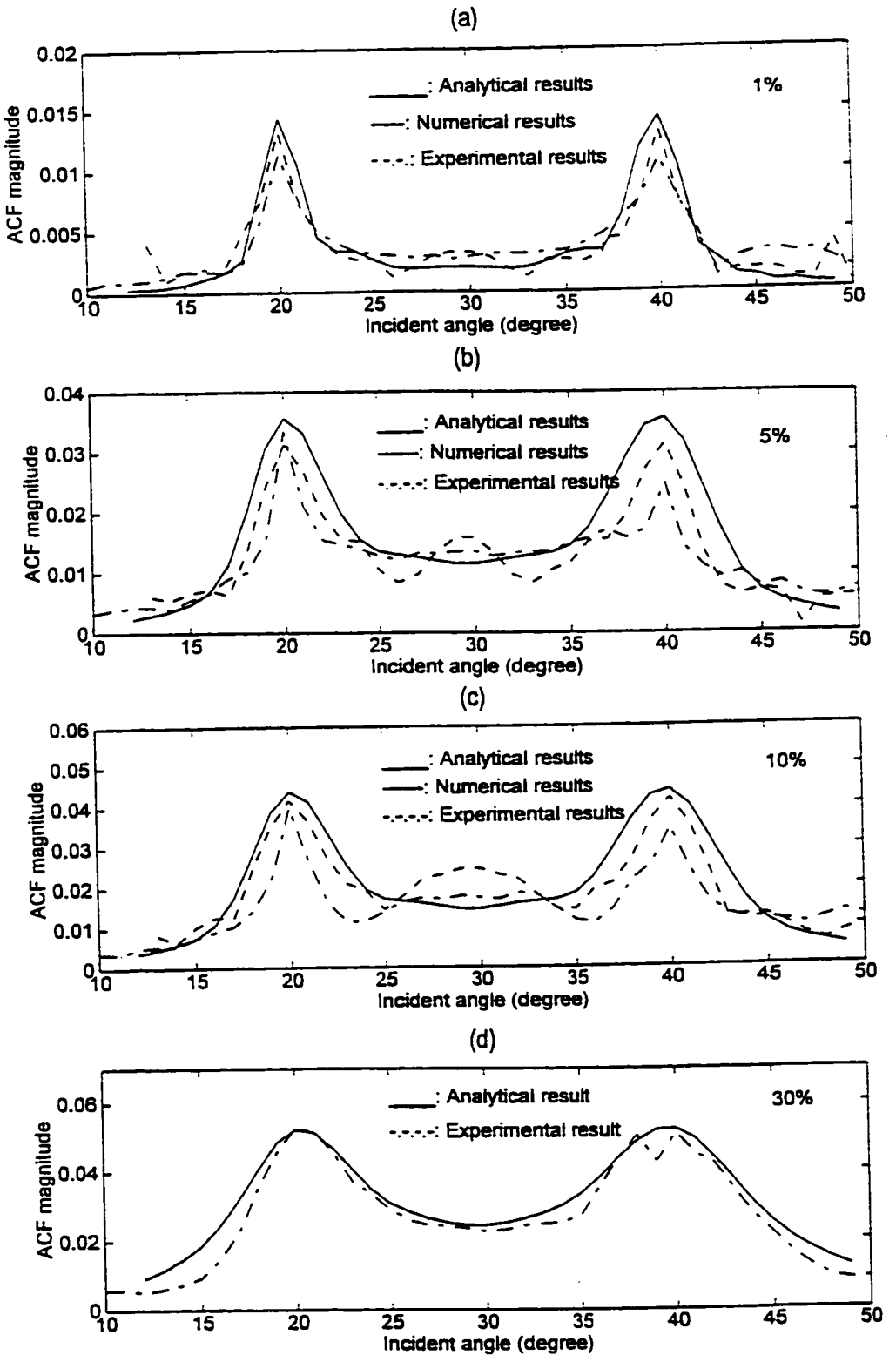


Figure 2.4: Comparison of analytical, numerical and experimental results for the volume density $f_v = 1, 5, 10$, and 30%. The reference angles are at $(\theta_{i1}, \theta_{s1}) = (20°, -40°)$.

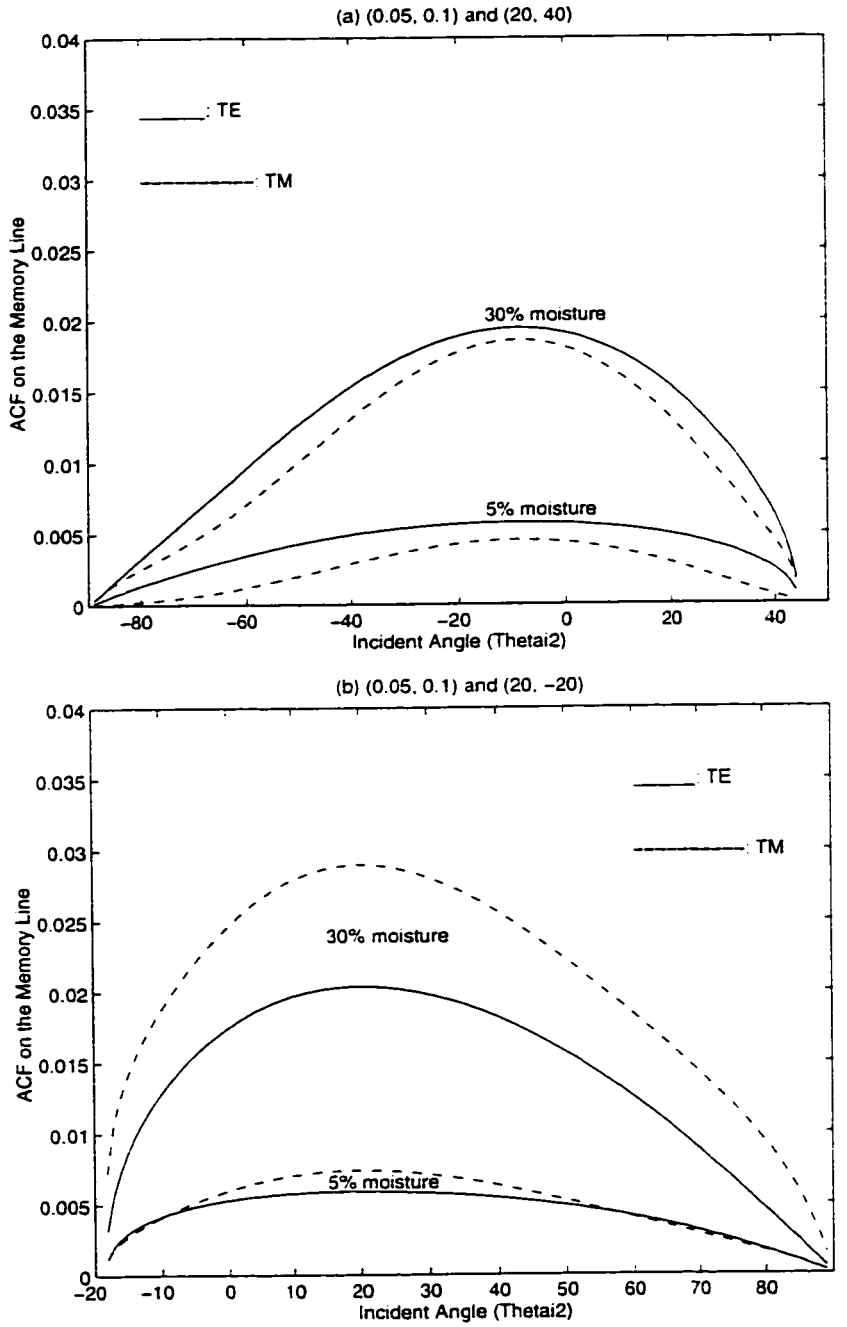


Figure 2.5: ACF along the memory line for TE and TM cases. $\epsilon_r = (16.67 + i1.15)$ for 30% moisture, $\epsilon_r = (3.71 + i0.13)$ for 5% moisture, $h = 0.05\lambda$, $l_o = 0.1\lambda$, and $L = 20\lambda$. (a): $\theta_{i1} = 20^o$, $\theta_{s1} = 40^o$. (b): $\theta_{i1} = 20^o$, $\theta_{s1} = -20^o$.

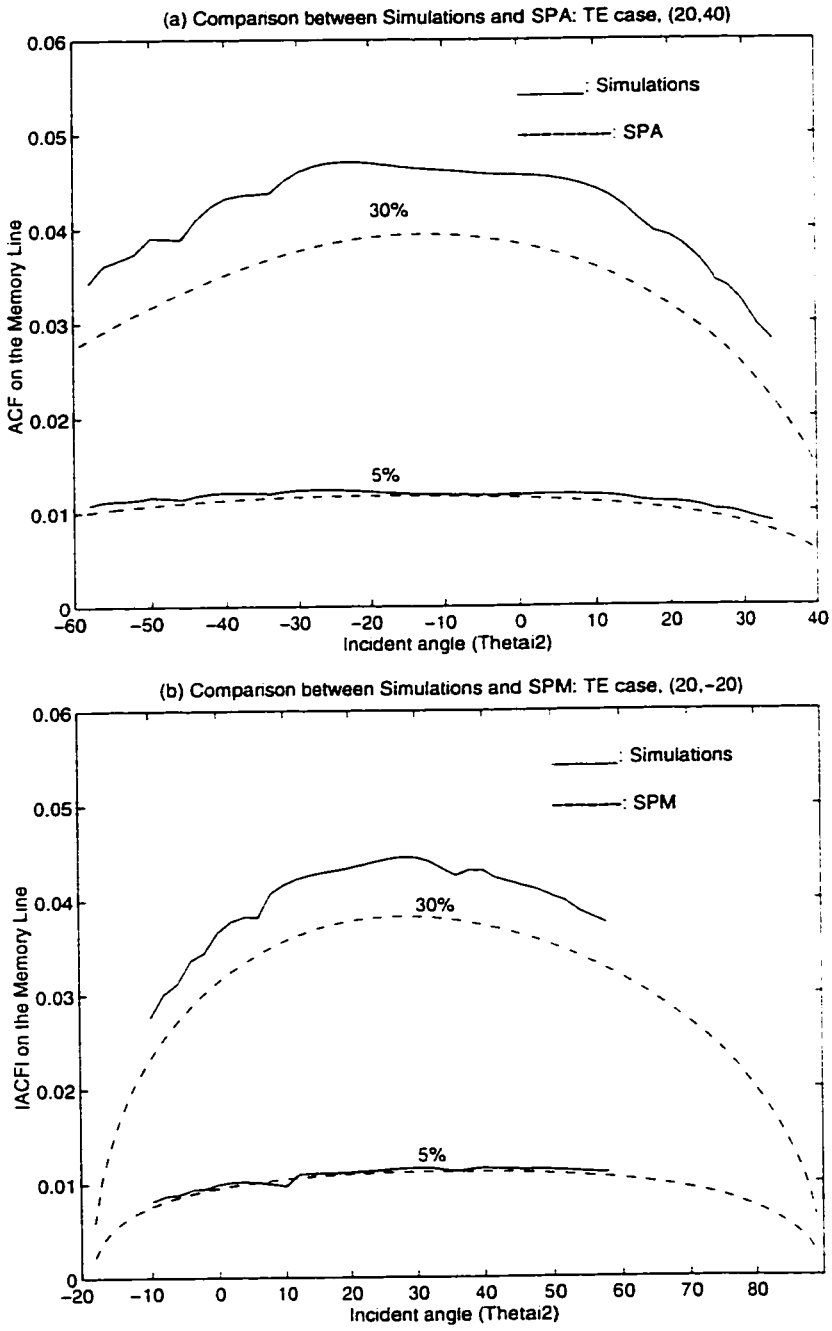


Figure 2.6: Comparison between SPM and numerical simulations for TE cese. $\epsilon_r = (16.67 + i1.15)$ for 30% moisture, $\epsilon_r = (3.71 + i0.13)$ for 5% moisture, $h = 0.05\lambda$, $l_o = 0.1\lambda$, and $L = 20\lambda$. (a): $\theta_{i1} = 20^o$, $\theta_{s1} = 40^o$. (b): $\theta_{i1} = 20^o$, $\theta_{s1} = -20^o$.

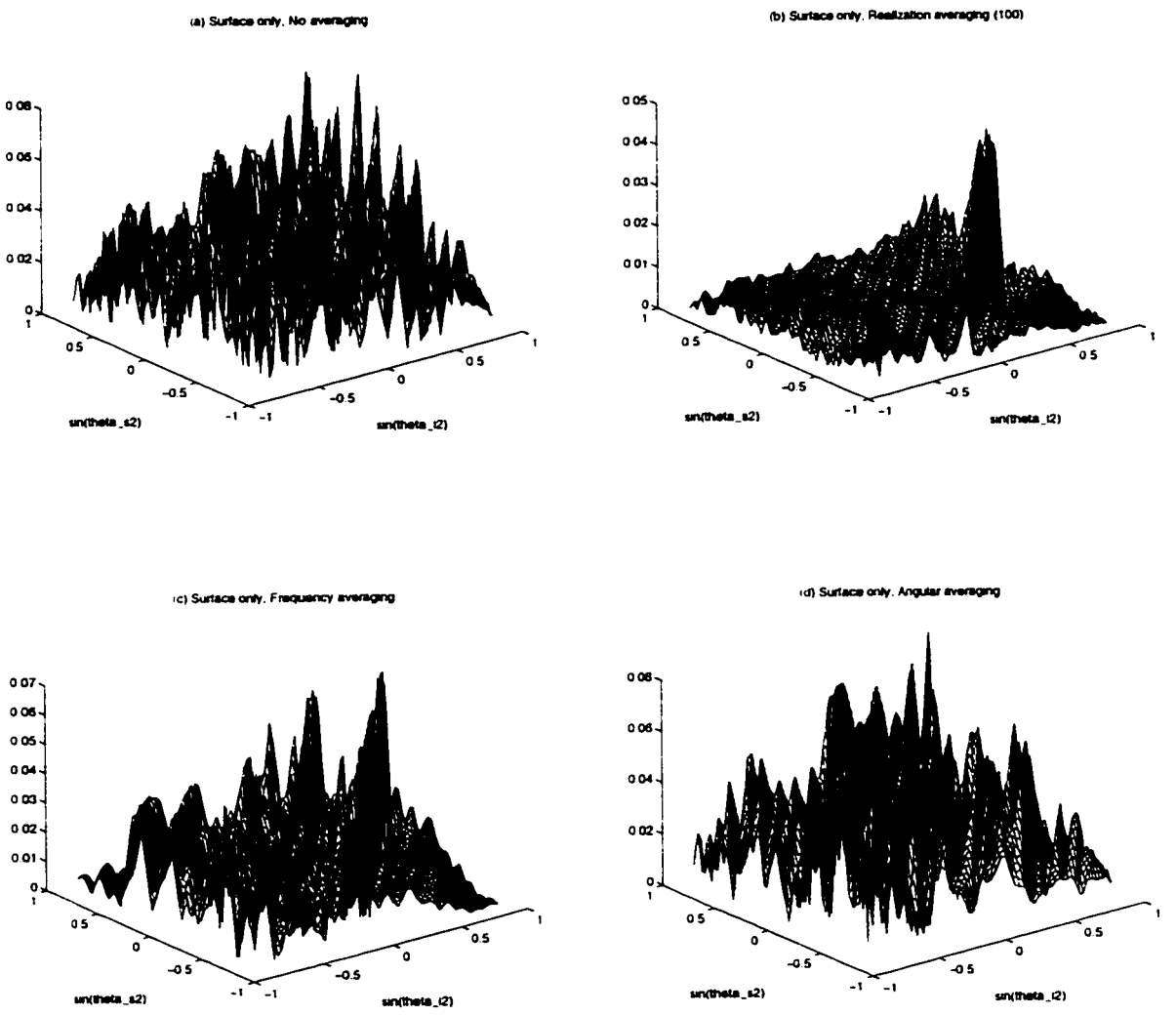


Figure 2.7: 3-D plots of ACF magnitude by different averaging methods. Reference angles are ($\theta_{i1} = 20^o$, $\theta_{s1} = -20^o$). Dielectric constant of region 1 is $\epsilon_r = 3.7 + i0.13$. $h = 0.35\lambda_o$, $l = 1.0\lambda_o$, $L = 40\lambda_o$, g = L/4. (a): One realization. (b): Realization averaging over 100 rough surfaces. (c): Frequency averaging over a frequency band of $0.5f_o$ to $1.5f_o$. (d): Angular averaging over an angular range of $\theta - 10^o$ to $\theta + 10^o$.

Chapter 3

THE DETECTION OF A BURIED OBJECT EMBEDDED IN RANDOM DISCRETE SCATTERERS UNDER A ONE-DIMENSIONAL RANDOM ROUGH SURFACE

3.1 Introduction

In this chapter, we study the angular correlation function for the detection of a target embedded in a clutter environment for two-dimensional scattering problem. The wave scattering by the target is often obscured by clutter caused by rough surfaces and random discrete scatterers such as rocks, ice grains, grass, etc. The problem is how to suppress clutter in radar images while keeping the target return unchanged. Conventional methods are based on radar cross section (RCS), including SAR and bistatic measurement.

In the last chapter, we have shown that the ACF of random media scattering only exhibit a strong correlation along the memory effect and the values of ACF are small away from the memoy line due to the statistically translational invariance of random media scattering. This property can be applied to the target detection problem since a single target scattering does have the translational invariance. Recently, we have been investigating a detection method based on ACF. We have applied the ACF technique for the detection of a buried object under a rough surface [13] [43] [55] and shown that the ACF technique is superior to the conventional RCS method for some cases.

For the targets buried under ground, the rough surface scattering can obscure the target. The volume scattering by vegetation, inhomogeneity of soil and snow,

and hydrometeors are also important in the target detection problems. To simulate the scattering, the deterministic buried object and 300 small cylinders representing random discrete scatterers are placed below a rough surface. The study presented in this chapter consists of two parts: (1) formulation and solution of the wave scattering problem, and (2) angular correlation processing of the simulated scattered wave fields using frequency averaging. For a fair comparison, the intensity is also calculated by using frequency averaging. It is found that the ACF can be used to distinguish scattering by random discrete scatterers, a random rough surface, and a deterministic buried object. The clutter return can be suppressed and the target becomes conspicuous. The ratio of ACF of the scattered wave with an object and that without an object can be 10 dB larger than the ratio of intensities.

3.2 Formulation

In this section, we will formulate the problem by using the integral equation method. The equations are solved numerically using method of moments giving an exact solution of Maxwell's equations. We assume the buried object and the surrounding scatterers are located below a rough surface as shown in Figure 3.1. The buried object is assumed to be perfect conducting. A tapered plane wave $\psi^i(\overline{r})$ is incident on the surface with height function $z = f(x)$.

$$\begin{aligned} \psi^i(\overline{r}) &= exp\left[i(kx\sin\theta_i - kz\cos\theta_i)(1 + w(\overline{r}))\right] \\ & \quad exp\left[-(x + z\tan\theta_i)^2/g^2\right] \end{aligned} \tag{3.1}$$

where g is the tapering parameter, and

$$w(\overline{r}) = \frac{2^{\frac{(x+z\tan\theta_i)^2}{g^2}} - 1}{(kg\cos\theta_i)^2} \tag{3.2}$$

Let ψ_o and ψ_1 be the fields in regions 0 and 1, respectively, and the boundary conditions are $\psi_o = \psi_1$, $\frac{\partial\psi_o}{\partial n} = \frac{\partial\psi_1}{\partial n}$ at the rough surface and $\psi_1 = 0$ on the surface of

the buried object. For a rough surface, we use surface integral equations. For discrete scatterers, we use a volume integral equation. The integral equations are as follows:

$$\frac{1}{2}\psi(\bar{r}_r) - \int_{s_r}\left[\psi(\bar{r}')\frac{\partial G_o(\bar{r}_r,\bar{r}')}{\partial n'} - G_o(\bar{r}_r,\bar{r}')\frac{\partial \psi(\bar{r}')}{\partial n'}\right]ds' = \psi^i(\bar{r}_r) \tag{3.3}$$

$$\begin{aligned}\frac{1}{2}\psi(\bar{r}_r) \;+\; &\int_{s_r}\left[\psi(\bar{r}')\frac{\partial G_1(\bar{r}_r,\bar{r}')}{\partial n'} - G_1(\bar{r}_r,\bar{r}')\frac{\partial \psi(\bar{r}')}{\partial n'}\right]ds' \\ &+ \int_{s_b} G_1(\bar{r},\bar{r}')\frac{\partial \psi(\bar{r}')}{\partial n'}ds' + \sum_{n=1}^{N} G_1(\bar{r}_r,\bar{r}_n)A_n\psi(\bar{r}_n) = 0\end{aligned} \tag{3.4}$$

$$\begin{aligned}&\int_{s_r}\left[\psi(\bar{r}')\frac{\partial G_1(\bar{r}_b,\bar{r}')}{\partial n'} - G_1(\bar{r}_b,\bar{r}')\frac{\partial \psi(\bar{r}')}{\partial n'}\right]ds' \\ +\; &\int_{s_b} G_1(\bar{r}_b,\bar{r}')\frac{\partial \psi(\bar{r}')}{\partial n'}ds' + \sum_{n=1}^{N} G_1(\bar{r}_b,\bar{r}_n)A_n\psi(\bar{r}_n) = 0\end{aligned} \tag{3.5}$$

$$\begin{aligned}&\int_{s_r}\left[\psi(\bar{r}')\frac{\partial G_1(\bar{r}_m,\bar{r}')}{\partial n'} - G_1(\bar{r}_m,\bar{r}')\frac{\partial \psi(\bar{r}')}{\partial n'}\right]ds' \\ +\; &\int_{s_b} G_1(\bar{r}_m,\bar{r}')\frac{\partial \psi(\bar{r}')}{\partial n'}ds' + \psi(\bar{r}_m) - \sum_{n=1,n\neq m}^{N} G_1(\bar{r}_m,\bar{r}_n)A_n\psi(\bar{r}_n) = 0\end{aligned} \tag{3.6}$$

where s_r is the rough surface, s_b is the surface of the buried object, G_o and G_1 are Green's functions in region 0 and region 1, respectively, A_n is the scattering amplitude of the nth scatterer, and $\bar{r}_n$ is the position of the nth discrete scatterer.

Equations (3.3)-(3.6) are coupled integral equations. We solve Eqs. (3.3)-(3.6) by the method of moments. After the matrix equation is solved, the far-field scattered field in region 0 is calculated by

$$\psi_s(\bar{r}) = \frac{e^{ikr}}{\sqrt{r}}\psi_s^N(\theta_s,\theta_i) \tag{3.7}$$

and

$$\psi_s^N(\theta_s,\theta_i) = \frac{i}{4}\sqrt{\frac{2}{\pi k}}e^{-i\pi/4} \int dx'\left[-F_2(x') + F_1(x')ik\left(\frac{df}{dx'}\sin\theta_s - \cos\theta_s\right)\right] e^{-ik(x'\sin\theta_s + f(x')\cos\theta_s)} \tag{3.8}$$

where

$$F_1(\bar{r}) = \psi(\bar{r})$$
$$F_2(\bar{r}) = \sqrt{1+(f(x))^2}\frac{\partial \psi(\bar{r})}{\partial n}$$

Then the angular correlation function can be calculated based on realization averaging as discussed in last chapter. However, realization averaging is not applicable in this case in which a target is surrounded by a single realization of a random medium under a rough surface. Other methods of taking averaging need to be used. We use frequency averaging which takes an average over a frequency bandwidth $2\triangle f$ centered at f_o, as given by

$$\Gamma_f(\theta_{s1},\theta_{i1};\theta_{s2},\theta_{i2}) = \frac{1}{N_f}\sum_{n=1}^{N_f} \psi_s^N(\theta_{s1},\theta_{i1};f_n)\psi_s^{N*}(\theta_{s2},\theta_{i2};f_n)/\sqrt{P_1 P_2} \qquad (3.9)$$

where N_f is the number of frequencies over the frequency range $f_o - \triangle f < f_n < f_o + \triangle f$, and n is the frequency index for f_n. P_1 and P_2 are the total power flux of the two incident waves, respectively.

When $(\theta_{i1} = \theta_{i2})$ and $(\theta_{s1} = \theta_{s2})$, the ACF becomes the well-known scattered intensity. To make a fair comparison, the intensity can also be defined based on the frequency averaging method,

$$I_f(\theta_s,\theta_i) = \frac{1}{N_f}\sum_{n=1}^{N_f} |\psi_s^N(\theta_s,\theta_i;f_n)|^2/P \qquad (3.10)$$

3.3 Numerical Results and Analysis

To show memory effects, numerical simulations are conducted for random discrete scatterers under a flat surface, scatterers under a rough surface, and scatterers under a rough surface with a buried object. Three hundred small cylinders with radius of $a = 0.05\lambda_o$ are used to simulate the random discrete scatterers and are randomly distributed in the region of $20\lambda_o$X$10\lambda_o$ below the surface (Figure 3.1). The relative dielectric constant is $\epsilon_1 = 3.7 + i0.13$ for the lower region and $\epsilon_p = 16.67 + i1.15$

for the small cylinders. The buried object is a perfect conducting elliptical cylinder. The rough surface is generated by using the spectrum method [35]. We use a surface length of $L = 40\lambda_o$ and a tapering parameter of $g = L/4 = 10$ so that the incident beam falls in the region of the scatterers.

We first show the memory effect for combined volume scattering and rough surface scattering in Figure 3.2. The results are obtained by frequency averaging within the frequency band from $0.5f_o$ to $1.5f_o$. Figure 3.2a shows the 3-D plot of the angular correlation function. The large ridges correspond to the w_2 measured in the specular direction ($\theta_{s2} = \theta_{i2}$). We see a line on $\sin\theta_{i2} - \sin\theta_{s2} = 2\sin\theta_{i1} = 0.684$. There is a maximum on the memory line at $\theta_{i2} = 20^o$ on which the phase matching condition is satisfied in both the horizontal and the vertical directions. It is called a memory dot. In chapter 2, we derive the memory dot for volume scattering and also an estimate of the angular width of the memory effect. The location of the maximum value of ACF is then determined by Eq.(2.33) and (2.34). It is found that the memory line is shorter compared with that by rough surface scattering [54]. We also plot the ACF along the memory line and perpendicular to the memory line in Figure 3.2b.

Figure 3.3 shows the results with a buried object. Figure 3.3a is that an elliptical object buried under a flat surface with small cylinders around it. The size of the elliptical cylinder is $a = 0.7\lambda_o$ in the horizontal direction and $b = 1.0\lambda_o$ in the vertical direction. It is placed at $x_b = 0.0\lambda_o$ and $z_b = 2.0\lambda_o$. Besides the fluctuation perpendicular to the memory line, the fluctuation along the memory line can also be seen. There is a sharp peak of the ACF magnitude along the memory line, which is not in Figure 2. That is due to the scattering pattern of the object. Figure 3.3b shows the results of the same object buried under a rough surface and with the same scatterers. It is different from that in Figure 3.3a. This means the wave scattering by the buried object is obscured by the clutter scattering of the rough surface and random discrete scatterers. Hence, the angle pairs along the memory line are avoided for the detection of a buried object.

To show the effectiveness of the ACF method in the detection of a buried object embedded in clutter, we compare the ratio of the ACF with and without a object and that of intensity. The ratios are defined as follows:

$$ratio\ of\ |ACF| = \frac{|ACF|\ with\ buried\ object}{|ACF|\ without\ buried\ object} = \frac{|\Gamma_{wb}(\theta_{s1},\theta_{i1};\theta_{s2},\theta_{i2})|}{|\Gamma_{nb}(\theta_{s1},\theta_{i1};\theta_{s2},\theta_{i2})|} \quad (3.11)$$

$$ratio\ of\ I = \frac{I\ with\ buried\ object}{I\ without\ buried\ object} = \frac{I_{wb}(\theta_s,\theta_i)}{I_{nb}(\theta_s,\theta_i)} \quad (3.12)$$

where wb = "with buried object" and nb = "no buried object".

Figure 3.4 shows ratios of ACF with and without a object and that of intensity. The ACF and its ratio are evaluated at back-scattering directions with $\theta_{s2} = -\theta_{i2}$. The reference angles are $(\theta_{s1},\theta_{i1}) = (-20^o, 20^o)$, $(-50^o, 30^o)$, and $(-60^o, 60^o)$, as shown in the figure. The intensity and its ratio are that for $(\theta_s,\theta_i) = (\theta_{s2},\theta_{i2})$ at back directions $(\theta_s = -\theta_i)$. Figure 3.4a shows the result for an object buried under a rough surface without the random discrete scatterers. The ratio of ACF is usually larger than that of intensity. Figure 3.4b shows the result of a object buried in small scatterers under a flat surface. We see that the ratio of intensity is always close to a unity, but the ratio of the ACF can be many dB.

To see the difference between the ACF method and intensity clearly, we take the angular averaging over the frequency averaging results as given by

$$\Gamma_{fa}(\theta_{s1},\theta_{i1};-\theta_{i2},\theta_{i2}) = \frac{1}{N_a}\sum_{n=1}^{N_a}\Gamma_f(\theta_{s1},\theta_{i1};-\theta_{i2}-\delta_n,\theta_{i2}+\delta_n) \quad (3.13)$$

After the angular averaging those shown in Figure 3.4, the results are shown in Figure 3.5. For an object embedded in small scatterers under a flat surface, the ratio of ACF magnitude is much larger than the ratio of intensity. On the average, the ACF ratio is about 5 dB while the intensity ratio is about 1.

Figure 3.6 shows results of wave scattering by a buried object with both the rough surface and random discrete scatterers included. Again, the ratio of intensity is much smaller than that of the ACF by 10 dB. The ACF ratio depends the reference angles

and is usually large for a pair of large reference angles. This is because the rough surface scattering is small at large angles.

In this chapter, we studied the angular correlation function of wave scattering by an object buried under a rough surface and surrounded by randomly distributed scatterers. It is found that the memory effect is different for rough surface scattering and volume scattering as well as target scattering. The ACF can be used to detect an object embedded in clutter of random discrete scatterers and a random rough surface.

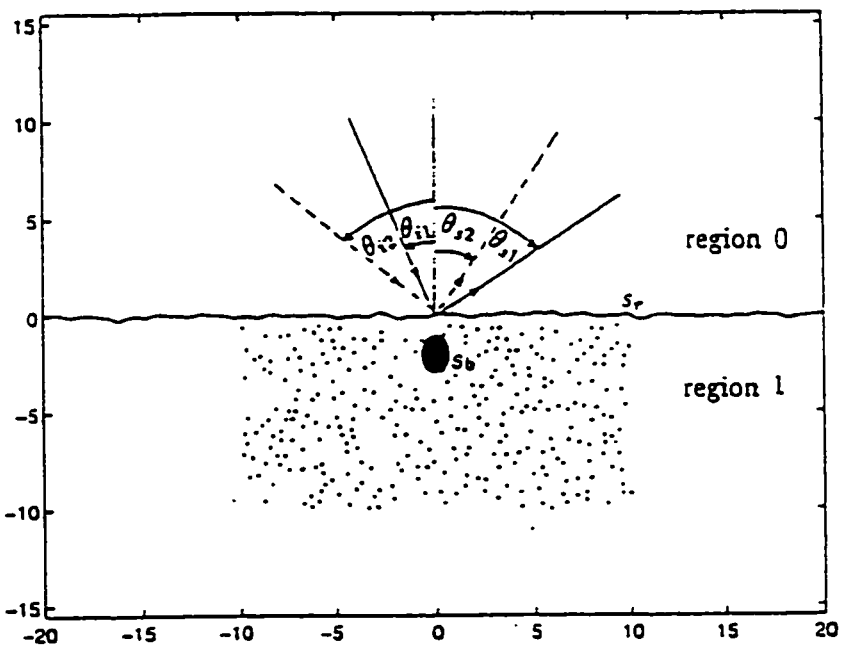


Figure 3.1: Configuration of angular correlation of wave scattering from a rough surface with a buried object embedded in small scatterers. The two incident angles are θ_{i1} and θ_{i2} with respective scattered angles θ_{s1} and θ_{s2}.

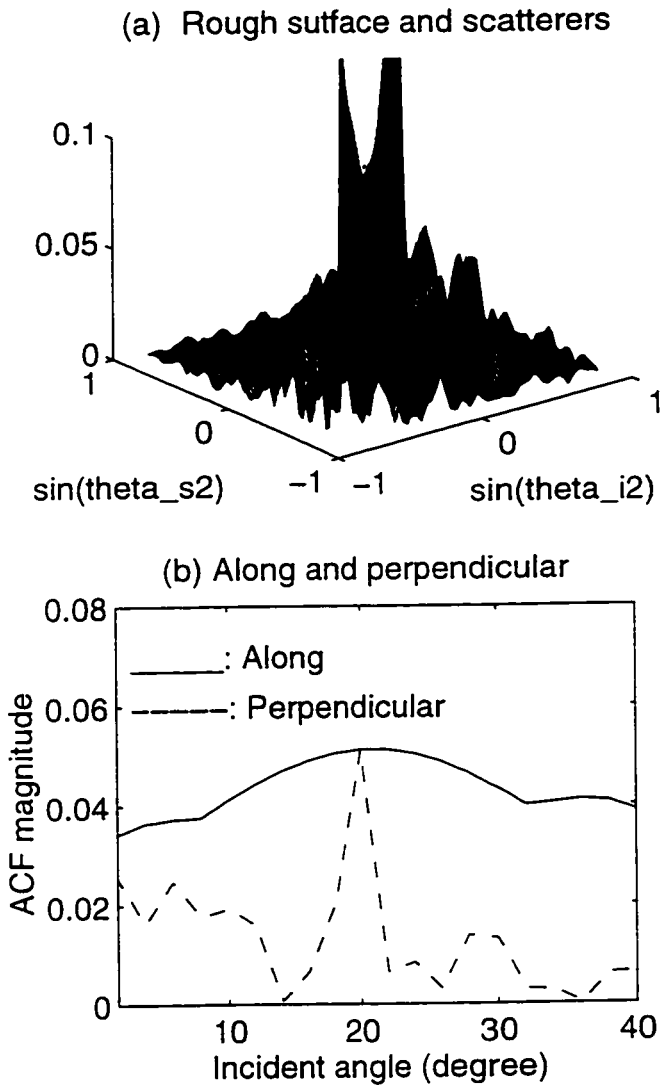


Figure 3.2: ACF for wave scattering by a rough surface and 300 scatterers based on frequency averaging. (a): 3-D plot of the ACF. (b): ACF magnitude along and perpendicular to the memory line.

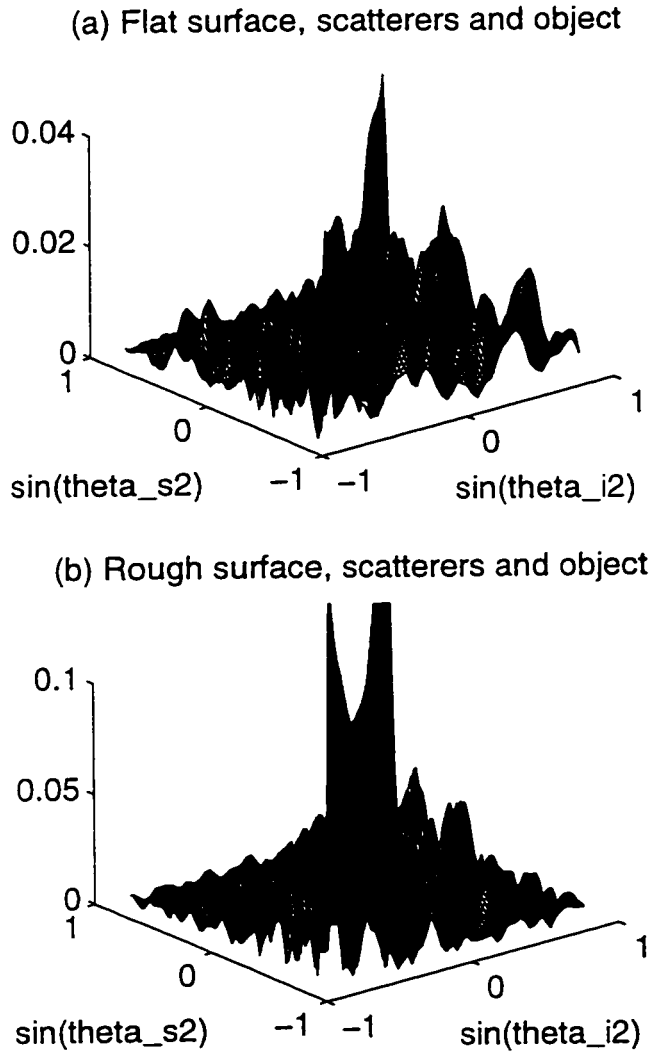


Figure 3.3: 3-D plots of ACF for an object embedded in scatterers under a surface. (a): Under a flat surface. (b): Under a rough surface ($h = 0.1\lambda_o, l = 0.5\lambda_o$).

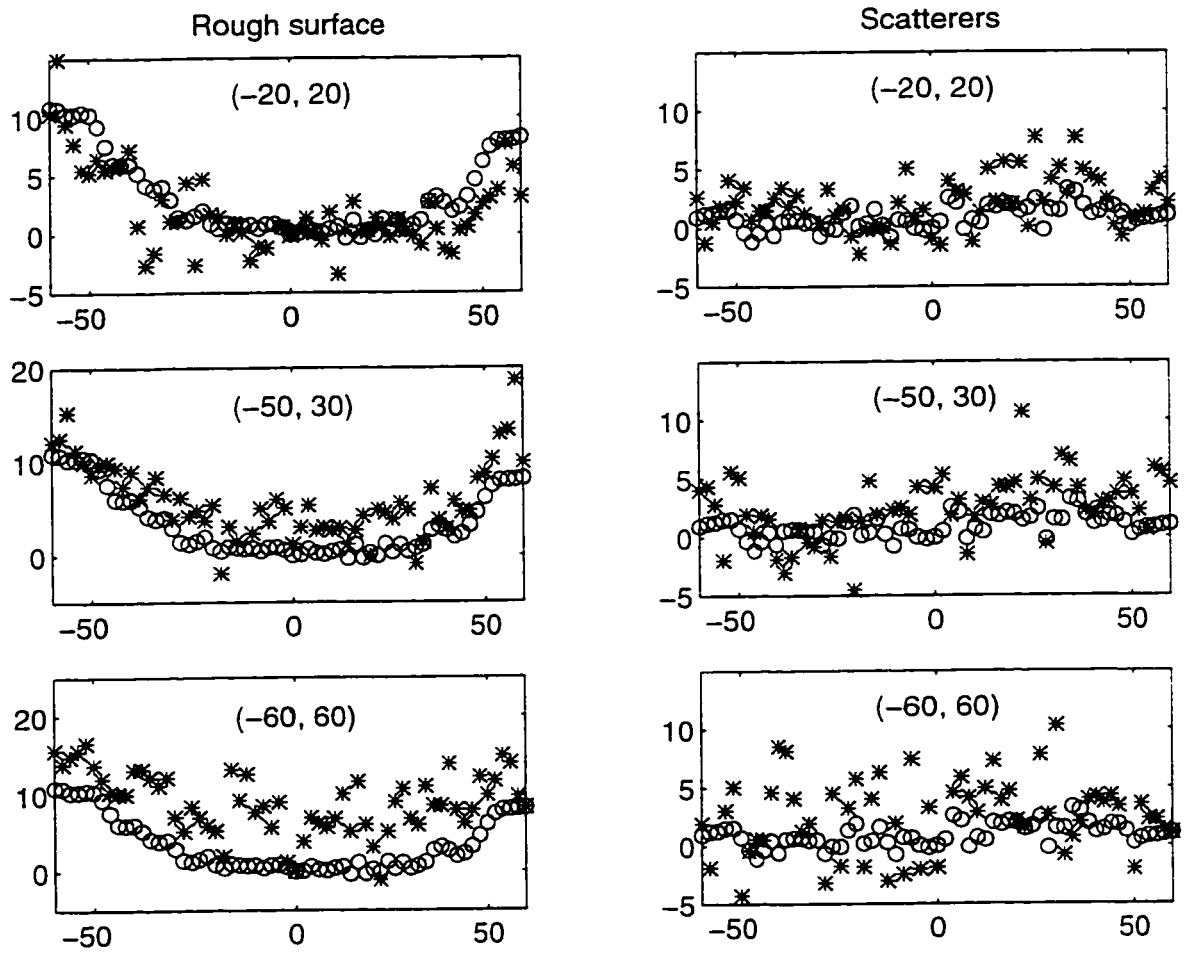


Figure 3.4: Comparisons between ratios (in dB) of the ACF ("*") with object that without object and ratios of intensity ("o"). (a) rough surface without scatterers ($h = 0.1\lambda_o$, $l = 0.5\lambda_o$) (b) flat surface with 300 scatterers.

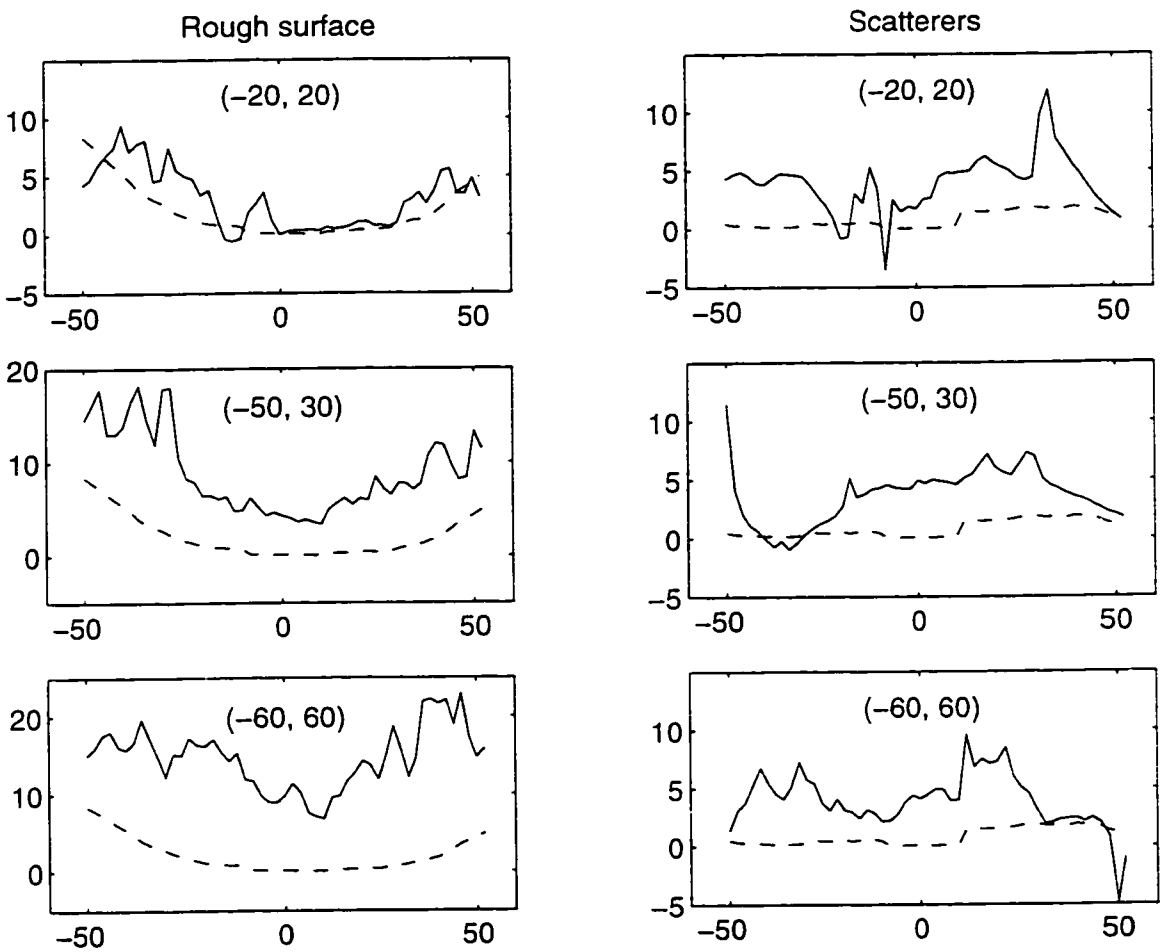


Figure 3.5: Results of Figure 3.4 after angular averaging.

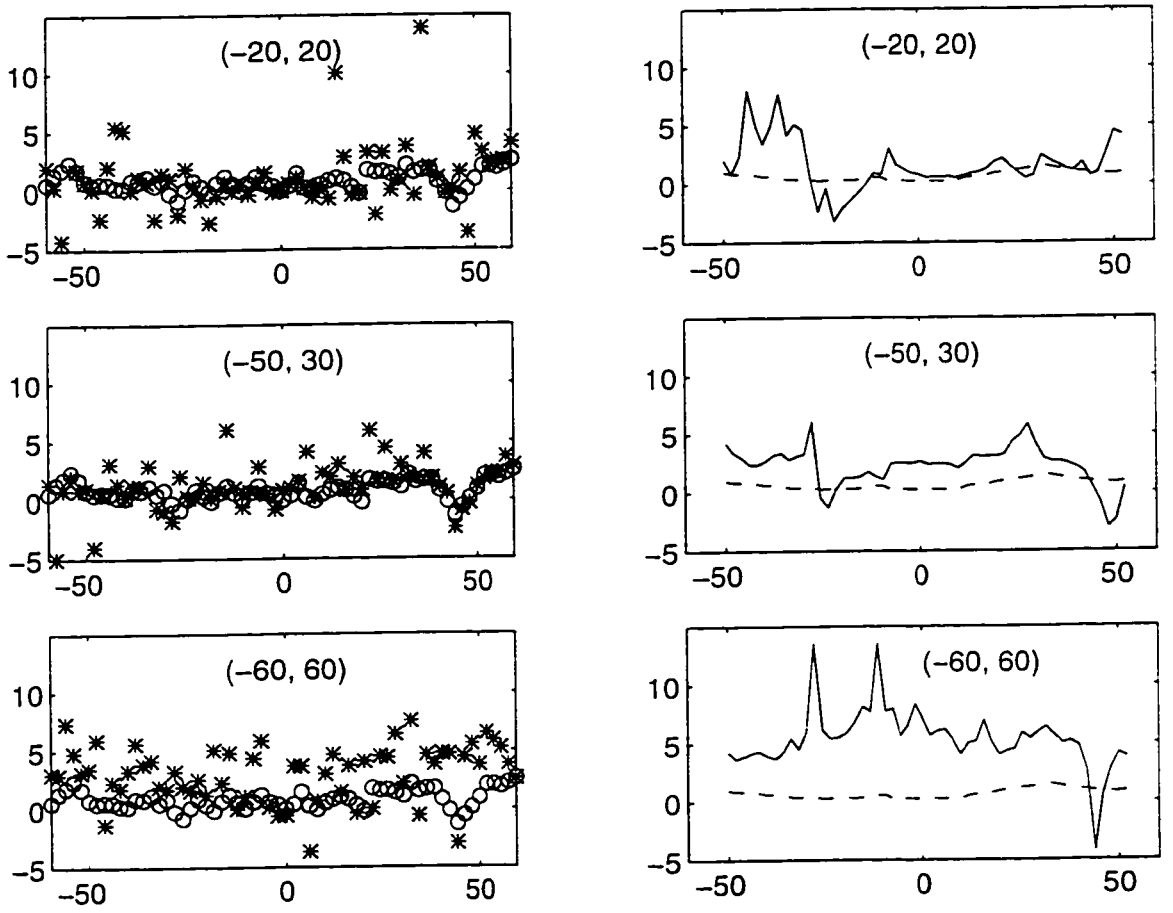


Figure 3.6: Comparisons between ratios (in dB) of the ACF with object that without object and ratios of intensity for the case in which both the rough surface and the random discrete scatterers are present. The left column is the results by frequency averaging only ("*": ACF, "o": Intensity). The right column is the results obtained by taking angular averaging over the frequency averaging results of the left column (—: ACF, —: Intensity).

Chapter 4

ANGULAR CORRELATION FUNCTION AND SCATTERING COEFFICIENT OF ELECTROMAGNETIC WAVES SCATTERED BY A BURIED OBJECT UNDER A TWO-DIMENSIONAL ROUGH SURFACE

4.1 Introduction

The wave scattering from a 2-D object buried under a 1-D rough surface has been numerically studied by using the method of moments (MoM) and illustrated in last chapter as well as other reference [43][25][16]. The wave scattering by a 3-D object above and below a flat surface is studied by T-matrix [44]. Since most practical problems involve a 3-D object buried under a 2-D rough surface, a solution of electromagnetic wave scattering from a 2-D random rough surface with a 3-D buried object is needed. Recently, the finite-difference time-domain method (FDTD) has been used in subsurface electromagnetic wave scattering problems [36][3]. The solution of surface integral equations obtained using MoM has been widely used in the study of rough surface scattering problems. The difficulty of the surface integral equation method with MoM for 2-D rough surface scattering problems is the large size of the impedance matrix of $6N \times 6N$, where N is the total number of points representing the surface. Furthermore, the solution time scales as $(6N)^3$ for the matrix inversion method.

To speed up the solution of surface integral equations with MoM, fast methods have been used. Fast multipole and FFT have been combined for 2-D rough surface scattering [46]. Recently, the sparse-matrix canonical-grid method (SMCG) has been

developed for large-scale rough surface scattering problems [26]. The SMCG was used to study backscattering enhancement of electromagnetic wave scattering by 2-D perfectly conducting and dielectric rough surfaces [27][28]. All of these studies have been concentrated on the calculation of scattering coefficients (RCS).

As shown in last chapter, angular correlation function (ACF) has advantages over RCS for the detection of targets in clutter. ACFs are obtained by taking averages over a product of two scattered wave fields corresponding two incident waves. Thus, a key step of calculating the ACF is taking averages. For random media scattering, the average is usually taken over realizations of random media or rough surfaces, which is not applicable for the detection of the object buried under a rough surface. For 2-D scattering problems of a target embedded in clutter, we have used frequency averaging and angular averaging to obtain the ACF [55][54][51]. Numerical results have shown that detection of targets by the ACF with frequency and angular averaging has advantages over the radar cross section (RCS). Since the scattering characteristics of the object is frequency dependent, the frequency averaging may smear the result. Therefore, we will use azimuthal angular averaging instead.

In this chapter, we study electromagnetic wave scattering by a 3-D buried object under a 2-D random rough surface. First, we formulate the problem based on the Stratton-Chu surface integral equations for the rough surface and the surface of the object. Then, the scattered wave fields from the object onto the rough surface are treated as additional incident fields on the rough surface. Then, the SMCG method is used for the solution of the matrix equation. Numerical results are calculated for a perfectly conducting sphere under the rough surface. Both the ACF and scattering coefficient are calculated. To obtain the statistical results, we use azimuthal averaging. It is found that ACF is more effective in suppressing the effects of the rough surface scattering. Also, the cross-polarization components of ACF can be more useful than co-polarization components for the detection of the buried object.

4.2 Formulation

4.2.1 Incident Wave

Consider an electromagnetic wave with fields $\vec{E}^i(x,y,z)$ and $\vec{H}^i(x,y,z)$ with time dependence $\exp(-i\omega t)$ impinging on a 2-D rough surface with a random height profile $z = f(x,y)$. Above the rough surface is a free space (region 0) while the subsurface is characterized by permittivity ϵ_1 and permeability μ_1 (region 1). The height function $z = f(x,y)$ is a random process with zero mean. The incident direction is $\hat{k}_i = \sin\theta_i\cos\phi_i\hat{x} + \sin\theta_i\sin\phi_i\hat{y} - \cos\theta_i\hat{z}$. The incident wave is tapered so that the illuminated rough surface can be confined to the surface area $L_x \times L_y$. The incident electric and magnetic fields are given as

$$\vec{E}^i_\alpha(x,y,z) = \int_{-\infty}^{+\infty} dk_x \int_{-\infty}^{+\infty} dk_y \exp(ik_x x + ik_y y - ik_z z) E_i(k_x,k_y)\hat{e}_\alpha(-k_z) \qquad (4.1)$$

and

$$\vec{H}^i_\alpha(x,y,z) = -\frac{1}{\eta_1}\int_{-\infty}^{+\infty} dk_x \int_{-\infty}^{+\infty} dk_y \exp(ik_x x + ik_y y - ik_z z) E_i(k_x,k_y)\hat{h}_\alpha(-k_z) \qquad (4.2)$$

where α denotes horizontal polarization (h) and vertical polarization (v) . as given as

$$\hat{e}_h(-k_z) = \frac{1}{k_\rho}(\hat{x}k_y - \hat{y}k_x) \quad , \quad \hat{h}_h(-k_z) = \frac{k_z}{kk_\rho}(\hat{x}k_x + \hat{y}k_y) + \frac{k_\rho}{k}\hat{z} \qquad (4.3)$$

$$\hat{e}_v(-k_z) = \frac{k_z}{kk_\rho}(\hat{x}k_x + \hat{y}k_y) + \frac{k_\rho}{k}\hat{z} \quad , \quad \hat{h}_v(-k_z) = \frac{1}{k_\rho}(-\hat{x}k_y + \hat{y}k_x) \qquad (4.4)$$

with $k_z = \sqrt{k^2 - k_x^2 - k_y^2}$ and $k_\rho = \sqrt{k_x^2 + k_y^2}$ with k being the incident wave number, η_1 the intrinsic impedance of free space, and $E_i(k_x,k_y)$ the spectrum of the incident wave. We use the following spectrum

$$E_i(k_x,k_y) = \frac{1}{4\pi^2}\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \exp(-ik_x x - ik_y y)\exp(i(k_{ix}x + k_{iy}y)(1+w))\exp(-t) \qquad (4.5)$$

where $t = t_x + t_y = (x^2+y^2)/g^2$ and

$$t_x = \frac{(\cos\theta_i\cos\phi_i x + \cos\theta_i\sin\phi_i y)^2}{g^2\cos^2\theta_i} \qquad (4.6)$$

$$t_y = \frac{(-\sin\phi_i x + \cos\phi_i y)^2}{g^2} \tag{4.7}$$

$$w = \frac{1}{k^2}[\frac{(2t_x - 1)}{g^2\cos^2\theta_i} + \frac{(2t_y - 1)}{g^2}] \tag{4.8}$$

The parameter g controls the tapering of the incident wave. The w and t terms are introduced to approximate the tapered wave solution.

4.2.2 Surface Integral Equations

We apply Stratton-Chu surface integral equations on the rough surface and the surface of the buried object. For a rough surface with the height function $z = f(x, y)$, we have the unit normal vector as

$$\hat{n} = \frac{1}{\sqrt{(\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2 + 1}}\left(-\frac{\partial f}{\partial x}\hat{x} - \frac{\partial f}{\partial y}\hat{y} + 1\hat{z}\right) \tag{4.9}$$

The boundary conditions for the dielectric rough surface are

$$\begin{aligned} \hat{n}\times\vec{E}_0 = \hat{n}\times\vec{E}_1 \quad , \qquad & \hat{n}\times\vec{H}_0 = \hat{n}\times\vec{H}_1 \\ \epsilon_0\hat{n}\cdot\vec{E}_0 = \epsilon_1\hat{n}\cdot\vec{E}_1 \quad , \qquad & \mu_0\hat{n}\cdot\vec{H}_0 = \mu_1\hat{n}\cdot\vec{H}_1 \end{aligned} \tag{4.10}$$

By noting $\mu_0 = \mu_1$, we use Strantton-Chu equations approaching from both sides and the above boundary conditions. The integral equations on the rough surface are

$$\begin{aligned} \hat{n}\cdot\vec{E}^i(\vec{r}) &= \frac{\hat{n}\cdot\vec{E}(\vec{r})}{2} - \hat{n}\cdot[\int \hat{n}'\times\vec{H}(\vec{r'})i\omega\mu_0 g_0 dS' \\ &+ P\int\{(\hat{n}'\times\vec{E}(\vec{r'})\times\nabla' g_0) + \nabla' g_0\hat{n}'\cdot\vec{E}(\vec{r'})\}dS'] \end{aligned} \tag{4.11}$$

$$\begin{aligned} \hat{n}\times\vec{H}^i(\vec{r}) &= \frac{\hat{n}\times\vec{H}(\vec{r})}{2} - \hat{n}\times[\int -i\omega\hat{n}'\times\vec{E}(\vec{r'})\epsilon_0 g_0 dS' \\ &+ P\int\{\hat{n}'\times\vec{H}(\vec{r'})\times\nabla' g_0 + \hat{n}'\cdot\vec{H}(\vec{r'})\nabla' g_0\}dS'] \end{aligned} \tag{4.12}$$

$$\begin{aligned} -\hat{n}\times\vec{E}^s_b(\vec{r}) &= -\frac{\hat{n}\times\vec{E}(\vec{r})}{2} - \hat{n}\times[\int i\omega\hat{n}\times\vec{H}(\vec{r'})\mu_1 g_1 dS' \\ &+ P\int\{(\hat{n}'\times\vec{E}(\vec{r'})\times\nabla' g_1) + (\hat{n}'\cdot\vec{E}(\vec{r'})\frac{\epsilon_1}{\epsilon_0}\nabla' g_1)\}dS'] \end{aligned} \tag{4.13}$$

$$
\begin{aligned}
-\hat{n}\cdot\vec{H}_b^s(\vec{r}) &= -\frac{\hat{n}\cdot\vec{H}(\vec{r})}{2} - \hat{n}\cdot[\int -\hat{n}'\times\vec{E}(\vec{r'})i\omega\epsilon_1 g_1 dS' \\
&+ P\int\{(\hat{n}'\times\vec{H}(\vec{r'})\times\nabla' g_1) + \nabla' g_1\hat{n}'\cdot\vec{H}(\vec{r'})\}dS']
\end{aligned}
\tag{4.14}
$$

where the above integral $P\int$ represents the principal-value integral and g_0 and g_1 are the scalar Green's function in region 0 (air) and region 1 (lossy dielectric medium), respectively

$$
g_{0,1} = \frac{\exp(ik_{0,1}R)}{4\pi R} \tag{4.15}
$$

The distance between a field point $\vec{r}$ and a source point $\vec{r'}$ is
$R = \sqrt{(x-x')^2+(y-y')^2+(f(x,y)-f(x',y'))^2}$.

In Eqs. (4.13) and (4.14), $\vec{E}_b^s$ and $\vec{H}_b^s$ are scattered fields from the buried object onto the rough surface. They are calculated as follows.

We assume that the buried object is a perfectly conducting sphere. It is convenient to use the magnetic field integral equation (MFIE) to solve the surface current on the object. For an exciting field of $\vec{H}_b^e$, the MFIE for the surface current $\vec{J}_b = \hat{n}_b\times\vec{H}_b$

$$
\hat{n}_b\times\vec{H}_b^e(\vec{r}) = \frac{\vec{J}_b(\vec{r})}{2} - \hat{n}_b\times\int_{S_b}\vec{J}_b(\vec{r'})\times\nabla' g_1 dS' \tag{4.16}
$$

where the unit normal vector of the surface of the buried object is $\hat{n}_b$, the exciting field of $\vec{H}_b^e$ for the buried object is the scattered field from the rough surface, i.e.

$$
\begin{aligned}
\hat{n}_b\times\vec{H}_b^e(\vec{r}) &= -\hat{n}_b\times\int_{S_r}[-\hat{n}'\times\vec{E}(\vec{r'})i\omega\epsilon_1 g_1 \\
&+ (\hat{n}'\times\vec{H}(\vec{r'})\times\nabla' g_1) + \nabla' g_1\hat{n}'\cdot\vec{H}(\vec{r'})]dS'
\end{aligned}
\tag{4.17}
$$

Then, the scattered fields from the buried object are expressed in terms of the surface current as

$$
\vec{H}_b^s(\vec{r}) = \int_{S_b}\vec{J}_b(\vec{r'})\times\nabla' g_1 dS' \tag{4.18}
$$

$$
\vec{E}_b^s(\vec{r}) = \frac{-i}{\omega\epsilon_1}\nabla\times\int_{S_b}\vec{J}_b(\vec{r'})\times\nabla' g_1 dS' \tag{4.19}
$$

Eqs. (4.11) - (4.14) combined (4.16) -(4.19) constitute the coupled integral equations between the rough surface and the buried object.

4.2.3 Matrix Equation and Its Solution

To solve the coupled integral equations, we discretize the rough surface and the surface of the object into small patches. For a patch at the rough surface, we use 6 knowns for the surface fields as follows:

$$J_1(\vec{r}) = \vec{n} \times \vec{E}_r(\vec{r}) \cdot \hat{x} \tag{4.20}$$

$$J_2(\vec{r}) = \vec{n} \times \vec{E}_r(\vec{r}) \cdot \hat{y} \tag{4.21}$$

$$J_3(\vec{r}) = \vec{n} \cdot \vec{E}_r(\vec{r}) \tag{4.22}$$

$$J_4(\vec{r}) = \vec{n} \times \vec{H}_r(\vec{r}) \cdot \hat{x} \tag{4.23}$$

$$J_5(\vec{r}) = \vec{n} \times \vec{H}_r(\vec{r}) \cdot \hat{y} \tag{4.24}$$

$$J_6(\vec{r}) = \vec{n} \cdot \vec{H}_r(\vec{r}) \tag{4.25}$$

Using the above definitions for unknowns and the MoM, Eqs. (4.11) - (4.14) become a matrix equation as

$$\overline{\overline{Z}}\ \overline{J} = \overline{b}_i + \overline{b}_b \tag{4.26}$$

where the impedance matrix of rough surface is $\overline{\overline{Z}}$, $\overline{b}_i$ represents incident fields, and $\overline{b}_b$ represents scattered fields from the buried object. We discretize the surface of the object and use the following definition for the surface current on the object

Let $\overline{\overline{Z}}_b$ be the impedance matrix for the buried object, $\overline{\overline{Z}}_{rb}$ be impedence matrix from buried object to rough surface, and $\overline{\overline{Z}}_{br}$ be that from the rough surface to the buried object. Then, it follows from (4.16)-(4.19)

$$J_{b1}(\vec{r}) = \hat{n}_b \times \vec{H}_b(\vec{r}) \cdot \hat{x} \tag{4.27}$$

$$J_{b2}(\vec{r}) = \hat{n}_b \times \vec{H}_b(\vec{r}) \cdot \hat{y} \tag{4.28}$$

$$J_{b3}(\vec{r}) = \hat{n}_b \times \vec{H}_b(\vec{r}) \cdot \hat{z} \tag{4.29}$$

and the Eqs. (4.17) - (4.20) , we have

$$\overline{b}_b = \overline{\overline{Z}}_{rb}\ \overline{J}_b = \overline{\overline{Z}}_{rb}\ \overline{\overline{Z}}_b^{-1} \overline{J}_b^e = \overline{\overline{Z}}_{rb}\ \overline{\overline{Z}}_b^{-1} \overline{\overline{Z}}_{br}\ \overline{J} \tag{4.30}$$

Substituting Eq. (4.30) into (4.26), we get

$$(\overline{\overline{Z}} - \overline{\overline{Z}}_{rb}\,\overline{\overline{Z}}_{b}^{-1}\overline{\overline{Z}}_{br})\overline{J} = \overline{b}_i \tag{4.31}$$

In principle, Eq. (4.31) is then ready for solution by matrix inversion. The problem is that the matrix inversion for a large number of unknowns becomes computationally formidable. Thus, the following efficient method, sparse-matrix canonical-grid method is used.

To speed up the solution of Eq. (4.31), we decompose the impedance matrix for the rough surface $\overline{\overline{Z}}$ into three parts: a block Toeplitz flat-surface part $\overline{\overline{Z}}^{FS}$, strong interaction part $\overline{\overline{Z}}^{S}$, and the weak remainder $\overline{\overline{Z}}^{W}$. With the weak remainder part moved to the right-hand side, we have

$$(\overline{\overline{Z}}^{FS} + \overline{\overline{Z}}^{S} - \overline{\overline{Z}}_{rb}\,\overline{\overline{Z}}_{b}^{-1}\overline{\overline{Z}}_{br})\overline{J} = \overline{b}_i - \overline{\overline{Z}}^{W}\overline{J} \tag{4.32}$$

Eq. (4.32) is then solved by the conjugate gradient method (CGM) with an updated right-hand side.

For a small object, the solution of Eq. (4.32) can be further speeded up with the buried object term moved to the right-hand side, as given by

$$(\overline{\overline{Z}}^{FS} + \overline{\overline{Z}}^{S})\overline{J} = \overline{b}_i + \overline{\overline{Z}}_{rb}\,\overline{\overline{Z}}_{b}^{-1}\overline{\overline{Z}}_{br}\overline{J} - \overline{\overline{Z}}^{W}\overline{J} \tag{4.33}$$

The new matrix equations (4.32) and (4.33) are solved using CGM [2]. The product of $\overline{\overline{Z}}^{(FS)}$ with $\overline{J}$ can be computed using a 2-D fast FFT algorithm which makes conjugate gradient iteration efficient. Updating the right-hand side is also quickly calculated. The iteration is terminated when the error norm criterion is less than 0.2%.

4.2.4 Statistics of Scattered Fields

After the surface currents are solved, the scattered fields in medium 0 can be calculated. The scattering amplitudes for both the co-polarized and cross-polarized

polarizations $F_{\beta\alpha}$ are respectively

$$\begin{aligned} F_{h\alpha} &= \frac{ik}{4\pi\sqrt{2\eta P_\alpha^i}} \int_{dS'} dx'dy' \exp(-ik\gamma')[\{J_1(x',y')\cos\theta_s\cos\phi_s \\ &+ J_2(x',y')\cos\theta_s\sin\theta_s - J_1(x',y')\frac{\partial f(x',y')}{\partial x'}\sin\theta_s - J_2(x',y')\frac{\partial f(x',y')}{\partial y'}\sin\theta s\} \\ &- \eta\{J_4(x',y')\sin\phi_s - J_5(x',y')\cos\phi_s\}] \end{aligned} \tag{4.34}$$

and

$$\begin{aligned} F_{v\alpha} &= \frac{ik}{4\pi\sqrt{2\eta P_\alpha^i}} \int_{dS'} dx'dy' \exp(-ik\gamma')[\{J_1(x',y')\sin\varphi_s - J_2(x',y')\cos\phi_s\} \\ &+ \eta\{J_4(x',y')\cos\theta_s\cos\theta_s + J_5\cos\theta_s\sin\phi_s - J_4(x',y')\frac{\partial f(x',y')}{\partial x'}\sin\theta_s \\ &- J_5(x',y')\frac{\partial f(x',y')}{\partial y'}\sin\theta_s\}] \end{aligned} \tag{4.35}$$

where $\gamma' = x'\sin\theta_s\cos\phi_s + y'\sin\theta_s\sin\phi_s + f(x',y')\cos\theta_s$. It is noticed that the scattering amplitudes are normalized by the square root of the incident power $2\eta P_{i\alpha}$ as

$$P_\alpha^i = \frac{2\pi^2}{\eta}\int\limits_{k_\rho<k} dk_x dk_y |E_{i\alpha}(k_x,k_y)|^2\frac{k_z}{k} \tag{4.36}$$

Since the buried object is under a single random rough surface, the realization averaging is not applicable. Both the bistatic scattering coefficient (normalized RCS) and ACF are calculated based on azimuthal averaging as follows

$$\sigma_{\beta\alpha}(\theta_s,\theta_i) = \frac{1}{N\phi}\sum_{n=1}^{N_\phi}|F_{\beta\alpha}(\theta_s,\phi_{sn};\theta_i,\phi_{in})|^2 \tag{4.37}$$

and

$$\Gamma_{\beta\alpha}(\theta_{s2},\theta_{i2};\theta s1,\theta i1) = \frac{1}{N\phi}\sum_{n=1}^{N_\phi}F_{\beta\alpha}(\theta_{s2},\phi_{s2n};\theta_{i2},\phi_{i2n})F^*_{\beta\alpha}(\theta_{s1},\phi_{s1n};\theta_{i1},\varphi_{i1n}) \tag{4.38}$$

where ϕ_{in} and ϕ_{sn} are incident and scattering azimuthal angles. In the scattering plane, they are related to each other by the relations of (i) $\phi_{sn} = \phi_{in}$ for θ_s having the same sign as θ_i, (ii) $\phi_{sn} = \phi_{in} + 180^o$ for θ_s having a opposite sign of θ_i.

Eq. (4.38) can be extended to calculate the polarization-angular correlation function as follows

$$\Gamma_{\beta_1\alpha_1\beta_2\alpha_2}(\theta_{s1},\theta_{i1};\theta s2.\theta i2) = \frac{1}{N\phi}\sum_{n=1}^{N_\phi} F_{\beta_1\alpha_1}(\theta_{s1},\phi_{s1n};\theta_{i1},\phi_{i1n})F^*_{\beta_2\alpha_2}(\theta_{s2},\phi_{s2n};\theta_{i2},\phi_{i2n}) \tag{4.39}$$

4.3 Results and Discussions

The numerical simulation is conducted for a perfectly electrical conducting (PEC) sphere buried under a 2-D Gaussian random rough surface. The rough surface is generated by using the spectrum method with an assumption of a Gaussian spectrum. The sizes of the rough surface in the x and y directions are $L_x = L_y = 8.0\lambda$. The surface rms heights are $h_x = h_y = 0.02\lambda$, and the correlation lengths are $l_x = l_y = 0.5\lambda$. The relative dielectric constant of the lower medium is $\epsilon_r = (2.0 + i0.2)$. The surface is sampled at 64 points per λ^2 to give 24576 surface unknowns. The neighborhood distance $r_d = 3.5\lambda$. The sphere of radius of $a = 0.3\lambda$ is buried under the rough surface at a depth of $d = 0.6\lambda$. The sphere surface is discretized into 80 triangle patches as shown in Figure 4.2, for which the surface currents are represented by 240 unknowns. The impedance matrix of the sphere $\overline{\overline{Z}}_b$ is calculated and tested by calculating the scattering cross section of a sphere in free space. The numerical results agree with that obtained by Mie scattering as shown in Figure 4.3.

We solve the matrix equations with the buried object contribution on the left-hand side (4.32) and that on the right side (4.33). Both give the same result. The scattering coefficients are shown in Figure 4.4. The CPU with the target term on the right-hand side is five times faster than the CPU with the target term on the left-hand side.

We calculate the scattering amplitudes for 10 azimuthal angles at 0^o, 36^o,..., and 324^o, respectively. The RCS and ACF are calculated by using azimuthal angular averaging as given in Eq. (37) and (38). We plot the results as functions of the

scattering angle θ_{s2}. Parameters for other angles are $\theta_{i1} = 20°$, $\theta_{s1} = -40°$, and $\theta_{i2} = 20°$. Figure 4.5 shows the results for hh polarization component. Both the results with and without the target sphere are shown for comparison. Figure 4.5b is for RCS. As expected, there is a peak in the specular direction, which is due to the slightly rough surface. The difference of RCS with and that without the target is large only for large scattering angles, since the rough surface scattering is small in these cases. As shown in Figure 4.5a, however, the difference of ACF can be 7dB, even for angles closed to the nadir direction. This is because the memory effect is avoided and rough surface scattering is minimized.

The fully polarimetric results of RCS and ACF are calculated and shown in Figures 4.6 and 4.7. Figure 4.6 shows the results of RCS. We see that the differences of RCS with and without the target for co-polarizations are larger than those for cross-polarizations. This is because the cross-polarization components are mainly due to the rough surface scattering. Since the target is a sphere, it has only a small cross-polarization contribution. It can also be found that there are larger differences for the vv component than the hh component, since the vertical polarization wave has better penetration through the surface. Figure 4.7 shows the results of ACF. We can see the big difference of ACF with and without a target in both the co-polarization and cross-polarization result. This is because the random phase of rough surface scattering which is included in the ACF calculation. The results of polarization-angular correlation function (PACF) are also calculated and shown in Figure 4.8. Figure 4.8a is the ACF between hh components and vh components. Figure 4.8b is that between vv components and hv components. We can see an even larger difference up to 10 to 20dB.

4.4 Conclusions and Discussions

In this chapter, we studed electromagnetic wave scattering from a 3-D buried object under a 2-D random rough surface based on the formulation of surface integral equations. The surface fields are then solved by the method of moments. The scattered wave field from the object is represented by the rough surface field so that the matrix equation can be solved efficiently using the sparse-matrix canonical-grid method. Numerical simulations are illustrated for a PEC sphere buried under a 2-D rough surface. Both RCS and ACF are calculated based on azimuthal angular averaging. Results show that ACF is more effective in suppressing the clutter due to rough surface scattering than RCS, and the cross-polarization components of ACF can also be used for target detection. PACF can be even more effective for target detection.

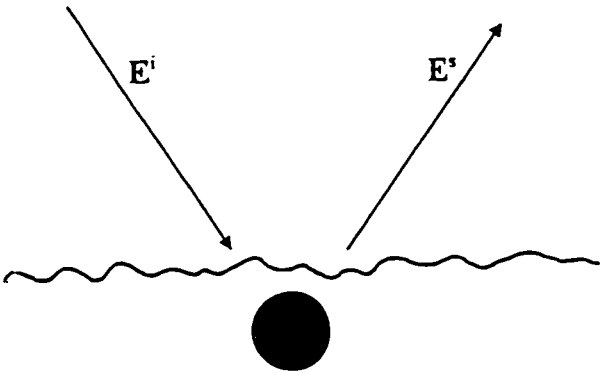


Figure 4.1: Configuration of EM wave scattering by a 3-D object buried under a 2-D rough surface.

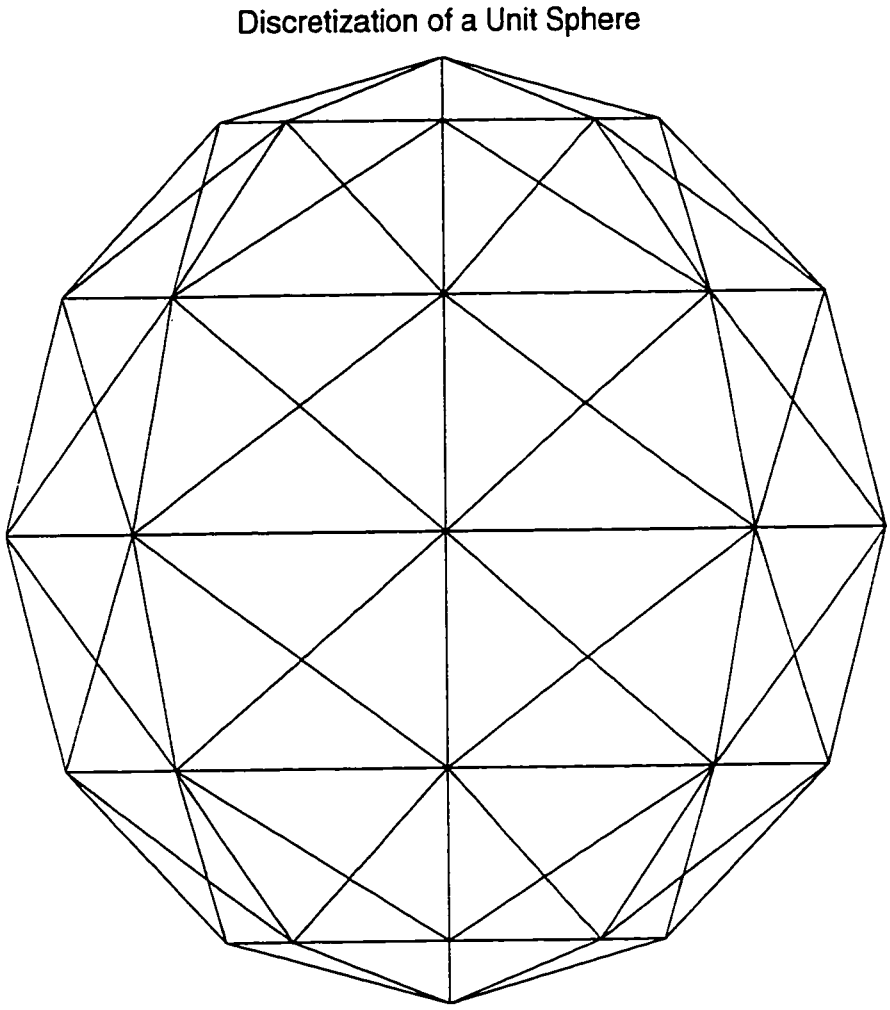


Figure 4.2: Discretization of a sphere into 80 triangle patches.

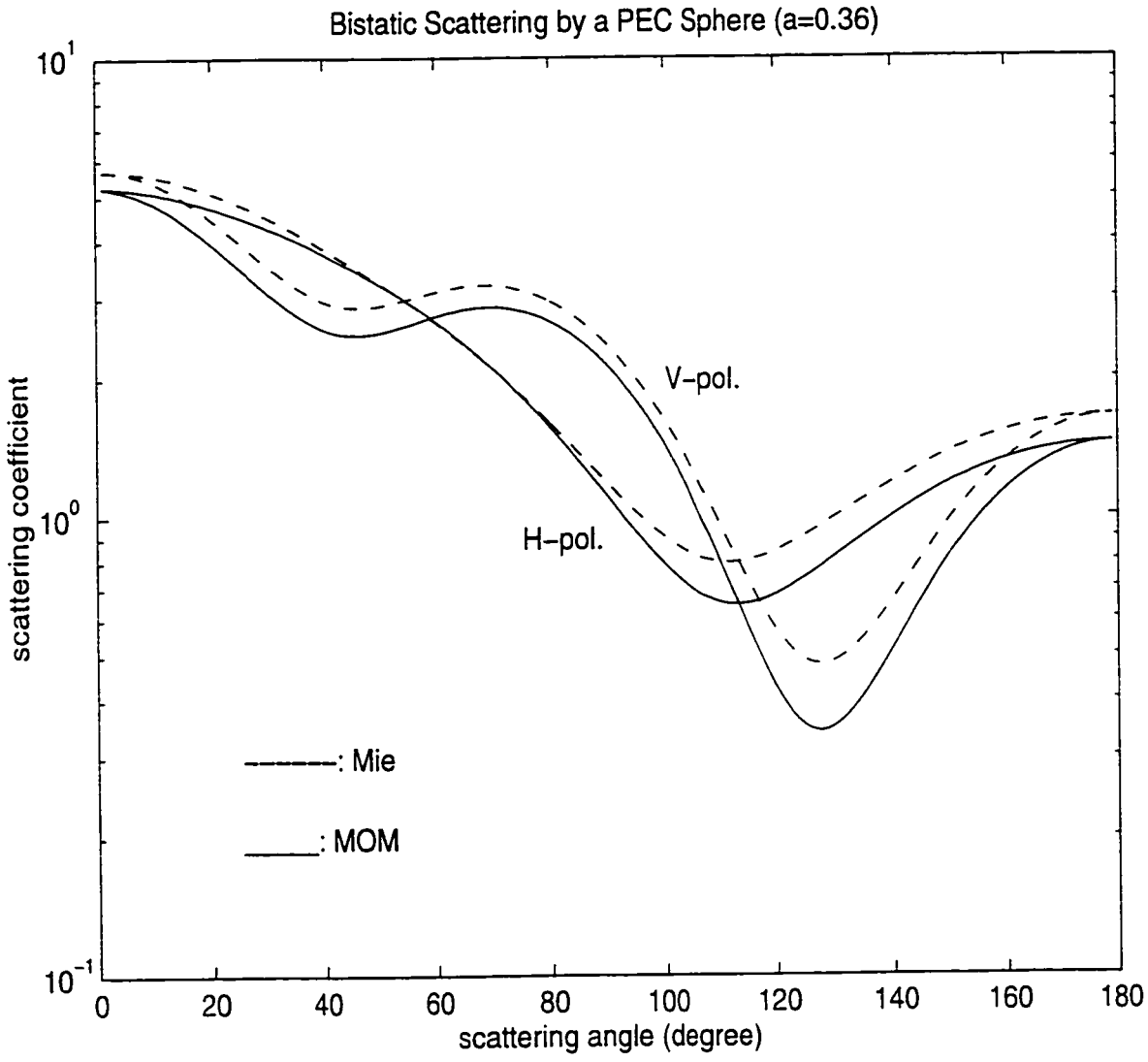


Figure 4.3: Comparison of MoM and Mie scattering for the radar cross section of a PEC sphere.

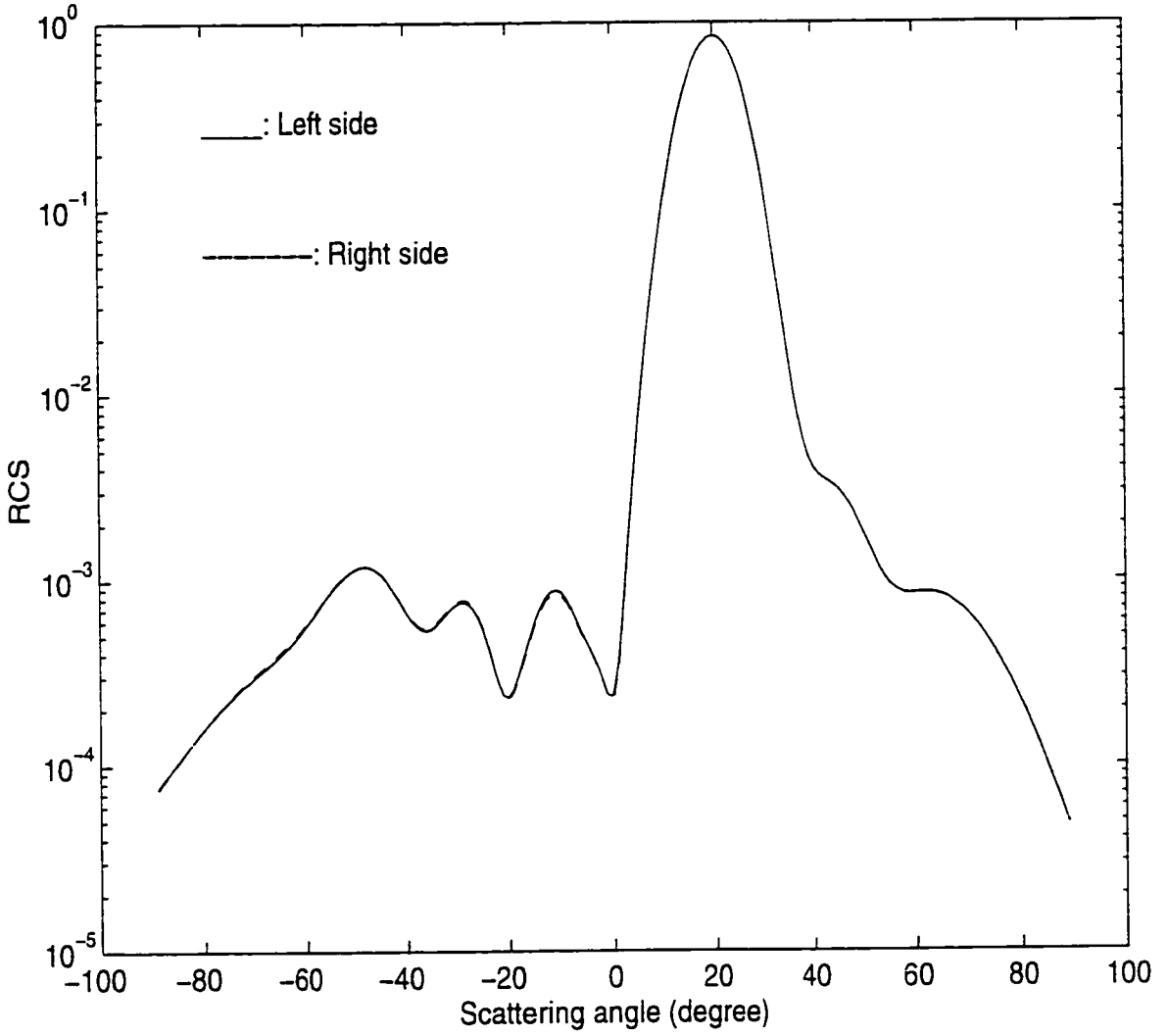


Figure 4.4: Comparison between the solution of the matrix equation for the object on left-hand side and that on right-hand side.

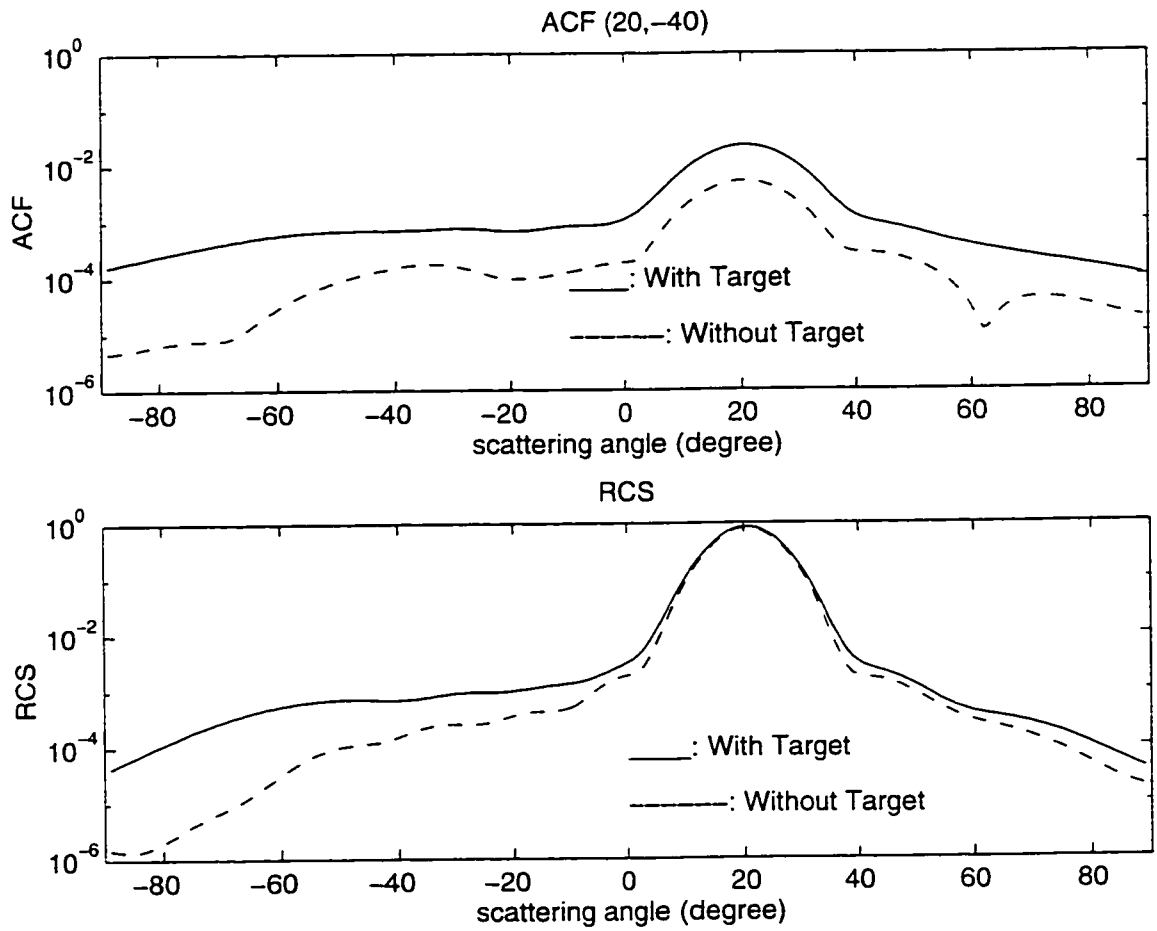


Figure 4.5: ACF and RCS of EM wave scattering by a PEC sphere buried under a 2-D rough surface for hh polarization component. Parameters are: $a = 0.3\lambda$, $d = 0.6\lambda$;$L_x = L_y = 8.0\lambda$, $h = 0.02\lambda$, $l_x = l_y = 0.5\lambda$, $r_d = 3.5\lambda$, $\theta_{i1} = 20°$, $\theta_{s1} = -40°$, and $\theta_{i2} = 20°$.

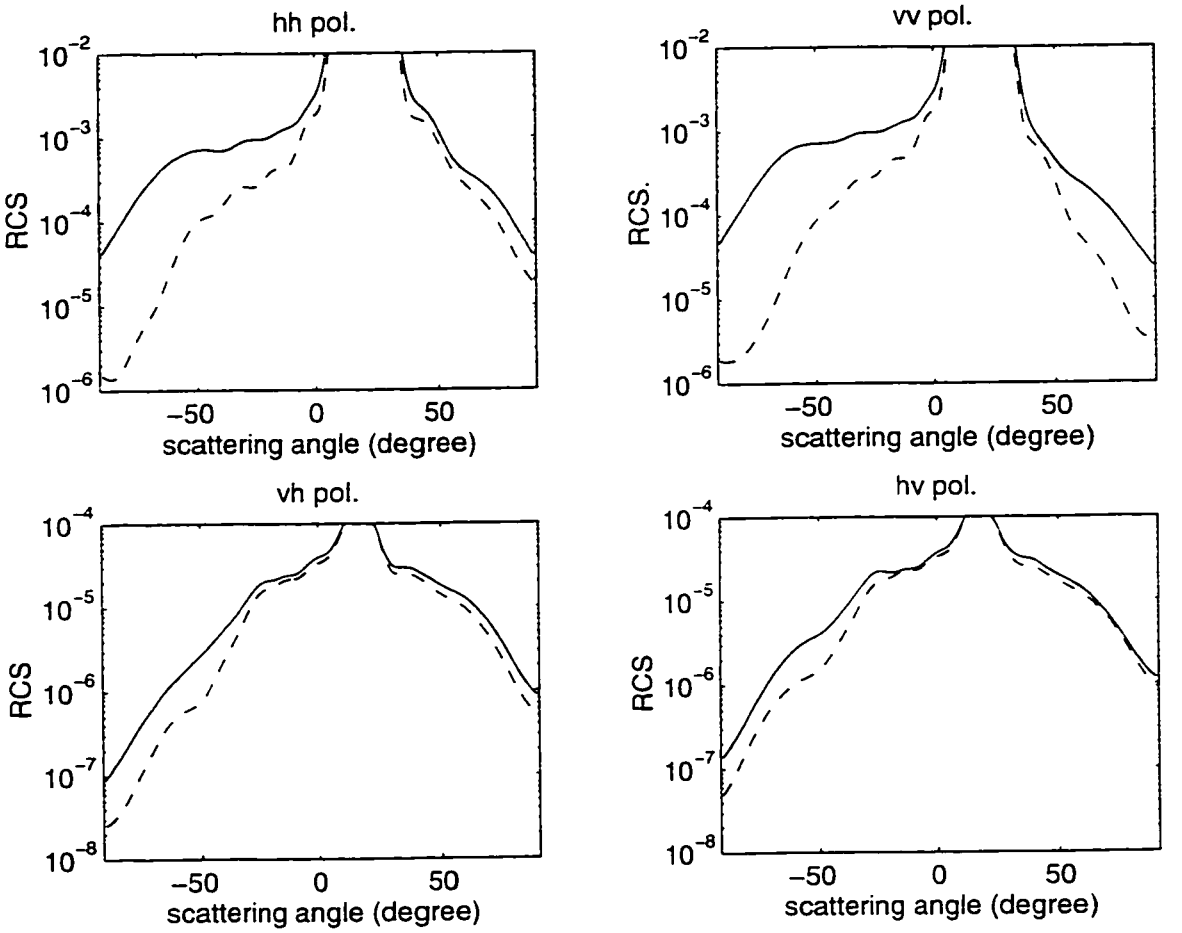


Figure 4.6: RCS of EM wave scattering by a PEC sphere buried under a 2-D rough surface for co-pol. and cross-pol. components. Parameters are: $a = 0.3\lambda$, $d = 0.6\lambda$;$L_x = L_y = 8.0\lambda$, $h = 0.02\lambda$, $l_x = l_y = 0.5\lambda$, $r_d = 3.5\lambda$, $\theta_{i1} = 20°$, $\theta_{s1} = -40°$, and $\theta_{i2} = 20°$.

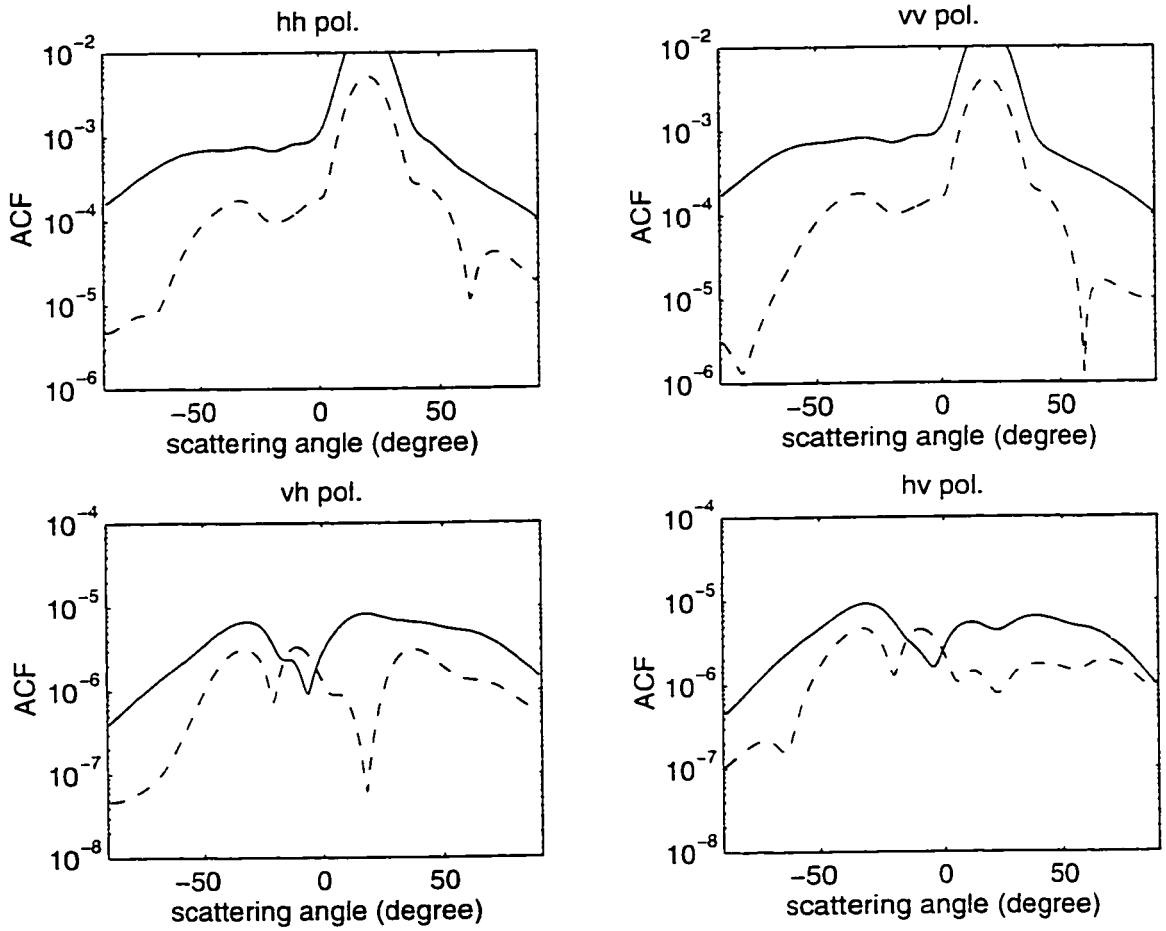


Figure 4.7: ACF of EM wave scattering by a PEC sphere buried under a 2-D rough surface for co-pol. and cross-pol. components. Parameters are: $a = 0.3\lambda$, $d = 0.6\lambda$;$L_x = L_y = 8.0\lambda$, $h = 0.02\lambda$, $l_x = l_y = 0.5\lambda$, $r_d = 3.5\lambda$, $\theta_{i1} = 20°$, $\theta_{s1} = -40°$, and $\theta_{i2} = 20°$.

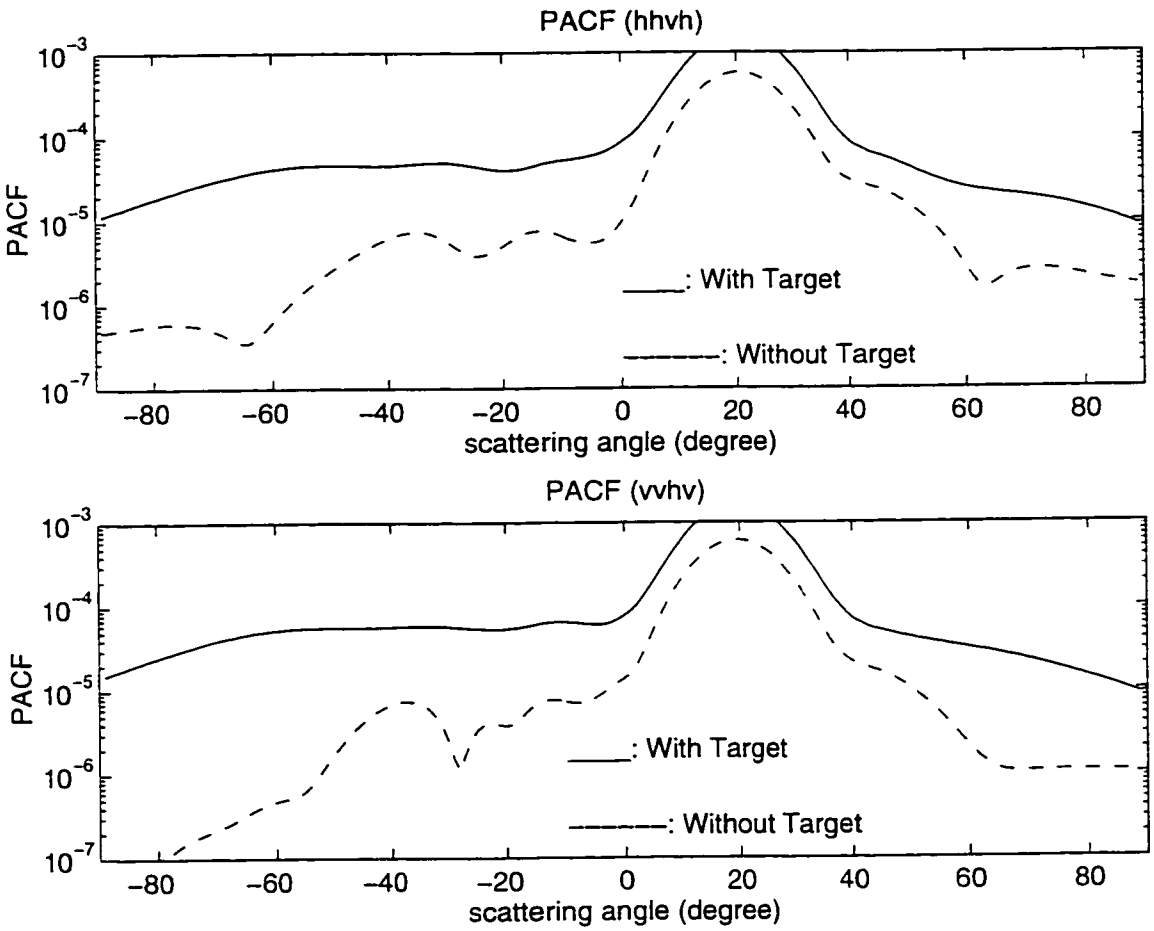


Figure 4.8: PACF of EM wave scattering by a PEC sphere buried under a 2-D rough surface between co-pol. and cross-pol. components. Parameters are: $a = 0.3\lambda$, $d = 0.6\lambda$;$L_x = L_y = 8.0\lambda$, $h = 0.02\lambda$, $l_x = l_y = 0.5\lambda$, $r_d = 3.5\lambda$, $\theta_{i1} = 20°$, $\theta_{s1} = -40°$, and $\theta_{i2} = 20°$. (a): hhvh, (b) vvhv

Chapter 5

APPLICATION OF ANGULAR CORRELATION FUNCTION OF CLUTTER SCATTERING AND CORRELATION IMAGING IN TARGET DETECTION

5.1 Introduction

In previous chapters, we have studied the angular correlation function of EM wave scattering from random media and rough surfaces with and without a target. It has been shown that ACF is more effective in supressing clutter effects than RCS. For target detection, other methods have been proposed and applied to supress clutter. Wavelet transform techniques have been used to separate targets from clutter [49]. Polarimetric measurements have been made for the detection of targets buried in a natural snowpack to enhance the visibility [48]. SAR imaging has attracted great interest [1][33][48]. Resolution and signal-noise ratio are two important criteria in image processing. A radar system of fine resolution usually requires a high operating frequency. At high frequencies, however, waves are also scattered by clutter such as rough surfaces and random media, causing a low signal-noise ratio. Therefore, research has been conducted on developing data processing methods which give finer resolution and suppress clutter.

Image processing is to obtain detailed information from wide-band and large angular range measurements. The conventional SAR imaging method can be called field imaging, in which the target function is obtained from the measured fields by inverse Fourier transform [33] or by focusing [1][21]. The focusing method is also called correlation imaging since the field is correlated with a reference signal. The field focusing

method can be improved by introducing a filter function such as matched filtering if the scattering property of the target is known [21]. However, if the target scattering function is not known, the matched filtering method can not be applied. In this chapter, we study a different imaging method that is based on the calculation of the mutual correlation function of two received signals.

In last two chapters, we have studied the angular correlation function for the detection of targets embedded in clutter environments. Correlation functions are obtained by taking averages over a product of two signals. Thus, a key step of calculating the ACF is taking averages. For random media scattering, the average is usually taken over realizations of random media or rough surfaces [42]. In the studies of the detection of targets embedded in the clutter environments, we have used frequency averaging and angular averaging to obtain the ACF [43][54]. Numerical results have shown that detection of the targets by the ACF with frequency and angular averaging has advantages over the radar cross section (RCS).

Correlation imaging calculates the correlation function with focusing on every position in a region giving an image is obtained for the region. The memory effect of the mutual correlation function for random media scattering is avoided in image processing. Correlation imaging uses the phase difference (or sum) of two wave propagation paths, while field imaging uses the phase of one path. Therefore, correlation imaging uses larger spectrum domain information, and can give a finer resolution and suppress clutter. Correlation imaging can be implemented by calculating the angular correlation function (ACF), frequency correlation function (FCF), and frequency-angular correlation function (FACF) of the scattered fields.

In this chapter, we perform numerical simulations of SAR imaging. In section II, we first develop the basic equations of correlation imaging. The similarities and differences between field imaging and correlation imaging are discussed. In section III, we describe how the SAR scattered fields of targets embedded in clutter are generated by using Monte Carlo simulations. In section IV, the angular correlation

imaging (ACF imaging) method is illustrated for circular SAR. which gives finer resolution and better signal-clutter ratio of 100%. The decrease of range resolution due to the frequency dependence of the scattering is compensated for by the cross-range resolution. In section V, the correlation imaging method is studied for linear SAR. It is found that the frequency-angular correlation imaging is needed. The imaging results are also shown by numerical simulations. The advantages of using correlation is that the clutter can be suppressed by avoiding the memory effect of random scattering. Avoiding the memory effect in spectral domain is equivalent to the gating in the time or space domain.

5.2 Formulation

Imaging is to obtain detailed information of a large region by measuring the scattered field from many view angles and frequencies. Under the far-field approximation. the received signal is a set of data $E(\overline{k})$ which can be written as

$$E(\overline{k}) = \int f(\overline{r}, \overline{k}) e^{i\overline{k}\cdot\overline{r}} d\overline{r} \tag{5.1}$$

where $\overline{k} = \overline{k}_i - \overline{k}_s$ is the difference between the incident wave vector $\overline{k}_i$ and the scattered wave vector $\overline{k}_s$. Thus, $\overline{k}$ is a function of frequency and view angle. The target function $f(\overline{r}, \overline{k})$ is frequency and angular dependent. In the following derivation. the dependence is ignored for simplicity. From (5.1), the target function is the inverse Fourier transform of received signals.

$$f(\overline{r}) = \frac{1}{(2\pi)^3} \int E(\overline{k}) e^{-i\overline{k}\cdot\overline{r}} d\overline{k} \tag{5.2}$$

A second way to obtain the target function is by focusing. The received signal $E(\overline{k})$ is correlated with a reference response focused on $\overline{r}_o$, i. e.

$$C_F(\overline{r}_o) = \frac{1}{(2\pi)^3} \int d\overline{k} E(\overline{k}) E_o^* e^{-i\overline{k}\cdot\overline{r}_o} \tag{5.3}$$

Substituting (5.1) in (5.3) and using $E_o = 1$, we have

$$\begin{aligned} C_F(\overline{r}_o) &= \frac{1}{(2\pi)^3}\int d\overline{k}\int d\overline{r} f(\overline{r}) e^{i\overline{k}\cdot\overline{r}} E_o^* e^{-i\overline{k}\cdot\overline{r}_o} \\ &= f(\overline{r}_o) \end{aligned} \tag{5.4}$$

We will use the notation that

$$< w(\overline{k}) >_{\overline{k}} = \frac{1}{(2\pi)^3}\int d\overline{k} w(\overline{k}) \tag{5.5}$$

where $< ... >_{\overline{k}}$ denotes an averaging over $\overline{k}$ space. This operation can be interpreted as spectrum averaging. Eq. (5.3) is also called correlation imaging with E_o named as the filter function.

Both methods (inverse Fourier transform and focusing) of finding the target function in (5.2) and (5.3) are based on the received field, which we shall call field imaging.

In correlation imaging, we calculate the correlation function of two received signals with focusing by spectrum averaging as

$$\begin{aligned} C_\Gamma(\overline{r}_o) &= < E(\overline{k}_1) e^{-i\overline{k}_1\cdot\overline{r}_o} E^*(\overline{k}_2) e^{i\overline{k}_2\cdot\overline{r}_o} >_{\overline{k}_1,\overline{k}_2} \\ &= < \int f(\overline{r}_1) e^{i\overline{k}_1\cdot(\overline{r}_1-\overline{r}_o)} d\overline{r}_1 \int f^*(\overline{r}_2) e^{-i\overline{k}_2\cdot(\overline{r}_2-\overline{r}_o)} d\overline{r}_2 >_{\overline{k}_1,\overline{k}_2} \\ &= |f(\overline{r}_o)|^2 \end{aligned} \tag{5.6}$$

Thus the correlation imaging of (5.6) gives the target function $|f(\overline{r}_o)|^2$. The correlation imaging can be related to frequency-angular correlation function of scattering as follows.

First, we let $\overline{k} = \overline{k}_1$ and $\overline{k}_d = \overline{k}_2 - \overline{k}_1$. Then

$$C_\Gamma(\overline{r}_o) = \frac{1}{(2\pi)^6}\int d\overline{k}_d \int d\overline{k} E(\overline{k}) E^*(\overline{k}+\overline{k}_d) e^{i\overline{k}_d\cdot\overline{r}_o} \tag{5.7}$$

The frequency angular correlation function is

$$\begin{aligned} \Gamma(\overline{k}_d) &= < E(\overline{k}) E^*(\overline{k}+\overline{k}_d) >_{\overline{k}} \\ &= \frac{1}{(2\pi)^3}\int d\overline{k} E(\overline{k}) E^*(\overline{k}+\overline{k}_d) \end{aligned} \tag{5.8}$$

Using (5.8) in (5.7), we have

$$C_\Gamma(\overline{r}_o) =< \Gamma(\overline{k}_d)e^{i\overline{k}_d\cdot\overline{r}_o} >_{\overline{k}_d} \tag{5.9}$$

The quantity $\Gamma(\overline{k}_d)$ is known as the frequency-angular correlation function (FACF) of scattering. The angular correlation function (ACF) of scattering by random media (clutter) has been studied extensively [6][18][19][10]. It has been shown that the scattering by random media does not contribute to ACF except along the memory directions of $\overline{k}_d = 0$.

For imaging purposes, random media scattering can be minimized by avoiding the memory effect by $\overline{k}_d \neq 0$. We also define a general expression for the correlation imaging function

$$\begin{aligned} C_\Gamma(\overline{r}_o) &= \frac{1}{(2\pi)^6}\int d\overline{k}_1 \int_{k_d\neq 0} d\overline{k}_2 E(\overline{k}_1)e^{-2ik_1|\overline{R}(\overline{k}_1)-\overline{r}_o|} \\ &\quad E^*(\overline{k}_2)e^{2ik_2|\overline{R}(\overline{k}_2)-\overline{r}_o|}W(\overline{k}_1;\overline{k}_2) \end{aligned} \tag{5.10}$$

where $\overline{R}(\overline{k}_1)$ and $\overline{R}(\overline{k}_2)$ are distances between the focusing position and the receiver positions. In Eq. (5.10), $W(\overline{k}_1;\overline{k}_2)$ is a weighting function, which can be chosen to further suppress clutter effect if the clutter scattering characteristics are known. In this paper, however, it is assumed to be unity for simplicity. We will use equation (5.10) to process (simulated) SAR data. In practice, the integration is discritized and approximated by finite sums.

The correlation imaging can be obtained by calculating angular correlation function (ACF), frequency correlation function (FCF) and frequency-angular correlation function (FACF). The central idea of correlation imaging is to use a large phase difference for focusing and to avoid the memory effect so as to suppress clutter. The large phase difference can not be obtained by ACF imaging for linear SAR. In this case, frequency-angular correlation imaging (FACF imaging) is used instead.

5.3 Monte Carlo Simulations of SAR data

In this section, we describe Monte Carlo simulations of wave scattering that are used to simulate SAR data.

Consider an incident wave impinging upon targets embedded in a random medium (Figure 5.1). Let the incident wave be given by

$$E_i(\overline{k}_i, \overline{r}) = e^{ik|\overline{R}_t - \overline{r}|} \tag{5.11}$$

where $\overline{R}_t$ is antenna position, $\overline{r}$ is field position. We have only included the phase factor and neglected the divergence factor in (5.11).

In the numerical simulation, the random medium is modeled by a collection of randomly distributed small scatterers. We have used 80,000 small scatterers in our simulations.

By assuming single scattering, the scattered field at the receiver is given by (with the divergence factor $1/(\overline{R} - \overline{r})$ ignored)

$$E(\overline{k}, \overline{R}) = \sum_{n=1}^{N_t} e^{ik|\overline{R}(\hat{k}_i) - \overline{r}_n| - |\overline{R}(\hat{k}_s) - \overline{r}_n|} F_{tn}(\hat{k}_s, \hat{k}_i) + \sum_{n=1}^{N_p} e^{ik|\overline{R}(\hat{k}_i) - \overline{r}_n| - |\overline{R}(\hat{k}_s) - \overline{r}_n|} f_n(\hat{k}_s, \hat{k}_i) \tag{5.12}$$

where the received field is expressed in terms of the scattering amplitude of targets $F_{tn}(k)$ and the scattering amplitudes of randomly distributed particles $f_n(k)$ (n=1,2,$\cdots$,N_p).

In this manner, we generate the scattered field for many combinations of incident and scattered field directions and frequencies. These will be the simulated SAR data of scattered fields. In the subsequent two sections we describe how these simulated data are processed in circular SAR and linear SAR.

5.4 Circular SAR Imaging

Let a mono-static radar be moving along a circular path of radius of R_o and at height H above the target region (Figure 1a). The radar position is a function of the

azimuthal angle ϕ. Thus the vector $\overline{R}$ in (5.12) becomes $\overline{R}(\phi) = R_o cos\phi\hat{x} + R_o sin\phi\hat{y} + H\hat{z}$. We write the total scattered field at the radar as

$$E(k,\phi) = \sum_{n=1}^{N_t} e^{2ik|\overline{R}(\phi)-\overline{r}_n|} F_{tn}(k) + \sum_{n=1}^{N_p} e^{2ik|\overline{R}(\phi)-\overline{r}_n|} f_n(k) \tag{5.13}$$

where the received signal is expressed in terms of the scattering amplitude of targets $F_{tn}(k)$ and the scattering amplitudes of randomly distributed particles $f_n(k)$ (n=1.2.$\cdots$.N_p). The factor of 2 in the phase accounts for the round trip phase shift of a mono-static radar.

The conventional SAR imaging can be realized by correlating the received signal with a reference signal E_o and then summing over frequency and angle.

$$C_F(\overline{r}_o) = \sum_{m=1}^{N_k} \sum_{n=1}^{N_\phi} E(k_m,\phi_n) E_o^* e^{-2ik_m|\overline{R}(\phi_n)-\overline{r}_o|} \tag{5.14}$$

The reference signal E_o is chosen to be a constant. The correlation defined in (5.14) has well defined peaks when $\overline{r}_o$ approaches the location of the scatterers as discussed in section 5.2.

The image can also be obtained by the angular correlation imaging (ACF imaging) method. The ACF imaging correlates two signals received at different angles and sums over frequency and angles. It is given by

$$\begin{aligned} C_\Gamma(\overline{r}_o) \quad = \quad & \sum_{m=1}^{N_k} \sum_{n_1=1}^{N_\phi} \sum_{n_2=1,|n_2-n_1|>N_o}^{N_\phi} E(k_m,\phi_{n_1}) e^{-2ik_m|\overline{R}(\phi_{n_1})-\overline{r}_o|} \\ & E^*(k_m,\phi_{n_2}) e^{2ik_m|\overline{R}(\phi_{n_2})-\overline{r}_o|} \end{aligned} \tag{5.15}$$

Note that the integrand in (5.15) is the angular correlation function. The memory effect of random scattering is avoided by excluding $n_2 = n_1$. In the simulations. we let $|n_1 - n_2| > N_o$ so that the angular difference $\phi_d = |\phi_{n_1} - \phi_{n_2}|$ is larger than the correlation angle of the angular correlation function. The choice of N_o depends on the correlation angle of random scattering. The correlation angle of random scattering is usually small, which is in the order of λ/L with the size of the medium L. Only a

small portion of information is taken out for the target. Therefore, target information is preserved while the clutter due to random scattering is suppressed.

In the numerical simulations, 80,000 small particles with radius of $a = 0.04\lambda_o$ are used to model clutter. The particles are randomly distributed in a layer region of $40\lambda_o$X$40\lambda_o$X $0.5\lambda_o$. Four target spheres with radius of $a = 0.3\lambda_o$ are placed at positions of $(10,14.0)\lambda_o$, $(10,28,0)\lambda_o$, $(30,10,0)\lambda_o$, and $(30,32,0)\lambda_o$. The dielectric constant of both targets and small particles are set as (3.23+i0.36). Distances are expressed in terms of the wavelength λ_o at the center frequency. The circular path of the receiver has a radius of $R_o = 1732\lambda_o$ and at a height of $H = 1000\lambda_o$. The back-scattering amplitudes for targets and particles are calculated based on Mie scattering. The received signal is calculated by Eq. (5.13). The frequency band is from $0.5f_o$ to $1.5f_o$ with an increment of $0.01f_o$. Thus $N_k = 100$, $k_1 = 0.5f_o$, $k_2 = 0.51f_o$, ... The azimuth angle is from 0^o to 360^o at an interval of 3.6 degree. Thus $N_\phi = 100$, $\phi_1 = 0^o$, $\phi_2 = 3.6^o$, $\phi_3 = 7.2^o$, etc.

The simulated data is processed by field imaging and ACF imaging. In ACF imaging, we calculate Eq. (5.15). Note that (5.15) is the product of fields. For a fair comparison, we calculate $|C_F(\bar{r}_o)|^2$ in Figure 5.2. The results are normalized by the maximum value which is the value at the target. The normalized results are shown in Figures 5.2 and 5.3. Figure 5.2 shows field imaging, which is the square of Eq. (5.14). The four targets are spread out to a two wavelength size and are obscured by the back ground clutter. Figure 5.3 shows the results of ACF imaging. We see that the spot size of targets is smaller, which is a result of finer resolution. The background clutter is significantly suppressed. We use the condition of $|n_1 - n_2| > 5$, i.e. $N_o = 5$ to ensure that the angular pairs ϕ_{n_1} and ϕ_{n_2} are away from the memory dot. The better result of ACF imaging is because (1) the clutter effect is minimized by avoiding the memory effect of random scattering, and (2) the spreading due to the frequency dependence of scattering is compensated for by the cross-range resolution in ACF imaging. The angular correlation function with focusing is also shown in Figure 5.4.

The ACF focusing $C_\Gamma(\overline{r}_o, \phi_d)$ is obtained by using Eq. (5.15) with the summation over $\sum_{n_2=1}^{N_\phi}$ replaced by just one value of $n_2 = n_1 + n_d$, where n_d is a fixed number

$$C_\Gamma(\overline{r}_o, \phi_d) = \sum_{m=1}^{N_k} \sum_{n=1}^{N_\phi} E(k_m, \varphi_n) e^{-2ik_m|\overline{R}(\phi_n)-\overline{r}_o|} E^*(k_m, \phi_{n+n_d}) e^{2ik_m|\overline{R}(\phi_{n+n_d})-\overline{r}_o|} \tag{5.16}$$

In (5.16), $\varphi_d = |\phi_n - \phi_{n+n_d}|$ and we set $\phi_d = 180^o$ in Figure 5.4. Therefore, ACF imaging is the summation over many images of ACF focusing. The ACF focusing is a random function of correlation angle ϕ_d when $\overline{r}_o$ is not at a target. Hence, clutter is further suppressed by taking averages of (5.16) to obtain (5.15).

5.5 *Linear SAR Imaging*

As shown in Figure 5.1b, a mono-static radar moves along a linear path at height H above the target region. The radar position is a function of the azimuthal position x as $\overline{R}(x) = x\hat{x} - d\hat{y} + H\hat{z}$. The received signals can be written as

$$E(k, x) = \sum_{n=1}^{N_t} e^{2ik|\overline{R}(x)-\overline{r}_n|} F_{tn}(k) + \sum_{n=1}^{N_p} e^{2ik|\overline{R}(x)-\overline{r}_n|} f_n(k) \tag{5.17}$$

where the received signal is a function of wave number k (frequency) and antenna position x. The scattering amplitude of targets is $F_{tn}(k)$ and the scattering amplitudes of randomly distributed particles are $f_n(k)$ (n=1.2,$\cdots$.N_p).

The field imaging correlates the received signal with a reference signal E_o and then sums over frequency and azimuthal position.

$$C_F(\overline{r}_o) = \sum_{m=1}^{N_k} \sum_{n=1}^{N_x} E(k_m, x_n) E_o(k_m)^* e^{-2ik_m|\overline{R}(x_n)-\overline{r}_o|} \tag{5.18}$$

The reference signal E_o is chosen to be 1 for simplicity.

In circular SAR the angular correlation imaging can give clear images. In linear SAR, however, spatial correlation imaging gives a fine resolution in cross-range, and

frequency correlation imaging gives range resolution. In general, frequency spatial correlation is needed in correlation imaging for linear SAR data. It is given by

$$C_\Gamma(\overline{r}_o) = \sum_{m_1=1}^{N_k}\sum_{n_1=1}^{N_x}\sum_{m_2=1}^{N_k}\sum_{n_2=1,k_d>k_o/16}^{N_x} E(k_{m_1},x_{n_1})e^{-2ik_{m_1}|\overline{R}(x_{n_1})-\overline{r}_o|} \quad (5.19)$$
$$E^*(k_{m_2},x_{n_2})e^{2i(k_{m_2})|\overline{R}(x_{n_2})-\overline{r}_o|}$$

The memory effect is avoided in the summation by excluding $\overline{k}_d \neq 0$. In actual simulations, we choose $k_d > \frac{k_o}{16}$, where k_o is the wave number of center frequency. The magnitude of the difference of the two wave vectors k_d is given by

$$\begin{aligned} k_d &= [(k_{m_1}\cos\phi_1\sin\theta_1 - k_{m_2}\cos\phi_2\sin\theta_2)^2 \quad (5.20)\\ &+ (k_{m_1}\sin\phi_1\sin\theta_1 - k_{m_2}\sin\phi_2\sin\theta_2)^2 + (k_{m_1}\cos\theta_1 - k_{m_2}\cos\theta_2)^2]^{1/2} \end{aligned}$$

where the angular pairs (θ_1,ϕ_1) and (θ_2,ϕ_2) are the orientation angles for the vectors $\overline{R}(x_{n_1}) - \overline{r}_o$ and $\overline{R}(x_{n_2}) - \overline{r}_o$, respectively. Eq.(5.19) reduces to FCF imaging with $n_2 = n_1$ and to ACF imaging by letting $m_2 = m_1$. FCF imaging is thus given by

$$C_\Gamma^{(FCF)}(\overline{r}_o) = \sum_{n=1}^{N_x}\sum_{m_1=1}^{N_k}\sum_{m_2=1,k_d>k_o/16}^{N_k} E(k_{m_1},x_n)e^{-2ik_{m_1}|\overline{R}(x_n)-\overline{r}_o|} \quad (5.21)$$
$$E^*(k_{m_2},x_n)e^{2i(k_{m_2})|\overline{R}(x_n)-\overline{r}_o|}$$

ACF imaging is thus written as

$$C_\Gamma(\overline{r}_o) = \sum_{m=1}^{N_k}\sum_{n_1=1}^{N_x}\sum_{n_2=1,k_d>k_o/16}^{N_x} E(k_m,x_{n_1})e^{-2ik_m|\overline{R}(x_{n_1})-\overline{r}_o|} \quad (5.22)$$
$$E^*(k_m,x_{n_2})e^{2i(k_m)|\overline{R}(x_{n_2})-\overline{r}_o|}$$

For the numerical simulations of linear SAR imaging, 80,000 small particles are used to simulate clutter. They are randomly distributed in a layer region of $40\lambda_o$X$40\lambda_o$X $0.5\lambda_o$. Four targets are placed at positions of $(10,14,0)\lambda_o$, $(10,28,0)\lambda_o$, $(30,10,0)\lambda_o$, and $(30,32,0)\lambda_o$. The dielectric constant of targets and scatterers are equal to (3.23+i0.36). The back-scattering amplitudes for targets and particles are calculated based on Mie

scattering. The receiver moves from $x_s = -d/2$ to $x_s = d/2$ with an increment of $d/100$. The horizontal position is $y_s = d$ with $d = 1732\lambda_o$ and the height is $z_s = H = 1000\lambda_o$. The received signal is calculated over a frequency band of $0.5f_o$ to $1.5f_o$ with an increment of $0.01f_o$ and 100 azimuthal positions with equal space.

The simulated data is processed by field imaging and FACF imaging. The results are normalized by the maximum value which is the value at the target. The normalized results are shown in Figures 5, and 6. Figure 5 shows field imaging. The four targets are obscured by the back ground clutter. Figure 6 shows the results of FACF imaging with the memory effect avoided, which has a lower clutter level than that in Figure 5. The results of FCF imaging and ACF imaging are also shown in Figures 7 and 8. We see that ACF imaging does not have good range resolution and FCF does not have good cross-range resolution. Therefore, FACF imaging is needed for linear SAR.

To compare results quantitatively, we define the visibility of targets in clutter as a ratio of target signal and the average signal strength for the whole region. The visibility is calculated by

$$v = \frac{Max(I(\bar{r}_o))}{\int d\bar{r}_o I(\bar{r}_o) / \int d\bar{r}_o} \tag{5.23}$$

where $I(\bar{r}_o)$ is the intensity. $I(\bar{r}_o)$ is $|C_\Gamma(\bar{r}_o)|$ for correlation imaging and is $|C_F(\bar{r}_o)|^2$ for field imaging. In Eq. (5.23), the numerator corresponds to targets, and the denominator is the average of background clutter.

The visibility of targets in clutter for images shown in Figures 5.2, 5.3, 5.5 and 5.6 are calculated and tabulated in table 5.1. We find that correlation imaging gives better visibility than the conventional field imaging. The image in Figure 5.3 obtained by ACF imaging has a visibility of 2 times that by field imaging shown in Figure 5.2. The visibility of the image in Figure 5.6 is 1.3 times that in Figure 5.5.

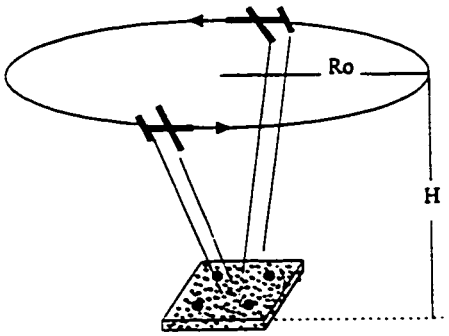
5.6 *Conclusions and Discussions*

From the formulation and numerical results, we found that correlation imaging gives better results for the detection of targets embedded in clutter than field imaging when the size of the randomly distributed small particles are smaller than a wavelength. For the same received signals, the image obtained by correlation imaging has finer resolution and larger signal-clutter ratio than that by field imaging. This is because the clutter effect is minimized in correlation imaging by avoiding the memory effect of random media scattering. Avoiding the memory effect in the spectral domain is equivalent to space-domain gating which can be used to suppress clutter. Also, ACF imaging in circular SAR has less frequency dependence than the conventional imaging. The decrease range resolution due to the frequency dependence is compensated for by the cross-range resolution.

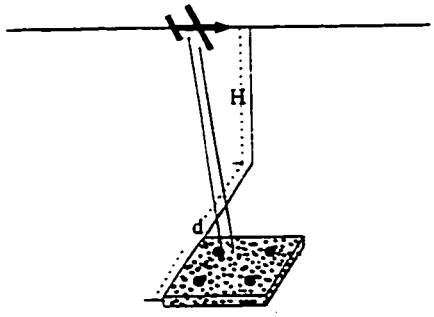
In this chapter, we have discussed the general formulation of correlation imaging. The ACF imaging is illustrated for circular SAR, and FACF imaging is used in linear SAR. ACF imaging in a circular configuration overcomes the spreading due to frequency dependence (decrease of range resolution) by the compensation of angular correlation (giving fine resolution in cross-range). In linear SAR, however, the range resolution and cross-range resolution are almost independent. FACF is needed to obtain a fine resolution image. The clutter effect can be substantially reduced for both circular and linear configurations by avoiding the memory effect of random media scattering. In a similar way, a 3-D image can be obtained with focusing on different layers. The frequency dependence can be compensated by using a weighting function. The CPU time for image processing can be reduced by using the far-field approximation for the phase and the summation then computed by FFT and the product of the images obtained using different sub-portion of the data. The propagation effects of random media on image can be reduced.

TABLE 5.1 Comparison of Visibility of Target in Clutter

Imaging	Circular		Linear	
Method	Field	ACF	Field	FACF
Visibility	16.5	33.8	15.2	19.2



(a)



(b)

Figure 5.1: Configurations of SAR imaging for targets embedded in small scatterers. (a) Circular SAR, (b) linear SAR

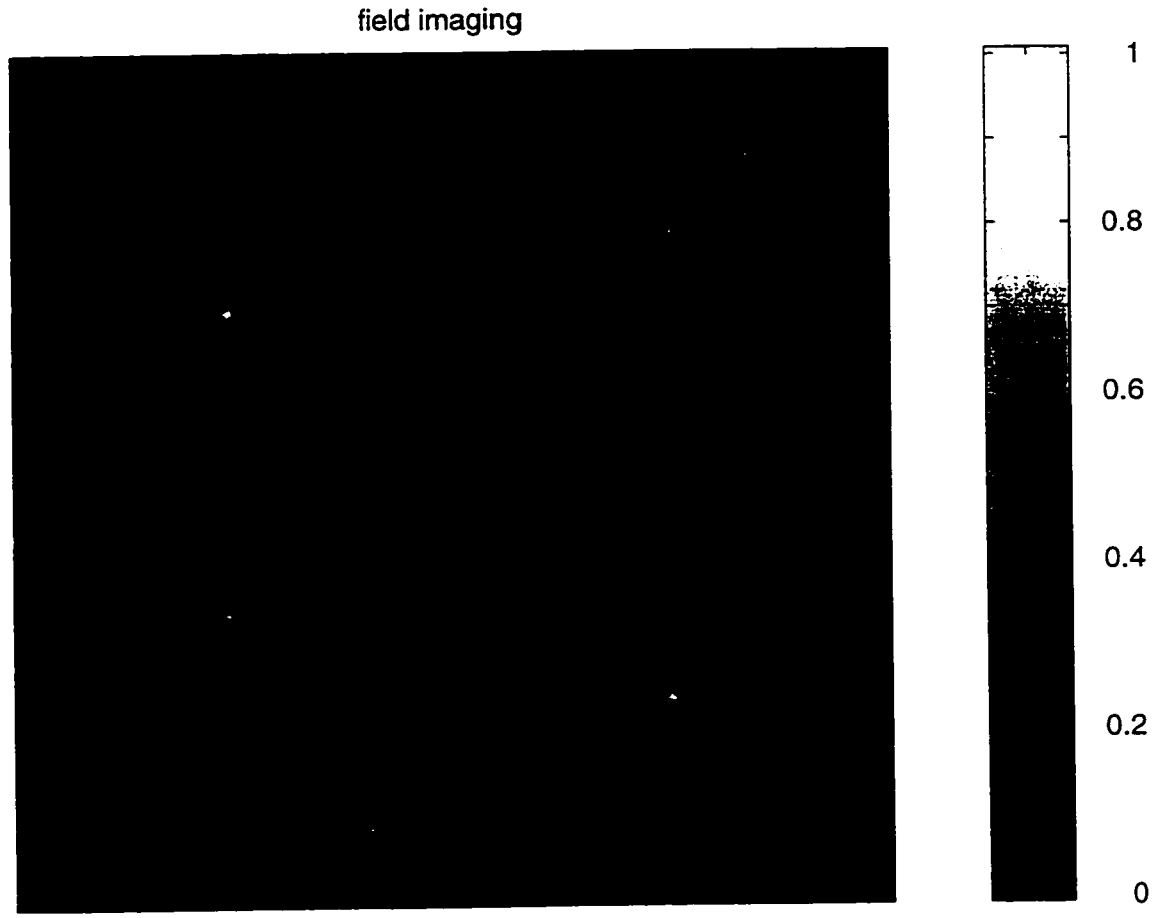


Figure 5.2: Simulated image of targets embedded in clutter by field imaging for circular SAR

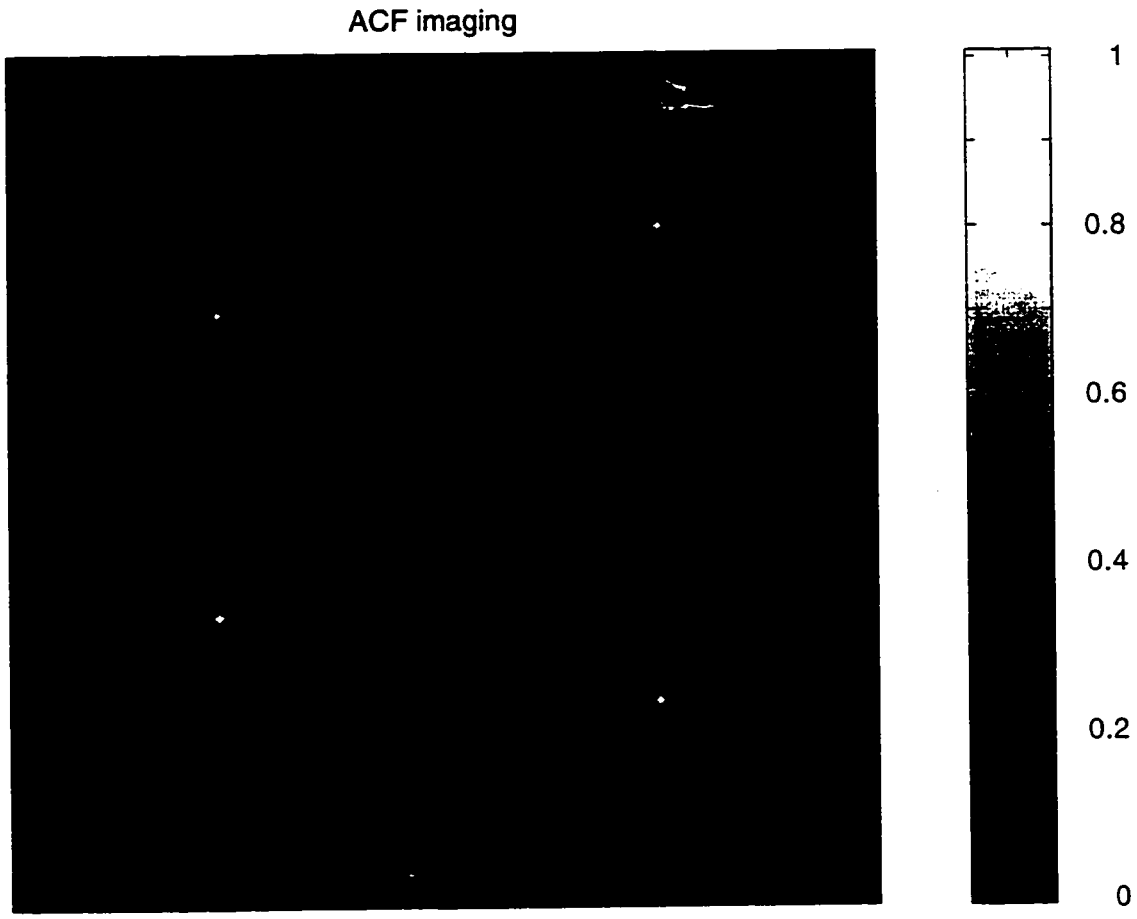


Figure 5.3: Simulated image of targets embedded in clutter by ACF imaging for circular SAR.

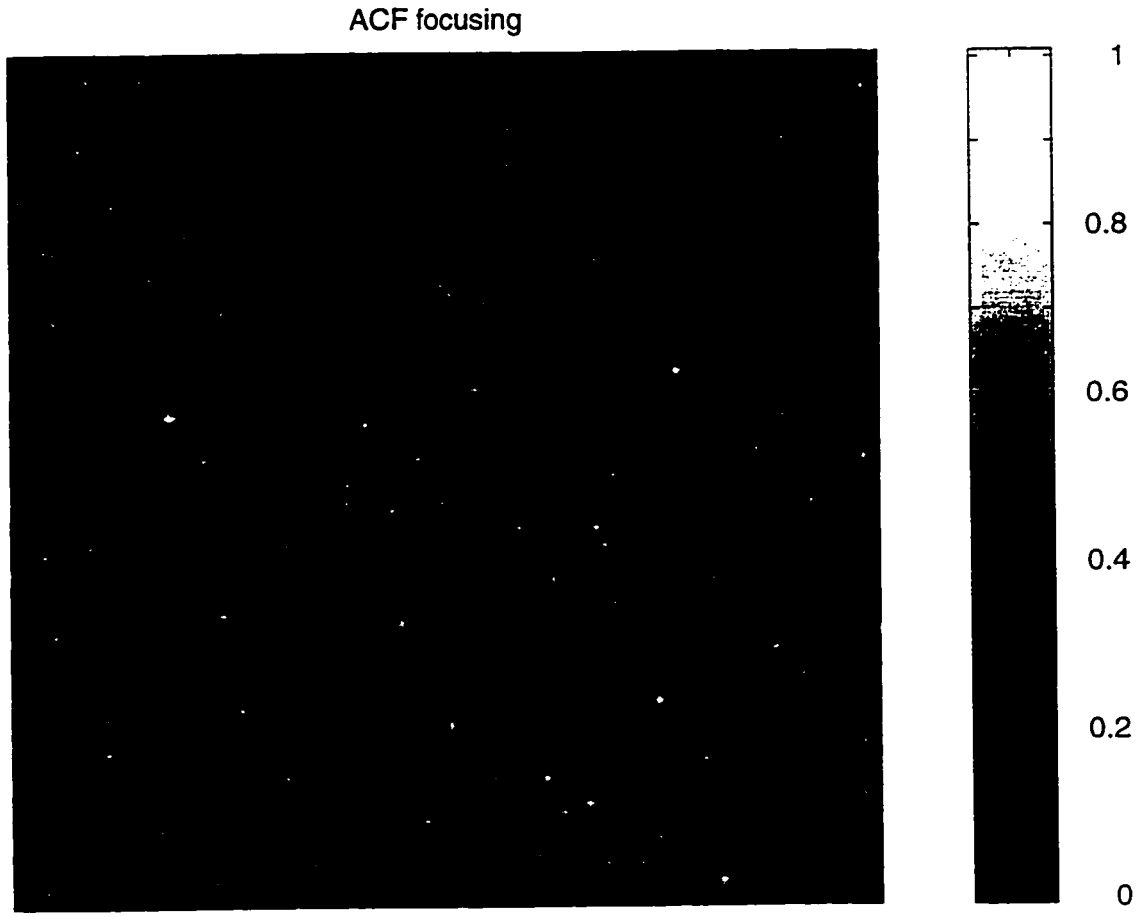


Figure 5.4: Simulated image of targets embedded in clutter by ACF focusing for circular SAR.

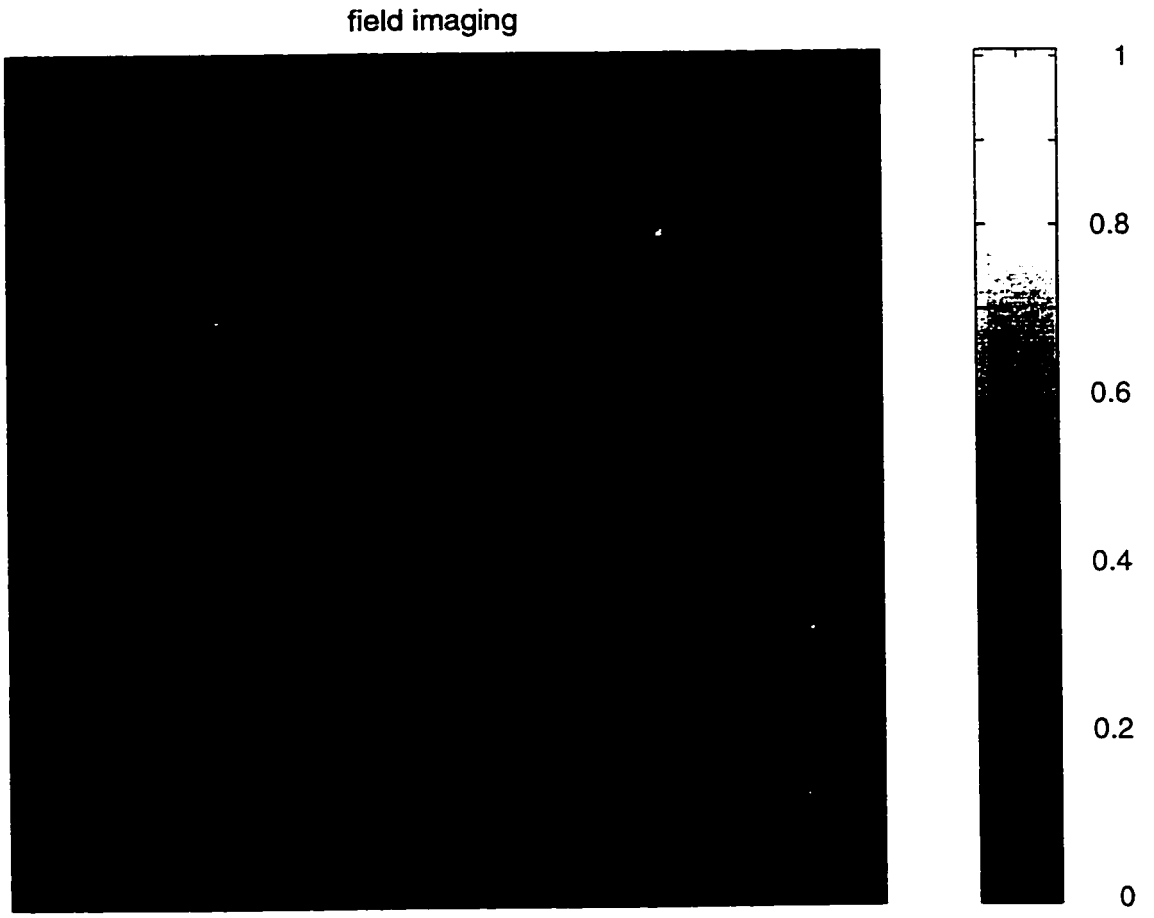


Figure 5.5: Simulated image of targets embedded in clutter by field imaging for linear SAR.

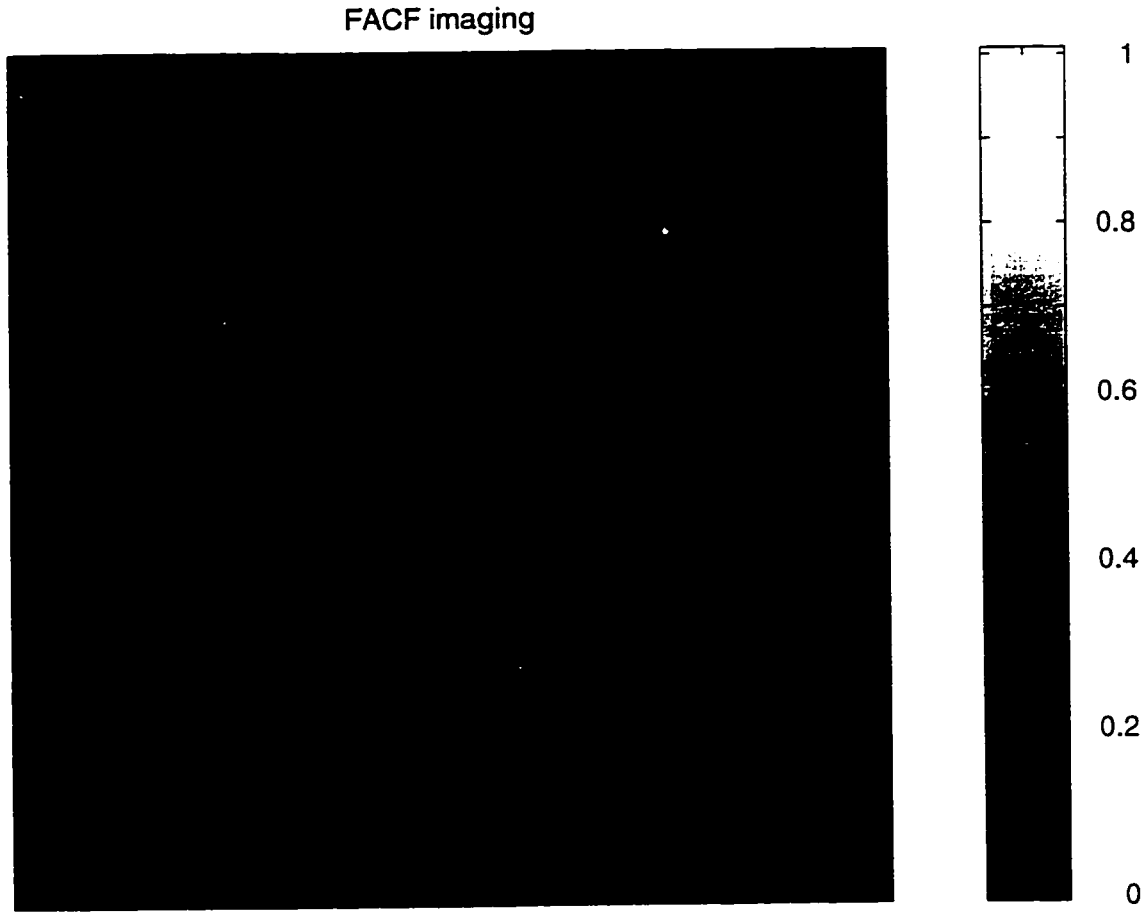


Figure 5.6: Simulated image of targets embedded in clutter by FACF imaging for linear SAR.

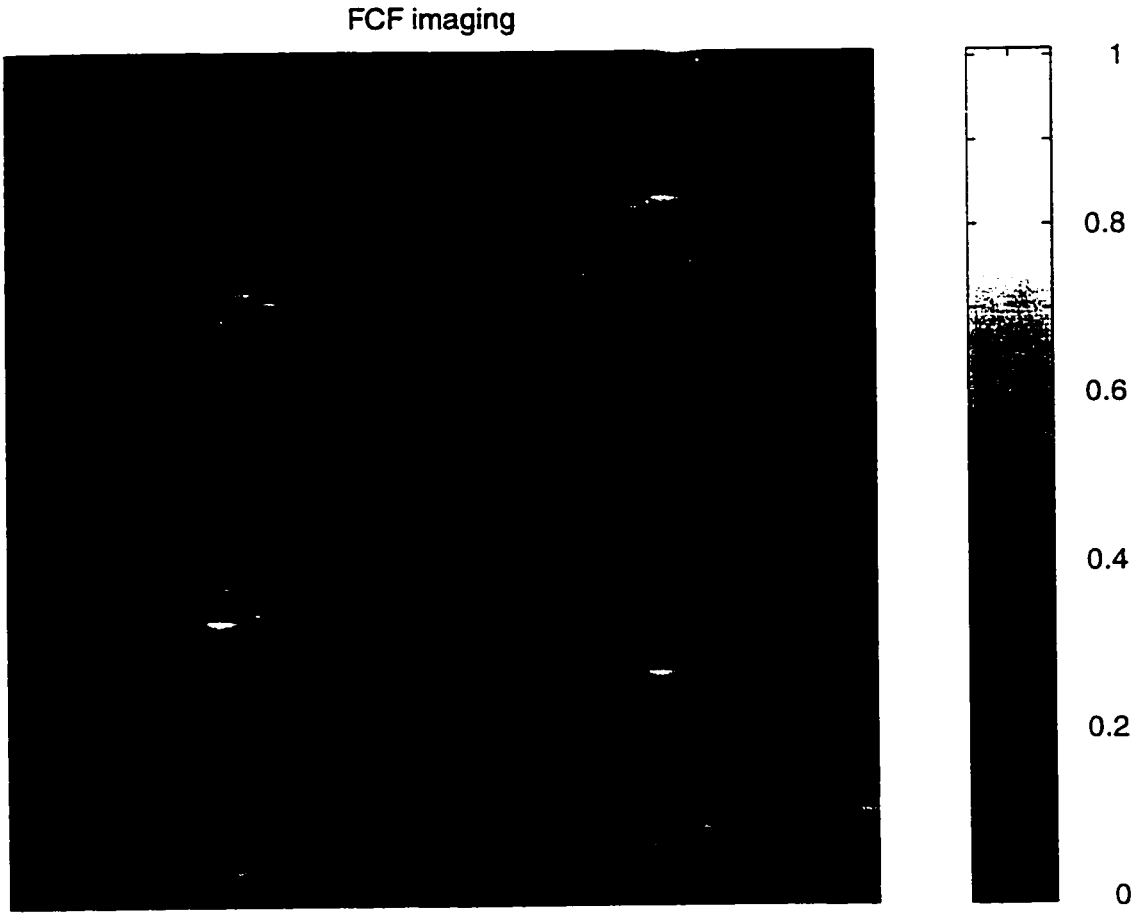


Figure 5.7: Simulated image of targets embedded in clutter by FCF imaging for linear SAR.

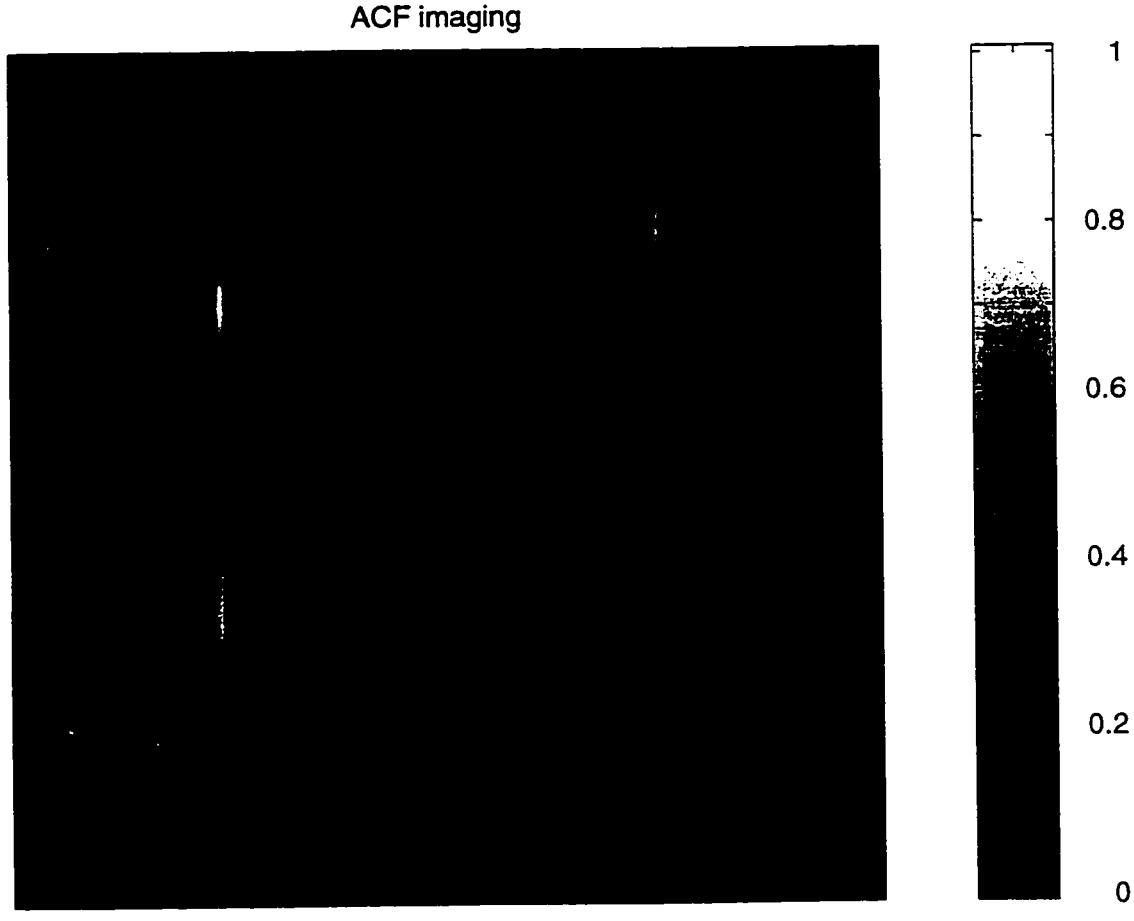


Figure 5.8: Simulated image of targets embedded in clutter by ACF imaging for linear SAR.

Chapter 6

SUMMARY AND FUTURE RESEARCH

6.1 Summary

In this dissertation, we have studied the detection and imaging of targets embedded in clutter environments based on the calculation of angular correlation function. ACF of random media scattering was first studied. We have derived general expressions, and studied ACF for volume scattering, rough surface scattering, and averaging techniques. It has been shown that the memory effect is a result of the statistically tranlational invariance of random scattering. Random rough surface scattering exhibits a memory line because of the horizontally translational invariance. Volume scattering exhibits two memory dots corresponding to self-correlation and reciprocity path. At the memory dots, both horizontal and vertical phase-matching conditions are satisfied. For the purpose of target detection, the memory effect is suggested to be avoided to suppress the clutter. Averaging technigues are discussed and illustrated for the target detection. In this case, realization averaging is not applicable, frequency averaging and angular averaging are needed.

Numerical studies of target detection consist of two steps: (1) calculation of the scattered wave fields, and (2) data processing. To obtain wave fields scattered by a target embedded in clutter, we formulated problems exactly by using surface and volume integral equations. The equations are then solved numercally with the use of the method of moment and the sparse-matrix canonical-grid method, which give exact solution of Maxwell's equations. Coherent addition approximation has also been used to generate data for SAR and ACF processing. To process the simulated

data, we calculated the ACF and did ACF imaging. It has been found that ACF is more effective in suppressing the clutter due to the random media and rough surface scattering than RCS, and ACF imaging gives larger signal-clutter ratios and finer resolution than the conventional SAR imaging. That is because, in ACF processing, the memory effect of random media scattering has been avoided and the frequence dependence of target scattering is overcome in ACF imaging.

6.2 Future Research

There is no need to repeat the importance of studies for the detection and imaging of targets embedded in clutter. They are practical and fundamental problems. In this dissertion, the angular correlation function of scattered wave fields have been studied for target detection and ACF imaging has been illustrated for imaging. However, some other research still need further studies. They are as follows:

6.2.1 High-order Moments

In this dissertation, we have been concentrating studies of the second-order moment: ACF and its memory effect. As shown in chapter 2, high-order moments also have memory effect as long as the phase-matching conditions are satisfied. High-order moments give more detailed information about clutter and target, which need further studies.

6.2.2 Multiple scattering effect on imaging

As we know, conventional image processing methods are based on first-order scattering approximation(Born or distorted Born approximation). However, in practical problems such as targets embedded in clutter, multiple scattering effect can be important. Figure 6.1 shows an example of multiple scattering effect on imaging for periodically distributed cylinders, in which we calculate the scattered fields by using

coherent addition approximation for isotropic single scattering and exact single scattering, and exact solution of Maxwell's equations. We obtain very different images. When multiple scattering occurs, how to image the target region without solving the inverse problem is a very challenging problem and needs to be further studied.

6.2.3 Bistatic imaging at a single frequency

Conventional imaging is to process the backscattered wave fields measured at many frequencies and angles by moving a mono-static wide frequency band radar. However, the back-scattering amplitude of a target is frequency dependent, especially for the target with a comparable size to wavelength. This frequency dependence can smear target image in conventional imaging. This effect can be avoided if we use bistatic scattering wave fields and image target at a single frequency. Figure 6.2 shows images of three spheres at different frequency using bistatic imaging. The radii of these spheres are chosen to be $a_1 = 0.175\lambda_o$, $a_2 = 0.235\lambda_o$ and $a_3 = 0.35\lambda_o$. We see different sphere is lit up for the image at different frequence. The sphere is brighter than others when it satisfies the resonant condition of $ka = 1.17$. This property can be used for target detection by choosing the frequency for imaging such that the resonance condition satisfied, which needs further studies.

The bistatic imaging has also been applied to 2-D scattering problems as shown in figure 6.3 and 6.4. The configuration for figure 6.3 is similar to that shown in figure 6.1a. The only difference is that the cylinder on 2nd raw and 6th column is the target with radius of 5cm and the radia of other cylinders are 1.67cm. It has been found that conventional imaging does not show the target, but bistatic imaging can show the target. The bistatic imaging at low frequency shows the target clearly and that at high frequency high light the cylinder right above the target due to mutual interaction.

Figure 6.4 shows the bistatic image for a target cylinder embedded in randomly distributed small cylinders under a rough surface (Figure 3.1). The rough surface is

shown at both frequencies of 1.5GHz and 2.0GHz. The target can be seen in 1.5GHz image, but not in the 2.0GHz image.

6.2.4 Applications of non-Gaussian random field

Currently, most theories and results of wave scattering and propagation in random media are based on the assumption of Gaussian random field for the media so that the wave statistics can be easily calculated. In the studies of rough surface scattering, the rough surfaces are usually assumed to be Gaussian distribution so that the surface profiles can be easily generated based on the spectrum method. However, nature media and surface may not obey the Gaussian distribution [51]. Non-Gaussian random rough surfaces scatter wave differently from Gaussian rough surfaces with the same rms height and correlation length. Figure 6.5 shows the comparison of bistatic scattering coefficients of the random rough surfaces with Gaussian, Rayleigh, and exponential distributions. It is interesting to conduct further studies of wave scattering and propagation in non-Gaussian random media and apply them to remote sensing of nature environments.

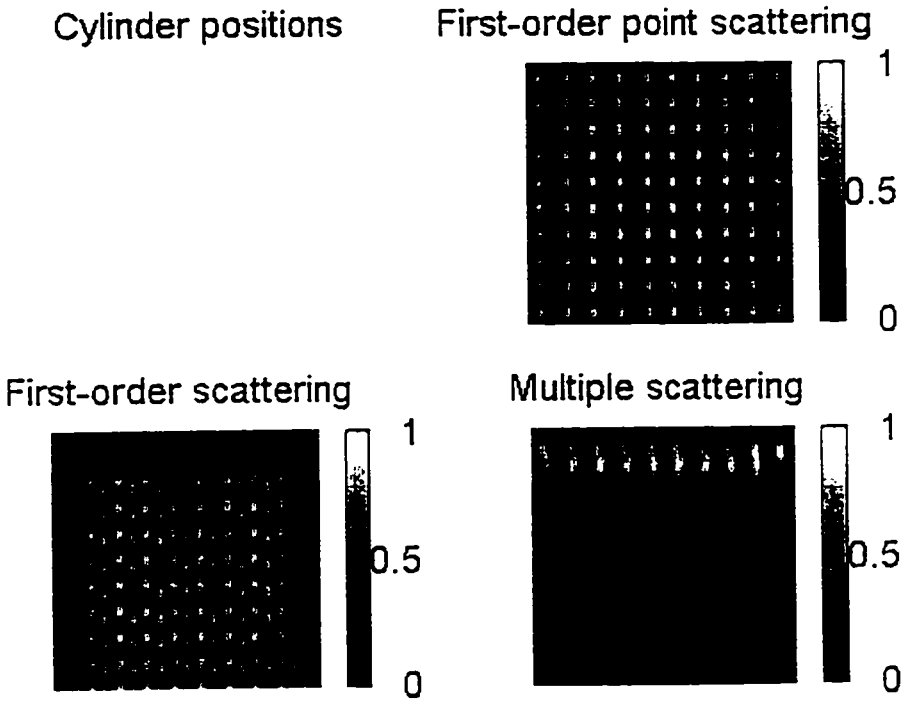


Figure 6.1: Multiple scattering effect on imaging for volume scattering

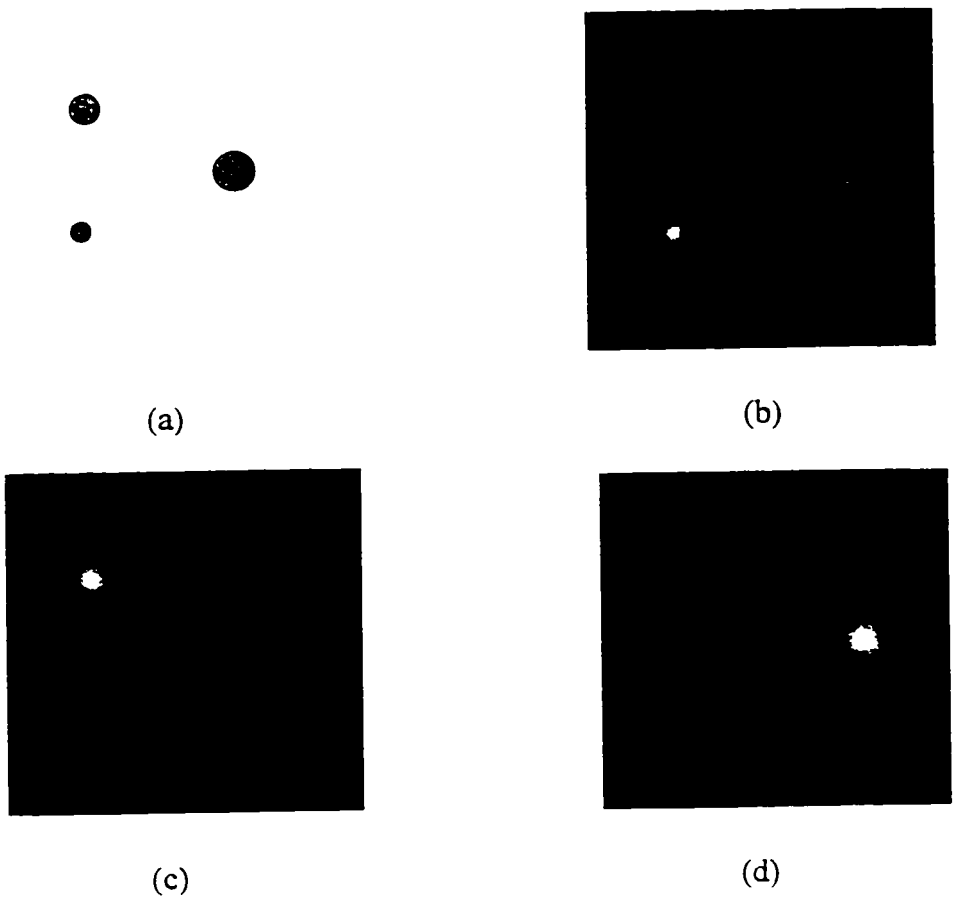


Figure 6.2: Bistatic imaging of three spheres at different frequency. (a): Three spheres with radius of $a_1 = 0.175\lambda_o$, $a_2 = 0.235\lambda_o$ and $a_3 = 0.35\lambda_o$. (b): Imaging at frequency of $f_1 = f_o$, $k_o a_1 = 1.17$. (c): Imaging at frequency of $f_2 = 0.75 f_o$, $k_2 a_2 = 1.17$. (d): Imaging at frequency of $f_3 = 0.5 f_o$, $k_3 a_3 = 1.17$.

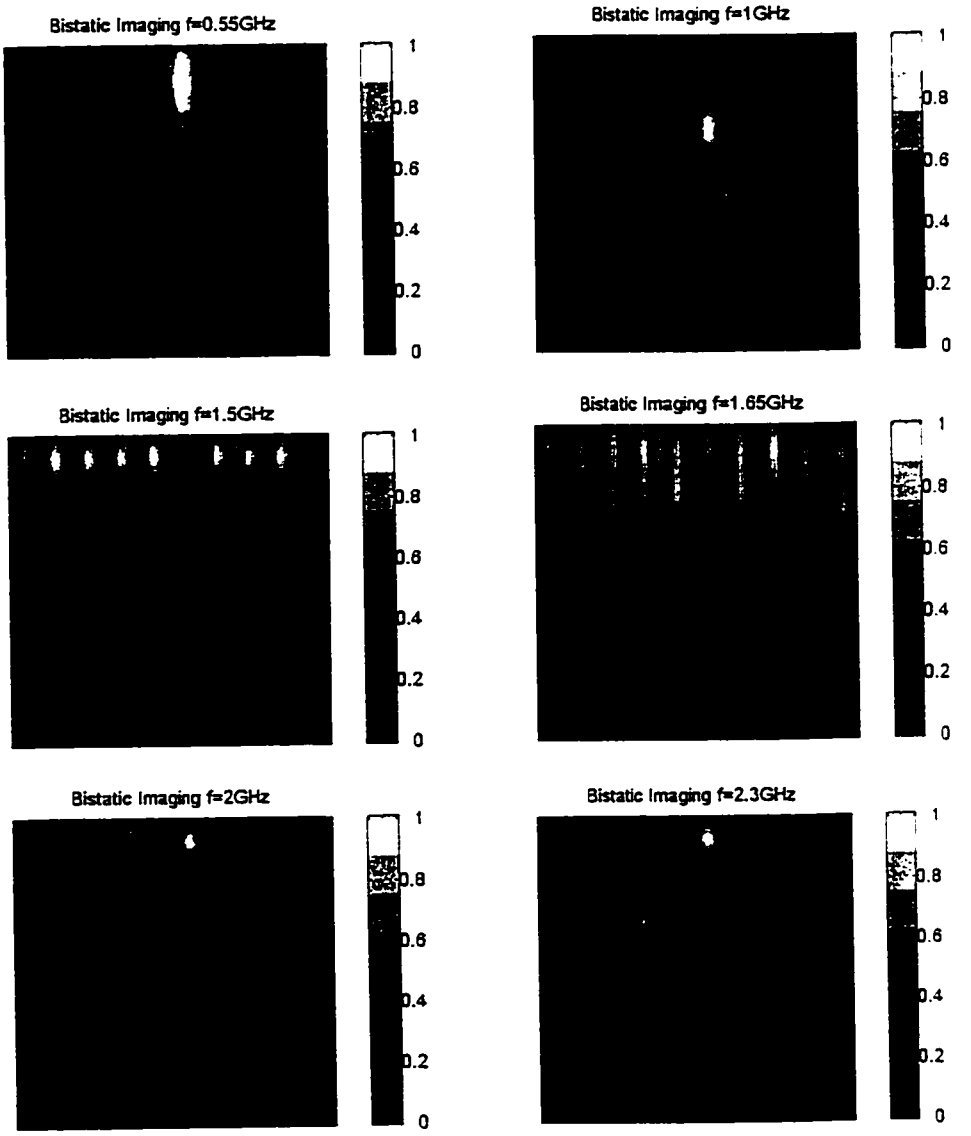


Figure 6.3: Bistatic imaging for a target cylinder embedded in small cylinders. Parameters are $a_t = 5cm$, $a_c = 1.67cm$, and $L_x = L_z = 200cm$.

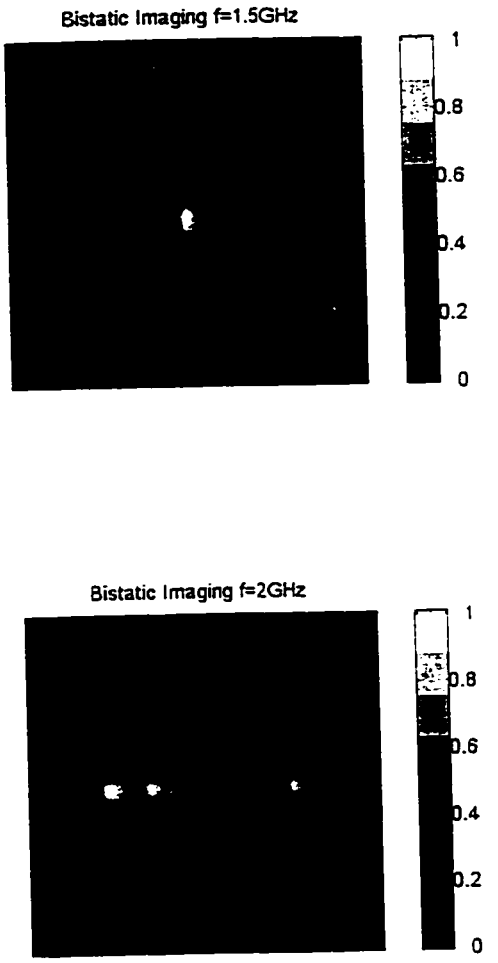


Figure 6.4: Bistatic imaging for a target cylinder embedded in small cylinders under a rough surface. Parameters are $a_t = 10cm$, $a_c = 0.5cm$, $N = 300$, and $L_x = 800cm$.

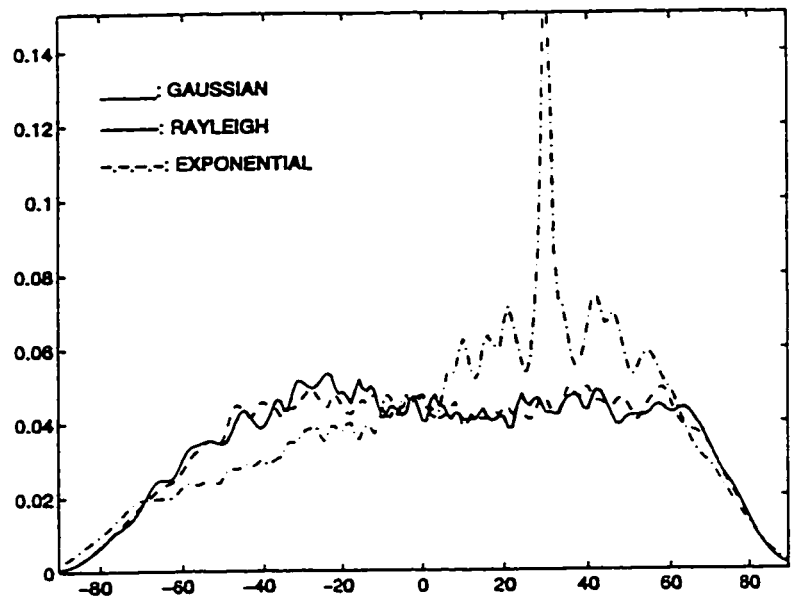


Figure 6.5: Comparison of bistatic scattering coefficients of the random rough surfaces with Gaussian (solid line), Rayleigh (dash line), and exponential (dotdash line) distributions. Results are obtained by averaging over 100 realizations. Parameters are $h = 0.5\lambda$, $l = 1.0\lambda$, $\theta_i = 30^o$, and $\epsilon_r = 3.7 + 0.13i$.

BIBLIOGRAPHY

[1] S. Axelsson. Frequency and azimuth variation and their influence upon low-frequency sar imaging. *IEEE Trans. on Geosci. Remote Sensing*, 33:1258–1265, 1995.

[2] C.H. Chan and R. Mittra. Some recent developments in iterative techniques for solving electromagnetic boundary value problems. *Radio Science*, 22:929–934, 1987.

[3] Y. H. Chen, W. C. Chew, and M. L. Oristaglio. Application of perfertly matched layers to the transient modeling of subsurface em problems. *Geophysics*, 62:1730–1736, 1997.

[4] E. Conte and M. Lops. Clutter-map cfar detection for range-spread targets in non-gaussian clutter. *IEEE Trans. on AES*, 33:432–454, 1997.

[5] P. G. Cottis, C. N. Vazouras, C. Kalamatianos, and J. D. Kanellopoulos. Scattering of tm waves from a cylindrical scatterer buried inside a two-layer lossy ground with sinusoidal surface. *Journal of Electromagnetic Waves and Applications*, 10:1005–1021, 1996.

[6] S. Feng, C. Kane, P. A. Lee, and A. D. Stone. Correlations and fluctuations of coherent wave transmission through disordered media. *Phys. Rev. Lett.*, 61:834–837, 1988.

[7] P. Gloersen and F.T.Barath. A scanning multichannel microwave radiometer for nimbus-7 and seasat-a. *IEEE J. Oceanic Eng., vol. OE-2*, pages 172–180, 1977.

[8] S. Haykin. *Neural Networks: A Comprehensive Foundation.* MacMillan College Publishing Co. Inc., New York, 1994.

[9] A. Ishimaru. *Wave Propagation and Scattering in Random Media.* volume 1 and 2. Academic Press, New York, 1978.

[10] A. Ishimaru, C. Le, Y. Kuga, L. Ailes-Sengers, and T.K. Chan. Polarimetric scattering theory for high slope rough surfaces. *Progress in Electromagnetics Research*, 14:123–192, 1996.

[11] K. P. Kirdyashev, A. A. Chukhlantsev, and A. M. Shutko. Microwave radiation of the earth's surface in the presence of a vegetation cover. *Radio Eng. Electron Phys.*, 24:37–44, 1979.

[12] Y. Kuga and A. Ishimaru. Retroreflectance from a dense distribution of spherical particles. *J. Opt. Soc. Am. A*, 1:831–835, 1996.

[13] Y. Kuga, C. Le, A. Ishimaru, and L. Ailes-Sengers. Analytical experimental and numerical studies of angular memory signatures of waves scattered from one-dimensional rough surfaces. *IEEE Trans. Geosci. Rem. Sens.*, 34(6):1300–1307, 1996.

[14] Y. Kuga, G. Zhang, J. Yea, and A. Ishimaru. Analytical and numerical studies of angular correlation function of waves scattered from randomly distributed cylinders. *to appear in Waves in Random Media*, 6, 1998.

[15] M. Lax. Multiple scattering of waves. *Rev. Mod. Phys.*, 23:287–310, 1951.

[16] A. Madrazo and M. Nieto-Vesperinas. Scattering of light and other electromagnetic waves from a body buried beneath a highly rough random surface. *J. Opt. Soc. Am. A*, 14:1859–1866, 1997.

[17] C. Mandt. *Multiple Scattering in Random Media: Backscattering Enhancement in a Sparse Distribution of Large Scatterers and Monte Carlo Simulations of Extinction Rate in Dense Media.* Ph.D. Dissertation, University of Washington, April, 1992.

[18] A. A. Maradudin, M. Nieto-Vesperinas, and E. Thorsos. Enhanced backscattering of light from randomly rough surfaces and related phnonomena i: One-dimensional surfaces and angular correlation function of scattered field. *Comments on Condensed Matter Physics*, 17(1):13–37, 1994.

[19] T.R. Michel and K.A. O'Donnell. Angular correlation functions of amplitudes scattered from a one-dimensional, perfectly conduction rough surface. *J. OPt. Soc. Am.*, 9(8):1374–1384, 1992.

[20] T. Mo, T. J. Schmugge, and J. R. Wang. Calculations of the brightness temperature of rough soil surfaces: Bare field. *IEEE Trans. Geosci. Remote Sensing*, 25:47–54, 1987.

[21] T. G. Moore. A new algorithm for the formation of isar images. *IEEE Trans. on Aerospace and Electronic Systems*, 32(2):714–721, 1996.

[22] M. Nieto-Vesperinas and J. A. Sanchez-Gil. Intensity angular correlations of light multiply scattered from random rough surfaces. *J. Opt. Soc. Am. A*, 10:150–157, 1993.

[23] E. G. Njoku, E. J. Christensen, and R. E. Cofield. The seasat scanning multichannel microwave radiometer (smmr): Antenna pattern corrections - development and implementation. *IEEE J. Oceanic Eng.,vol. OE-5*, pages 125–137, 1980.

[24] E. G. Njoku, J. M. Stacey, and F. T. Barath. The seasat smmr: Instrument

description and performance. *IEE J. Oceanic Eng., vol OE-5*, pages 100–115, 1980.

[25] K. O'Neill, R. Lussky, and K. D. Paulsen. Scattering from a metallic reflector embedded in a lossy dielectric beneath random rough surface. *IEEE Trans. on Geoscience and Remote Sensing*, 34:367–376, 1996.

[26] K. Pak. *Study of Large-Scale Random Rough Surface Scattering problems based on Monte Carlo Simulations with Efficient Computational Integral Equation Methods.* Ph.D. Dissertation, University of Washington, July, 1996.

[27] K. Pak, L. Tsang, C.H. Chan, and J. Johnson. Backscattering enhancement of vector electromagnetic waves from two-dimensional perfectly conducting random rough surface based on monte carlo simulations. *Journal of Optical Society of America*, 12:851–859, 1995.

[28] K. Pak, L. Tsang, and J. Johnson. Numerical simulations and backscattering enhancement of scattered waves from two-dimensional dielectric random rough surface with sparse-matrix canonical grid method. *Journal of Optical Society of America*, 14:1515–1529, 1997.

[29] L. Peters, J. Daniels, and J. Young. Ground penetrating radar as a subsurface environmental sensing tool. *Proceedings of the IEEE*, 82:1802–1822, 1994.

[30] P. Phu, A. Ishimaru, and Y. Kuga. Controlled millimeter wave experiments and numerical simulations on the enhanced backscattering from one-dimensional very rough surfaces. *Radio Science*, 28:533–548, 1993.

[31] P. Prusinkiewicz and A. Lindenmayer. *The Algorithmic Beauty of Plants.* Springer-Verlag, New York, 1990.

[32] E.J.M. Rignot, C.L. Williams, J. Way, and L.A. Viereck. Mapping of forest types in alaskan boreal forests using sar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 32(5), 1994.

[33] M. Soumekh. Reconnaissance with slant plant circular sar imaging. *IEEE Trans. Image Processing*, 35:1252–1265. 1996.

[34] V. I. Tatarskii. *The Effects of Turbulent Atmosphere on Wave Propagation.* National Tech. Information Service. Springfield. 1971.

[35] E. I. Thorsos. The validity of the kirchoff approximation for rough surface scattering using a gaussian roughness spectrum. *J. Acoust. Soc. Amer.*. 83:78–92, 1988.

[36] S. Tjuatja. A. K. Fung, P. Zhou S. Wu, and Z. Li. Remote sensing of buried objects: An analysis using fd-td simulation. *IGARSS'97*, 1997.

[37] P. Tran and A. A. Maradudin. The scattering of electromagnetic waves from 2d metallic surface. *Optics Communications*. 110:269–273. 1994.

[38] L. Tsang. Passive remote sensing of dense nontenuous media. *J. Electromagnetic Waves and Applications*, 1(2):159–173. 1987.

[39] L. Tsang, C.H. Chan, and K. Pak. Backscattering enhancement of a two-dimensional random rough surface (three-dimensional scattering) based on monte carlo simulations. *Journal of Optical Society of America*, 11:711–715, 1994.

[40] L. Tsang, K. H. Ding, G. Zhang, C. Hsu, and J. A. Kong. Backscattering enhancement and clusering effects of randomly distributed dielectric cylinders overlying a dielectric half space based on monte carlo simulations. *IEEE Transactions on Antenna and Propagation*, 43(5):488–499, 1994.

[41] L. Tsang and A. Ishimaru. Backscattering enhancement of random discrete scatterers. *Journal of Optical Society of America*, 1:836–839, 1984.

[42] L. Tsang, J. A. Kong, and R. T. Shin. *Theory of Microwave Remote Sensing*. Wiley Interscience, New York, 1985.

[43] L. Tsang, G. Zhang, and K. Pak. Detection of a buried object under a single random rough surface with angular correlation function in em wave scattering. *Microwave and Optical Technology Letters*, 11(6):300–304, 1996.

[44] G. Videen. Light scattering from a sphere behind a plane. *J. Opt. Soc. Am. A*, 14(1):110–117, 1993.

[45] S. Vitebskiy, L. Carin, M. A. Ressler, and F. H. Le. Ultra-wideband short-pulse ground-penetrating radar: simulation and measurement. *IEEE Trans. on Geosci. Remote Sensing*, 35:763–772, 1997.

[46] R. L. Wagner, Jiming Song, and W. C. Chew. Monte carlo simulation of electromagnetic scattering from two-dimensional random rough surfaces. *IEEE Transactions on Antenna and Propagation*, 45(2):235–245, 1997.

[47] J. R. Wang and T. J. Schmugge. An empirical model for the complex dielectric perimittivity of soils as a function of water content. *IEEE Trans. Geosci. Remote Sensing*, 8:288–295, 1980.

[48] Y. Yamaguchi and T. Moriyama. Polarimetric detection of objects buried in snowpack by a synthetic aperture fm-cm radar. *IEEE Trans. on Geosci. Remote Sensing*, 34:45–51, 1996.

[49] X. Yang, D. A. Gregory, and P. S. Erbach. Clutter reduction and target detection enhancement using wave let transform techniques. *Proc. of SPIE- The International Society for Optical Engineering*, 2490:125–139, 1995.

[50] J. Yea, G. Zhang, Y. Kuga, and A. Ishimaru. Millimeter wave experiments on angular correlation function of waves scattered from randomly distributed cylinders. *to appear in Radio Science*, 33, 1998.

[51] G. Zhang and L. Tsang. Angular correlation function of wave scattering by a random rough surface and random discrete scatterers and its application in the detection of a buried object. *Wave in Random Media*, 7:467–478, 1997.

[52] G. Zhang and L. Tsang. Application of angular correlation function of clutter scattering in target detection. *to appear in Special Issue on IGARSS'97 of IEEE Trans. on Geoscience and Remote Sensing*, 36, 1998.

[53] G. Zhang, L. Tsang, and Z. Chen. Collective scattering effects of trees generated by stochastic lindenmayer systems. *Microwave and Optical Technology Letters*, 11(2):107–111, 1996.

[54] G. Zhang, L. Tsang, and Y. Kuga. Angular correlation function of wave scattering by a buried object embedded in random discrete scatterers under a rough surface. *Microwave and Optical Technology Letters*, 14(3):144–151, 1997.

[55] G. Zhang, L. Tsang, and Y. Kuga. Studies of the angular correlation function of scattering by random rough surfaces with and without a buried object. *IEEE Trans. on Geoscience and Remote Sensing*, 35(2):444–453, 1997.

[56] G. Zhang, L. Tsang, and Y. Kuga. Numerical studies of the detection of targets in clutter by using angular correlation function and angular correlation imaging. *Microwave and Optical Technology Letters*, 8:467–478, 1998.

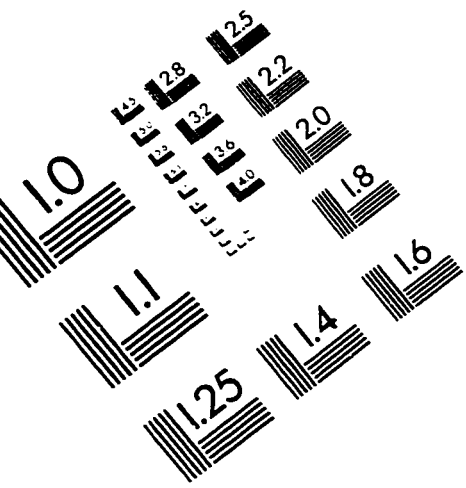
[57] G. Zhang, L. Tsang, and K. Pak. Angular correlation function and scattering coefficient of electromagnetic waves scattered by a buried object under a two-dimensional rough surface. *submitted to Journal of Optical Society of America (A)*, 15, 1998.

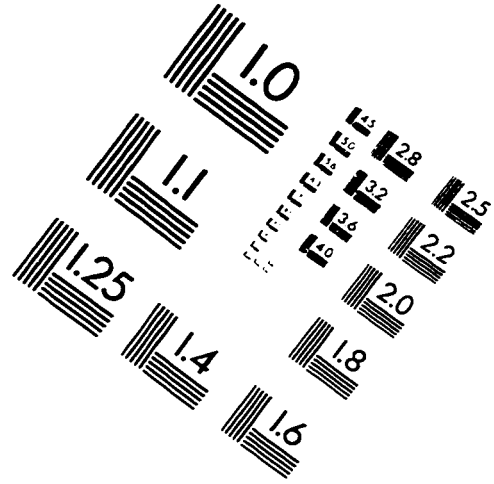
Vita

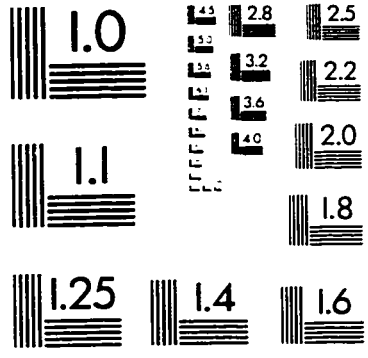
Guifu Zhang was born in China on July 16, 1962. He received a B. S. in physics from Anhui University in 1982 and a M. S. in radio physics from Wuhan University in 1985, respectively. In 1998, he received his Ph. D. in Electrical Engineering from University of Washington.

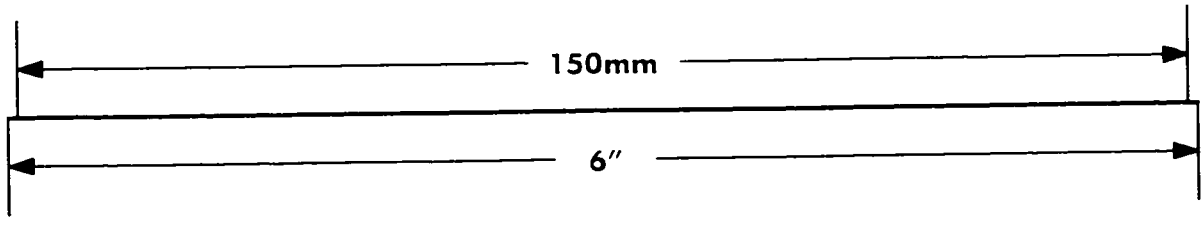
From 1985 to 1993. he had been teaching in the Space Physics Department at Wahan University as an assistant professor and associate professor. In 1989, he was a visiting scholar working with Dr. Tomohiro Oguchi at the Communication Research Laboratory, Japan, with the award of a followship of Matsmae International Foundation. From 1993 to 1998, he studied and worked in the Department of Electrical Engineering, University of Washington. where he first was an visiting scientist and later became a Ph. D. student. His research interests are wave propagation and scattering in random media and from rough surface as well as applications in environmental remote sensing and target detection.

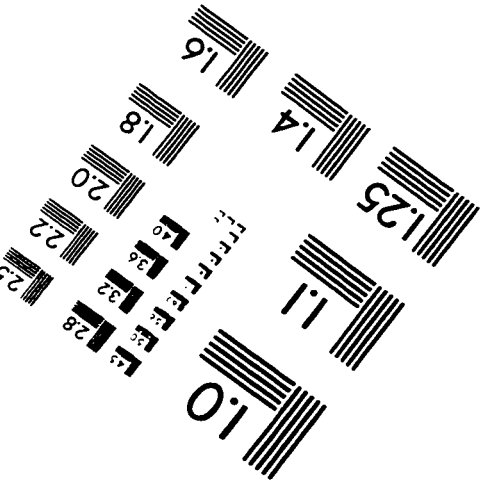
IMAGE EVALUATION
TEST TARGET (QA-3)

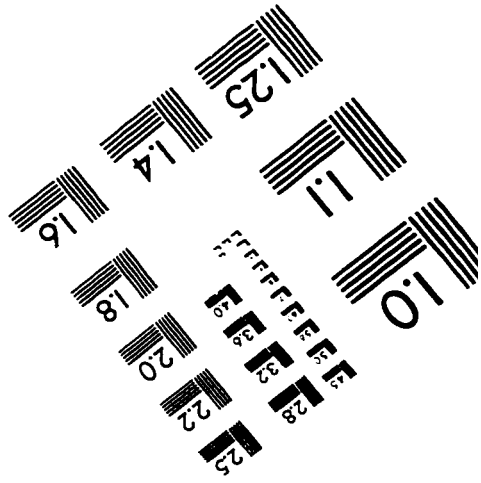












APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved

99

07905

UMI

MICROFILMED 1998

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

PHYSIOLOGIC DEVELOPMENT OF SPEECH MOTOR CONTROL: ARTICULATORY COORDINATION OF LIPS AND JAW

by

Jordan R. Green

A dissertation submitted in partial fulfillment
of the requirement for the degree of

Doctor of Philosophy

University of Washington

1998

Approved by _____
Chairperson of Supervisory Committee

*

Program Authorized to Offer Degree Department of Speech and Hearing Sciences

Date 7/14/98

UMI Number: 9907905

Copyright 1998 by
Green, Jordan R.

All rights reserved.

UMI Microform 9907905
Copyright 1998, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized
copying under Title 17, United States Code.

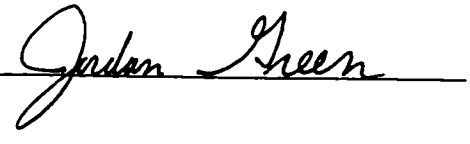
UMI
300 North Zeeb Road
Ann Arbor, MI 48103

© Copyright 1998

Jordan R. Green

Doctoral Dissertation

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date 7/15/98

University of Washington

ABSTRACT

Physiologic Development of Speech Motor Control: Articulatory coordination of lips and jaw

by Jordan Green

Chair of the Supervisory Committee: Associate Professor Christopher A. Moore, Ph.D.
Department of Speech and Hearing Sciences

The objective of this investigation was to describe development of lip and jaw coordination during speech. The potential influence of labiomandibular coordination on phonologic acquisition was also considered. A computer-based movement tracking system was used to transduce movement of the upper lip, lower lip, and jaw. Productions of syllables containing bilabial consonants were obtained from four age groups (i.e., one-, two-, six-years, and young adult). Two complementary analytic techniques were used to quantify interarticulator coordination, one reflecting spatial and temporal coupling of articulatory pairs and the other isolating each articulator's contribution to oral closure.

The coordinative organization of these articulatory gestures was shown to change dramatically during the first several years of life and to continue to undergo refinement past age six. At one year of age, jaw displacement contributed the most to oral closure. The contribution of the lower lip increased gradually with age, whereas the contribution of the upper lip was greater for two-year-olds than for any other group.

Spatial and temporal coupling of movement of the upper lip, lower lip, and jaw were shown to increase with maturation. A similar developmental trend was exhibited for each measure. Coupling of upper lip and lower lip movement was rigid in even the

youngest subject groups. In contrast, coupling of lip and jaw pairs was initially weak and gradually increased with age.

The present results can be interpreted as representing three primary phases in the development of lip and jaw coordination for speech: integration, differentiation, and refinement. Each of these developmental processes entails the existence of distinct coordinative constraints on early articulatory movement. It is suggested that these constraints will have predictable consequences for the sequence of phonologic development.

TABLE OF CONTENTS

LIST OF TABLES IV

LIST OF FIGURES V

INTRODUCTION 1

Physiologic Constraints on Speech Acquisition 2

The Problem of Early Speech Development 3

Processes in the Development of Motor Control 6

Differentiation 7

Integration 9

Sequence of Oromotor Development 10

Development of Motor Skills for Speech 12

Physiologic Interpretations 14

Orofacial Control for Speech 18

Development of Lip and Jaw Coordination for Speech 20

Purpose of Study and Statement of Problem 25

EXPERIMENTAL DESIGN AND METHODS 28

Subjects 28

Speech Samples 29

Data Collection and Recording Conditions 31

Digitization and Signal Conditioning 32

Quantitative Analyses of the Kinematic Traces 34

Articulatory Coupling and Synchrony 34

Articulatory Contribution to Oral Closure 37

Statistical Treatment 41

Reliability of Measurement 42

RESULTS 44

Data Corpus 44

Phoneme Effects 45

Qualitative Observations 46

Crosscorrelation Analysis 48

Spatial Coupling 48

Movement Synchrony 52

Contribution to Oral Closure 55

DISCUSSION 60

The Development of Articulatory Coordination 60

The Mature Pattern 61

One-Year-Olds 61

Two-Year-Olds 64

Six-Year-Olds 67

Mechanisms: Data, theory, and speculation 68

Developmental Sequences and Changes in Neural Substrates 69

Developmental Sequences and Changes in Biomechanics 72

Changes in Coordinative Organization 74

Sequential Motor Learning 74

Simplification Control Strategy 75

Negative Transfer of Learning 76

Physiologic Constraints and Phonologic Acquisition 77

Methodological Limitations 80

Clinical Implications 83

Future Research 83

Summary 85

REFERENCES 87

APPENDIX 103

Code Generated for Experiment 103

Programs for Crosscorrelation Analysis 103

Programs for Contribution for Oral Closure 113

LIST OF TABLES

Number *Page*

1. Stages of normal speech development within the first year of life 13

2. Test-retest reliability for measurement of coefficient and lag from crosscorrelation analysis..... 43

3. Mean coefficient values for articulator pairs across age groups..... 50

4. Results of pairwise comparisons of age by articulatory pair 51

5. Mean absolute lags (in milliseconds) for different articulator pairs as a function of age 52

6. Results of pairwise comparisons for articulator pair for lag values (ms) 54

7. Mean percentage contribution to oral closure as a function of age and articulator..... 56

8. Results of pairwise comparisons for articulator for relative contribution to oral closure (%)..... 58

LIST OF FIGURES

Number *Page*

1. The treated kinematic traces from upper lip (UL) , lower lip (LL), and jaw (J) produced by an adult subject saying "baba." 33

2. Pairwise crosscorrelation functions computed on the signals presented in Figure 1 36

3. Calculation of relative contribution to oral closure 40

4. Number of utterances per subject in each age group 45

5. Representative kinematic records from a subject in each age group based on a single trial 47

6. Average coefficients and standard deviations obtained from pairwise crosscorrelations for upper lip and lower lip, upper lip and jaw, and lower lip x jaw by age 49

7. Average absolute lag values and standard deviations obtained from pairwise crosscorrelations for UL x LL, UL x J, and LL x J for each age group 53

8. Relative contribution to oral closure for each articulator by age 57

ACKNOWLEDGMENTS

I wish to acknowledge the many individuals who have contributed to this study and have supported me throughout the doctoral degree. Special gratitude is expressed to Dr. Christopher Moore. His continued support and assistance made the doctoral program an exceptional experience. It has been an honor to work with an academic of his caliber. I would like to extend thanks to the members of the supervisory committee Dr. Fred Minifie, Dr. Carol Stoel-Gammon, Dr. Susan Herring, and Dr. Frank Smoll. Dr. Minifie provided continual support throughout my doctoral study. He was always eager to entertain new ideas and provide thoughtful feedback. Dr. Stoel-Gammon's and Dr. Herring's comments on experimental design and on earlier drafts were helpful. I would like to also thank Dr. Smoll for being an excellent GFR.

Additional thanks to Dr. Masahiko Higashikawa, Roger Steeve, Becca Culver and Nicole Engel not only for their assistance in data analysis and collection, but for their camaraderie.

Special acknowledgments to Dr. Margaret Rogers, Dr. Rebecca Leonard, and the deceased Bob Hall and Dr. Jack Palmer for their encouragement along the way.

This project was supported in part by a research grant from the National Institutes of Health/NIDCD (R01 DC00822) and a personnel training grant form the National Institutes of Health/NIDCD (T32 DC00033) at the Department of Speech and Hearing Sciences at the University of Washington, Seattle, WA.

DEDICATION

This dissertation is dedicated to my family. I am indebted to my wife Kimber, who has demonstrated much love, support, and patience, throughout my doctoral studies. My parents Nancy and Stewart Green have always encouraged me to follow my passions, to trust myself, and to only learn from my mistakes. Their friendship and love is gratefully appreciated.

INTRODUCTION

The transition from prelinguistic vocalizations to adult speech represents mastery of coordination of multiple speech subsystems. This remarkable behavioral accomplishment emerges in the context of rapid changes in musculoskeletal growth and neuromotor development (Enstrom, 1982; Kent, 1984; Kent & Murray, 1982; Lenneberg, 1967; Lieberman, 1977; 1984). Despite the sizable literature on the development of speech and language, there is a paucity of data describing the development of speech motor coordination and its influences on the sequence of phoneme acquisition. A better understanding of the sequence of motor acquisition for speech is essential before we can begin to comprehend delayed and abnormal speech development.

A complete description of speech acquisition will detail the mechanisms underlying phonologic development. Formulation of this description has been impeded by the logistically difficult task of collecting articulatory kinematic and electromyographic data from young children. Consequently, most of the existing literature on speech development has been based on phonetic transcription and quantitative acoustic analyses (e.g., Kent, 1976; Kent & Bauer, 1985; Nittrouer, 1993; Nittrouer, Studdert-Kennedy, & Neely, 1996; Stoel-Gammon, 1988; Stoel-Gammon & Otomo, 1986). The comments by Kent (1992) emphasize the importance of including physiologic descriptors in an account of speech motor development: "A proper understanding of motor events may be clouded by a premature phonetic attribution to infant's sound productions.... A conservative approach is to emphasize the nascent motor functions rather than the apparent (but arguable) phonetic representation underlying

infant sound productions." (p. 84). Recently, investigators have begun to develop methods for acquiring stable physiologic measurements from the speech apparatus during the first years of life (Barlow, Finan, Bradford, & Andreatta, 1993; Boliek, Hixon, Watson, & Morgan, 1996; Moore & Ruark, 1996; Ruark & Moore, 1997).

Physiologic Constraints on Speech Acquisition

The universality of speech sound acquisition across languages (Locke, 1983) suggests that the infant has a propensity for certain articulatory dynamics and configurations, and is incapable of producing later developing sounds. Locke (1993) recognized the predispositions in vocal development: "The gestures of babbling are so robust they survive retardation, neonatal brain damage, and congenital deafness" (p. 179). Contrary to this notion, Jakobson (1949; as cited in Vihman, 1996) suggested that each infant produces the complete inventory of the world's languages and that the essential process of phonemic development is the shedding of nonambient sounds. This hypothesis was later refuted by observations that babble consists of only a small repertoire of consonants and vowels (Mitchell & Kent, 1990; Smith, Brown-Sweeney, & Stoel-Gammon, 1989; Stoel-Gammon, 1985; Stoel-Gammon & Otomo, 1986).

The universals in early speech may reflect the shared experiences among humans while developing speech proficiency. Speakers share similar anatomic structures and engage in similar oral motor experiences during early development (i.e., sucking, chewing, early vocalizations). Moreover, speakers must solve common motor problems including management of reactive forces (intrinsic and extrinsic) and establishment of

supporting neuromotor organization. The question becomes how can processes in speech motor development be characterized and how do they influence the sequence of phonologic development.

Of course, the development of speech motor control entails more than these biologic influences. Motor processes of speech are shaped by multiple intrinsic (e.g., cognitive/linguistic and sensorimotor maturation) and environmental (e.g., auditory and visual stimulation, learning, social experience) forces. Accordingly, verbal communication can be modeled as a dynamic system (Kelso, Saltzman, & Tuller, 1986). The evolution of a dynamic system will be retarded by its slowest developing components (i.e., " rate limiting" factors; Thelen, Ulrich, & Jensen, 1989). The following literature review focuses on potential anatomic and physiologic factors that may limit the rate of speech development. The focus of discussion is not intended to minimize the role of other influences, such as environmental stimulation and cognitive readiness, but is intended to motivate further investigation into the directly observable aspects of the speech developmental process.

The Problem of Early Speech Development: Anatomic and physiologic constraints on early sound production

The infant is challenged by a unique set of motor control problems during the acquisition of mature speech. A well recognized problem in early speech production is that of normalization (Kent, 1984; Lieberman, 1977). Initially, the child must (1) contend with a vocal tract that is not capable of generating the range of sounds

characteristic of adult speech (Lieberman, Crelin, & Klatt, 1970) and (2) develop motoric and "speech signal" stability as vocal tract structures undergo dramatic changes in mass and geometry (Kent, 1984; Lieberman, 1977; Lieberman, et al., 1970; Sasaki, Levine, Laitman & Crelin, 1977). Vocal tract growth yields shifts in resonances, which challenge the infant's attempts to acquire a "target acoustic output." The impact of vocal tract expansion on both emerging phonetic forms and consistency of speech motor performance is poorly understood.

The size and geometry of a newborn's supralaryngeal vocal tract contributes to its inability to produce adult sounds. In comparison to the adult vocal tract, the infant's has several distinguishing features:

1. The hyoid, epiglottis, and larynx are positioned high in the pharynx (Lieberman, 1977).
2. The tongue is completely housed in the oral cavity (Lieberman, 1977).
3. The supralaryngeal cavity between the dorsum of tongue and larynx is small (Lieberman, 1977).
4. Adenoids are inverted (Subtelny & Baker, 1953).

Combined these features make the infant vocal tract more like a "single tube system." In contrast is the adult human vocal tract where the "cross-sectional area of the oral and pharyngeal cavities can be independently manipulated ... while a midpoint constriction is maintained." (Lieberman, 1977, p. 77). The continuous nature of the newborn vocal tract considerably limits the acoustic variations available to the infant (Lieberman, 1977).

During the first six months of life, the anatomic configuration of the infant's vocal tract limits its ability to produce oral sounds (Kent & Murray, 1982). Between birth and four–to-six months of age, the epiglottis and the velum nearly form a contiguous structure (Sasaki, et al., 1977). When the child is upright, this bridge presents high resistance to oral airflow and lowers nasal resistance. This observation parallels the common report that the earliest vocalizations are predominately nasal (Oller, 1978). Thus, along with neuromuscular maturation, the repositioning of the velum in the nasopharynx assists the child in producing oral speech.

Of even greater significance to sound acquisition than the maturation of the vocal tract is the formation of the neuromuscular framework that coordinates the articulators (Lieberman, 1984). Many of the most basic questions concerning the development of speech motor control have yet to be addressed: How does the sequence of neuromotor maturation influence the sequence of sound acquisition? How does growth interact with coordination of the articulators during speech acquisition? What is the role of reflexes and other extant neural circuits in the development of oral motor control for speech? What are the motor milestones of speech?

A number of experimental findings provide indirect support for the rate limiting effects of physiologic constraints on the development of speech:

1. The sequence of acquisition of sounds and sound classes is similar across cultures (Locke, 1983).

2. Some aspects of auditory perception are established long before corresponding speech emerges (e.g., perception of the voicing contrast; Eimas, Siqueland, Jusczyk, & Vigorito, 1971).
3. Infants attempt to imitate speech as early as three to five months of age (Kuhl & Meltzoff, 1996), suggesting that production precedes linguistic need.
4. Speech performance does not stabilize until adolescence (Eguchi & Hirsh, 1969; Kent, 1976; Tingley & Allen, 1975), maturing much later than underlying cognitive/perceptual substrates.

Thus, during the first year of life, anatomic constraints limit the sound producing capabilities of the infant. Physiologic constraints may also influence the course of speech development, but little is known about the potential consequences of such limits. Accordingly, a firmer understanding of physiologic constraints on early articulatory control may change the way we conceptualize speech development.

Processes in the Development of Motor Control

Sequences in neuromotor development may differentially constrain articulatory movement, thereby influencing the sequence of phonologic development. Universalities in the sequence of motor skill development led early researchers to focus their attention on the involvement of neuromotor maturation in motor development (Gesell & Ames, 1940; McGraw, 1940, cited in Haywood, 1993). These investigators observed the skills for posture, sitting, walking, and grasping tended to develop in the same order in all infants despite significant variations in their experiences (Haywood, 1993). The parallel

question of whether speech motor skills are attained in a consistent sequential order has not been addressed directly.

The developmental processes observed in the sequence of coordination for walking and reaching may be similarly operative in early speech. The challenges the child must overcome to grasp and walk are similar to those posed by coordination of vocal tract structures for sound modulation. For instance, for effective bilabial closure, the child must learn to coordinate lip and jaw movement within narrow time constraints. The developmental sequence of motor skills has been characterized as requiring *differentiation* (i.e., the modification of a pre-existing behavior into more specialized ones) and/or *integration* (i.e., integration of new behaviors with previously stabilized ones). This question of the ability of the oromotor system to exploit extant coordinative structures is central to the understanding of speech development.

Differentiation: Speech motor control may emerge from undifferentiated to differentiated organization

Gabbard (1992) has defined differentiation in motor control:

> Differentiation is the process by which structure, function, or forms of behavior become more specialized. In general, it is the progression of motor control from gross, poorly controlled movements (like those displayed by young infants) to more precise, complex forms of motor behavior commonly exhibited by older individuals. (p. 6)

The involvement of differentiation in motor development may be manifested behaviorally as comodulation of non-target muscles or at the level of movement, the

presence of extraneous movements accompanying an intentional movement. Such associative movements have been observed to decrease with age and specific training (Connolly & Stratton, 1968; Lazarus & Todor, 1987). These associative movements have also been termed "motor overflow" and "mirror movements," and have been observed at various anatomical sites (e.g., ears, fingers, and limbs) and levels of organization (e.g., limbs, Lazarus & Todor, 1987; motor units, see Provins, 1997). The presence of these movements in early motor development have led a number of investigators to suggest that neuromotor maturation may limit the young child's ability to selectively activate some neuromuscular elements (Jeannerod, 1988).

The learning of grasping conforms to this general sequence of motor development. When a child first learns to grasp an object, the arm segments move as a unit and the hand is primarily transported by rotation of proximal joints (Jeannerod, 1988). Gradually, the child works toward gaining independent movement of the arm, hand, and fingers (Schuster & Ashburn, 1992). Trevarthen (1984) described the whole body response in pre-reaching:

> Pre-reaching, since it does not involve guidance or redirection of movements during the cycle, can be described as 'ballistic' or 'open loop'. However, a succession of pre-reaching movements, possibly involving both hands, may show tracking in conjunction with whole body orientations... Such oriented activities also involve the feet which move in quadripedal synchrony with the hands. (p. 242)

Although mature speech production requires the ability to independently move vocal tract structures, it is not known if early speech exhibits such widely distributed motor patterns. In other words, tendencies in early speech may reflect the presence of obligatory coordinative linkages that give way to simultaneous activation of the involved articulators. In this case, advancement to mature speech would require the emergence of selective discrimination of neuromuscular elements. Moreover, the development of articulatory control would follow the general-to-specific sequence that would be manifested behaviorally as gains in independent control of vocal tract structures.

Integration: Speech motor control may require the assembly of existing skills

Motor control does not develop uniformly across the various motor systems. Along with structural growth and myelination (Schuster & Ashburn, 1992), motor control generally emerges cephalocaudally and proximodistally (Stallings, 1973). Postural control, for example, exhibits this developmental progression. For posture, control is first demonstrated in the head and neck and later becomes apparent in the trunk and lower limbs. This gradual acquisition of coordination reflects the integration of new behaviors with previously stabilized behaviors. Development of speech motor control may exhibit a similar progression, where gains in articulatory control are sequential.

It is probable that the organization of coordination for speech involves both the *integration* and the *differentiation* of vocal tract components (Fentress, 1984; Kent, 1992). Lenneberg (1967) emphasized the involvement of both maturational forces in determining the sequence of motor development:

> First, the order in the emergence of sensory and motor capacities, reflexes, and spontaneous motoric events is constant and predictable for a given species. The sequence of events is not dependent upon experience. Second, the embryological emergence of behavior in mammals cannot be characterized by one simple scheme such as gradual differentiation of a whole or assembly of independently arising component reflexes. Although some aspects of behavior do emerge as undifferentiated patterns with subsequent individuation (for instance, the movements of limbs and digits and, to some extent, of head and trunk), other aspects of behavior are best regarded as a lawful but gradual integration of reflexes that can be observed to function independently before total integration takes place. (p. 9)

Sequence of Oromotor Development

It is known that, prenatally, the control of oral structures emerges at different times (Herring, 1985). For instance, while the lip musculature is still in the premyoblast stage at eight weeks gestation (Gasser, 1967), the fetus is already opening the jaw (Humphrey, 1964). Herring (1985) has speculated that the sequence of early oromotor development is orderly and driven by neuromuscular development. Prenatal studies of both animals and humans suggest that the order of appearance of oral movements tends to be jaw opening, jaw closing, tongue movement, and lip movement (see Herring, 1985; Humphrey, 1970; 1971). Description of the postnatal development of orofacial control is

confounded by the varying context of the oral functions. For instance, lip control may develop adequately for sucking much earlier than for speech.

The ontogeny of oral function includes the neurophysiologic control of sucking, chewing, and speech. Although these tasks share the same motor effector structures, they emerge during different periods of postnatal development (Bosma, 1985). Bosma (1985) suggests that the appearance of new oral functions coincides with the survival needs of the infant. Bosma (1973) further suggests that pharyngeal and laryngeal functions develop ahead of oral functions, citing the early appearance of swallowing and crying in the premature infant in comparison to sucking, which does not mature until the 37th week of gestation. The suck-swallow is one of the earliest oral behaviors to appear, usually close to term (Bosma, 1985). The shift from sucking to chewing coincides with the reorganization of central motor pathways (Iriki, Nozaki, & Nakamura, 1988). For chewing, oral coordination must evolve considerably, in comparison to sucking, for the child to advance to solid foods. Although the basic coordinative infrastructure for chewing is well established as early as 12 months of age, toddlers continue to make refinements until approximately three years at which time the chewing pattern appears adult like (Green, Moore, Ruark, Rodda, Morvée, & VanWitzenberg, 1997). In contrast, children typically do not master the sounds in their ambient language until eight years (Sanders, 1972). Moreover, adult-like stability in the acoustic signal is not achieved until adolescence (Eguchi & Hirsh, 1969).

Alimentary functions such as sucking and chewing have been frequently cited as antecedent behaviors from which speech motor coordination emerges (Darley, Aronson,

& Brown, 1975; Grillner, 1982; Ling, 1976; Mysak, 1980). This notion has intuitive appeal on several grounds. First, these oral functions share the same peripheral structures (Lund, Appenteng, & Séguin, 1982). Second, alimentary functions, such as chewing, have been historically grouped with the "earliest appearing" oral behaviors both ontogenetically and phylogenetically (MacNeilage, 1997). However, the disparities among the coordinative patterns for alimentary functions and speech have prompted a number of investigators to question the plausibility of a shared neural structure (Goffman & Smith, 1994; Green et al., 1997; Moore & Ruark, 1996). The coordination demands for speech probably exceed those of alimentary functions because (1) alimentary functions involve only a subset of the oral structures engaged for speech production (Bosma, 1985), and (2) the requirements for speech coordination are non-stereotypical and highly time specified (Gracco, 1994).

Development of Motor Skills for Speech: Current models and issues

A frequently cited sequence of speech acquisition is the one developed by Oller (1978). His basic taxonomy of early speech development is presented in Table 1. Note that "speech like" syllables do not appear until four to six months, Oller's stage III. Prior to this time the infant produces a great variety of sounds, some of which have little resemblance to speech (e.g., trills, coos; see Kent & Murray, 1982). At about six months, syllable primitives emerge as the infant engages the articulators to modulate the glottal tone. During Oller's stages III-V, rhythmic oscillations of the mandible produce primitive syllable trains commonly referred to as "reduplicated" or "canonical"

babbling. A number of researchers have suggested that such repetitive oscillatory movements, which are observed across many parts of the effector system between six and nine months, may assist in the transition from uncoordinated to coordinated movement (Kent, 1984; 1992; Thelen, 1991).

Table 1. Stages of normal speech development within the first year of life.

Stage (age in months)	I (0-1)	II(2-3)	III(4-6)	IV-V (7-12)
Vocal output	Quasi Resonant Nucleus "reflexive vocalizations"	GOO "Cooing"	Fully Resonant Nuclei and Marginal Babble	Reduplicated and Variegated Babbling

Adapted from Oller, 1978

A model proposed by MacNeilage and Davis (1990a; 1990b; Davis & MacNeilage, 1994; 1995) complements Oller's taxonomy, linking the pattern of sound acquisition to the underlying physiologic processes. The core of their model of speech development is the "Frames, then Content" metaphor, which was designed to explain how primitive syllables evolve into mature speech. In their model, "frames" are the most primitive articulatory pattern and are produced by the cyclic depression and elevation of the mandible. The emergence of speech-like vocalizations or babble coincides with the successive repetition of these primitive frames. Vocalizations become increasingly more complex or "adult like" as subsyllable organization develops within the frame structure.

These authors posit that infants produce variegated babbling by modulating the basic frame, a first step toward development of segmental properties. "Content" refers to this more complex subsyllable organization. In this view, old forms (i.e., "frames") are modified to create forms that are more complex. MacNeilage and Davis (1990a) draw support from the observation that voiceless phonemes appear to develop from alterations of the voiced form. MacNeilage and Davis' (1990a; 1990b) and Oller's (1978) proposals provide a conceptualization of speech motor development though lack empirical physiologic testing.

Physiologic Interpretations

MacNeilage and Davis' (1990a; 1990b) and Oller's (1978) characterizations of speech development can be viewed as integrative models of articulatory control. Each model proposes that early speech is produced primarily with a limited set of articulators and that development of mature speech coincides with increases in the number of involved articulators. Part of MacNeilage and Davis' (1990a) Frame-Content hypothesis proposes that early speech is produced exclusively using mandibular movement: ".... in the course of developing the capacity for simulating the alternation of sound levels that is characteristic of speech, the child develops *one* behavior that is a universal motor base for subsequent speech - the open-close alternation of the mandible." (p. 461) and "babbling episodes with labial stop consonants [p] and [m] may be produced without any specific positioning of the lips themselves. The fact that the lower lip 'rides' on the mandible may be sufficient to produce lip closure when the mandible is elevated." (p. 461).

It is alternatively possible that MacNeilage and Davis' "frame" is a manifestation of an undifferentiated motor program distributed across multiple articulators; other articulators may be equally responsible for producing early syllable primitives. Limited independence of functional components is a hallmark of early skilled movement (Provins, 1997) and may constrain early speech attempts as well. Anecdotal support for the presence of shared or widely distributed motor organization in early speech comes from the infant's emergent phonemic inventory. First, there is the frequent occurrence of voiced stops in early speech (Stoel-Gammon, 1985). Inter-articulator coordination for a voiced stop is "ballistic," unlike those classes of speech sounds that require a high degree of spatial and temporal specification (e.g., aspirated voiceless stops; Kewley-Port & Preston, 1974) and appear much later (Macken & Barton, 1980). Second, nasal consonants occur infrequently in early speech (Stoel-Gammon, 1985). This preference for oral sounds might be taken to suggest that velar closure is part of a motor program that is widely distributed across the speech apparatus. Alternatively, a developmental process might be hypothesized in which there is a more gradual increase in oral consonants in babble, reflecting a learning period with progressive gains in coordination of the velum and the supralaryngeal articulators. The infant's motivation to produce oral sounds might be attributed to a number of factors: oral sounds are acoustically rich, perceptually salient, and likely predominant in the infant's ambient language.

From studies of orofacial reflexes comes support for the notion that some of the neural pathways in the infant's oral region become more specified with maturation

(Barlow et al., 1993; Humphrey, 1970). Barlow and colleagues (1993) have demonstrated that tactile stimulation of the vermilion border in infants evokes a large reflex response (perioral reflex) across the quadrants of the orbicularis oris superior (OOS) and orbicularis oris inferior (OOI), indicating an absence of specificity between site of stimulus and effector activation. In contrast, adult subjects exhibited a localized ipsilateral response with the strongest effect recorded in the homonymous muscle. These findings raise the question: Does the development of voluntary orofacial control parallel perioral reflex development? In addition, these results provide further motivation for the exploration of widely distributed movement patterns in early speech.

Several mechanisms may limit the independent control of orofacial structures in early speech: *a widely distributed central motor program, a common source of neural drive, a simplification control strategy, or a pre-established motor pattern*. The wide distribution of the central motor plan for speech entails that the command signals do not contain the details necessary to activate only those muscles or articulators required for the task. According to Bernstein's (1996) notion of "motor constraints," motor learning involves the elimination of superfluous and redundant forces. This optimization process is an essential feature of skilled movement development (Bernstein, 1996) and can develop over a period of years (Provins, 1997). Thus, the organization of speech motor control may require the paring down of grossly defined motor programs.

Alternatively, a common source of neural drive may exert control over multiple articulators limiting the potential for independent movement of these anatomically distinct structures. To this author's knowledge, the prevalence of shared neural structures

in early speech motor control has not been considered. However, several researchers have begun to examine the possibility of shared loci of control for articulators in mature speakers (Goffman & Smith, 1994; Smith & Denny, 1990).

Another possibility is that an immature speaker may compensate for poor control by reducing the degrees of freedom of control (i.e., simplification control strategy). This compensatory strategy represents an initial solution to a motor problem and differs from the previous notion of reducing degrees of freedom for achieving optimization of a skilled movement (Bernstein, 1996). Simplification control strategies are observed, for instance, from the kinematic differences in grasping between preferred vs. non-preferred hand during the transportation of objects (Steenbergen, Marteniuk, Kalbfleisch, 1995). When grasping with the non-preferred hand, subjects tend to reduce the number of variables that needed to be independently controlled by decreasing angular motion of distal joints (shoulder and elbow) and increasing displacement of proximal segments (trunk). It is also possible that an immature speaker would reduce the complexity of articulatory control during early speech by linking portions of the speech apparatus.

Finally, the developing child may need to overcome pre-existing coordinative patterns (i.e., negative transfer) employed for tasks that share the same peripheral structures such as chewing and sucking. Transfers of learning effects occur when a pre-existing skill influences the learning of a new skill. These effects may facilitate (positive transfer) or impede (negative transfer) the learning of a new skill. Evidence for this process in motor learning is exemplified when a trained tennis player first attempts to play badminton (Magill, 1993). Individuals who play tennis will often erroneously

stiffen their wrist when learning badminton, which requires a wrist snap. Traditionally, extant behaviors such as chewing and sucking have been viewed as facilitating speech motor development (see Moore & Ruark, 1996). However, if negative transfer effects are operative, the advancement to mature speech may require that the young child overcome such well-ingrained oromotor patterns.

In summary, if any of these mechanisms are operative in the immature motor system, the development of coordinative organization for speech may require increases in the ability to control vocal tract structures independently. Moreover, the presence of linked articulatory control may have predictable effects on the child's ability to produce certain phonemes under these coordinative constraints. Additional physiologic data may provide evidence of these mechanisms in early speech development.

Orofacial Control for Speech

Functionally, the upper and lower lips are distinct. Unlike the upper lip, the movement of the lower lip must be coordinated with respect to the position of the mandible. During speech, the lower lip produces greater force (Barlow & Netsell, 1986) and movement than the upper lip (Gay, 1977; Sussman, MacNeilage, & Hanson, 1973). The flexibility in articulatory control exhibited by adult speakers reflects several organizational features in the neural infrastructure mediating the perioral musculature. The perioral musculature is finely regulated with the number of muscle fibers innervated by each motor neuron estimated to be between 3:1 and 5:1 (Laurenson, 1968). This innervation ratio is much lower in contrast to other human muscles such as the medial

head of gastrocnemius where one motor neuron innervates up to 2600 fibers (Feinstein, Lindegard, Nyman, Wohlfart, 1955).

Several researchers have suggested that the upper and lower lips have different loci of neural control (Abbs & Gracco, 1984; Goffman & Smith, 1994; Smith, 1992; Wohlert & Goffman, 1994) and that there are biomechanically distinct compartments within the orbicularis oris (Abbs, Gracco, & Blair, 1984). Support for the existence of separate neural inputs to various regions of the orbicularis oris comes from (1) studies showing non-overlapping sub-nuclei in the facial nucleus of several species (e.g., Welt & Abbs, 1990, rhesus monkey; Courville, 1966; cat) and (2) studies that have reported an absence of coherence in lip EMG recorded from different quadrants of the orbicularis oris (Goffman & Smith, 1994).

Despite functional partitioning of the labial musculature, a speaker's capacity to control these muscles independently is limited (Folkins, 1978; Folkins, Linville, Garrett, & Brown, 1988). Folkins and colleagues (Folkins et al., 1988) examined the interactions in muscle activity of upper lip and lower lip during speech. Muscle activity was sampled from four different sites on the lips: midline orbicularis oris superior (OOS), midline orbicularis oris inferior (OOI), lateral OOI, and midline mentalis. Subjects were provided visual feedback reflecting activation level from one of the electrode sites (either midline OOI or midline OOS) and were instructed to manipulate the amplitude of muscle activity at that site while speaking. A volitional change in the level of activity in the target muscle produced similar modulation in the muscle activity of the other three muscles. Thus, for the speech task that was studied, elements of the perioral musculature

behaved as an interdependent, linked system. Because the correlations between OOI and OOS were not significantly higher than those of OOS and mentalis, it was suggested that these coordinative linkages are not restricted to orbicularis oris. Of interest was that only positive correlations were observed between upper lip and lower lip electrode sites, which is the opposite of what would have been predicted if the perioral musculature obeyed principles of complementary variation.

In summary, the coordinative organization of the perioral musculature allows oral structures to perform the varied actions required for speech, sucking, and chewing. Because coupled activation is obligatory among some perioral muscles, the extent to which humans can control these muscles independently is limited. However, the functional significance of these coordinative linkages may be inconsequential because isolated activation of perioral muscles is probably not essential to speech.

Development of Lip and Jaw Coordination for Speech

Few studies have described developmental changes in articulatory coordination. Research has been impeded by the absence of viable methods for obtaining physiologic measures from the articulators of young children. Several investigators (Sharkey & Folkins, 1985; Smith & Gartenberg, 1984; Smith & McLean-Muse, 1987; Watkin & Fromm, 1984) have successfully employed strain gauge devices to transduce lip and jaw movement in children, but because this method requires considerable cooperation from the subject, it is not amenable for use with very young children. Consequently, articulatory kinematic data are nonexistent from children under the age of 4 years. More

recently, several investigators have been able to record electromyographic activity from the orofacial musculature in very young children (e.g., Barlow et al., 1993; Moore & Ruark, 1996; Ruark & Moore, 1997).

By comparing the activation patterns of perioral musculature across oral tasks (e.g., chewing, speech), Moore and colleagues (Moore & Ruark, 1996; Ruark & Moore, 1997) have described traces of the coordinative infrastructure for speech in children as young as 15 months of age. Moore and Ruark (1996) studied differences in mandibular coordination during chewing, sucking, babbling, and speech in a group of normally developing 15-month-old children. They observed distinct coordinative patterns between non-speech and speech tasks. Specifically, chewing was characterized by a pattern of reciprocal activation of antagonist muscle pairs whereas speech was characterized by co-contraction of antagonist muscle pairs. Thus, as early as 15 months of age, children produce coordinative strategies for mandibular control that resemble those of the adult (Moore, Smith, & Ringel, 1988).

Ruark and Moore (1997) compared changes in upper and lower lip activity in two-year-old children across a variety of speech and non-speech tasks (e.g., chewing, rhythmic lip pursing). The degree of coupling varied across the different tasks. For instance, upper and lower lip activity was highly correlated during pursing and syllable repetitions, and less correlated during lip rounding and non-labial speech. These coordinative patterns were similar to those previously observed in adult speakers (Wohlert & Goffman, 1994). The authors interpreted these findings to suggest that, at

age two, children exhibited differential control of upper and lower lips for these different oral tasks.

Smith and Gartenberg (1984) examined differences in lip and jaw movements during speech in four children (ages: 4,6; 5,4; 6,11; and 7,0) and three adults. Unlike the adult speakers, the children did not produce systematic increases in peak velocity of lip and jaw movement from /p/ to /b/ to /m/. In both children and adults, the lower lip exhibited the greatest extent of movement, followed by jaw then upper lip. Children's speech movements were consistently slower and more variable than adults', corroborating previous observations of longer and more variable segment durations in children as compared to adults (Eguchi & Hirsh, 1969; Kent & Forner, 1980). Because children and adults produced similar magnitudes of movement across articulators, differences in movement velocity, for example, could not be explained by anatomical differences alone.

Smith and McLean-Muse (1987) examined potential developmental changes in motor equivalence. These authors recognized that few studies have empirically supported the existence of motor equivalence in speech (Hughes & Abbs, 1976), but reasoned that motor equivalence should increase with age if it is a characteristic of skilled motor abilities. Using a strain gauge apparatus, the experimenters recorded upper lip, lower lip, and jaw movements from four age groups (i.e., five-, eight-, and eleven-years-old, and adult) with six subjects in each group. Subjects produced CVC stimuli in one "normal" condition (i.e., normal speech rate) and in two perturbed conditions (i.e., fast rate, bite block condition). The data were subjected to several analytic procedures;

however, motor equivalence was not supported in any of the age groups studied. In one method, for example, Pearson product moment correlations were performed on the displacement values of articulator pairs collected from repetitions of the same utterance. Motor equivalence would have been supported if significant negative correlations between articulator pairs were observed, indicating a trading relationship among articulators. However, few subjects exhibited correlations that were significant from zero suggesting that motor equivalence was not operative in any of the age groups.

Sharkey and Folkins (1985) examined developmental changes in the variability of lower lip and jaw movement across several age groups (i.e., four-, seven-, and ten-years, and adult). These investigators measured the duration of lip opening, jaw opening, lip-open posture, jaw-open posture, and timing between the onset of lower lip lowering, and jaw lowering. Although adults' articulatory movements were significantly less variable than those of children, few significant differences in variability were found among the child groups. The variability in lower lip displacement decreased significantly between ages four and seven years, while the variability in jaw displacement did not change across any of the ages studied. This finding supports an earlier mastery of jaw control than for the lips. In contrast to the findings reported by Smith and Gartenberg (1984), no phoneme effects were found for any measure across groups.

Watkin and Fromm (1984) studied differences in the extent of upper and lower lip separation in vowels produced by eight subjects ranging in age from four to ten years. These investigators found that the relative contribution of the upper lip to labial separation decreased significantly with age from 19.4% to 8.5%. Successive repetitions

of the same stimuli became less variable with age with the greatest decrease occurring between seven and ten years of age.

Smith and Goffman (1998; Goffman & Smith, in press) studied differences in lower lip plus jaw movement among four- and seven-year-old children and a group of young adults. Smith and Goffman (1998) found no differences in the amplitude of articulatory movement among the groups, but found a significant increase in movement velocity and a decrease in overall movement duration. In contrast, Goffman and Smith (in press) reported that both movement displacements and velocity increased with age. These experiments were similar in their method with the exception that they used different speech stimuli. Therefore, the difference in their findings with respect to movement displacement may be related to differences in the intrinsic properties of the phonemes that were studied. Both studies reported a decrease in variability across all kinematic parameters with age. Calculation of the spatiotemporal index (STI) for each subject revealed that the children's movement patterns were more variable on repeated tasks than the adult subjects.

These investigators (Goffman & Smith, in press; Smith & Goffman, 1998) also designed their experiments specifically to ask the question: Do different phonemes develop from one primary form or "coordinative template?" This developmental process would be supported empirically if movement patterns for different phonemes appear similar initially but become more differentiated with maturation. Smith and Goffman (1998) found differentiated patterns in their youngest subjects, suggesting that children as young as age four produced movement trajectories that were distinguishable among

the different phonemic categories examined. Therefore, their data did not support the suggestion that the diverse movement patterns observed in adult speech are derived from a single underlying gesture or "articulatory primitive."

In summary, although the neural infrastructure for speech coordination is evident as early as 15 months, significant changes continue to occur well after age five. The findings of decreased variability and increased speed of movement with age is common across all studies and suggests that articulatory control undergoes progressive stabilization. Although the developmental sequence of articulatory control was never directly examined, several of these studies might be interpreted to suggest that (1) jaw control stabilizes earlier than lip control, and that (2) upper lip movement decreases while the lower lip assumes a more dominant role.

Purpose of Study and Statement of Problem

Although, the dominant models of early speech development (Oller, 1978; MacNeilage & Davis, 1990a; 1990b) predict specific changes in articulatory control, the developmental sequence of motor speech control has not been studied directly. Prior studies of speech motor development have been based on children aged four years and older, presumably after the coordinative organization for speech has been largely formed (Moore & Ruark, 1996, Ruark & Moore, 1997). An improved understanding of the developmental course of articulatory coordination will broaden our understanding of some of the organizational features of speech production:

1. the coordinative organizational features of mature articulatory motor control [e.g., velocity and displacement, Kuehn & Moll, (1976); velocity profiles, Gracco (1988); motor equivalence, Abbs & Gracco, (1984), speed accuracy tradeoff, Lindblom (1963); spatiotemporal stability, Smith & Goffman (1998)]
2. the sequence of speech motor development
3. the physiologic processes underlying developmental speech disorders
4. the sequence of phonemic acquisition

A reasonable beginning toward understanding the development of speech motor control is to study changes in articulatory movement. The purpose of the present investigation was to describe to describe developmental processes in labiomandibular coordination for speech. A central concern was to identify processes of development that may reflect integration and/or differentiation. The functional significance of these processes in early motor development is that they may have predictable effects on the child's ability to produce certain phonemes under these coordinative constraints, thereby accounting for preferences observed in early phonologic development.

Upper lip, lower lip, and jaw movements were recorded during the production of syllables containing bilabial consonants across age groups spanning the developmental continuum from babble to mature speech. The movement signals were subjected to two complementary analyses to quantify developmental changes in articulator coordination. One technique compared similarities in the spatial aspects of articulatory movement between articulator pairs (spatial coupling) and the degree of movement synchrony between articulatory pairs (temporal coupling). The other technique described

developmental changes in each articulator's contribution to closing the oral aperture for bilabial closure.

Beyond providing a general description of developmental shifts in coordination, these techniques provided a method to examine the presence of integration and differentiation in speech motor development. If differentiation is operative in the developmental course of articulatory control, we would expect infants to demonstrate a high degree of spatial and temporal coupling among the articulators. In addition, we might anticipate a decrease in an articulator's *contribution to oral closure* as it becomes disassociated from a tightly coupled ensemble. In this case, gains in independent control of the articulators would accompany speech motor development. Alternatively, if integration is operative, we would expect to observe the sequential appearance of articulatory movement for oral closure with age. Finally, in the absence of one of these distinct processes, we would anticipate (1) gradual increases in spatial and temporal coupling among the articulators with age, and (2) no dramatic shift in the role of each articulator for oral closure. One probable outcome is that these processes of motor skill development coexist but demonstrate differential degrees of involvement depending on the stage of speech motor development.

EXPERIMENTAL DESIGN AND METHODS

Subjects

A cross-sectional approach was employed to study several stages in speech development. Forty-six subjects comprised four subject groups: 6 infants (3 females, 3 males) between 11 and 13 months of age, 10 toddlers (5 females, 5 males) between 23 and 29 months, 10 children (5 females, 5 males) between 6 and 7 years, and 10 adults (5 females, 5 males) between 27 and 35 years. Seventeen additional subjects (15 infants and 2 two-year-olds) failed to produce the target utterances during the experiment.

The age groups studied were selected based on published findings on the development of the voicing contrast. Advances in the voicing contrast are expected to grossly reflect the development of speech motor control. One-year-old subjects were intended to represent an early phase in speech development that is predominated by voiced phonemes. Two-year-old subjects were intended to represent the period when the voicing contrast emerges (Engstrand & Williams, 1996; Macken & Barton, 1980). The six-year-old subjects were intended to represent the period when voice onset times (VOT) for voiceless and voiced stops are produced with little or no overlap, but with elevated levels of variability (Eguchi & Hirsh, 1969; Gandour, Holasuit Petty, Dardarananda, Dechongkit, & Mukngoen, 1986; Kent, 1976). Finally, the adult group reflected the mature pattern of coordination.

All participants were monolingual, native speakers of American-English and were pre-screened by interview of either the adult subject or the child's parent. All

participants had negative histories of speech, language, hearing, or vision problems, and developmental or neurological disorders.

Speech Samples

The target speech utterances sampled were "baba," "papa," and "mama." Sampling was limited to bilabial consonants for several reasons:

1) Phonetic inventories of early vocalizations demonstrate that the voiced bilabial consonants occur frequently in early speech (Stoel-Gammon, 1988; Stoel-Gammon & Otomo, 1986).
2) The production of bilabial consonants involves the coordination of articulators that are accessible for movement transduction and therefore amenable to study.
3) The production of a bilabial consonant involves a rigidly specified goal (bilabial closure) with relatively high degree of spatial and temporal coupling among the lips and jaw.

Elicitation techniques were designed to accommodate the different age groups. Samples from the young children were elicited during play activities involving the child, the caretaker, and the experimenter. Adult and six-year-old subjects were asked to read the target words from a poster in a pseudo-random order with a slight pause between each production at their normal conversational rate and loudness. The experimenter provided verbal exemplars throughout each experimental session to encourage the subjects to stress the first syllable of the target utterance and to speak at a normal conversational rate and loudness.

Approximately 45 speech samples were obtained (15 repetitions x 3 phonemes) from the adult and six-year-old subjects. The younger subjects (infant and two-year-old) produced only a subset of these utterances because children this young (1) vary in their willingness to speak in an unfamiliar environment, (2) vary in their vocal imitative skills, and (3) do not typically produce the voiceless bilabial stop (i.e., /p/) until around age two (Stoel-Gammon, 1985). Consistent with those characteristics, the two-year-old subjects in the present experiment varied considerably in their ability to produce the voicing contrast accurately, the predominant error being the substitution of [b] for /p/.

The utterances produced by the infants and two-year-old subjects included both spontaneous and imitative tokens. The two-year-old subjects had little difficulty imitating the models provided by the experimenter or parent. In contrast, only two of the six infant subjects unmistakably imitated utterances modeled by the experimenter or parent during data collection. The inclusionary criteria for the analysis of the infant's utterances specified (1) that complete lip closure was observed on the videorecording during the consonant embedded in a CV with the mouth initially open, or a VCV or CVCV utterance, or a sequence of canonical babble, and (2) that the utterance was produced during a normal mode of phonation (e.g., not during whisper) and at conversational loudness level, as judged by the experimenter. Additionally, utterances associated with "normal" dysfluencies (i.e., block or hesitation), coughs, or mispronunciations were not included in the study. Few tokens were rejected for these reasons.

Data Collection and Recording Conditions

Experimental sessions lasted about one hour for the one- and two-year-old subjects and approximately twenty minutes for the six-year-old and adult subjects. Data were collected in a large sound-treated booth equipped for audio and video recording. Subjects two-years-old and under were seated in either the caretaker's lap or in a highchair with the tray pushed in to minimize extraneous movement. Each child's utterances were transduced using a wireless remote microphone (Telex, FMR-25) attached to his/her shirt collar, and were recorded by a digital audio recorder (Panasonic, SV-3700). A full-face video recording of each subject was used to capture lip and jaw movement using an infrared camera (Burle, TC351A) coupled to a videorecorder (Panasonic, AG-1980).

Three single flat, circular reflective markers (~2 mm in diameter) were placed at the midline on the margin of the vermilion border of the upper lip (UL), lower lip (LL), and just superior to the mental protuberance of the mandible (J). Two reference markers (~2 mm in diameter), also placed midline, one on the tip of the nose and one on the nasion, were used to correct for head movement that would otherwise be included in the articulatory movement signals. These two markers translated the origin to the nasion marker and aligned the axis to the line defined by these two markers. A 4-cm reference marker, placed on the subject's forehead, was used to calibrate the movement signals. If a child exhibited any concern with this large reference marker, it was replaced with a smaller one (~ 1 cm).

Several precautionary measures were taken to reduce optical distortion while videorecording. Distortion due to the shape of the camera lens was minimized by positioning the subject's face in the center of the field of view with the camera zoom at maximum. The experimenter encouraged younger children to position their face perpendicular to the camera lens by holding a toy directly above the camera while eliciting samples. Speech samples accompanied by significant forward or backward rotation of the head were excluded from this analysis because significant rotation about the Z plane (sagittal) distorts the relative position between objects in a two dimensional coordinate system (i.e., x, y).

Digitization and Signal Conditioning

Bilabial consonants were identified from the combined audio and video recordings. Upper lip, lower lip, and jaw positions were extracted automatically from the videorecordings using a computer-based movement tracking system (Motus, version 2, Peak Performance). Vertical positions of the upper lip, lower lip and jaw were sampled at 60 samples per second.

The accuracy of the movement tracking system was evaluated by measuring the position of a single marker attached to the end of a micrometer. Vertical displacement of the marker in 16 successive steps of 5 mm each was measured for the experimental conditions (e.g., videocamera, approximate zoom factor, lighting, and reflective stickers). The average error for the position extracted by the movement tracking system was .1mm (SD = .05), which was within the limits of the experimenter's ability to

position the micrometer manually. The accuracy of this movement tracking system was assessed to be better than .1mm.

The displacement signals were digitally low-pass filtered (f_{lp} = 15 Hz) using a zero-phase shift forward and reverse digital filter (Butterworth, 8 pole). The lower lip signal was derived by subtracting the lower lip displacement signal from the jaw signal. An example of a kinematic record from an adult subject is presented in Figure 1.

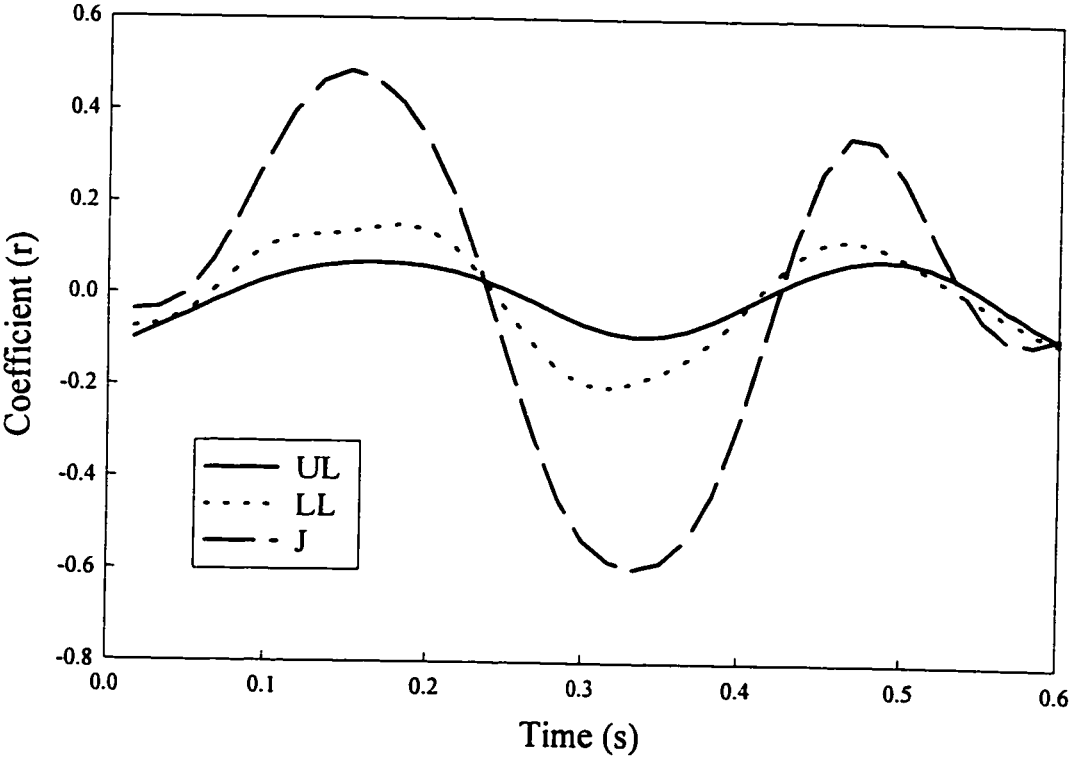


Figure 1. The treated kinematic traces from upper lip (UL), lower lip (LL), and jaw (J) produced by an adult subject saying "baba." For ease of interpretation, each signal has been centered about its mean and the UL signal has been inverted.

Quantitative Analyses of the Kinematic Traces

The kinematic tracings from upper lip, lower lip, and jaw were subjected to two complementary analytic techniques. Custom routines written for Matlab (The MathWorks Inc., version 5.1) were developed (1) to compute crosscorrelation functions across movement records and (2) to measure each articulator's contribution to oral closure during speech. These two separate analyses yielded three indices of coordination:

(1) *maximum or minimum coefficient from crosscorrelation function* - an index of the similarity in spatial aspects of movement trajectories (spatial coupling)

(2) *lag to coefficient* - an index of the degree of movement synchrony (temporal coupling)

(3) *percentage contribution to oral closure* – an index reflecting the relative contribution of each articulator to closing the oral aperture during bilabial closure for speech.

Articulatory Coupling and Synchrony

Peak coefficients (negative or positive) and their associated lags were derived from the crosscorrelation functions computed between the treated displacement traces of all possible articulatory pairs (i.e., UL x LL, UL x J, LL x J). One strength of this method for quantifying inter-articulator coordination is that it is inherently normalized to differences in magnitude of the movement; thus, changes in inter-articulator

coordination are reflected independent of differences in relative vocal tract size among subjects.

The movement waveforms were trimmed prior to analysis with the assistance of a custom algorithm. The waveforms from each articulator and the velocity trace of the jaw (the time derivative of the jaw position) were displayed simultaneously on a computer monitor. An algorithm automatically indicated utterance boundaries as instances of zero velocity occurring in the jaw velocity signal. Jaw velocity was used because jaw movements tended to be more predictable and well-defined (i.e., characterized by two rising and falling gestures across the CVCV utterance) than were upper or lower lip movements across the age groups. If the subject's jaw was lowered before speaking, the initial zero velocity would be associated with the beginning of the closing phase for a CVCV utterance. If the subject's jaw was elevated before producing the utterance, the initial zero velocity point was associated with the release of the initial bilabial consonant. The end of the segment was the zero velocity occurring at the termination of jaw lowering for the vowel. In the event that zero velocity was not apparent in the jaw velocity signal, the experimenter would identify the onset and offset manually based on a visual display of the movement.

Figure 2 shows a single crosscorrelation function computed on the movement traces displayed in Figure 1. For ease of interpretation, the upper lip signal was inverted prior to analysis and each signal was centered about its mean. The most prominent peak (positive or negative) was identified from each crosscorrelation function using a custom algorithm written for Matlab. This routine required the investigator to place a cursor at

the most prominent peak or trough in the crosscorrelation function in a ~ 200 ms window centered on zero lag. The algorithm identified the maximum or minimum *coefficient* value in this window. The *coefficient* and its *lag* value were automatically exported to a database.

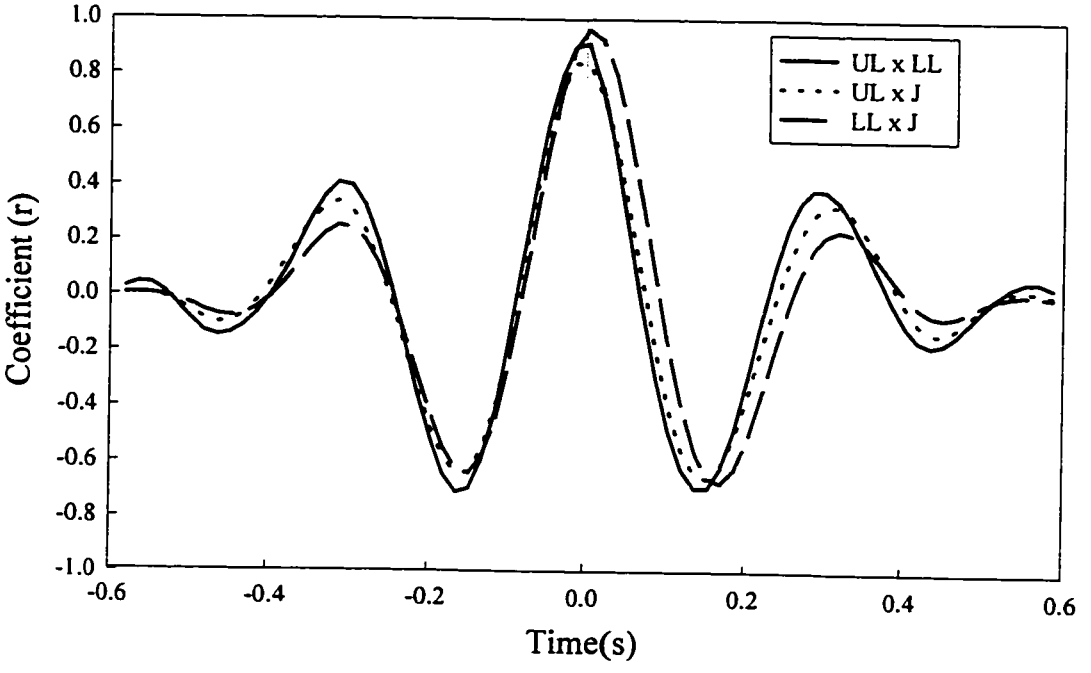


Figure 2. Pairwise crosscorrelation functions computed on the signals presented in Figure 1. The dashed vertical line denotes zero lag.

If the crosscorrelation function did not contain a prominent peak within the 200 ms window, the *coefficient* and *lag* for that articulatory pair were omitted from the final data corpus. This precautionary measure reduced the possibility of erroneously selecting peaks in the crosscorrelation function that were greater in duration than a unidirectional movement (i.e., lip elevation for /p/). For instance, it would be erroneous to select a

prominent negative peak in the crosscorrelation function that represents the correlation between the opening gesture of one signal and the closing gesture of another. One drawback of this approach is that differences in inter-articulator timing that are greater than half the window width (approximately 100 ms) cannot be detected. Approximately 8% of all tokens were rejected because of this criterion. This proportion did not differ significantly across age groups.

In this analysis, articulators that exhibit rigid temporal and spatial coupling yield high correlations and near-zero *lag* values. The magnitude of the correlation coefficient reflects the degree of spatial coupling and the value of the *lag* reflects the degree of temporal coupling (e.g., movement synchrony) within the articulatory pair. For the speech stimuli being examined, adult subjects usually produce spatially and temporally coupled articulatory movements. If the articulators differ significantly in their direction and pattern of movement, the signals will not be correlated. Thus, a significant deviation from the "mature pattern" would reflect inter-articulator dyscoordination and would produce coefficients closer to zero and/or greater *lag* values.

Articulatory Contribution to Oral Closure

Developmental shifts in articulatory control were examined by comparing changes in the relative *contribution* of the upper lip, lower lip and jaw for oral closure across age groups. This analysis was designed to complement the crosscorrelation methods by providing information on age-related changes in the extent of movement displacement. The information provided by the crosscorrelation analysis is limited because small articulatory displacements can be highly coupled although their functional

significance may be minimal. For instance, the upper lip and lower lip may be highly coupled but their movement may be relatively unimportant to the goal of oral closure (i.e., jaw movement may predominate oral closure).

Custom algorithms were designed to compute each articulator's relative *contribution to oral closure*. These routines required only that the experimenter grossly identify the onset and offset of the movements associated with each syllable. Once the syllable was roughly identified, calculating each articulator's relative *contribution to oral closure* involved four steps:

Step 1: Identification of each articulator's position during oral closure: (e.g., $UL_{closure}$)

For each syllable, the positions of the upper lip, lower lip, and jaw were recorded when the distance between the lips was at a minimum (i.e., position during oral closure).

Step 2. Identification of each articulator's position during maximum performance task: (e.g., $UL_{reference}$)

The position of each articulator was recorded in its maximum open position. These "reference" positions were intended to represent the functional boundaries of the oral aperture in their maximum open position. To capture these positions, the experimenter verbally and/or gesturally cued each subject to produce a smile and a yawn like gesture. The lower lip and jaw reference positions were recorded during the yawn and the upper lip's position was recorded during the smile. In some instances, the maximum open positions were recorded from spontaneous yawns, loud cries, or smiles in the younger children. Five of the younger subjects

(3 one- and 2 two-year-olds) were excluded from this analysis because they did not produce acceptable maximum open positions during the session. If available, several maximum opening positions were recorded and the greatest opening excursion observed was deemed the reference position for each articulator.

Step 3. Calculate each articulator's position in the oral aperture: (e.g., UL_{co})

For each syllable, the position of each articulator during the reference posture (i.e., maximum opening) was subtracted from its position during oral closure. These values represented the extent that each articulator occluded the oral aperture. See Equation 1.

Equation (1)

$$UL_{co} = UL_{reference} - UL_{closure}$$

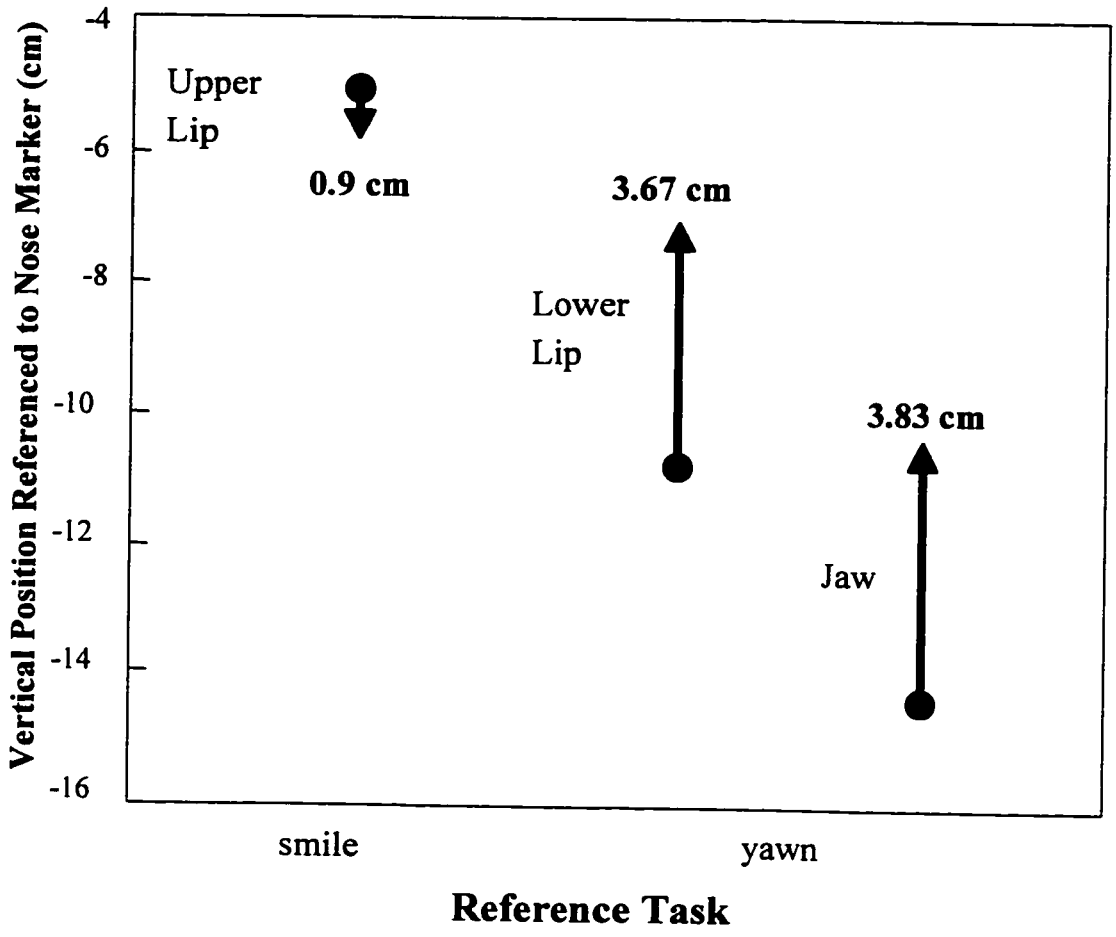
Step 4. Calculate each articulator's relative contribution to oral closure: (e.g., $\%UL_{co}$)

Finally, to calculate each articulator's relative *contribution to oral closure*, each value calculated in Step 3 was individually divided by the summed positions of upper lip, lower lip, and jaw (also calculated individually in Step 3). See Equation 2.

Equation (2)

$$\%UL_{co} = UL_{co} \div (UL_{co} + LL_{co} + J_{co}) \times 100$$

This transformation was performed on each syllable. Figure 3 demonstrates graphically how each articulators relative *contribution to oral closure* was calculated. This technique had several advantages over more traditional measures of movement displacement because (1) it minimizes jaw movement variability related to



Calculations of Relative Contribution to Oral Closure:

$\%UL_{co} = UL_{co} / (UL_{co} + LL_{co} + J_{co}) \times 100$
$= 0.9cm / (0.9cm + 3.67cm + 3.83cm) \times 100 = \mathbf{11\%}$

$\%LL_{co} = LL_{co} / (UL_{co} + LL_{co} + J_{co}) \times 100$
$= 3.67cm / (0.9cm + 3.67cm + 3.83cm) \times 100 = \mathbf{44\ \%}$

$\%J_{co} = J_{co} / (UL_{co} + LL_{co} + J_{co}) \times 100$
$=3.83cm / (0.9cm + 3.67cm + 3.83cm) \times 100 = \mathbf{46\%}$

Figure 3. Calculation of relative contribution to oral closure. The length of each vector corresponds to each articulator's contribution to closing the oral aperture during oral closure. The end of the arrow represents the position of that articulator during oral closure. The circle represents the position of that articulator during its open position (maximum performance task). For each syllable, each articulator's contribution was computed by referencing its position during oral closure to its maximum opening position. To calculate relative contribution to oral closure, each articulators value was divided by the sum of UL, LL and J values for each syllable and multiplied by 100.

vowel context (Sussman et al., 1973), and (2) it did not require precise identification of onset and offset of each articulatory gesture, which can be unreliable in the irregular movement traces exhibited in young children

Statistical Treatment

Developmental changes in upper lip, lower lip, and jaw coordination for speech were characterized by three coordinative indices: maximum or minimum *coefficient*, *lag to coefficient*, and *relative contribution to oral closure*. Before collapsing the data, *lags* were converted to absolute values because their direction was not of interest in the present design. Developmental trends were examined by computing the average of each coordinative index for each subject. The subjects' averages were combined in each age group and subjected to a three-way ANOVA (gender x age x articulator pair). Because the coordinative indexes are expected to vary with age and/or gender, multiple pairwise comparisons were performed on all significant main effects and interactions using the Fisher LSD method (alpha level = .05).

Additionally, the data from the adult and six-year-old group were subjected to a three-way ANOVA (gender x age x phoneme) to test for potential phoneme effects. This analysis was restricted to these groups because the one- and two-year-old subjects did not produce all the phonemes.

Reliability of Measurement

One subject in each group was selected randomly for the reliability analysis. The same experimenter remeasured all the utterances produced by these subjects for the three coordinative indices, which together comprised approximately 10% of the entire set. Test-retest reliability was assessed by calculating the Pearson product moment correlation and average deviation between the first and second measurement for each coordinative index (i.e., *coefficient*, *lag*, and percent *contribution to oral closure*). Table 2 reports average differences between the first and second measurements for *coefficients* and *lags*. The average absolute difference between first and second measurements of *coefficient* and *lag* was .012 and .003 s, which was acceptable for the present analysis. It was anticipated that measurement reliability would be worse for the younger subjects because of increased instability in their movement traces leading to increased difficulty in determining token boundaries. Contrary to this expectation, the difference between measurements of *coefficients* and *lags* tended to be small across all groups, although there was a slight decrease in average deviation between measurements with age for both *coefficient* and *lag* values. Pearson product moment correlations between the first and second measurement for each of the three indices ranged from 0.96 to 0.99, indicating that the difference between the two measurements was negligible.

Measures of percent *contribution to oral closure* relied heavily on computer algorithms and required that the experimenter only grossly define the syllable boundary.

This measure was reproducible with 100% accuracy. Generally, the implementation of computer algorithms wherever possible enhanced measurement reliability.

Table 2. Test-retest reliability for measurement of coefficient and lag from crosscorrelation analysis.

Age Group	Average Deviation	
	Coefficient (r)	Lag (s)
One	0.023	0.0074
Two	0.010	0.0023
Six	0.013	0.0017
Adult	0.003	0.0004
Average for Study	0.012	0.0030

RESULTS

The primary purpose of the present study was to provide a description of developmental changes in lip and jaw coordination during production of syllables containing bilabial consonants. Developmental changes in articulatory coordination were characterized by two distinct analyses examining (1) the strength of spatial and temporal coupling among the upper lip, lower lip, and jaw movements, and (2) the relative contribution of each articulator for oral closure. Beyond providing a general description of age-related changes in coordination, these data were evaluated to elucidate the coordinative constraints on early speech motor control.

Data Corpus

A total of 1098 utterances were analyzed, including: 54 from the one-year-olds, 256 from the two-year-olds, 429 from the six-year-olds, 422 from the adults. The number of bilabial utterances produced by each subject varied considerably depending on their age (see Figure 4). All utterances were CVCV combinations produced in isolation with the exception of nine from the one-year-old group. Five of these utterances were VCV (or CV with mouth in initially in open position) combinations and four were CVCV combinations extracted from continuous canonical babble.

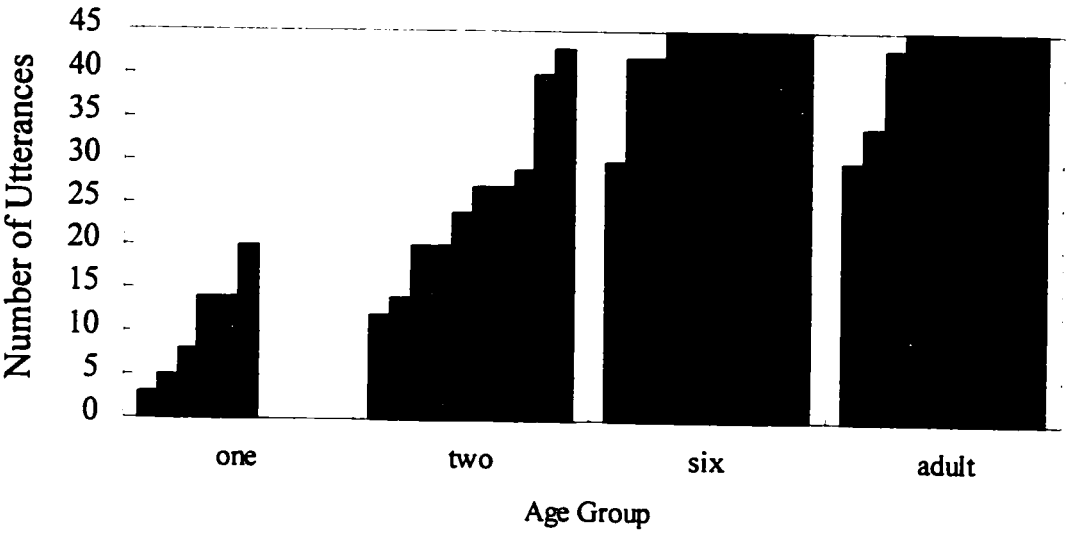


Figure 4. Number of utterances per subject in each age group. Each bar represents the number of utterances for one subject. The data are ordered from smallest to largest number of utterances.

Phoneme Effects

Phonemes were not evenly represented among the age groups. None of the one-year-olds produced utterances that contained a [p] exemplar. In addition, two of the two-year-olds did not produce examples of the /p/, and half of the children in this group produced five or fewer of these utterances. These findings were anticipated, as the voiceless bilabial stop (i.e., /p/) emerges around age two (Stoel-Gammon, 1985).

This imbalance in the data set required evaluation of potential phoneme effects on the three coordinative indices (i.e., *coefficient*, *lag*, and *contribution to oral closure*).

Only the data from the six-year-olds and adults could be included in this analysis because only these subjects consistently produce all three bilabials (i.e., /b/, /p/, /m/). The results of a three-way analysis of variance on repeated measures (phoneme x pair x gender) indicated that there were no statistically significant phoneme effects for *coefficient*, *lag*, or *contribution to oral closure*. Based on these results, the data for each subject were collapsed across phonemes to yield a single average for each coordinative index.

Qualitative Observations

Figure 5 includes a kinematic record from one subject from each age group producing "baba." For ease of interpretation, each kinematic signal was centered about its mean and the upper lip signal was inverted. While generalizations cannot be made based on single productions, these examples illustrate differences in the coordinative organization exhibited among age groups that are supported by findings revealed in the quantitative analyses. Adult subjects uniformly produced the movement sequences with high levels of inter-articulator coupling. Movement trajectories in these subjects were characterized by a predominant single rising and falling pattern for each syllable.

In contrast to the adult pattern, one-year-old children typically exhibited pronounced jaw movements accompanied by excessive compression of lip tissues during oral closure. As displayed in Figure 5, this compression was associated with oppositional movements (180 degrees out of phase) between the lips and jaw. These

deflections in the movement trajectories at oral closure were much larger in one-year-olds than in any other age group. In two-year-old subjects, jaw movement appeared to decrease as upper and lower lip movement increased. The upper and lower lip movement trajectories were often similar in form (e.g., "mirror movements") and frequently characterized by a single rising-falling across both syllables. The movement patterns of six-year-olds were similar to those of adults, but were generally more variable.

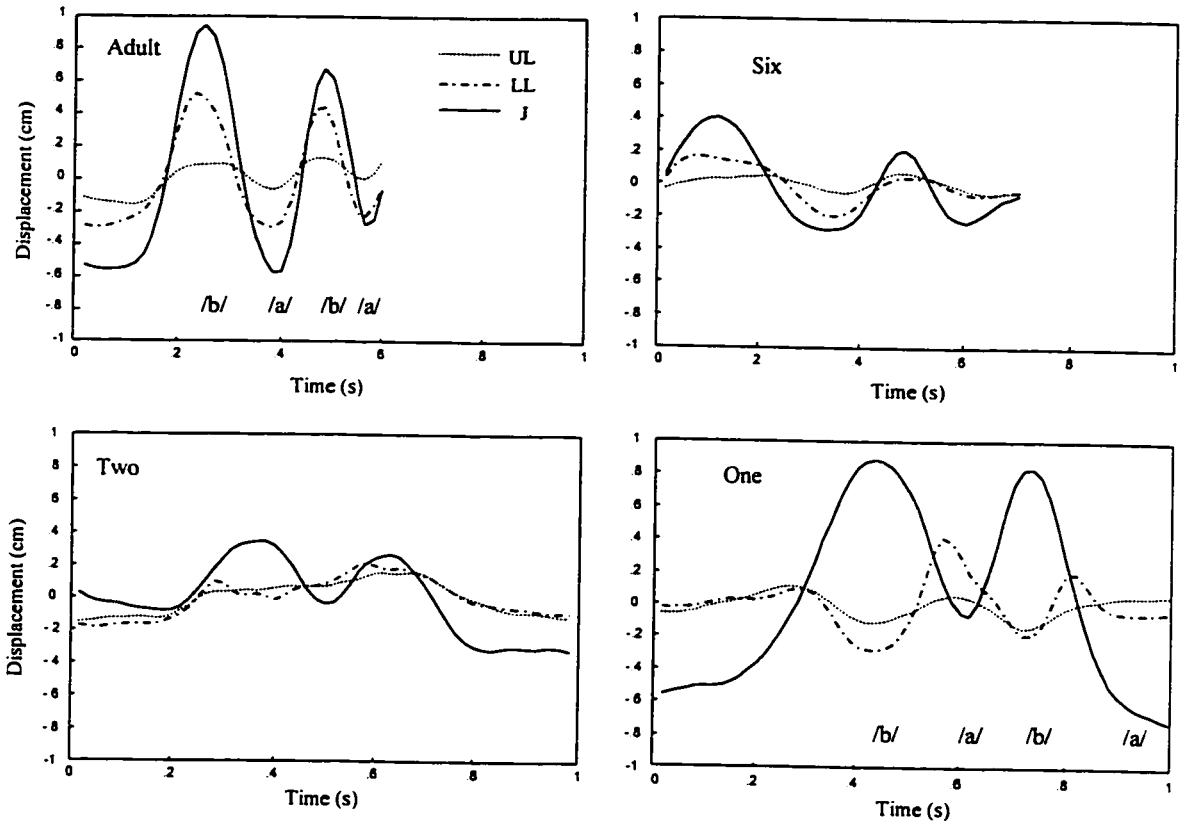


Figure 5. Representative kinematic records from a subject in each age group based on a single trial. Traces are centered about their means and the upper lip is inverted for ease of interpretation.

Crosscorrelation Analysis

Developmental changes in spatial and temporal coupling among upper lip, lower lip, and jaw were quantified by subjecting the movement traces to a pairwise crosscorrelation analysis. Peak *coefficients* were abstracted from each crosscorrelation function and were interpreted to reflect spatial coupling between articulatory pairs. In addition, the *lag of the peak coefficient* was recorded and indicated the degree of movement synchrony between articulatory pairs. The peak *coefficients* and *lag* values exhibited by the youngest subjects were of special interest. If spatial and temporal coupling are low (low *coefficients* and high *lag* values) in these groups, we might conclude that movement synergies are not features of early speech motor organization.

Spatial Coupling

Spatial coupling increased significantly with age [$F(3,84) = 28.41, p < .001$]. Figure 6 shows the mean peak *coefficients* values obtained at each age for each articulator pair. The mean *coefficients* and standard deviations are presented in Table 3. Gender was the only main effect that was not significant. The only significant interaction was articulator pair by age [$F(6,84) = 3.0, p < .01$].

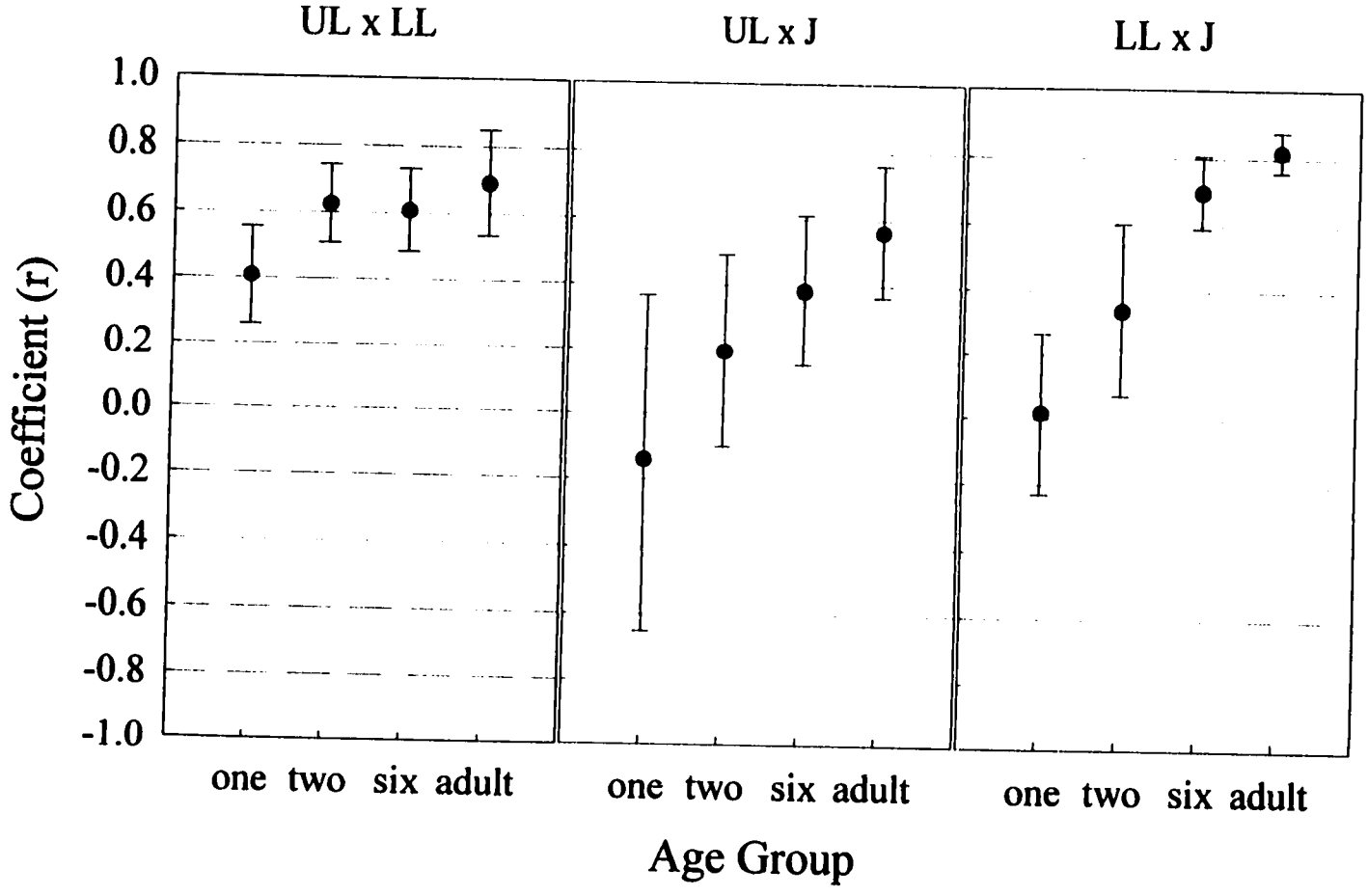


Figure 6. Average coefficients and standard deviations obtained from pairwise crosscorrelations for upper lip and lower lip (UL x LL), upper lip and jaw (UL x J), and lower lip x jaw (LL x J) by age. Error bars represent average standard deviation between subjects in each age group.

Table 3. Mean coefficient (r) values for articulator pairs across age groups. Standard deviations of the means are in the parentheses.

	Pair		
Age Group	UL x LL	UL x J	LL x J
One	0.41 (0.15)	-0.18 (0.55)	0.00 (0.27)
Two	0.63 (0.12)	0.20 (0.30)	0.34 (0.26)
Six	0.61 (0.12)	0.38 (0.23)	0.73 (0.09)
Adult	0.69 (0.16)	0.56 (0.20)	0.82 (0.06)

Note: (UL = upper lip, LL = Lower Lip, and J = Jaw)

Multiple comparisons of the interaction of articulator pair and age revealed different developmental progressions for UL x LL, UL x J, and LL x J[1]. Only the one-year-olds exhibited significantly less UL x LL coupling than the adults. As demonstrated in Figure 6, UL x LL coupling was relatively high for the younger age groups. In contrast, UL x J and LL x J *coefficients* were centered near zero at age one year, reflecting the unpredictability of coordination between these articulators. Coupling between these articulators increased gradually with age, although several

---

[1] All multiple pairwise comparisons were performed using the Fisher LSD method (alpha level = .05).

adjacent age groups did not differ significantly in these measures. Specifically, six-year-olds did not differ significantly from adults for LL x J, and from two-year-olds and adults for UL x J.

Age-related coordinative characteristics were revealed by differences in the relative degree of spatial coupling among articulator pairs within each age group. Table 4 highlights the age-related changes in spatial coupling that occurred for all three articulator pairs based on the results of the multiple comparisons analysis. One- and two-year-old children exhibited greater spatial coupling between the lips than between the lips and jaw (i.e., UL x LL > UL x J and LL x J). In contrast, spatial coupling between the UL x LL was not significantly different from that of UL x LL and LL x J for adult subjects. In six-year-olds and adults, UL x J coupling was lower than for LL x J coupling (i.e., UL x J < LL x J).

Table 4. Results of pairwise comparisons of age by articulatory pair.

	Age			
Comparison	One	Two	Six	Adult
UL x LL vs. UL x J	0.4 > -0.2	0.6 > 0.2	0.6 > 0.4	0.7 ≈ 0.6
UL x LL vs. LL x J	0.4 > 0.0	0.6 > 0.3	0.6 ≈ 0.7	0.7 ≈ 0.8
LL x J vs. UL x J	0.0 ≈ -0.2	0.3 ≈ 0.2	0.7 > 0.4	0.8 > 0.6

Movement Synchrony

Overall, the developmental trends in movement synchrony paralleled those observed for spatial coupling. Lags between articulatory pairs, as measured by the *lag-to-peak coefficient*, became shorter with age [$F(3,84) = 5.43, p < .01$]. Figure 7 displays the averages and standard deviations of *lag* values across the age groups for UL x LL, UL x J, and LL x J, respectively. The mean values and standard deviations are presented in Table 5. Temporal resolution was ± 8.8 ms, which was determined by the videorecording rate (i.e., 60 frames per second). Similar to the coefficient analysis, gender was the only main effect that did not exhibit statistically significant differences. No interactions between variables were observed.

Table 5. Mean absolute lags (in milliseconds) for articulator pairs as a function of age. Standard deviations of the means are in the parentheses.

	Pair		
Age	UL x LL	UL x J	LL x J
One	18 (6)	29 (31)	23 (12)
Two	12 (6)	21 (6)	18 (7)
Six	14 (5)	20 (5)	11 (4)
Adult	14 (2)	18 (8)	8 (4)

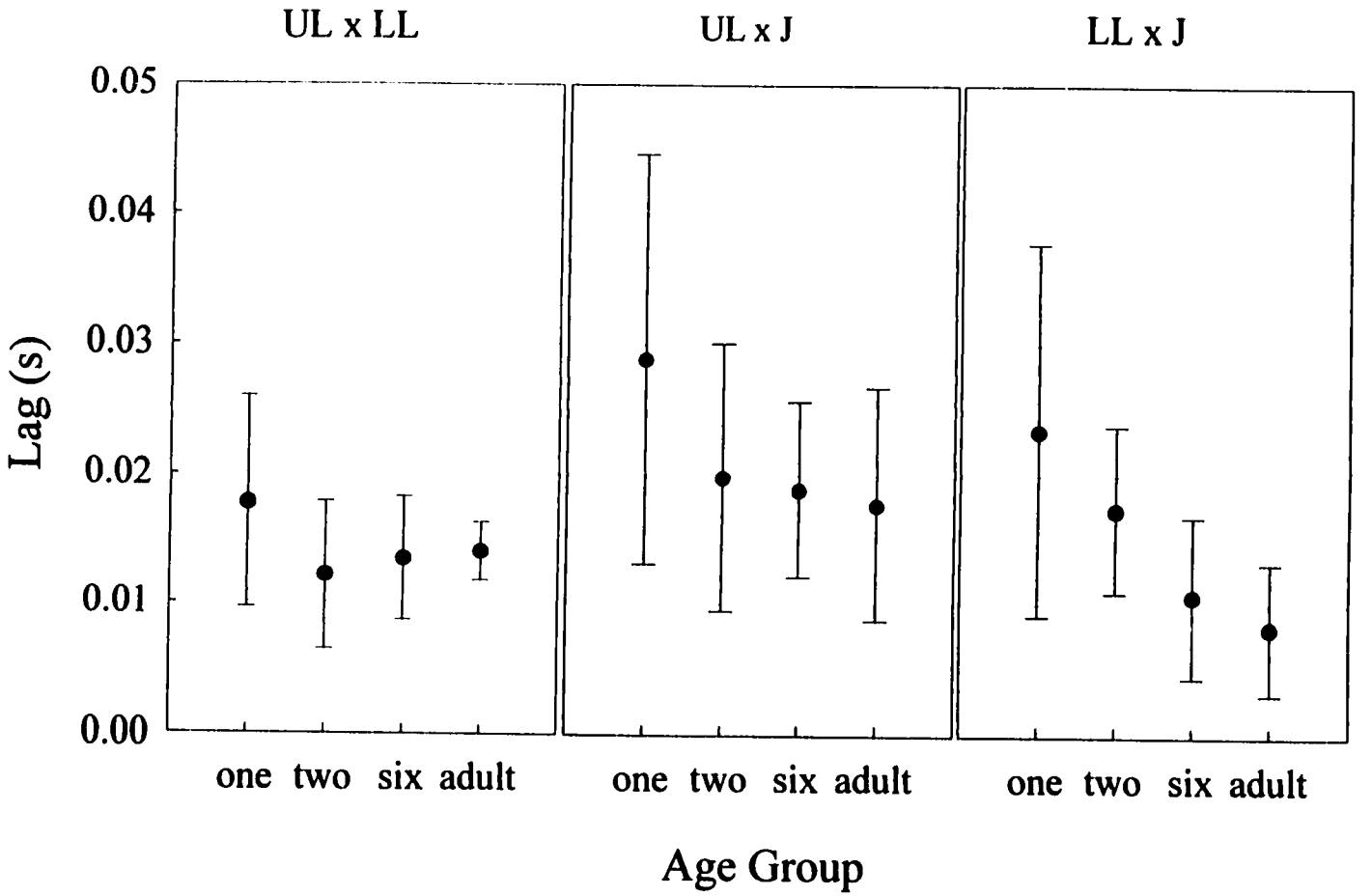


Figure7. Average absolute lag values and standard deviations obtained from pairwise crosscorrelations for UL x LL, UL x J, and LL x J for each age group. Smaller lag values represent increased interarticulator synchrony. Error bars represent average standard deviation between subjects in each age group.

Overall, age related differences in *lag* values were relatively small. Multiple comparisons revealed that there was no age effect for UL x LL. In contrast, UL x J and LL x J articulatory *lags* decreased with age. The multiple comparisons revealed longer *lags* for one-year-olds than for six-year-olds and adults. In addition, LL x J *lags* were significantly longer for two-year-olds than for adults.

Movement synchrony differed among articulator pairs [$F(2, 84) = 7.54$, $p < .001$]. As in the coefficient analysis, age-related coordinative preferences were revealed by differences in the relative degree of synchrony among articulator pairs within each age group. Table 6 highlights the age-related changes in movement synchrony that occurred for all three articulator pairs based on the results of the multiple comparisons analysis. Both one- and two-year-old subjects exhibited greater synchrony for UL x LL than for UL x J and LL x J (i.e., UL x LL > UL x J and LL x J). In contrast,

Table 6. Results of pairwise comparisons for articulator pair for lag values (ms). Smaller values indicate greater movement synchrony.

	Age			
Comparison	One	Two	Six	Adult
UL x LL vs. UL x J	18 < 29	12 < 21	14 ≈ 20	14 ≈ 18
UL x LL vs. LL x J	18 ≈ 23	12 ≈ 18	14 ≈ 11	14 ≈ 8
LL x J vs. UL x J	23 ≈ 29	18 ≈ 21	11 < 20	8 < 18

six-year-old and adult subjects exhibited significantly greater movement synchrony for UL x LL and LL x J than for UL x J (i.e., UL x LL and LL x J > UL x J).

Taken together, these results demonstrate two distinct processes in the development of lip and jaw coordination for speech. First, young children exhibit relatively high levels of coupling for UL and LL. The magnitude of this coupling was surprising, especially when compared to the poor coordination of lip and jaw movement exhibited at these young ages. These findings are consistent with the possibility of linked control between these articulators. In contrast to the changes in UL x LL coordination, coupling of lip and jaw pairs (i.e., UL x J and LL x J) gradually improved with age. This sequence reflected improvements in inter-articulatory coordination in the absence of early movement synergies in lip and jaw articulator pairs.

The relatively high degree of spatial and temporal coupling between upper and lower lips observed in one-year-olds might suggest synchronous neural drive. However, as demonstrated in Figure 5, the jaw produced these passive displacements of the lips which probably yielded *coefficients* that were spuriously high and *lags* that were spuriously low at this age.

Contribution to Oral Closure

Developmental changes in each articulator's *contribution to oral closure* were examined. As in the crosscorrelational analysis, the coordinative patterns of the young children were of special interest. Integration of motor capabilities in speech motor

development would be supported if the jaw was shown to contribute most to oral closure in early speech with the lips becoming more involved with age. This sequence is consistent with a hypothesis advanced by MacNeilage and Davis (1990a, 1990b). On the other hand, differentiation would be supported by increases in the independent control of individual articulators. In the present design, this shift in coordination may be accompanied by a decrease in an articulator's *contribution to oral closure* as it becomes disassociated from tightly coupled ensemble.

Several developmental changes in labiomandibular coordination for oral closure were exhibited. The percentage *contribution to oral closure* differed significantly for each articulator with age [Articulator x Age, $F(6, 92) = 11.34, p < .001$]. Figure 8 displays the means and standard deviations by age group for UL, LL, and J. The mean values and standard deviations are presented in Table 7. There were no statistically significant gender effects for these measures.

Table 7. Mean percentage contribution to oral closure as a function of age and articulator. Standard deviations of the means are in the parentheses.

	Articulator		
Age	UL	LL	J
One	8 (5)	26 (13)	66 (15)
Two	16 (10)	34 (9)	50 (8)
Six	9 (4)	44 (6)	47 (4)
Adult	8 (3)	46 (3)	46 (2)

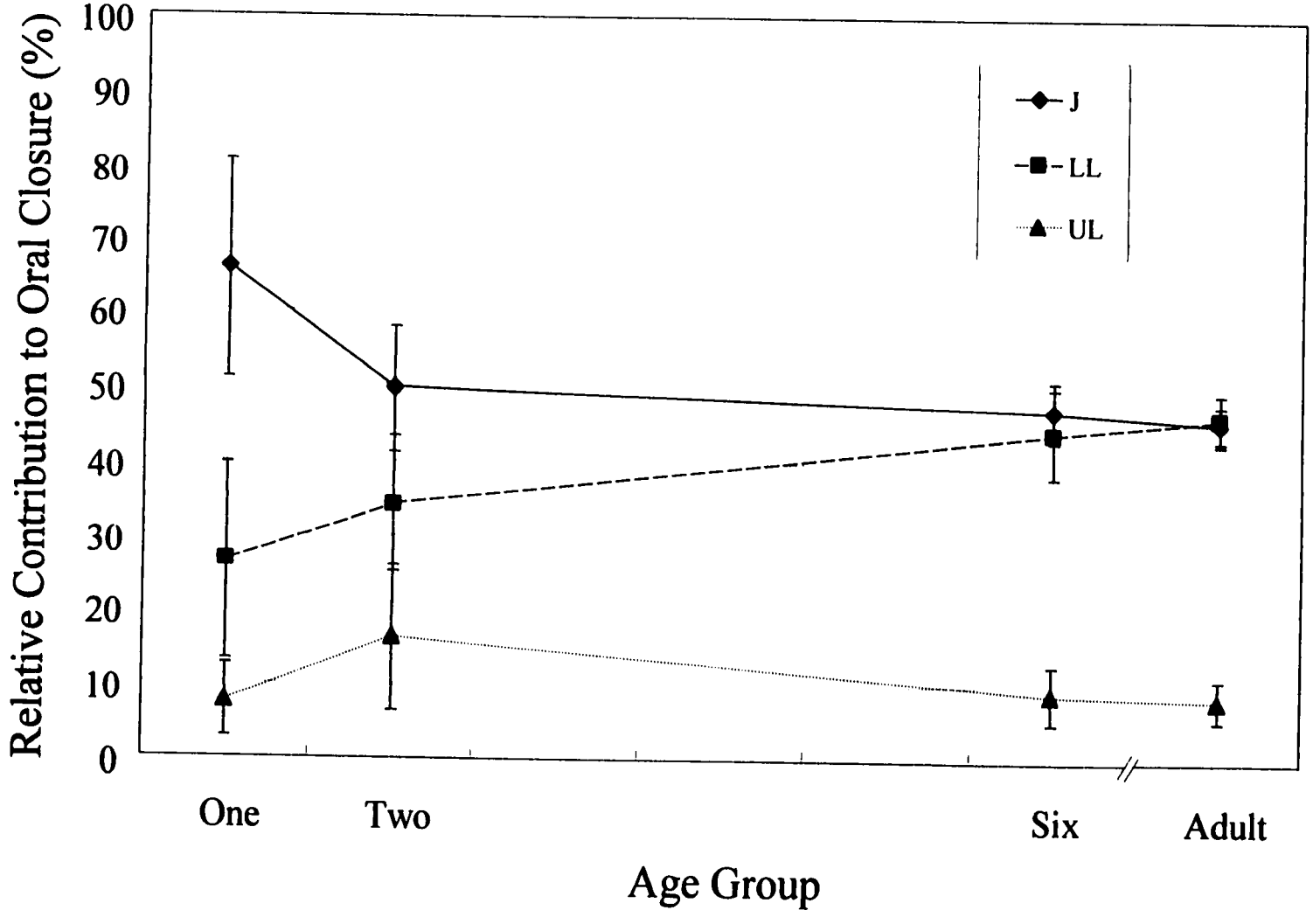


Figure 8. Relative contribution to oral closure for each articulator by age. Error bars represent average standard deviation between subjects in each age group.

Multiple comparisons of the articulator by age interaction revealed specific age-related changes in the relative *contribution* of each articulator. *Contribution* of J to oral closure decreased significantly between one and two years of age. This decrease was associated with a significant increase in *contribution* from UL and LL in two-year-old subjects. The *contribution* of UL decreased after age two, but LL *contribution* continued to increase. The adult and six-year-old subjects exhibited a similar pattern for oral closure. For these groups, LL and J *contributions* were comparable, and the involvement from UL was small in comparison to either of these articulators.

Age-related coordinative preferences in articulatory movement were revealed by differences in the *contribution to oral closure* within each age group. Table 8 highlights the age-related changes that occurred for all three articulators based on the results of the multiple comparisons analysis. In one- and two-year-old children, jaw movement contributed most to oral closure, followed by LL, then UL. In contrast, six-year-old children and adults used the LL and J equally to close the oral aperture, and the UL contributed significantly less than either of these articulators.

Table 8. Results of pairwise comparisons for articulator for relative contribution to oral closure (%).

Comparison	**Age**			
	One	Two	Six	Adult
J vs. LL	66 > 26	50 > 34	47 ≈ 44	46 ≈ 46
J vs. UL	66 > 8	50 > 16	47 > 9	46 > 8
LL vs. UL	26 > 8	34 > 16	44 > 9	46 > 8

In summary, these findings suggest that an essential aspect of speech motor development for these movement sequences is increasing lower lip independence of movement. The predominant role of the jaw in early articulation and the gradual increase in lower lip *contribution* suggests an integrative process operating in the development of speech motor control. On the other hand, the high degree of UL x LL coupling at two years of age, and the decrease in the upper lip's involvement between two and six years suggests a differentiating process in the development of speech motor control.

DISCUSSION

The Development of Articulatory Coordination: Integration, differentiation, and refinement

The objective of this investigation was to describe age-related changes in lip and jaw coordination for bilabial closure during speech. Movement of the upper lip, lower lip, and jaw were recorded during the production of syllables containing bilabial consonants in four age groups (i.e., one-, two-, and six-years, and young adult). The results revealed that the coordinative organization of articulatory gestures shifts dramatically during the first several years of life and continues to be refined past age six.

The findings might be interpreted to support three primary phases in the development of lip and jaw coordination for speech. Each phase is characterized by processes in the development of coordination that exhibit integration, differentiation, and refinement. Although distinct developmental changes occurred at each hypothetical phase, they probably overlap considerably or coexist. An integrative process in the organization of early speech might entail the sequential development of articulatory control, which in the present experiment reflects the assimilation of lip movement into an established movement pattern predominated by jaw movement. Alternatively, differentiation of articulatory coordination of upper and lower lips may be revealed as a unitary assembly in early speech, with motor development requiring increasingly independent control of these anatomically distinct structures. Finally, refinement of

motor organization may be seen in the gradual refinement of spatial and temporal control even after the age of six years. The coordinative constraints imposed by each of these mechanisms will have predictable consequences for phonologic development.

The Mature Pattern

Movement sequences by adults exhibited several features characteristic of skilled movement (see Figure 5). First, the articulators exhibited near synchronous movements and well-formed movement trajectories (i.e., characterized by a single, predominant, rising and falling pattern for each CV syllable). These features yielded high correlations and low *lags* in the present analysis, and were consistent with previous descriptions of adult articulatory control for bilabial stops (Gracco, 1988; Löfqvist & Gracco, 1997). Additionally, the lower lip and jaw were comparably involved in closing the oral aperture in adult subjects, and the upper lip contributed significantly less than either of these articulators.

One-Year-Olds

The results of the oral closure and qualitative analysis suggested that the infant exerts greater control over the jaw than the lips for speech. The jaw contributed to closing the oral aperture twice as much as the lower lip did at this age. Although the spatial characteristics of jaw movement by one-year-olds were similar to those exhibited by adults, lip movements were characterized by reduced extent of movement and frequent reversals in direction unrelated to syllable boundaries. A hypothesis of delayed achievement of lip control is controversial given the wide array of lip configurations

observed during early oromotor behaviors (e.g., vocalizations, facial expressions, sucking, and chewing). However, coordination of spontaneous movements or extant behaviors may not reveal the underlying control structures for novel, more complex tasks (see Moore and Ruark, 1996 for discussion on the role of extant behaviors on speech motor development).

The developmental sequence of integration is supported in the present investigation by the augmentation of oral closure by lip movement with established jaw movement at age two. This sequence is consistent with the earlier appearance of jaw control observed in prenatal studies of orofacial movement (Humphrey, 1970; 1971). The sequential development of articulatory control for speech has received some theoretical consideration (MacNeilage & Davis, 1990a; 1990b; Nittrouer, 1993), although physiologic data are lacking.

The present observations of the predominance of jaw movement during early speech received further support from studies showing the coordinative organization for jaw control for speech to be adult-like by 15 months of age (Moore and Ruark, 1996). These observations provide some physiologic support for the model of speech development proposed by MacNeilage and Davis (1990a; 1990b) who suggest that the earliest articulations are predominated by jaw movements with little or no contribution from the lips.

Nittrouer (1993, 1995) suggested that children master some speech gestures earlier than other gestures. She derives her use of the term "gestures" from Browman and Goldstein's (1986) definition: " an articulatory structure or, more commonly, a

constellation of structures working together to achieve a linguistically significant pattern of action (i.e., the formation and release of a vocal-tract constriction)." (Nittrouer, 1995, p. 520). The present results suggest that in early development, the formation of articulatory gestures must operate within the coordinative constraints imposed on individual articulators by the motor system. Therefore, one important step in accounting for the emergence of speech gestures will be the description of the developmental sequence of motor control for individual articulators.

The low *coefficient* and high *lag* values obtained for lip and jaw pairs suggest that the young child is not endowed with early movement synergies (e.g., a widely distributed central motor program) among these articulators. However, to definitively rule out the existence of early movement synergies would require observation of lip muscle activation patterns. Specifically, the lip musculature may be activated during oral closure, but the underlying motor organization may not be sufficient to produce the trajectories characteristic for these movement sequences. The one-year-old subjects exhibited a variety of lip configurations for oral closure within a single data collection session. In some instances, the lips appeared to be at rest, but in others, they were tense with the lower lip elevated and the upper lip depressed. These varied patterns yielded a wide range of *coefficients* in the crosscorrelation analysis between and within subjects.

Although the infants generally produced well-formed jaw movements, the lips were often 180 degrees out of phase with the jaw, yielding negative *coefficients* in the crosscorrelation analysis. The amount of lower lip and upper lip deflection (i.e., interlabial compression) was much greater in this group than in any other. It is possible

that this pattern of interlabial compression reflects the generation of poorly controlled mandibular movement. Excessive displacement of movement in early speech may be related to a more general characteristic of immature motor control as, for instance, overshooting of the hand and arm is a feature of immature grasping (Jeannerod, 1988). This notion coincides with Bernstein's (1996) suggestion that one essential aspect in motor control development is the reduction of superfluous movements.

Kent (1992) suggested that early articulatory movements are rapid and ballistic (i.e., movements that are characterized by high velocity, and exhibit rapid acceleration and deceleration). He differentiated these types of movements from those produced with constant velocity over a relatively long duration, which characterize, for instance, the /w/. A limited ability to regulate jaw control may explain why complete closing and opening gestures are so common in early vocalizations (Locke, 1983).

Two-Year-Olds

In two-year-old subjects, the movement of the jaw assimilated with movement of the lips. Upper lip and lower lip *contribution to oral closure* became more prominent. Further support for the hypothesis of sequential development of articulatory control comes from the observations that the two-year-old children tended to exhibit poorly defined lip movements across both syllables (see Figure 5). In some instances, upper and lower lip movements were characterized by a single rising and falling trajectory across both syllables. In contrast were the relatively well-defined movement trajectories exhibited in the jaw. Another distinguishing feature of this age group's productions was

the increased *contribution* of upper lip to oral closure. This observation parallels previous reports of decreases in the relative contribution of the upper lip to labial separation for vowel opening with age (Watkin & Fromm, 1978).

Movement of upper and lower lips appeared to be tightly coupled at age two. As a group, two-year-old children exhibited rigid spatial and temporal coupling of upper and lower lips in comparison to that of lip and jaw coupling. The functional significance of these high *coefficients* and short *lags* may vary depending on the age of the subject. In adult speakers, the upper lip, lower lip, and jaw are capable of producing highly independent movements. The high degree of coupling and synchrony of articulatory movement observed in adults, therefore, reflects highly specified, coordinated movement. In young speakers, a comparably high degree of coupling may indicate a lack of coordinative plasticity. Further support for the suggestion that these anatomically distinct structures may behave as a unit comes from the parallel increases in upper and lower lip *contribution to oral closure* at age two years (see Figure 8). Moreover, as demonstrated in Figure 5, lip movement trajectories at this age could be remarkably similar in shape and amplitude, especially when compared to the movement trajectories of the jaw.

Linked upper lip and lower lip control may be related to a more general feature of motor skill development that has appeared in the literature as *associative movements* (Todor & Lazarus, 1986) and *motor overflow* (Cohen, Taft, Mahadeviah, & Birch, 1967). Instances of linked motor control are commonly observed in early motor development where symmetrical muscles (homologous) and asymmetrical

(heterologous) muscles tend to produce associative movements (Lazarus & Todor, 1987). Dramatic examples of associative movements occur when movement can be observed in the resting contralateral extremity while its counterpart is performing a motor act (Cohen et al., 1967). Associative movements have been observed at various anatomical sites (e.g., ears, fingers, and limbs) and levels of organization (e.g., limbs, motor units; see Provins, 1997). These movements have been reported to decrease with age and with differential practice (Provins, 1997). In some clinical tests, the presence of such movements has been used as an indicator of neuromotor immaturity or motor impairment (Connolly & Stratton, 1968).

The putative coupling of lip movement is consistent with progressive differentiation with development. Development of lip control for speech using this mechanism would coincide with gains in the selection of discrete neuromuscular elements of the labial apparatus. Differentiation was supported by the present results in two-year-olds which revealed increased levels of spatial and temporal coupling for upper and lower lip, and elevated involvement of upper lip *contribution to oral closure*. The subsequent decrease in upper lip's *contribution to oral closure* may reflect its disassociation from lower lip. Linkage of upper and lower lip control may have neuroanatomic (i.e., shared neural structures), functional (i.e., simplification control strategy) or experiential (i.e., negative transfer of learning) bases. These possibilities will be considered further in the following sections.

A more rigorous test of differentiation requires the observation of increased upper and lower lip coupling in speech tasks that specify independent control of those

structures (e.g., as during the pronunciation of /f/ in "food"). Ruark and Moore (1997) studied upper and lower lip coupling directly using electromyography in two-year-old children during production of a variety of nonspeech (e.g., lip pursing, chewing) and speech tasks (e.g., syllables containing bilabials). These investigators observed task-specific patterns of labial coordination across the varied tasks (Ruark & Moore, 1997), a finding that might be interpreted as failing to support the dependence of upper lip and lower lip motor control. However, examination of their results reveals a relatively high-positive degree of coupling for all the speech tasks performed (i.e., average Fisher's z scores approximately ranging from .50 to .77). Thus, the interdependence of upper and lower lip control remains viable in support of a differentiating hypothesis for addressing the coordinative patterns exhibited by the two-year-old subjects.

Six-Year-Olds

At six years of age, lip and jaw coordination for the movement sequences examined continued to undergo refinement prior to reaching the mature form. Between ages two and six years, lip and jaw spatiotemporal coupling continued to increase. The qualitative observations revealed that movement patterns exhibited by six-year-olds were similar to those of adults, but were often more variable. Generally, spatial and temporal coupling in six-year-olds were decreased in comparison to those observed in adults; however, differences between these groups were small and did not reach statistical significance. A significant decrease in upper lip involvement occurred between ages two and six years in the oral closure analysis. In contrast, the involvement of upper lip, lower

lip, and jaw for oral closure was similar between six-year-old and adult subjects. Together these findings give the impression that the period between six-years and adult reflects continued refinement of movement control and optimization of coordination. These findings parallel the continuous refinement of speech performance from mid-childhood to adolescence, including reduced token-to-token variability (Sharkey & Folkins, 1985; Smith & Goffman, 1998) and shortened segment durations (Kent & Forner, 1980).

In summary, the present experiment has revealed age-related shifts in the coordinative organization of lip and jaw movement for speech. Based on these observations a developmental sequence of labiomandibular coordination for speech is proposed to involve integration, differentiation, and refinement. The first phase is characterized by the predominance of jaw control. The second phase involves gaining independent control of the upper and lower lip. The final phase is characterized by optimization of the established coordinative pattern. This characterization of speech motor development should be viewed as speculative; nonetheless, a number of hypotheses can be generated regarding the potential influence of this sequence on phonologic development.

Mechanisms: Data, theory, and speculation

The observed sequences in labiomandibular coordination support the presence of integration, differentiation, and refinement in the development of speech motor control. A deeper understanding of the present results requires a consideration of developmental

changes in articulatory motor pathways and biomechanics. However, before these changes can be attributed to processes related to development, movement sequences exhibited during motor learning that are independent of maturation also need to be considered.

Developmental Sequences and Changes in Neural Substrates

Integration and differentiation in early development of motor control may reflect differential development of neural centers. Myelination (Schuster & Ashburn, 1992) and early motor control (Stallings, 1973) proceed cephalocaudally and proximodistally, processes that are also reflected in early motor skill development. Moreover, studies of non-human animals have linked the emergence of oromotor skills such as sucking, biting, and chewing to maturational changes in the neural populations mediating orofacial movements (Herring & Wineski, 1986; Iriki, et al., 1988; Kubota, Narita, Ohkubo, Shibania, Nagae, Kubota, Odagiri, & Kawamoto, 1988; Lakars & Herring, 1980).

Several investigators have suggested that the location of a neural center is a good predictor of when it matures. Jeannerod (1988) hypothesized that the early appearance of proximal control in the arm is associated with an inherent neural organization where proximal motor pathways have unique locations from those controlling distal segments. In addition, Kubota and colleagues (1988) have provided compelling evidence that sucking appears earlier than biting because facial motor pathways mature (e.g., myelination and cell area) prior to trigeminal motor pathways in mice. Thus, in early

development, the emergence of some oromotor behaviors appears to be closely tied to the maturation of their supporting neural pathways.

These findings support the suggestion that neural maturation differentially constrains articulatory control during early development, limiting the movement solutions available to the infant and yielding the process of integration observed in the present study. Specifically, the results of the present experiment support the earlier appearance of jaw control compared to that of lip for speech. This developmental sequence may arise from the fact that the motor neuron pools mediating these articulators have distinct locations in the brainstem. This possibility would receive empirical support if the motor pathways for speech in the trigeminal nucleus mature earlier than those of the facial nucleus.

A number of characteristics of the immature neuromotor system may limit independent control of upper and lower lip. Although there appear to be distinct sources of neural input to the upper and lower lips in mature speakers (Abbs & Gracco, 1984; Goffman & Smith, 1994; Smith, 1992; Wohlert & Goffman, 1994), the immature neuromotor system may not be endowed with this fine level of organization. That is, the subnuclei in the facial motor nucleus controlling upper and lower lip may be functionally indistinguishable in early development. This suggestion parallels the increases in specificity of perioral afferents with maturation observed by Barlow and colleagues (1993) and is consistent with the suggestion by Edelman and colleagues that neuronal selection is experience-driven (Edelman, 1987; Sporns & Edelman, 1993). Edelman's theory of neuronal group selection states that the formation of distinct

neuronal pathways requires specific experiences. Accordingly, speech maturation may require experience-related differentiation of subpopulations within the facial nucleus or higher neural centers.

Finally, a number of organizational features in efferent motor pathways crucial for skilled movement are absent in the young mammal, a fact which may underlie the immature motor patterns observed in the younger subjects in this study. In the immature motor system, there is an absence of the morphological diversity characteristic of most muscles. Muscle fibers are small (Sato, Mizuno, & Konishi, 1977), more homogeneous in fiber type (Rubenstein & Kelly, 1981), and innervated by multiple neurons (Navarrete & Vrbova, 1983). Together these features limit the graded recruitment of motor units and reduce the variety of movements and coordinative patterns generated.

In summary, several potential neural mechanisms in early oromotor development require consideration: (1) the effective neural centers mediating the articulators may mature at different times (i.e., integration), and (2) some neural centers may be functionally indistinguishable (i.e., differentiation). Because neural mechanisms cannot be identified from behavioral data (i.e., the problem of inverse kinematics), we can only speculate about their existence. Future developments in neural imaging techniques (i.e., functional magnetic resonance imaging) may afford the opportunity to specify the roles of neural maturation and experience in speech motor development.

Developmental Sequences and Changes in Biomechanics and Body Composition

The development of motor control may depend as much on the biomechanical properties of the articulators as on neuromotor development (Thelen, 1995). The neural organization subserving speech must adjust for developmental changes in the biomechanical properties of the articulators, including changes in tendon elasticity, muscle fiber characteristics, and increases in fat, bone, and muscle mass. Recent models of motor skill development have begun to emphasize the effects of such structural changes on early coordination (Thelen, 1995).

The intrinsic properties of body segments (e.g., mass) and environmental loads (e.g., gravity) may rate limit early motor skill development. In a series of experiments Thelen and colleagues (Thelen, 1983; 1986; Thelen, Bradshaw, & Ward, 1981; Thelen & Fisher, 1982) provided compelling evidence for the rate limiting effects of body mass on the development of locomotion. Their findings were in contrast to the long standing view that CNS maturation determines the age of onset for walking. The influence of biomechanics on emerging motor functions may vary considerably depending on the physiologic composition of the components involved and the task requirements (e.g., accuracy, speech, strength, endurance, and flexibility).

The upper lip, lower lip, and jaw exhibit different growth curves. The mandible is proportionately very small in the infant in relation to the skull. Growth of the mandible proceeds logarithmically while its characteristic shape remains relatively

unchanged (Kent & Vorperian, 1995). In contrast, the size, shape, and tissue composition of the lips change most dramatically during the first several years of life. The upper lip may even have a distinct developmental course from that of the lower lip (Kent & Vorperian, 1995).

It is difficult to specify how these anatomical changes along with changes in the biomechanical properties of muscle might differentially affect control of each articulator. There have been few investigations of the biomechanical properties of the articulators in adults (e.g., Abbs, 1973; Baragar & Osborn, 1984; Müller, Milenkovic, & MacLeod, 1984; Sanguineti, Laboissière, & Ostry, 1998), none in the context of speech motor development. It might be anticipated that changes in the absolute or proportional size of the child's articulators would dramatically alter the coordinative organization of speech. The potential for dramatic coordinative changes secondary to anatomical growth has been recognized by proponents of a dynamic systems approach to speech development: "even small changes in anatomy, for example, potentially can engender new attractor states." (Thelen, 1991, p. 343).

Surprisingly, the present findings and the available data do not support the view that changes in the relative size or mass of the lips and jaw alter articulatory control. For example, adults and children (as young as age four) exhibit only small differences in articulatory displacements despite dramatic differences in the sizes of their vocal tract structures (Smith & Gartenberg, 1984; Smith & Goffman, 1998). Although peak displacement was not measured directly in this study, Figure 5 shows that the jaw displacement of the one-year-old subjects was similar in magnitude to that of the adult

subjects. This observation is surprising given the proportionately small size of the mandible at one year of age. This mismatch of observations and expectations regarding biomechanical constraints suggests that additional studies are needed to better understand speech motor development.

Changes in Coordinative Organization Associated with Motor Learning

The observed changes in articulatory coordination may also reflect motor learning independent of maturation. Motor learning exhibits distinct phases (i.e., temporary motor solutions) with the accumulation of practice and experience. The adoption of a specific motor solution will depend on the complexity of the task and its relationship to pre-existing skills. Processes in motor learning such as *sequential motor learning*, *simplification control strategies*, and *negative transfer of learning* may account for some of the coordinative changes observed in the present study.

Sequential Motor Learning

The process of learning a complex motor skill typically requires proficiency with fundamental skills prior to advancing to higher levels of performance (Haywood, 1993). This view implies sequential learning of motor skills and may apply to the developmental sequences observed in the present data set. Specifically, the jaw-predominant gestures, which characterized the one-year-old's utterances, may constitute a fundamental skill underlying speech. This suggestion is consistent with the MacNeilage and Davis (1990a; 1990b; Davis & MacNeilage, 1994; 1995) "Frames, then Content" model of early speech production.

MacNeilage and Davis suggested that this jaw-predominate articulatory pattern has phylogenetic origins, and that it may be adapted from the coordinative framework previously established for chewing. However, the assertion that speech motor control is derived from alimentary functions has not received support by any of a series of recent investigations (see Moore & Ruark, 1996), which have demonstrated parallel development of the coordinative infrastructures of speech and nonspeech task. The present data do not resolve the question of the origin of these patterns, but support the suggestion that jaw oscillations are fundamental developmental movements in the sequence of speech motor control.

Simplification Control Strategy: Constraining the degrees of freedom

According to Bernstein (1996) novice performers of a complex motor task solve the degrees of freedom problem by "freezing" or "linking" some components to reduce the number of controlled elements. We have described this process as *simplification control strategies,* to distinguish it from the processes of motor skill acquisition in which constraint of redundant degrees of freedom reflects optimization of control. Bernstein's suggestion has received empirical support from studies showing that separate body segments act as a unit when learning a new motor skill (e.g., handwriting, Newell & van Emmerik, 1989; racquetball, Southard & Higgins, 1987). These studies have shown that the ability to control each segment separately is achieved through practice and is accompanied by improved skill performance (Southard & Higgins, 1987).

The upper lip, lower lip, and jaw form a redundant three-component system with respect to oral closure. Any one of these articulators can effectively occlude the oral

aperture. In the context of speech motor control, Bernstein's (1996) notion might be interpreted to suggest that during speech development children simplify an already existing set of articulatory goals to achieve more effective or efficient articulation. Thus, the rigid coordinative linkage of upper and lower lips in the present study may reflect the child's attempt to constrain the number of controlled elements. Conversely, the young child may have to rely on a reduced set of functional articulators and may recruit only those articulators over which he/she can achieve the greatest control (Kent, 1992). This possibility may especially apply to the early predominance of jaw movement in comparison to that of the lips.

Negative Transfer of Learning

The tightly coupled lip movements of early speech may be the result of negative transfer of learning. Transfer of learning effects occur when a pre-existing skill influences the learning of a new skill (Magill, 1993). Effective sucking requires high levels of interlabial coupling to form the oral seal. Therefore, it is possible that upper and lower lip coordinative patterns established for feeding influence initial attempts to coordinate these structures for speech. Although most negative transfer effects tend to be short lived and are easily overcome through practice, for example when learning a sport (Magill, 1993), this effect may be more persistent during motor skill development.

In summary, speech motor development entails the sequential emergence of articulatory control. The present results probably reflect extensive changes in the neuromotor pathways controlling the articulators and in their biomechanical composition. More general principles of skilled movement acquisition may also account

for some of the changes observed. Additional studies are required to reveal the relative importance of each of these factors.

Physiologic Constraints and Phonologic Acquisition

Several researchers have advanced a "physiological and human factors" orientation to phonology (Diver, 1979; Tobin, 1997), which suggest that constraints in the articulatory production and the auditory perceptual systems produce lawful relations in phonology. Universal phonologic patterns are viewed as emergents of the human communication system (Lindblom, 1992). One basic assumption of this approach is that some sounds are inherently more difficult to produce than others, and sounds or sound combinations that require less effort appear more frequently in the world's languages (Tobin, 1997). Tobin (1997) summarizes the human factors orientation to phonology:

> "The explanation of the general skewing- relying on the physiological and human factors orientations – which is based on the relative degree of ease or difficulty of the control of musculature, can even be applied further to explain the subskewings within the larger skewings, thus providing us with a single linguistic generalization that will account for all the data in the simplest way possible..." (p 37).

Tobin (1997) has extended this approach to speech development to explain the nonrandom distribution of phonemes produced by young children. The human factors orientation suggests that young children favor or disfavor certain phonemes because of inherent differences in ease of production (Tobin, 1997). Articulatory ease may account

for preferences in place, voice, and manner of articulation exhibited in young children (e.g., the prevalence of voiced bilabial stops in early speech; Stoel-Gammon, 1988). Unfortunately, there is no straightforward method to characterize sounds in terms of their articulatory complexity or ease or articulation. Such an attempt would have to include many factors relating both to perception and to production. Contemporary arguments are largely circular: later appearing phonemes must be more difficult because they appear later. From a developmental motor control perspective articulatory ease might be affected by a number of factors:

1. *The degree of specification of spatial and temporal movement parameters.* Phonemes are probably not equivalent in their coordinative demands. Some phonemes must be produced using a very narrowly specified articulatory gesture (e.g., /s/), while others can be specified quite generally (e.g., /b/).
2. *The biomechanical composition of the articulators.* The differences in the biomechanical properties of articulators interact with their function in an immature motor system. As illustrated by Thelen (1995), biomechanical properties can sharply constrain movement of some components of a motor system in early development. Furthermore, articulators with greater degrees of freedom will require either greater control specification or constraint from the motor system (i.e., tongue versus velum).
3. *The pre-existing neuromuscular organization.* There may be developmental predispositions in coordinative organization of oromotor control, that are determined by neuromotor maturation. Of course, the capacity of the system may facilitate or impede the learning of speech gestures.

Several investigators (Locke, 1993; MacNeilage & Davis, 1995) have suggested that early syllables are predominantly characterized by complete closing and opening gestures, and few narrow constrictions. These observations agree with Tobin's (1997) hypothesis that one important variable in determining the articulatory ease of a phoneme is the degree of constriction. Phonemes that require a narrow constriction (e.g., labiodental fricative /f/) may require greater control and sustained effort over time in comparison to those produced with a complete closure (i.e., stops). Similarly, Kent (1992) has suggested that early articulations might be produced with relatively rapid or "ballistic" articulatory movement, differentiating this class of phonemes from those that appear later and require "fine force regulation for frication" (p. 75). These observations form the impression that early speech motor organization is well adapted for producing stop consonants, but poorly adapted for producing phonemes that demand the exertion of graded muscle force (e.g., fricatives, liquids, affricates).

The findings of the present study agree with and extend this proposal. Several features of articulatory coordination observed in the one- and two-year-old children suggest that, from a physiologic perspective, articulatory ease may change as a function of age. The observed coordinative features that may limit sound producing capabilities during the first several years of life include: (1) the prevalence of jaw movement, (2) poor lip and jaw coupling, (3) poor lip control, and (4) the potential for linked control of upper and lower lips. These coordinative constraints may explain why voiced stops predominate in the first year of life, and why labiodental fricatives do not emerge until around two years (Stoel-Gammon, 1985), attaining mastery at age four (Sanders, 1972).

The coordinative requirements of voiced stops apparently do not exceed the capabilities of the immature articulatory system. Stops consonants can be produced using relatively ballistic jaw control without active contribution from the lips or tongue (MacNeilage & Davis, 1995). In contrast, articulation of the labiodental /f/, for instance, requires graded lower lip and jaw control to produce a slight constriction between the upper central incisors and the lower lip.

In summary, the observed developmental changes in articulatory control support the role of physiologic constraints in the sequence of phonologic development. Accordingly, immature speech may reflect the child's exploitation of the articulators over which they have the most control. If this is the case, the divergence from babble to speech may entail the breaking away from preferred coordinative patterns in favor of those in the ambient language.

Methodological Limitations

The present techniques have provided a means by which the spatial and temporal aspects of articulatory movement in young children can be quantified. These methods were developed to address these specific experimental questions, but will be further refined for future investigations and applications.

There are a number of methodological factors that may have influenced these results and require further attention. The use of a single skin-based marker to represent articulatory movement has a number of potential problems. Because the movement markers are attached to the skin, mechanical interactions may occur. This potential

interaction between movement markers can become especially problematic when lower lip movement is derived by subtracting the jaw signal from the lower lip signal, as was done in the present study. In addition, although jaw movements were "removed" during this transformation, jaw occlusal forces continued to influence the lower lip and upper lip signals. However, the varying coordinative patterns in the data suggest that mechanical interactions between movement markers were not extensive. Specifically, if mechanical linkages were significantly influencing the position of the movement markers, we would have expected the correlation values to be uniformly high in the crosscorrelation analysis (Löfqvist, & Gracco, 1997), which was in contrast to the wide range of correlation values observed.

Another possible weakness was the effects of rotation of the head in the Z plane on movement measurements. Forward rotation of the head decreases the relative amplitudes of movements among the markers with the jaw signal being most affected. Accordingly, utterances associated with significant forward or backward head rotation were not included in the present study (see Method). Forward rotations of the head were more common than backward rotations because children tended to orient toward toys place on their lap tray. Head rotation was only a concern for the one- and two-year-old subjects as older subjects were able to comply with instructions to maintain an upright position with minimal head rotation. Throughout each experimental session, the child's head position was guided to an orientation by placing a toy just above the camera while eliciting utterances. The present results did not suggest any systematic influence of head rotation. In fact, one-year-old subjects produced relatively greater jaw movements than

any other group, a trend opposite to that which would result from artifactual forward rotation.

The coordinate system derived for the analysis of oral closure was designed to reveal developmental shifts in each articulator's role during speech, while limiting the influence of contextual variation. In the future, this measure will provide a means to examine developmental changes in articulatory coordination across a variety of speech contexts. Of course, each coordinate system has its own weaknesses and strengths. The greatest disadvantage of the present system is that passive contributions of the articulators cannot be discerned from active displacements. For instance, the position of the upper lip at rest was calculated to be contributing to oral closure because its reference position obtained during a smile was higher than its rest position. Ideally, we would like to measure only the movement of active components, which will reflect control strategies for oral closure. Consequently, individual differences in orofacial morphology among the subjects may have been reflected in the measurement of relative *contribution to oral closure*.

Finally, another potential problem for this analysis was achieving the reference postures in the young subjects. We could not be entirely confident that the young children were producing the greatest degree of oral opening possible. Despite these limitations, we were encouraged by the observation that this measure reflected the age-related differences that were clearly observed in the raw kinematic traces.

Clinical Implications

A better understanding of the sequence of speech motor control is essential before we can begin to comprehend delayed and abnormal speech development. Unlike speech motor control, developmental milestones and critical periods have been identified for other motor skills and systems. These normative descriptions have been clinically indispensable (locomotion, Ames, 1937; Gesell & Ames, 1940; reaching, Halverson, 1931, all cited in Haywood, 1993). Similar descriptions are needed for the motor milestones of speech.

The developmental sequence observed in the present study may lead to a descriptive framework in which speech motor delays can be detected at an earlier stage of development. For instance, the present results might be taken to suggest that limited mandibular control in early speech is a negative prognostic factor for later speech motor delays. While it is premature to make such specific recommendations, an improved understanding of the fundamental motor patterns for speech will dramatically strengthen differential diagnosis and treatment of developmental speech disorders.

Future Research

These preliminary findings raise a host of new questions with respect to the development of speech motor control. One important next step will be to provide a more complete description of the one- and two-year-old's articulatory kinematics. Thus, future studies will focus on describing early movement characteristics of the jaw and lips in

terms of displacement, velocity, and acceleration to gain a better understanding of the control exerted in early articulation. Another avenue of inquiry includes a more detailed examination of the associative movements of the lips at around two years of age.

Finally, because the experimental approach adopted was only concerned with group differences, futures studies must consider individual differences in the sequence of speech motor development. Because children tend to vary considerably in their motor development, longitudinal designs will provide a more suitable means for describing individual differences in the developmental process.

SUMMARY

The objective of this investigation was to describe development of lip and jaw coordination for production of syllables containing bilabial consonants. Movement of the upper lip, lower lip, and jaw were recorded during the production of "baba," "papa," and "mama" in four age groups (i.e., one-, two-, six-years, and adult). Two complementary analytic techniques were used to quantify interarticulator coordination, one reflecting spatial and temporal coupling of articulatory pairs and the other isolating each articulator's contribution to oral closure.

The coordinative organization of articulatory gestures changes dramatically during the first several years of life and continues to undergo refinement past age six. The present results can be interpreted as representing three primary phases in the development of lip and jaw coordination for speech: integration, differentiation, and refinement. Each of these developmental processes entails the existence of distinct coordinative constraints on early articulatory movement. These constraints will have predictable consequences for the sequence of phonologic development.

Integrative processes in speech motor development were supported by the observation that one-year-olds relied heavily on jaw movement for oral closure. In contrast to the early predominance of jaw movement, lower lip's contribution to oral closure was initially small and increased gradually with age. This developmental sequence in early speech supports the hypothesis that speech motor development requires assimilation of lower lip movement into established mandibular synergies. The

predominance of jaw movement in early speech may explain the frequent occurrence of stops in early speech.

Differentiating processes may also be operative during speech motor development. The present data suggest that the attainment of independent control of the upper and lower lips might be one phase in early speech motor development. The presence of linked lip control will have predictable effects on the child's ability to produce certain phonemes under these coordinative constraints. For instance, young children may have difficulty isolating upper lip and lower lip movement to produce the labiodental fricative /f/.

Other developmental changes in interarticulator coordination reflected refinement of coordination. The qualitative observations revealed that movement patterns exhibited by six-year-olds were similar to those of adults, but were often more variable. Moreover, lip and jaw spatiotemporal coupling gradually increased with age.

Future studies will focus on further describing developmental sequences of articulatory control. One important step will be to gain a better understanding of the control exerted in early articulation by examining early movement characteristics of the jaw and lips in terms of displacement, velocity, and acceleration.

REFERENCES

Abbs, J.H. (1973). Some mechanical properties of lower lip movements during speech production. Phonetica, 28, 65 – 75.

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: orofacial muscle responses to load perturbations of lip during speech. Journal of Neurophysiology, 51, 705-723.

Abbs, J. H., Gracco, V. L., & Blair, C. (1984). Functional muscle partitioning during voluntary movement: facial muscle activity for speech. Experimental Neurology, 85, 469 - 79.

Ames, L.B. (1937). The sequential patterning of prone progression in the human infant. Genetic Psychology Monographs, 19, 409 – 60.

Baragar, F.A., & Osborn, J.W. (1984). A model relating patterns of human jaw movement to biomechanical constraints. Journal of Biomechanics, 17, 757 – 67.

Barlow, S.M., Finan, D.S., Bradford, P.T., & Andreatta, R.D. (1993). Transitional properties of the mechanically evoked perioral reflex from infancy through adulthood. Brain Research, 623, 181 - 88.

Barlow, S.M., & Netsell, R. (1986). Differential fine force control of the upper and lower lips. Journal of Speech and Hearing Research, 29, 163 - 169.

Bernstein, N.A. (1996) On dexterity and its development. In M.L. Latash and M.T. Turvey's (Eds.) Dexterity and its development (pp. 9 - 244). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Boliek, C.A., Hixon, T.J., Watson, P.J., & Morgan, W.J. (1996). Vocalization and breathing during the first year of life. Journal of Voice, 10, 1 - 22.

Bosma, J.F. (1973). Prologue to the symposium. In J.F. Bosma (Ed.), Fourth Symposium on Oral Sensation and Perception: Development in the Fetus and Infant (pp. 3 - 8). Bethesda, MA: U.S. Department of Health, Education, and Welfare. National Institutes of Health.

Bosma, J.F. (1985). Postnatal ontogeny of performance of the pharynx, larynx, and mouth. Am Rev Respir Dis, 131, S10 - S15.

Browman, C.P. & Goldstein, L. (1986). Articulatory phonology: An overview, Phonetica, 49, 155 – 80.

Cohen, H.J., Taft, L.T., Mahadeviah, M.S., & Birch, H.G. (1967). Developmental changes in overflow in normal and aberrantly functioning children. The Journal of Pediatrics, 71, 39 – 47.

Connolly, K., & Stratton, P. (1968). Developmental changes in associated movements. Developmental Medical Child Neurology, 10, 49 - 56.

Courville, J. (1966). The nucleus of the facial nerve: The relation between cellular groups and peripheral branches of the nerve. Brain Research, 1, 338 - 354.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1975). Motor Speech Disorders. W.B. Saunders Company: Philadelphia, PA.

Davis, B.L., & MacNeilage, P.F. (1994) Organization of babbling: A case study. Language and Speech, 37, 341 - 55.

Davis, B.L., & MacNeilage, P.F. (1995). The articulatory basis of babbling. Journal of Speech and Hearing Research, 38, 1199 - 211.

Diver, E. (1979). Phonology as human behavior. In D. Aaronson and J. Reiber (Eds.), Psycholinguistic research: Implications and applications (pp. 161-98). Hillsdale, N.J.: Lawrence Eribaum.

Edelman, G.M. (1987). Neural Darwinism. New York: Basic Books.

Eguchi, S., & Hirsh, I.J. (1969). Development of speech sounds in children. Acta Oto-Laryngologica, supp. 256.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. Science, 171, 303 - 06.

Engstrand, O., & Williams, K. (1996). VOT in stop inventories and in young children's vocalizations: preliminary analyses. Proceedings from the Swedish Phonetic Conference, Nässlingen, 29-31 May 1996.

Enstrom, D.H. (1982). Infant labial, apical and velar stop productions: a voice onset time analysis. Phonetica, 39, 47 - 60.

Feinstein, B., Lindegard, B., Nyman, E., & Wohlfart, G. (1955). Morphological studies of motor units in normal human muscles. Acta Anatomische, 23, 127-142.

Fentress, J.C. (1984). The development of coordination. Journal of Motor Behavior, 16, 99 – 134.

Folkins, J.W. (1978). Lower lip displacement during voluntary activation of individual labial muscles in man. Archives of Oral Biology, 23, 189 - 193.

Folkins, J.W., Linville, R.N., Garrett, J.D., & Brown, C.K. (1988). Interactions in the labial musculature during speech. Journal of Speech and Hearing Research, 31, 253 - 264.

Gabbard, C. (1992). Lifelong Motor Development. Dubuque, IA:W.C. Brown Publishers.

Gandour, J., Holasuit Petty, S., Dardarananda, R., Dechongkit, S., & Mukngoen, S. (1986). The acquisition of the voicing contrast in Thai: a study of voice onset time in word-initial stop consonants. Journal of Child Language, 13, 561-572.

Gasser, R.F. (1967). The development of the facial muscles in man. American Journal of Anatomy, 120, 357 - 76.

Gay, T. (1977). Articulatory movements in VCV sequences. Journal of Acoustical Society of America, 62, 183 - 93.

Gesell, A., & Ames, L.B. (1940). The ontogenetic organization of prone behavior in human infancy. Journal of Genetic Psychology, 56, 247-263.

Goffman, L., & Smith, A. (1994). Motor unit territories in the human perioral musculature. Journal of Speech and Hearing Research, 37, 975-84.

Goffman, L., & Smith, A. (in press). Development and phonetic differentiation of speech movement patterns. Journal of Experimental Psychology: Human Perception and Performance.

Gracco, V.L. (1988). Timing factors in the coordination of speech movements. The Journal of Neuroscience, 8, 4628 –39.

Gracco, V. (1994). Some organizational characteristics of speech movement control. Journal of Speech and Hearing Research, 37, 4 – 27.

Green, J.R., Moore, C.A., Ruark, J.L., Rodda, P.R., Morvée, W.T., & VanWitzenburg, M.J. (1997). Development of chewing in children from 12 to 48 months: longitudinal study of EMG patterns. Journal of Neurophysiology, 77, 2704 - 27.

Grillner, S. (1982). Possible analogies in the control of innate motor acts and the production of sound in speech. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), Speech motor control (pp. 217 - 230). New York: Pergamon Press.

Halverson, H.M. (1931). An experimental study of prehension in infants by means of systematic cinema records. Genetic Psychology Monographs, 10, 107 – 286.

Haywood, K.M. (1993). Life span and motor development (2nd edition). Champaign, IL: Human Kinetics.

Herring, S.W. (1985). The ontogeny of mammalian mastication. American Zoology, 25, 339 - 349.

Herring, S.W., & Wineski, L.E. (1986). Development of the masseter muscle and oral behaviors in the pig. Journal of Experimental Zoology, 237, 191 - 207.

Hughes, O.M. & Abbs, J.H. (1976). Labial-mandibular coordination in the production of speech: implications for the operation of motor equivalence. Phonetica, 33, 199 - 221.

Humphrey, T. (1964). Some correlations between the appearance of human fetal reflexes and the development of the nervous system. Progress in Brain Research, 4, 93 - 135.

Humphrey, T. (1970). The development of human fetal activity and its relation to postnatal behavior. Advances in Child Development and Behavior, 5, 1 - 57.

Humphrey, T. (1971). Development of oral and facial motor mechanisms in human fetuses and their relation to craniofacial growth. Journal of Dental Research, 50, 1428 - 41.

Iriki, A., Nozaki, S., & Nakamura, Y. (1988). Feeding behavior in mammals: corticobulbar projection is reorganized during conversion from sucking to chewing. Brain Res. Dev. Brain Res., 44, 189 - 96.

Jakobson, R. (1949). Les lois phoniques du langage enfantin et leur place dans la phonologie générale. In N.S. Trubetzkoy, Principes de phonolgie. (Tr. Into French by J. Cantineau.) Paris: Editions Kincksieck.

Jeannerod, M. (1988). The Neural and Behavioral Organization of Goal-Directed Movements: Oxford Psychology Series NO. 15. Oxford: Clarendon Press.

Kelso, J.A.S., Saltzman, E.L., & Tuller, B. (1986). The dynamical perspective on speech production: data and theory. Journal of Phonetics, 14, 29 - 59.

Kent, R.D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. Journal of Speech and Hearing Research, 19, 421 - 47.

Kent, R.D. (1984). Psychobiology of speech development: Coemergence of language and a movement system. American Journal of Physiology, 246, R888 - 94.

Kent, R.D. (1992). The biology of phonological development. In C.A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), Phonological Development: Models, research, implications (pp. 65 – 90). Timonium, MD:York Press.

Kent, R.D., & Bauer, H.R. (1985). Vocalizations of one-year-olds. Journal of Child Language, 12, 491 - 525.

Kent, R.D., & Forner, L.L. (1980). Speech segment durations in sentence recitations by children and adults. Journal of Phonetics, 8, 157 - 68.

Kent, R.D., & Murray, A. (1982). Acoustic features of infant vocal utterances at 3, 6, 9 months. Journal of Acoustical Society of America, 72, 353 - 65.

Kent, R.D., & Vorperian, H.K. (1995). Development of the craniofacial-oral-laryngeal anatomy: A review. San Diego, CA: Singular Publishing Group, Inc.

Kewley-Port, D., & Preston, M. (1974). Early apical stop production: a voice onset time analysis. Journal of Phonetics, 2, 195 - 210.

Kubota, K., Narita, N., Ohkubo, K., Shibania, S., Nagae, K., Kubota, M. Odagiri, N., & Kawamoto, T. (1988). Morphological studies of the neuromusculature shifting from sucking to biting of mice. Acta Anatomica, 133, 200 – 08.

Kuehn, D.P. & Moll, K.L. (1976). A cineradiographic study of VC and CV articulatory velocities. Journal of Phonetics, 4, 303 - 20.

Kuhl, P.K., & Meltzoff, A.N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. Journal of Acoustic Society of America, 100, 2425 - 38.

Lakars, T.C., & Herring, S. W. (1980). Ontogeny of oral function in hamsters. Journal of Morphology, 165, 364 – 88.

Laurenson, R.D. (1968). An introduction to clinical anatomy by dissection of the human body. Philadelphia: W.B. Saunders Company.

Lazarus, J.C. & Todor, J.I. (1987). Age differences in the magnitude of associated movement. Developmental Medicine and Child Neurology, 29, 726 – 33.

Lenneberg, E.H. (1967). Biological Foundations of Language. New York: John Wiley & Sons, Inc.

Lieberman, P. (1977). Speech Physiology and Acoustic Phonetics. New York, NY: McMillan Publishing Co., INC.

Lieberman, P. (1984). The Biology and Evolution of Language. Cambridge, MA: Harvard University Press:

Lieberman, P., Crelin, E.S., & Klatt, D.H. (1970). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal, man, and the chimpanzee. Haskins Laboratories Status Report of Speech Research, SR-24, 56 - 90.

Lindblom, B. (1963). Spectrographic study of vowel reduction. Journal of Acoustical Society of America, 35, 1773 – 81.

Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In C.A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), Phonological Development: Models, Research, Implications (pp. 131 – 164). Timonium, MA: York Press, Inc.

Ling, D. (1976). Speech and the hearing-impaired child: theory and practice. Washington, DC: Alexander Graham Bell Association for the Deaf, Inc,.

Locke, J.L. (1983). Phonological Acquisition and Change. San Diego: Academic Press.

Locke, J.L. (1993). The Child's Path to Spoken Language. Second edition. Cambridge, MA: Harvard University Press.

Löfqvist, A. & Gracco, V.L. (1997). Lip and jaw kinematics in bilabial stop consonant production. Journal of Speech, Language, and Hearing Research, 40, 877 – 893.

Lund, J. P., Appenteng, K., & Séguin, J. J. (1982). Analogies and common features in the speech and masticatory control systems. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), Speech Motor Control (pp. 231 - 245). New York: Pergamon Press.

Macken, M.A. & Barton, D. (1980). The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stops. Journal of Child Language, 7, 41-74.

MacNeilage, P.F. (1997). "The frame/content theory of evolution of speech production." http://www.cogsci.soton.ac.uk/bbs/Archive/bbs.macneilage.html

MacNeilage, P.F. & Davis, B.F. (1990a). Acquisition of speech production: frames, then content. In M. Jeannerod (ed.), Attention and Performance XIII: Motor Representation and Control (pp. 453 - 76). Hillsdale, N.J.: Larwrance Erlbaum.

MacNeilage, P.F. & Davis, B.L. (1990b). Acquisition of speech production: the achievement of segmental independence. In W.J. Hardcastle & A. Marchal (Eds.), Speech Production and Modeling (pp. 55 - 68). Dordrecht:Kluwer Academic.

MaGill, R.A. (1993). Motor learning: Concepts and applications (4th ed.). Bubuque, IA: WM. C. Brown. Communications, Inc.

Matlab, The MathWorks, Inc. (1997). 24 Prime Park Way, Natick, MA. 01760-1500.

McGraw, M.B. (1940). Neuromuscular development of the human infant as exemplified in the achievement of erect locomotion. Journal of Pediatrics, 17, 747 – 771.

Mitchell, P., & Kent, R. (1990). Phonetic variation in multisyllabic babbling. Journal of Child Language, 17, 247 - 65.

Moore, C.A., & Ruark, J.L. (1996). Does speech emerge from earlier appearing motor behaviors? Journal of Speech and Hearing Research, 39, 1034 -1047.

Moore, C. A., Smith, A., & Ringel, R. L. (1988). Task-specific organization of activity in human jaw muscles. Journal of Speech and Hearing Research, 31, 670-680.

Müller, E.M., Milenkovic, P.H., & MacLeod, G.E. (1984). Perioral tissue mechanics during speech production. In C. DeLisi and J. Eisenfeld (Eds.), Proceedings of Second IMAC International Symposium on Biomedical Systems Modeling. North-Holland: Amsterdam

Mysak, E. D. (1980). Neurospeech therapy for the cerebral palsied. New York: Teachers College Press.

Navarrete, R., & Vrbova, G. (1983). Changes of activity in slow and fast muscles during postnatal development. Developmental Brain Research, 8, 11 – 19.

Newell, K.M., & van Emmerik, R.E.A. (1989). The acquisition of coordination: Preliminary analysis of learning a rapid arm movement. Human Movement Science, 2, 255 – 69.

Nittrouer, S. (1993) The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. Journal of Speech and Hearing Research, 36, 959 – 71.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. Journal of Acoustical Society of America, 97, 520 – 530.

Nittrouer, S., Studdert-Kennedy, M., & Neely, S.T. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. Journal of Speech and Hearing Research, 39, 379 - 89.

Oller, D.K. (1978). Infant vocalization and the development of speech. Allied Health and Behavioral Sciences, 1, 525 - 49.

Provins, K.A. (1997). The specificity of motor skill and manual asymmetry: A review of the evidence and its implications. Journal of Motor Behavior, 29, 183 – 92.

Rubinstein, N.A., & Kelly, A.M. (1981). Development of muscle fiber specialization in the rat hindlimb. J-Cell-Biol., 90, 128-44.

Ruark, J.L., & Moore, C.A. (1997). Coordination of lip muscle activity by 2-year-old children during speech and nonspeech tasks. Journal of Speech and Hearing Research, 40, 1373 - 85.

Sanders, E.K. (1972). When are speech sounds learned? Journal of Speech and Hearing Disorders, 37, 55 - 63.

Sanguineti, V., Laboissière, R., & Ostry, D.J. (1998). Journal of Acoustic Society of America, 103, 1615 – 1627.

Sasaki, C.T., Levine, P.A., Laitman, J.T., & Crelin, E.S. (1977). Postnatal descent of the epiglottis in man. Archives of Otolaryngology, 103, 169 - 71.

Sato,M.; Mizuno,N.; Konishi,A. (1977). Postnatal differentiation of cell body volumes of spinal motoneurons innervating slow-twitch and fast-twitch muscles J.Comp.Neurol., 175, 27 – 36.

Schuster, C., & Ashburn, S.S. (Eds.). (1992). The Process of Human Development: A holistic life-span approach (3rd ed). Philadelphia: J.B. Lippincott Company.

Sharkey, S.G., & Folkins, J.W. (1985). Variability of lip and jaw movements in children and adults: Implications for the development of speech motor control. Journal of Speech and Hearing Research, 28, 8 - 15.

Smith, A. (1992). The control of orofacial movements in speech. Critical Reviews in Oral Biology and Medicine, 2, 233 - 67.

Smith, A., & Denny, M. (1990). High-frequency oscillations as indicators of neural control mechanisms in human respiration, mastication, and speech. Journal of Neurophysiology, 63, 745 - 58.

Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. Journal of Speech, Language, and Hearing Research, 41, 18 – 30.

Smith, B.L., & Gartenberg, T.E. (1984). Initial observations concerning development characteristics of labio-mandibular kinematics. Journal of Acoustical Society of America, 75, 1599 - 1605.

Smith, B.L., Brown-Sweeney, S., & Stoel-Gammon, C. (1989). A quantitative analysis of reduplicated and variegated babbling. First Language, 17, 147 - 53.

Smith, B.L., & McLean-Muse, A. (1987). An investigation of motor equivalence in children. Journal of Acoustical Society of America, 82, 837 - 42.

Southard, D., & Higgins, T. (1987). Changing movement patterns: Effects of demonstration and practice, Research Quarterly for Exercise and Sport, 58, 77 – 80.

Sporns, O., & Edelman, G.M. (1993). Solving Bernstein's problem: A proposal for the development of coordinated movement by selection. Child Development, 64, 960 – 81.

Stallings, L.M. (1973). Motor Skills Development and Learning. Dubuque, IA:W.C. Brown Company Publishers.

Steenbergen, B., Marteniuk, R.G., & Kalbfleisch, L.E. (1995). Achieving coordination in prehension: joint freezing and postural contributions. Journal of Motor Behavior, 27, 333 - 48.

Stoel-Gammon, C. (1985). Phonetic inventories, 15-24 months: a longitudinal study. Journal of Speech and Hearing Research, 28, 505 - 12.

Stoel-Gammon, C. (1988). Prelinguistic vocalizations of hearing-impaired and normally hearing subjects: a comparison of consonantal inventories. Journal of Speech and Hearing Research, 53, 302 - 15.

Stoel-Gammon, C., & Otomo, K. (1986). Babbling development of hearing-impaired and normally hearing subjects. Journal of Speech and Hearing Research, 51, 33 - 41.

Subtelny, D.J. & Baker, H.K. (1953). The significance of adenoid tissue in velopharyngeal function. Plastic and Reconstructive Surgery, 17, 235 - 50.

Sussman, H.M., MacNeilage, P.F., & Hanson, R.J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: preliminary observations. Journal of Speech and Hearing Research, 16, 397 - 420.

Thelen, E. (1983). Learning to walk is still an "old" problem: A reply to Zelazo. Journal of Motor Behavior, 15, 139 – 61.

Thelen, E. (1986). Treadmill-elicited stepping in seven-month-old infants. Child Development, 57, 1498 – 1506.

Thelen, E. (1991). Motor aspects of emergent speech: a dynamic perspective. In N. A. Krasnegor, D.M. Rumbaugh, R.L. Schiefelbusch, & M. Studdert- Kennedy (Eds.). Biological and Behavioral Determinants of Language Development (pp. 339 - 62). New Jersey: Lawrence Erlbaum Associates.

Thelen, E. (1995). Motor Development: A new synthesis. American Psychologist, 50, 79 – 95.

Thelen, E., Bradshaw, G., & Ward, J.A. (1981). Spontaneous kicking in month-old infants: Manifestations of a human central locomotor program. Behavioral and Neural Biology, 32, 45 – 53.

Thelen, E., & Fisher, D.M. (1982). Newborn stepping: An explanation for a "disappearing reflex." Developmental Psychology, 18, 750 – 75.

Thelen, E., Ulrich, B.D., & Jensen, J.L. (1989). The developmental origins of locomotion. In M.H. Woollacott & A. Shumway-Cook (Eds.), Development of posture and gait across the life span (pp. 25-47). Columbia, SC: University of South Carolina Press.

Tingley, B.M., & Allen, G.D. (1975). Development of speech timing control in children. Child Development, 46, 186 - 94.

Tobin, Y. (1997). Phonology as Human Behavior. London: Duke University Press

Todor, J.I., & Lazarus, J.C. (1986). Exertion level and the intensity of associative movements. Developmental Medicine & Child Neurology, 28, 205 - 12.

Trevarthen, C. (1984). How control of movement develops. In H.T.A. Whiting (Ed.) Human Motor Actions - Bernstein Reassessed. North-Holland: Elsevier Science Publishers B.V.

Vihman, M.M. (1996). Phonological Development: The Origins of Language in the Child. Oxford: Blackwell Publishers Inc.

Watkin, K.L., & Fromm, D. (1984). Labial coordination in children: preliminary considerations. Journal of Acoustic Society of America, 75, 629 - 32.

Welt, C., & Abbs, J. H. (1990). Musculotopic organization of the facial motor nucleus in Macaca fascicularis: a morphometric and retrograde tracing study with cholera toxin B-HRP. Journal of Comparative Neurology, 291, 621 - 36.

Wohlert, A.B., & Goffman, L. (1994). Human perioral muscle activation patterns. Journal of Speech and Hearing Research, 37, 1032 - 040.

APPENDIX

Code Generated for Experiments

Program for Crosscorrelation Analysis: Kincorr5

KINCORR5

```
%Kinematic analysis program for Peak Motus to perform
crosscorrelations
%kinematic signals of UL, LL, and J.
%UL signal is inverted prior to analysis and singals are
centered prior to correlation
%Designed for 3 movement and 2 reference markers
%Output Matrix is "data1" use with Excel-link
%Contains ULxLL Coef, ULxJ Coef, LLxJ Ceof, ULxLL Lag, ULxJ
Lag, LLxJ Lag.

close
clear

files = input (' Enter number of files to be measured: ',
's')
files = str2num(files);


KinTraces = [];
VelocityTraces = [];
Jaw_DAll =[];
KinJaw =[];
KinMatTemp = [];
CrossMAX = [];
CrossLag = [];
LagTotal = [];
CoefTotal = [];
AllMovement = [];

fs = 60; % sampling freqency
```

```
%%%
```

```
path4 = ['e:\jordan\dissert\data\1yr\hm_1\*.*'];
path5 = ['a:\']
path = path1;

[filename, pathname] = uigetfile(path,ii,100,100);

   file = [pathname filename];
        getfiles = (['DataMat',num2str(ii) ,'=
csvread(file);']);
   eval(getfiles);
        ii
   sizeit = (['[mrows chans] =
size(DataMat',num2str(ii),');']);
   eval(sizeit);


   %%  Format the file to create matrix with only the y
vectors
   %-------------------------------------------------------
------

   emptymat = (['KinMat',num2str(ii),' = [];']); % Create
empty matrix
   eval(emptymat);

   %Get all y vectors

        for iii = 1:5  %% Temporary
          place = (['place_hold =
DataMat',num2str(ii),'(:,(iii * 3)-1);']);
     eval(place);

       %Put all y vectors into one matrix

     Kin = (['KinMat',num2str(ii) ,'= [KinMat',
num2str(ii),', place_hold];']);
     eval(Kin);
   end

   DelRef = (['KinMat',num2str(ii),'(:,1:2) = [];']);
   eval(DelRef)      %Delete Reference Point coordinates

        DataMatCorrect = [];
```

```
        CenterKinMat = [];
        Segment = [];
        Peak_Time =[];
        Cent_Jaw = [];

%================  Subtract Lower Lip from Jaw to get
Lower Lip

centjaw = (['Center_Jaw = KinMat', num2str(ii),'(:,3) -
mean(KinMat',num2str(ii),'(:,3));']);
eval(centjaw);

Correction_Jaw_Mark = Center_Jaw;

%step1 = (['Correction_Jaw_Mark = Center_Jaw -
min(Center_Jaw);']); %Get correction factor for Jaw
%eval(step1);

step2 = (['KinMat',num2str(ii),'(:,2) =
KinMat',num2str(ii),'(:,2) - Correction_Jaw_Mark;']);
eval(step2);

%Invert all signals - need to adjust for Peak Motus
translation

step3 =(['KinMat',num2str(ii),'(:,:) =
KinMat',num2str(ii),' * invert;']);  %Invert all signals -
need to correct for Peak Motus translation
eval(step3);

%>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>> Define Segment

%%%% Plot Kinematic signal and associated velocity trace --
to get gesture boundaries

tempkin = (['tempk = KinMat',num2str(ii),'(:,:);']);
eval(tempkin);

tempjaw = (['tempj = KinMat',num2str(ii),'(:,3);']);
eval(tempjaw);

%Calculate derivative
```

```
tempjaw_d = (['tempj_d = diff(KinMat',
num2str(ii),'(:,3));']);
eval(tempjaw_d);

hold on

%>>>>>>>>>>>>>>>> PLOT INITIAL WINDOW for user to choose
movement of interest >>>>>>>>>
% Instructions Window
     instructions = uicontrol(gcf,...
        'Style','text',...
        'BackgroundColor', color(2, : ),...
        'String', 'IDENTIFY SEGMENT BOUNDARIES: based on
zero crossing of velocity trace',...
        'Position',[0 scrsz(4)-70 1300 100]);

%%%%%%%%%%     ZeroCrossing %%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Find Zero Crossing in Velocity Signal to estimate onset
and offset of movement %%%%%

TEMP = tempj_d; %rename for ZeroCrossing program
winsize_z = 1;
begwin_z = 1;
NumWindows = length(TEMP) - 1;
endwin_z = 2;
Movement = [];

for searchwin = 1:NumWindows  % L4 - Find zero crossings in
9 windows across signal

     AboveMat = find(TEMP(begwin_z:endwin_z)>0);
     BelowMat = find(TEMP(begwin_z:endwin_z)<0);

   if isempty(AboveMat) ==0 & isempty(BelowMat) == 0   % I2
-- If zero Crossing

      AboveMatTime = (begwin_z - 1) + AboveMat ;        %
Re-esablish the index
      AboveMatValue = TEMP(AboveMatTime);     % Get
corresponding value from data matrix
```

```
        BelowMatTime= (begwin_z - 1) + BelowMat;          %
Re-esablish the index
      BelowMatValue=TEMP(BelowMatTime);       % Get
corresponding value from data matrix


                if AboveMatTime > BelowMatTime, %%%%% I4
Positive Slope Zero Crossing --Closing Gesture

            Point1 = BelowMatValue;
            Point2 = AboveMatValue;

        if abs(Point1) < abs(Point2),
           ZeroPoint = BelowMatTime;
                else
           ZeroPoint = AboveMatTime;
                      end

                    else
                     Point1 = AboveMatValue;
            Point2 = BelowMatValue;

        if abs(Point1) > abs(Point2),
           ZeroPoint = BelowMatTime;
                else
           ZeroPoint = AboveMatTime;
                end


           end %I4

      Movement = [Movement,ZeroPoint];  %The estimated
movement onset and offset points

 end % I2

begwin_z = endwin_z;

endwin_z = begwin_z + winsize_z;


end % L4
```

```
AllMovement = [AllMovement Movement];

endloop = 0; %Initiate loop for measurements of beginning
and end points of movement

%%%%%%% END Zero Crossing Algorithm

plot(tempk);hold on; plot(tempj_d * 10);grid %Plot initial
window for boundary marking

%%%%%%%%%%%%%%%%%%%%%%%%%%%% Plot lines perpendicular to
zero crossings

hold on;
for plotzero = 1:length(Movement)
   v = axis;
   y = [v(3) v(4)];

   zero = Movement(plotzero);
   zero_x = [zero zero];

   plot(zero_x ,y,'-r');
end

%%%%%%%% User chooses gesture of interest - picks onset and
offset based on zero crossing of velocity trace

[onset junk] = ginput(1);
[offset junk] = ginput(1);

%%%%%%%%% Resize Kinematic and velocity traces

window_pad = 0; % Pad window 5 frames in front and 5 frames
in back - to get upper lip onset and offset

begwin_k = round(onset - window_pad);
endwin_k = round(offset + window_pad);

% In case chosen point for onset is less than 5.
if (begwin_k)<1,
begwin_k = 1;
end
```

```
% In case offset is greater than matrix length.
if (endwin_k)>length(tempj),
endwin_k = length(tempj)
end

ResizeK = (['KinMatGest',num2str(ii),' =
KinMat',num2str(ii),'(begwin_k:endwin_k,:);']);
eval(ResizeK);

close; %% Close Figure


end  % END for Loop1

%%%%%%%% Trial Loop - Post Data Conditioning %%%%%%%%%

for j = 1:files % Loop to Measure all trials (L1)

%%%%%%%%%  GRAPHICS  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%set(gcf, ...
        %'NumberTitle','off', ...
        %'Name','Articulatory Displacement Signals', ...
        %'backingstore','off',...
        %'Units','normalized');

KinTemp = (['KinMatTemp = KinMatGest', num2str(j),';']);
eval(KinTemp)

%%%%%%%%%%%%%%% CROSS CORRELATION %%%%%%%%%%%%%%%%%%%%

KinMatTemp(:,1) = KinMatTemp(:,1) .* (-1); %invert upper
lip signal



meanall = mean(KinMatTemp);
CenterKinMat = [];
for ii = 1:size(KinMatTemp,2)
   CenterKinMat(:,ii) = KinMatTemp(:,ii) - meanall(ii);
end

plot(CenterKinMat);  %pause
```

```
%%%%%%%%%%%%%  Cross Correlations %%%%%%%%%%%%%%%%%%%%%%%%%%

corout = xcorr(CenterKinMat,'coeff');      % cross-
correlation

croscorr = corout(:, [2 3 6]); %

middle = (length(croscorr) + 1)/2; % find middle element
gaxis = axis;

y = [gaxis(3) gaxis(4)];
x = [middle middle];


%%%%%%%%%%%%%%%%%%% Plot cross-correlations for testing
%%%%%%%%%%%%
[zz yy] = size(croscorr);
xsize = (-zz/2:zz/2-1);

for i = 1:cols,   % L1 = plot cross - correlation for each
channel

plot(croscorr(:,i)); axis([1 length(croscorr) -1 1]);hold
on;
vv = axis;
   yy = [vv(3) vv(4)];

   zero = Movement(plotzero);
   zero_x = [zero zero];


plot(x,yy,'-r');grid; plot(x + 6.25,yy,'-c'); plot(x -
6.25,yy,'-c'); %plot +- 100 ms border

[peak, y1, button]= ginput(1);  %%%%%%%% User identifies
peak with left mouse button

close

bgwin = round(peak - 2);
edwin = round(peak + 2);
```

```
   if button ==  1 ,
      [g h] = max(croscorr(bgwin:edwin,i));
      h = h + (bgwin-1);

   elseif button == 3,  % Find minimum by using right mouse
button

      [g h] = min(croscorr(bgwin:edwin,i));
      h = h + (bgwin-1);

   elseif button == 122        % If crosscorrelation does
not contain prominent peak
      g = nan;  % reject token for analysis - insert nan
      h = nan;
      end

   CrossMAX = [CrossMAX, g]
   CrossLag = [CrossLag, h]
  end %L1

            %  Main "for loop"

%%%%%%%%%%%%%%%% Lag from crosscorrelations %%%%%%%%%%%%

lagpoint = CrossLag - middle;
lagsecond = lagpoint/fs;

LagTotal = [LagTotal;lagsecond];

CrossLag
CrossLag =[];

CoefTotal = [CoefTotal;CrossMAX];

CrossMAX=[];

ZCoef = .5 * log(1 + CoefTotal);    % Transform all cross -
correlation coefficients
ZCoef2= ZCoef/(1 - CoefTotal);              % into Fisher Z.


MeanLag = Mean(LagTotal)
StdLag =  Std(LagTotal)
```

```
end

data1 = [CoefTotal, LagTotal];

FinalMeanMat = [MeanLag StdLag];

%filename = input (' Enter filename to be saved (i.e.,
blos_48): ', 's')

%file = (['save d:\jordan\chewing\lagdata\', filename,'.lag
data1 -ascii -tabs'])
%eval(file);

%file = (['save e:\lagdata\', filename,'.lag data1 -ascii -
tabs'])
%eval(file);
```

Programs for Contribution to Oral ClosureAnalysis: Oralclos and clostime

OralClos

```
%Kinematic analysis program for determining the
contribution of UL, LL and Jaw to ORAL CLosure
% This program calls the following functions: CLOSETIME.M
% Program runs multiple files but needs user to provide
path and name
%                  a reference file that was computed using
"openpos.m"

% Data for UL, LL, and J in "RelClosure"

close
clear

global path_to_file;

%TUNABLE VARIABLES

invert = 1; % Correct for Motus upgrade
```

```
path1 = 'c:\my documents\dissertation\data\adult\rj_a\p';
path2 = 'c:\my documents\dissertation\data\6yrs\lr_6\';
path3 = 'c:\my documents\dissertation\data\2yrs\mb_2\';

path_to_file = path3;

colordef black %Make background color "black" in all
figures

files = input (' Enter number of files to be measured: ',
's')
files = str2num(files);

%============  Get filename and load data
==============================

loadit = (['[filenameput, pathnameput] =
uigetfile(''',path_to_file,'\*.mat'',''Get Reference
File'',100,100);'])
eval(loadit)

ref_fileput = [pathnameput filenameput];

% Loads reference matrix (OpenMat) previously computed in
Openpos.m

getref = (['load ''', ref_fileput, ''';']);
eval(getref);

KinMatTemp = [];
Displacement = [];
RelClosure = [];
SyllableData = [];
syllable =[];
FileData = [];


fs = 60; %sampling rate

%%%
```

```
color = ['green  ';'red    '; 'cyan   ';'magenta';'yellow
'];

%_______GLOBAL VARIABLES DECLARATION _____________________

global ClosureTime Open DATAMAT BEGWIN ENDWIN syllable
directions ii chans mrows

%_____________________________________________________

winsize = fs * 10;

data = []; ydata = []; Movement = []; peak = []; Peak_Data
= []; Duration_Data = [];
Peak_Diff_Data = [];DataMat = []; Time_Data =
[];Peak_Diff_Open = [];
Peak_Diff_Close = [];Peak_Diff_Time_C = [];Peak_Diff_Time_O
= [];

for ii = 1:files % Loop 1

% Make RAW Data Matrix Global for "clostime.m".

   fileglobal = (['global DataMat', num2str(ii),';']);
   eval(fileglobal);

% Make TREATED Data Matrix Global

   getglobal =(['global KinMat',num2str(ii)]);
    eval(getglobal)


% Function call - getpeak5

DIR = (['directions = ''Get File ' num2str(ii),';''']);  %
Must specify for getpeak5.m
eval(DIR)

  getpeak5
%================  Subtract Lower Lip from Jaw to get
Lower Lip only ================
```

```
centjaw = (['Center_Jaw = detrend(KinMat',
num2str(ii),'(:,3));']);
eval(centjaw);

Correction_Jaw_Mark = Center_Jaw;

step2 = (['KinMat',num2str(ii),'(:,2) =
KinMat',num2str(ii),'(:,2) - Correction_Jaw_Mark;']);
eval(step2);

%Invert all signals - need to adjust for Peak Motus
translation

step3 =(['KinMat',num2str(ii),'(:,:) =
KinMat',num2str(ii),' * invert;']);  %Invert all signals -
need to correct for Peak Motus translation
eval(step3);

>>>>>>>>>>>>>>>>>> Define Segment >>>>>>>>>>>>>>>>>>>

%%%% Plot Kinematic signal -- to get gesture boundaries

tempkin = (['tempk = KinMat',num2str(ii),'(:,:);']);
eval(tempkin);

%%%%%%%%%  filter signal %%%%%%%%%%%%%%%%%%%%%%%%

fc = 25;

fcc = fc/((fs)/2);
[fb,fa] = butter(5,fcc);

FiltMat = [];
KinMatTemp = [];

for ff = 1:3
        home;
        KinMatTemp(:,ff) = filtfilt(fb,fa,tempk(:,ff));
end

%%%%%%%%%%%%%%%%%%%%%%% Invert UL%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%KinMatTemp(:,1) = KinMatTemp(:,1) .* (-1); %invert upper
lip signal

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

ITL = 0;
syllable = 1;
MultiSyllable = 1;

while MultiSyllable == 1; % MS Loop - if token contains
multiple syllables
while ITL == 0 % ITL (Intra Trail Loop)

%>>>>>>>>>>>>>>>> PLOT INITIAL WINDOW for user to choose
movement of interest >>>>>>>>>
% Instructions Window
     instructions = uicontrol(gcf,...
        'Style','text',...
        'BackgroundColor', color(2, : ),...
        'String', 'IDENTIFY SEGMENT THAT CONTAINS THE PEAKS
OF ALL MOVMENTS',...
        'Position',[0 scrsz(4)-70 1300 100]);

plot(KinMatTemp); grid on

%%%%%%%% User chooses gesture of interest - picks onset and
offset based on zero crossing of velocity trace

[onset y_on] = ginput(1);
[offset y_off] = ginput(1);

%%%%%%% Draw Window Around Selection

getaxis = axis;

x_window = [onset onset offset offset];
y_window = [getaxis(3) getaxis(4) getaxis(4) getaxis(3)];

line(x_window,y_window)

%%%%%%%%% Resize Kinematic and Velocity traces
```

```
window_pad = 2; % Pad window 2 points in front and 2 points
in back - to get upper lip onset and offset

BEGWIN = round(onset - window_pad);
ENDWIN = round(offset + window_pad);

% In case chosen point for onset is less than 5.
if(BEGWIN) < 1
BEGWIN = 1;
end

% In case offset is greater than matrix length.
if(ENDWIN) > length(KinMatTemp)
ENDWIN = length(KinMatTemp);
end

KinMatGest =  KinMatTemp(BEGWIN:ENDWIN,:);

%close; %% Close Figure

forward = menu('WHAT UP?','Accept - Next' ,'Re-Do', 'Accept
- Stay '); %%%%% If 'no' overwrite last cycle measured if
not satisfied
  if   forward == 1


     MultiSyllable = 0; %Token only contains one syllable
     ITL = 1; % End intratoken loop

     syl_position = menu('Postion','Final','Medial',
'Initial'); %%%%% If 'no' overwrite last cycle measured if
not satisfied
       if   syl_position == 3
         syllable = 'I';

       elseif  syl_position == 2
         syllable = 'M';

       else    syl_position == 1
         syllable = 'F';

       end
```

```
      %% Function call

      clostime  %%% Function to get time of minimum in
Interlabial Distance

   elseif forward == 2 % A chance to redefine segment

      syllable = [];
   else forward == 3

            % Indicate Syllable Position in Utterance

      syl_position = menu('Postion','Final','Medial',
'Initial'); %%%%% If 'no' overwrite last cycle measured if
not satisfied
         if   syl_position == 3
            syllable = 'I';
            MultiSyllable = 1; %Continue to loop
         elseif  syl_position == 2
            syllable = 'M';
            MultiSyllable = 1; %Continue to loop
          else    syl_position == 1
             syllable = 'F';
             MultiSyllable = 0; %End MultiSyllable Loop
         end

         %% Function call

         clostime  %%% Function to get time of minimum in
Interlabial Distance


      end

SyllableData = [SyllableData; syllable];

num2str(ii + 1)

OralClosure = OpenMat - KinMatGest(ClosureTime,:);

if OralClosure(1) < 0 % If position of articulator is more
in the open position than reference
```

```
   OralClosure(1) = 0; % ,due to labial compression, make
position contribute "0" to oral closure.
   else
   for i = 2:3
      if OralClosure(i) > 0
         OralClosure(i)= 0;
      elseif OralClosure(i) < 0;
         OralClosure(i) = abs(OralClosure(i));

      end
   end
end


SumOfClose = sum(OralClosure);

RelCloseMov = (OralClosure/SumOfClose) * 100;

Displacement = [Displacement; OralClosure]; % Store results
from multiple files
RelClosure = [RelClosure; RelCloseMov]; % Store relative
closure computations from multiple files

FileData = [FileData,ii];
     end % (ITL) Intra trial loop
  end % END MultiSyllable Loop
end  % END for Loop1

   data1 = [RelClosure];
   SyllableData;
```

Clostime

```
%
% Finds time of Maximum Oral Closure (MOC) by computing the
interlabial distance function and finding
% the minimum. Point(time) of MOC is stored in
"ClosureTime".
% No User Input Required
```

```
%
% For no interpolation set "Interp_Factor" to "1".
% Modified from LIPDIS
%
% Written by Jordan Green 6/27/97
% Modified from LIPDIS to compute artiuclatory
displacements 9/10/97

%%%
```

```
    CloseKinMat = []; % Create empty matrix

    %Get all y vectors

         for jj = 1:5  %% Temporary

      getcolumn = (jj * 3) - 1;
      place_hold = WorkMat(:,getcolumn);

        %Put all y vectors into one matrix

      CloseKinMat = [CloseKinMat place_hold];

    end

    %Delete Reference Point coordinates

    CloseKinMat(:,1:2) = [];

%%%%%%%%%  filter signal %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

for cc = 1:3
fc = 30;
fcc = fc/(SampRate/2);
[fb,fa] = butter(5,fcc);

CloseKinMat(:,cc) = filtfilt(fb,fa,CloseKinMat(:,cc));
end

%=============================================================
=

%Invert all signals - just for graphic - need to adjust for
Peak Motus translation

CloseKinMat(:,:) = CloseKinMat * (-1); %    all signals -
need to correct for Peak Motus translation

%=============================================================
=
Jaw_D = diff(CloseKinMat(:,3));  %Calculate derivative and
pad with zero
```

```
ul_i = CloseKinMat(:,1);

ll_i = CloseKinMat(:,2);

InterLabDis_I = abs(ul_i - ll_i);

InterLabDis_I = InterLabDis_I(BEGWIN:ENDWIN,:);

%%%%%% Within Trial loop - Articulator %%%%%%%%%%%%%%%%%%

     KinPlot = plot(InterLabDis_I);
     xlabel('Points');ylabel('Displacement (cm)');grid on

     i = 1; %For graphics

     %Create frame at top of graph to indicate channel
being measured

     channel = ['InterLabial Distance'];
     number = ['File ' num2str(j)];
     syllable_G = ['Syllable Position: ' syllable];

     % Label articulator being measured
     ft_chan = uicontrol(gcf,...
        'BackgroundColor', color(i, : ),...
        'Style','text',...
        'String', channel(i , : ),...
        'Position',[0 scrsz(4)-70 500 100]);
     hold on

              ft_chan = uicontrol(gcf,...
        'BackgroundColor', color(i, : ),...
        'Style','text',...
        'String', number,...
        'Position',[700 scrsz(4)-70 100 75]);



         ft_chan = uicontrol(gcf,...
        'BackgroundColor', color(i, : ),...
        'Style','text',...
        'String', syllable_G,...
```

```
        'Position', [0 .80 .10 .1],...
        'Units', 'Normalized');

%================ Get Minimum and Max Oral Closure
=============================================================

     % Get MIN in InterLabial Distance

     [Min_Data min_x] = min(InterLabDis_I);

     [Peak_Data peak_x]= max(InterLabDis_I);

 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

ClosureTime = min_x;  % Global Variable for OralClos.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear DataMat

close
```

General Purpose Programs

```
% Getpeak5 - Get Peak Motus file that contains 5 markers (2
reference and 3 movement)
% This program loads and formats Peak Motus files
% User needs to specify
% 1) the number of files to load in "files"
% 2) directions for menu
% Output matrix is KinMat# - '#' indicating the file number
if multiple files are loaded

global files directions mrows chans path_to_file

%============  Get filename and load data
==============================

loadit = (['[filenameget, pathnameget] =
uigetfile(''',path_to_file,'\*.csv'',directions,100,100);']
)
eval(loadit)
```

```
   file = [pathnameget filenameget];
    getfiles = (['DataMat',num2str(ii) ,'=
csvread(file);']);
   eval(getfiles);

   % Make RAW Data Matrix Global for "clostime.m".

   fileglobal = (['global DataMat', num2str(ii),';']);
   eval(fileglobal);

   % Make TREATED Data Matrix Global

   getglobal =(['global KinMat',num2str(ii)]);
        eval(getglobal)

        ii
   sizeit = (['[mrows chans] =
size(DataMat',num2str(ii),');']);
   eval(sizeit);

   %%  Format the file to create matrix with only the y
vectors
   %-----------------------------------------------------
------

   emptymat = (['KinMat',num2str(ii),' = [];']); % Create
empty matrix
   eval(emptymat);

   %Get all y vectors

        for jjj = 1:5  %% Temporary

     getcolumn = (jjj * 3) -1;
          place = (['place_hold =
DataMat',num2str(ii),'(:,getcolumn);']);
     eval(place);

       %Put all y vectors into one matrix

     Kin = (['KinMat',num2str(ii) ,' = [KinMat',
num2str(ii),', place_hold];']);
     eval(Kin);
```

```
end

DelRef = (['KinMat',num2str(ii),'(:,1:2) = [];']);
eval(DelRef)      %Delete Reference Point coordinates
```

Jordan R. Green, M.A. CCC-SLP

406 N 45th St.
Seattle, WA 98103
206-685-2205 (wrk) 206-362-2721(hm) 206-543-1093 (fax)
E-mail: greenj@u.washington.edu

EDUCATION

University of Washington, Seattle	predoctoral		*Sept '93 - Present*
University of Iowa, Iowa City	predoctoral		*Jan '95 - May '95*
California State University, Chico	M.A.	Speech-Language Pathology	*1991*
California State University, Chico	B.A.	Psychology	*1988*

EXPERIENCE

Children's Hospital and Health Center, San Diego, CA *April '91 - June '93*
SPEECH-LANGUAGE PATHOLOGIST/Researcher:

Clinical responsibilities included diagnosis and treatment of a variety of pediatric cognitive/communication disorders, and supervision of clinical fellows. Served as primary speech-language pathologist on Cleft Palate, Dysphagia, and Pediatric Rehabilitation Teams.

University of California, Davis Medical Center *June '91 - Dec '91*
Department of Otorhinolaryngology
TRAINEE AND RESEARCH ASSISTANT:

Selected to participate in therapy and research efforts of Dr. Rebecca Leonard Director of Speech Language Division of U.C. Davis Otorhinolaryngology Clinic.

PREDOCTORAL ACADEMIC POSITIONS

NIH Research Training Fellowship	*Sept '93 - May '95*
Research Assistantship	*June '95 - May '96*
NIH Research Training Fellowship/Teaching Assistantship	*Sept '96 – June '97*
Research Assistantship	*June '97 - Present*

PUBLICATIONS

Green, J.R., Moore, C.A., Ruark,J.L., Rodda, P.R., Morvee, W. & VanWitzenburg, M. (1997). Development of Chewing in Children from 12 to 48 Months: a longitudinal study of EMG patterns. *Journal of Neurophysiology, 77,* 2704-16.

Green, J.R., Buder, E.H., Rodda, P.R, & Moore, C. A. (1997). Reliability of measurement across several acoustic voice analysis systems. In M. Cannito, K. Yorkston, & D. Beukelman (Eds.) *Neuromotor Speech Disorders: Nature, Assessment, & Management.* Baltimore: Paul H. Brookes Publishing Co.

Minifie, F.D., Green, J.R., Smith, J., Huang, D.Z. Inter-judge variability in perception of voice qualities. *Proceedings of The International Congress of Phonetic Sciences, Stockholm, Sweden, 1995, 758-761.*

Minifie, F.D., Huang, D.A., & Green, J.R. Relationship between acoustic measures of vocal perturbation and perceptual judgments of breathiness, harshness, and hoarseness. *Proceedings of The International Conference on Spoken Language Processing. Yokohama, Japan, 1994, 1999-2002.*

CONTRIBUTED PRESENTATIONS

Green, J.R., Moore, C.A., Higashikawa, M. & Steeve, R. (1998). Development of Lip and Jaw Coordination for Production of Bilabial Consonants. Conference on Motor Speech, Tucson, AZ, Jan. 29 – 4.

Green, J.R., Moore, C.A., & Rodda, P.R. (1997). Longitudinal Study of Chewing in Children Aged 12 to 48 Months. ASHA's Annual Convention, Boston, MA, November 20-23.

Green, J.R., Buder, E.H., Rodda, P.R., & Moore, C.A. (1996). Reliability of measurement across several acoustic voice analysis systems. Conference on Motor Speech. Amelia Island, Florida, February 22-25.

Minifie, F.D., Green,J.R., Smith, J., Huang, D.Z. (1995). Inter-judge variability in perception of voice qualities. International Congress of Phonetic Sciences, Stockholm, Sweden, Aug. 13-19.

Minifie, F.D., Huang, D.A., & Green, J.R. (1994). Relationship between acoustic measures of vocal perturbation and perceptual judgments of breathiness, harshness, and hoarseness. International Conference on Spoken Language Processing. Yokohama, Japan, September 18-22.

RESEARCH AWARDS

Children's Hospital and Health Center. " Speech Duration Variances Between Cleft Palate Children With Nasal Emissions and Non Cleft Palate Children" (Total direct costs: $ 4104.00) principle investigator, *1992-1993.*

TEACHING EXPERIENCE

Teaching Assistantship (University of Washington) *Sept '96 - June '97*
Speech Science (SPHSC 300)

Departmental Tutor (California State University Chico) *Aug '89 - Dec '89*
Provided assistance to students in areas of speech science, speech perception, and acoustics.

Guest Lectures

University of Washington

Topic: *Source Filter Theory* *1998*
Class: *Speech Science (Instructor: Chris Moore)*

Topic: *Cleft lip and palate* *1996*
Class: *Neurological Bases of Speech and Language (Instructor: Edythe Strand)*

Topic: *Speech acoustics* *1996*
Class: *Speech Science (Instructor: Chris Moore)*

Topic: *Signal processing for speech analysis* *1996*

Class: *Speech Science (Instructor: Chris Moore)*

Topic: *Central nervous system demonstration* *1995*

Class: *Neurological Bases of Speech and Language (Instructor: Edythe Strand)*

Topic: *Experimental design (Instructor: Fred Minifie)* *1994 & 1995*

Class: *Research Methods in Speech and Hearing Sciences*

Topic: *Lectured to graduates and undergraduates on the role of the basal ganglia in motor control.* *1994*

Class: *Neurological Bases of Speech and Language (Instructor: Edythe Strand)*

Children's Hospital and Health Center, San Diego

Topic: Speech Duration Variances Between Cleft Palate Children With Nasal Emissions and Non Cleft Palate Children *1993*

Topic: Treatment of Speech Motor Disorders in Children *1992*

California State University Chico

Topic: Clinical applications of Mac Speech Lab *1990*
Class: Organic Disorders (Instructor: Bob Hall)

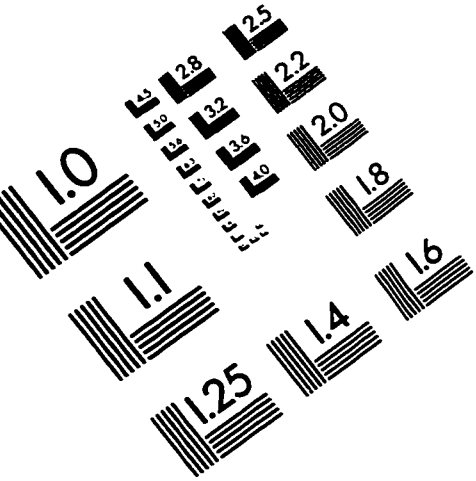
CREDENTIALS

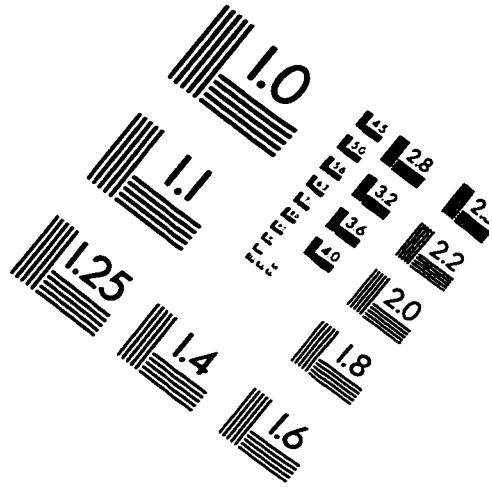
Certificate of Clinical Competence

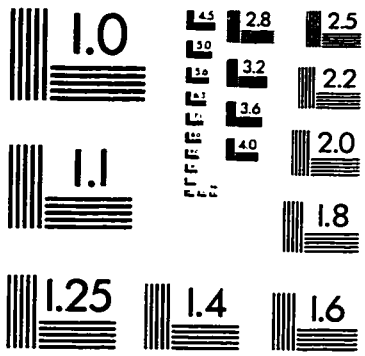
ORGANIZATIONS

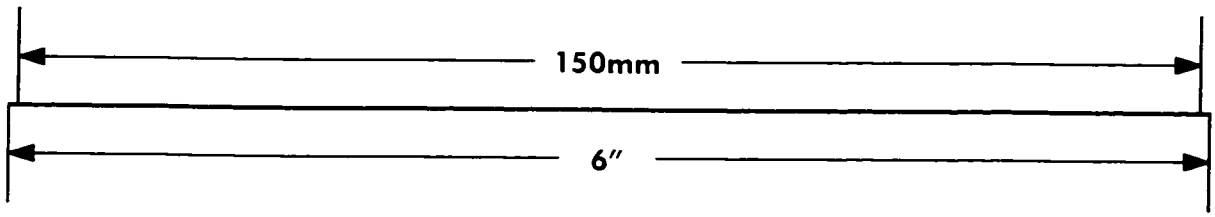
American Speech and Hearing Association
Acoustical Society of America

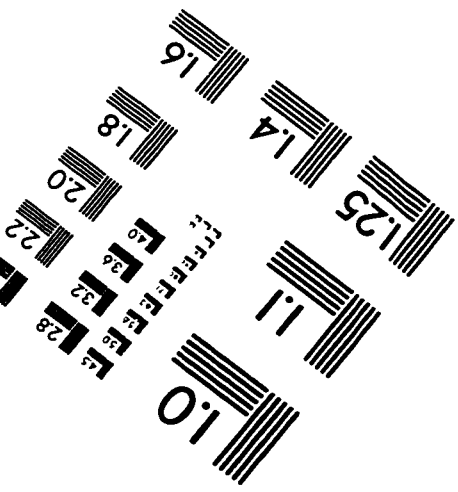
IMAGE EVALUATION TEST TARGET (QA-3)

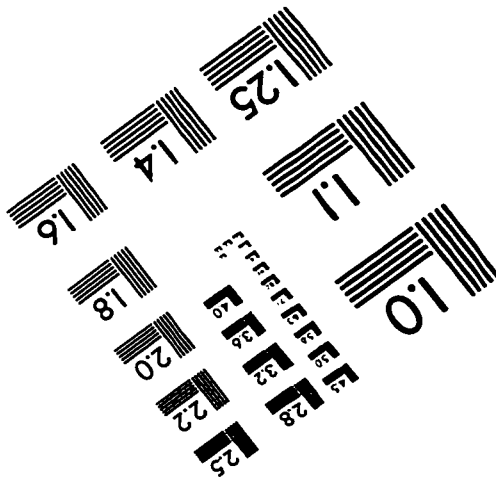












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved

99

07907

UMI

MICROFILMED 1998

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

The Composition of Near-Earth Objects

by

Mark Hammergren

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

1998

Approved by Donald B Brownlee
Chairperson of Supervisory Committee

Program Authorized
to Offer Degree ASTRONOMY

Date AUGUST 20, 1998

UMI Number: 9907907

Copyright 1998 by
Hammergren, Mark

All rights reserved.

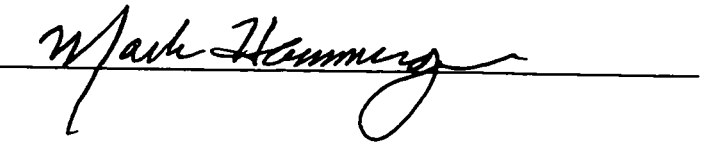
UMI Microform 9907907
Copyright 1998, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized
copying under Title 17, United States Code.

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

© Copyright 1998
Mark Hammergren

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable for scholarly purposes, consistent with "fair use" as prescribed in the U. S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P. O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date August 20, 1998

University of Washington

Abstract

The Composition of Near-Earth Objects

by Mark Hammergren

Chairperson of the Supervisory Committee
Professor Don Brownlee
Department of Astronomy

I present reflectance spectra of 34 near-Earth objects (NEOs), 6 main-belt asteroids, and four non-NEO cometary candidates, all obtained with the Apache Point Observatory 3.5m telescope + Double Imaging Spectrograph. The spectra cover the wavelength range 3800 - 10,000 Å, encompassing regions of mineralogically important absorption features. Nearly all of the NEOs observed display ultraviolet and near-infrared absorptions characteristic of rock-forming silicate minerals. Of the 27 NEOs belonging to the S or Q taxonomic classes observed in this study, 15 are spectrally indistinguishable from ordinary chondrite meteorites. I perform extensive Monte Carlo simulations of the NEO and main belt populations aimed at quantifying the severe biases affecting observed taxonomic distributions. The bias-corrected NEO population in the 1 - 10 km diameter range is composed of $67 \pm 13\%$ S- or Q-type objects, and $30 \pm 7\%$ C-types objects, with the remainder being primarily of the spectrally degenerate X-class. The NEO population resembles most closely that of the inner main belt near the 3:1 mean motion resonance, and is consistent with that region being the sole source for NEOs. If extinct comet nuclei resemble the primitive taxonomic classes C, P, or D, the cometary component of the NEOs is constrained at $\lesssim 30\%$. I investigate trends of S-type spectral characteristics with size. The strength of the 1 μm absorption increases with decreasing size. There is a possible trend towards shorter-wavelength band centers with decreasing size, possibly reflecting a decreasing olivine abundance in the optically active surface fraction. For sizes below about 6 km, the spectral continuum reddens with increasing size. Above

6 km, this trend reverses, and larger objects have on average bluer continua. For the smaller objects, these trends are all consistent with predictions of the "space weathering" hypothesis. The bluer continua of the larger objects remains unexplained. All of these trends appear to be systematically dependent on size; specifically, no distinct separation exists between the larger S-type objects and the smaller ordinary chondrite-like bodies.

Contents

List of Figures iii
List of Tables vi

1 Introduction **1**
1.1 NEOs: Unanswered Questions 1
1.1.1 Origins 1
1.1.2 The Meteorite / Asteroid Connection 2
1.1.3 Space Weathering and the Ordinary Chondrite Problem . 2
1.1.4 Planetary and Biological Evolution 3
1.1.5 Future Hazards and Resources 4
1.2 Reflectance Spectroscopy 4
1.3 Asteroid Taxonomy 5

2 Data Acquisition and Processing **7**
2.1 Object Selection 7
2.2 Observations 7
2.3 Data Reduction 11
2.4 Solar Analog Stars 12
2.5 Possible Sources of Systematic and Random Error 13
2.5.1 Bias Subtraction 14
2.5.2 Flat Fielding 14
2.5.3 Poor Subtraction of Night Sky Lines 15
2.5.4 Poor Choice of Solar Analog? 17
2.5.5 Misregistration of Object and Solar Analog Spectra 18

2.5.6 Improper Correction for Telluric Absorption 20
2.6 Other Factors Which May Affect Spectra 20
2.6.1 Differential Refraction 20
2.6.2 Phase Reddening 21

3 Summary of Results **22**
3.1 Taxonomic Classification 31

4 Discussion of Results for Individual Objects **35**
4.1 NEOs 35
4.1.1 Spectra similar to ordinary chondrite meteorites 35
4.1.2 Unusual spectra with UV- and 1-μm absorptions 36
4.1.3 3103 Eger 36
4.2 Mars Crosser – 2078 Nanking 37
4.3 Non-NEO Cometary Candidates 37

5 Spectral Trends Among S-type Objects **38**
5.1 Space Weathering 38
5.2 Modified Gaussian Model Fits 39
5.3 Spectral Trends with Diameter 41

6 Debiasing Asteroid Taxonomic Distributions **45**
6.1 Introduction 45
6.2 Debiasing Methodology 47
6.3 The Detection Simulation 48

7 Discussion of Debiasing Results **64**

8 Conclusions **104**

References **106**

List of Figures

2.1 Plot of DIS blue CCD bias "ripple" 14
2.2 DIS red CCD flat field image 16
2.3 Reflectance spectrum of 1627 Ivar displaying fringing "ripples" . . 17
2.4 Ratios of Solar Analog Stars 18
2.5 Spurious features introduced by misregistration of spectra 19

3.1 NEO reflectance spectra 23
3.2 NEO reflectance spectra 24
3.3 NEO reflectance spectra 25
3.4 NEO reflectance spectra 26
3.5 NEO reflectance spectra 27
3.6 NEO reflectance spectra 28
3.7 Main-belt asteroid reflectance spectra 29
3.8 Reflectance spectrum of Mars-crossing asteroid 2078 Nanking . . . 30
3.9 Non-NEO cometary candidate asteroid reflectance spectra 31

5.1 Modified Gaussian Fits to NEO Spectra 40
5.2 1-μm Band Strength vs. Size 42
5.3 1-μm Band Center vs. Size 43
5.4 Visual Slope vs. Size 44

6.1 Probability of potential observation 49

6.2 Dependence of potential observation on orbital elements for Spacewatch-type detection criteria 50
6.3 Taxonomic completenesses for the Atens, Apollos, and Amors. . . 56
6.4 Taxonomic completenesses for the Hungaria Group, Mars Crossers, and Flora Family. 57
6.5 Taxonomic completenesses for the Phocaea Group, Nysa Family, and Main Belt I. 58
6.6 Taxonomic completenesses for the Pallas Zone, Main Belt IIa, and Main Belt IIb. 59
6.7 Taxonomic completenesses for the Koronis Zone, Eos Zone, and Main Belt IIIa. 60
6.8 Taxonomic completenesses for the Themis Zone, Griqua Group, and Main Belt IIIb. 61
6.9 Taxonomic completenesses for the Cybele Group, Hilda Group, and Trojan Group. 62
6.10 Taxonomic completenesses for the Vesta Family. 63

7.1 Cumulative Size-Frequency Diagram for the Atens 65
7.2 Cumulative Size-Frequency Diagram for the Apollos 66
7.3 Cumulative Size-Frequency Diagram for the Amors 67
7.4 Cumulative Size-Frequency Diagram for the Hungaria Group . . . 68
7.5 Cumulative Size-Frequency Diagram for the Mars Crossers 69
7.6 Cumulative Size-Frequency Diagram for the Flora Family 70
7.7 Cumulative Size-Frequency Diagram for the Phocaea Group . . . 71
7.8 Cumulative Size-Frequency Diagram for the Nysa Family 72
7.9 Cumulative Size-Frequency Diagram for the Vesta Family 73
7.10 Cumulative Size-Frequency Diagram for Main Belt Zone I 74
7.11 Cumulative Size-Frequency Diagram for the Pallas Zone 75
7.12 Cumulative Size-Frequency Diagram for Main Belt Zone IIa . . . 76
7.13 Cumulative Size-Frequency Diagram for Main Belt Zone IIb . . . 77

7.14 Cumulative Size-Frequency Diagram for the Koronis Zone 78
7.15 Cumulative Size-Frequency Diagram for the Eos Zone 79
7.16 Cumulative Size-Frequency Diagram for Main Belt Zone IIIa . . . 80
7.17 Cumulative Size-Frequency Diagram for the Themis Zone 81
7.18 Cumulative Size-Frequency Diagram for the Griqua Group 82
7.19 Cumulative Size-Frequency Diagram for Main Belt Zone IIIb . . . 83
7.20 Cumulative Size-Frequency Diagram for the Cybele Group 84
7.21 Cumulative Size-Frequency Diagram for the Hilda Group 85
7.22 Cumulative Size-Frequency Diagram for the Trojan Group 86
7.23 Cumulative Size-Frequency Diagram for the NEOs 87
7.24 Taxonomic Gradient Across the Main Belt 89
7.25 Wave-like Oscillation in the Size-Frequency Distribution of S-type NEOs 103

List of Tables

2.1 Observational Circumstances for NEOs 9
2.2 Observational Circumstances for non-NEOs 11
2.3 Selected Properties of Solar Analog Stars 13

3.1 Physical parameters for the 34 NEOs observed 32
3.2 Physical parameters for the non-NEOs observed 33

6.1 Selected Average Parameters for 12 Taxonomic Classes 51
6.2 Orbital Element Zones 54

7.1 Bias-corrected Taxonomic Fractions 90
7.2 Bias-corrected Taxonomic Fractions for the NEOs (excluding X-types) 101
7.3 Bias-corrected Taxonomic Fractions for the NEOs (if all X-types are P-type) 102

Acknowledgements

I would like to thank everyone, everywhere, who helped me get to this place in my life. Folks of particular note include the following: my mother and brother, for their support throughout my life; Bill Baum, the consummate gentleman scientist; Don Brownlee, for his advice and for those great conversations on science, rockhounding, explosions, and beer; and Paul Hodge, for the example of his comfortable competence. Thanks too to Chris Stubbs and Craig Hogan for helping me out. My fellow students over the years, friends all, deserve special thanks for making this an enjoyable experience. To mention any names here would run the risk of leaving others off – and I can not bear to insult any by omission. Hail and well-met to my compatriots from Amnesty International Group 94! And finally, I must thank that most special institution whose threshold I crossed every time with joyful, thirsty anticipation: the College Inn Pub. Long may it serve astronomy grads!

To all my friends. You know who you are.

Chapter 1

Introduction

It is a cursed evil to any man to become so absorbed in any subject as I am in mine.

— Charles Darwin

On August 13, 1898, Gustav Witt in Berlin discovered the first of a new class of solar system objects. Later named 433 Eros, this apparently asteroidal object was notable for its orbit, which, unlike the more numerous population of main-belt asteroids which remain between the orbits of Mars and Jupiter, occasionally brings the object very close to the Earth. In the succeeding century, more than 500 planet-crossing peers of 433 Eros have been discovered. Known collectively today as the Near-Earth Objects (or "NEOs"), they remain the subject of intense scientific and popular interest for a simple and dramatic reason: NEOs have in the past, and will again in the future, strike the terrestrial worlds.

1.1 NEOs: Unanswered Questions

Knowledge of the distribution and composition of NEOs is becoming of increasing importance for addressing questions in many areas, including the following:

1.1.1 Origins

Due to interactions and impacts with the terrestrial planets, NEOs have a dynamical lifetime of $\sim 10^7$ - 10^8 years (Wetherill 1974) - much shorter than the age of the solar system. Since the cratering rate - and by implication the number

of NEOs – has remained nearly constant or even increased over the past 3.2 billion years (Grieve & Shoemaker 1994), the numbers of NEOs must continuously be replenished from some longer-lived reservoir located elsewhere in the solar system.

Two of the most commonly suggested sources include main-belt asteroids propelled by collisions into orbital resonances with Jupiter, and then perturbed into planet-crossing orbits, or cometary nuclei which have exhausted their supplies of volatiles due to repeated passages through the inner solar system ("extinct" comets). Most main-belt asteroids and comets appear to be compositionally distinct (Luu & Jewitt 1990). Compositional information on individual NEOs, and the distribution of taxonomic types among the NEOs, should provide constraints on the fraction of NEOs that may be extinct comets, as well as allow comparisons to main belt asteroids (Xu et al. 1995, Luu & Jewitt 1990).

1.1.2 The Meteorite / Asteroid Connection

The investigation of the relationship between meteorites and asteroids, their probable parent bodies, is a strong reason for studying the composition of asteroids. Meteorites have great scientific value since they may be studied in detail in a laboratory, and except for samples of the Moon and interplanetary dust particles, they represent the only extraterrestrial material we have ever examined. Detailed chronologies of chemical and physical processes in the early solar system have been derived from meteorites. Knowledge of specific meteorite parent bodies would then provide us with spatial information, a "map" to the locations of these physical processes. Since NEOs are already in planet-crossing, Earth-approaching orbits, they are in a particularly favorable location for the delivery of meteorites to the Earth. Dynamical studies have further suggested that a few well-placed parent bodies may dominate the flux of meteorites on the Earth (Greenberg & Nolan 1989). NEOs may thus be a very important connection in the transport of meteorites to the Earth.

1.1.3 Space Weathering and the Ordinary Chondrite Problem

The extreme rarity of spectral analogs for the ordinary chondrite meteorites, which make up more than 75% of terrestrial falls, is one of the most significant

and puzzling problems in solar system astronomy. More perplexing still, only one main-belt asteroid out of an observed sample of nearly a thousand appears to match an ordinary chondrite, while even though far fewer have been characterized, at least eight NEOs of similar colors have previously been observed (Xu et al. 1995, Binzel et al. 1996, Hicks et al. 1998).

Space weathering has been invoked as one way of solving the ordinary chondrite dilemma. In this hypothesis, the parent bodies of ordinary chondrites have been disguised by some unknown surface processing to appear as the very common S-type asteroids (cf., Lipschutz et al. 1989). NEOs are typically smaller than main-belt asteroids; additionally, since NEOs pass much nearer to the Earth than main-belt asteroids, very much smaller objects are occasionally observable. Smaller objects will have lower surface gravities and are generally predicted to retain less regolith (impact-generated "soil"). If some regolith process is responsible for diguising asteroid spectra, it should be less effective on NEOs than main-belt asteroids. Studies of NEO surface composition should thus contribute to our understanding of space weathering on all asteroids.

1.1.4 Planetary and Biological Evolution

Craters resulting from NEOs and other planet-crossing objects are the dominant landforms on the Moon, Mercury, Mars, and many other solar system bodies (Melosh 1980), and more than 140 impact sructures have been identified on the Earth (Grieve & Shoemaker 1994). It is thought that most of the volatiles present on the surface of the Earth (i.e., the atmosphere and oceans) were delivered after the formation of the planet by the accretion of asteroids and comets (Hunten 1993, Chyba 1990). It is possible that much of the Earth's storehouse of organic material – the basis for life – arrived in the same manner (Marcus & Olsen 1991). These theories are highly dependent on our understanding of NEO composition.

There is increasing evidence that the Cretaceous-Tertiary mass extinction was caused by the impact of an asteroid or comet (Grieve & Shoemaker 1994). At least one composition–dependent extinction mechanism – poisoning of the biosphere by meteoritic heavy metals – has been explored (Davenport et al. 1990).

1.1.5 Future Hazards and Resources

The recent collision of Comet Shoemaker-Levy 9 with Jupiter underscored the fact that planetary impacts still occur. Studies have been undertaken to determine the threat that such impacts pose to the Earth (e.g., "the Spaceguard Survey," Morrison et al. 1992). The probability that an as-yet undiscovered NEO will hit the Earth is estimated at one chance in a thousand during the lifetime of an average American. Knowledge of a potential impactor's composition would allow us to better estimate its mass, and thus destructive potential, and its likelihood to penetrate the Earth's atmosphere (Shoemaker et al. 1995, Hills & Goda 1993).

NEOs are also important resources for the future exploration of space. Pieces of asteroids, obtained on the ground as meteorites, are seen to be composed of a wide variety of materials including iron–nickel alloy, small amounts of precious metals, organic chemicals, water of hydration, and chemically-bound oxygen. Furthermore, many NEOs are accessible with less energy than required by a Moon landing, making them the cheapest targets beyond the Moon for robotic exploration, and by far the easiest for human exploration (Shoemaker et al. 1995).

1.2 Reflectance Spectroscopy

Reflectance spectroscopy has emerged as a powerful tool for probing the surface compositions of small solar system objects. Pioneering work in the remote sensing of solar system body composition by McCord et al. (1970) was paralleled by fundamental developments in the crystal field theory of mineral absorption features (Burns 1970a) and in the laboratory determination of diagnostic mineral absorption features (Adams & Filice 1967; Adams 1974, 1975; Hunt & Salisbury 1970). Pieters & McFadden (1994) recount a detailed history of the use of reflectance spectroscopy for meteorite and asteroid studies. These related lines of research paved the way for the utility of reflectance spectroscopy in remote sensing.

The slope of the spectral continuum, and the presence or absence of certain solid-state absorption features can provide specific mineralogical information (Burns 1983). For example, the common rock-forming silicate minerals pyroxene

and olivine show strong absorption features near 1 μm. Pyroxene has another strong feature further in the infrared near 2 μm, while such a feature is weak or nonexistent in spectra of olivine. Furthermore, Adams (1974) showed that the wavelength positions of the silicate absorption band minima are dependent on Fe, Mg, and Ca composition. Gaffey et al. (1993) provide a review of progress in the field, and summarize several mineralogical properties accessible to reflectance spectroscopy.

While crystal field theory can provide some background for the interpretation of reflectance spectra, analyses are more often dependent on an empirical comparison to laboratory spectra of material analogs or to a study of the general variations in spectra across an object (eg., studies of the Moon by Charette et al. 1974) or within a population of objects (e.g., differences among the S-type asteroids by Gaffey et al. 1993). A common practice in asteroid studies is to compare asteroid reflectance spectra to those of meteorites obtained in the laboratory, though arguments based on such comparisons are extremely controversial (see, for example, Section 1.1.3 on the ordinary chondrite dilemma, and Gaffey et al. 1993 for more information).

Although compositional surveys of NEOs are of great scientific value, these fast-moving, faint objects pose considerable observational difficulties, and few concerted studies have been attempted (Howell 1994, Hicks et al. 1998). Spectroscopic surveys of main-belt asteroids occasionally pick up some NEOs as targets-of-opportunity, when individual NEOs have particularly favorable apparitions near the Earth (Xu et al. 1995, Binzel 1996). Other programs have returned data of low signal-to-noise (McFadden et al. 1984), low spectral resolution (Zellner et al. 1985), or limited spectral range (Luu & Jewitt 1989).

Since ground-based remote sensing will always be far less expensive than space missions, reflectance spectroscopy will remain the most common method of studying the compositions of a population of objects such as the NEOs.

1.3 Asteroid Taxonomy

Taxonomy is the process by which a population of objects or organisms are split into unique categories, on the basis of perceived or actual similarities and

differences between and among these objects. It is important to remember that *perceived* similarities or differences are not necessarily indicative of genetic or compositional relationships. (This is akin to the distinction in the biological sciences between phenotyping and genotyping.)

The asteroid taxonomic system in most widespread current use is the "Tholen system" (Tholen & Barucci 1989), in which a cluster and principal component analysis were applied to Eight Color Asteroid Survey (ECAS) data along with corresponding IRAS asteroid albedos. The 14 classes in the original "Tholen system" have been supplemented by other researchers with at least four other classes, some of which contain only a single member (Tholen & Barucci 1989; Gaffey et al 1993). Other classification schemes, including the "G-mode analysis" (Barucci et al. 1987), the "Three-Parameter system" (Tedesco et al . 1989), and a neural network analysis (Howell et al. 1994) have suffered from a disadvantage common with the Tholen system, namely, that in order to formally classify a single new object, one must re-run the analysis on the entire set. All of these taxonomies use the relative fluxes of asteroids at a small number of wavelength bands as inputs; none intelligently apply information on solid-state absorptions gained from reflectance spectroscopy theory, and none make full use of the higher resolutions offered by reflectance spectroscopy. It is obvious that asteroid taxonomy needs to be refined to make it relevant to current technologies and theory, but it is beyond the scope of this work to carry out this task.

* * *

In this work I present moderate resolution reflectance spectra of 34 NEOs. The data acquisition and reduction of these spectra are described in Chapter 2. The spectra are presented in Chapter 3. In Chapter 4, I discuss individual objects and the implications of the spectral observations. In Chapter 5, I compare this work's sample of 34 NEO reflectance spectra with previously published NEO and main-belt asteroid reflectance spectra, with a particular view towards describing spectral trends with object size. In Chapter 6, I describe a new model for bias-correcting observed taxonomic distributions among the various NEO and asteroid populations. The results of the application of this bias correction are presented in Chapter 7. Chapter 8 summarizes the conclusions of this work.

Chapter 2

Data Acquisition and Processing

It seems to me that those sciences which are not born of experience, the mother of all certainty, and which do not end in known experience - that is to say, those sciences whose origin or process or end does not pass through any of the five senses - are vain and full of errors.

— Leonardo da Vinci

2.1 Object Selection

The NEOs (and other objects) observed during the course of this survey were selected from a catalog of $\sim 30,000$ asteroids for which reasonably good orbits were known (Bowell 1998). The main criteria for inclusion in the NEO survey were brightness and visibility at the Apache Point Observatory (APO). The faint magnitude limit used in the selection process was typically $V \sim 18.5$. Brighter objects were usually given a higher priority, since these objects were often those making particularly close approaches to the Earth. Objects with solar elongations less than 60° and those at minimum airmasses greater than 2.0 were generally excluded, although exceptions to these guidelines were sometimes made for high priority targets. On any given night, about 10 - 20 objects passed all the above selection criteria. No further attempt was made to define a strictly absolute-magnitude-limited sample, since it was felt that this would result in an impractically small number of candidates.

2.2 Observations

Spectra of a total of 34 near-Earth objects, 6 main-belt asteroids, 1 Mars-crosser, 4 non-NEO cometary candidates, and three comets were obtained during the

course of this research. The comet spectra will be discussed in another work. All of the spectra were taken using the Double Imaging Spectrograph (DIS) on the 3.5 m telescope at APO. Table 2.1 presents the observational circumstances for these objects. Table 2.2 presents the same data for the non-NEOs.

The DIS uses a dichroic filter with a transition wavelength of 5350 Å to split incoming light into separate red and blue beams. For spectroscopy, these beams are fed to two different gratings which are optimized for the appropriate wavelength regime, whereas for imaging, mirrors are put in place of the gratings. The red light is then sent to a 800 × 800 pixel TI CCD, and the blue light to a thinned, UV-coated 512 × 512 pixel SITe CCD. In low-resolution mode, the blue side utilizes a 150 line mm^{-1} grating resulting in a dispersion of about 6.3Å per pixel, and the red side a 300 line mm^{-1} grating with a dispersion of about 7.0 Å per pixel.

The DIS is described in more detail in Kundic et al. 1995 and on the APO DIS homepage at `http://www.apo.nmsu.edu/Intruments/DIS/`.

For the observations made prior to 08/19/96, a $6' \times 3''$ wide acetate slit was used. During these exposures, no type of guiding was possible, and individual exposures were limited to ~5–15 minutes to limit telescope drift. For later observations, a $6' \times 2''$ aluminized glass slit was used to allow the employment of a slit camera. Longer exposure times were then possible since the slit camera could be used to ensure the object did not drift out of the slit.

The predicted positions of these objects and the telescope pointing accuracy were generally not good enough for immediate identification on the basis of their apparent locations in images. Most objects were instead identified by their apparent motion between successive images. Both the amount and direction of apparent motion were used in ensuring that the moving object seen was indeed the intended target. The identification image exposures were short enough that no trailing was evident even for the fastest-moving NEOs observed.

During spectral exposures, the telescope was driven at the predicted apparent rates of the target objects. After the installation of a slit camera in DIS and the use of the $2''$ aluminized slit, occasional slit camera images were taken during spectroscopic exposures in an effort to compensate for telescope drift.

The spectrograph slit was either maintained at the parallactic angle during

Table 2.1: Observational Circumstances for the 34 NEOs observed

Object	Date(s) Observed	V (mag)	Δ (AU)	r (AU)	ϕ (deg)
433 Eros	95/09/21	11.6	0.60	1.57	16.6
	96/01/22	12.6	0.82	1.19	54.8
1627 Ivar	95/04/26	14.5	0.81	1.70	22.9
1864 Daedalus	97/01/31	18.2	1.41	2.30	13.7
2062 Aten	95/01/01	13.8	0.15	1.12	18.3
	96/04/17	18.4	0.52	0.80	96.9
	96/04/18	18.4	0.52	0.80	96.7
2063 Bacchus	96/04/17	14.1	0.13	1.12	24.8
	96/04/18	14.2	0.14	1.13	24.9
2102 Tantalus	95/05/05	18.2	0.83	1.07	62.5
	95/06/27	15.5	0.34	1.32	22.5
2201 Oljato	96/01/22	16.2	0.43	0.93	83.6
	96/01/23	16.2	0.43	0.92	85.2
2212 Hephaistos	95/01/01	18.9	2.44	2.59	22.3
	97/01/31	19.0	2.39	2.86	19.2
2368 Beltrovata	95/05/05	18.1	1.25	2.20	12.2
3103 Eger	96/05/22	17.3	0.78	1.47	40.6
3122 Florence	96/10/29	17.1	1.23	1.42	43.2
3199 Nefertiti	96/09/18	16.8	0.83	1.13	59.4
3200 Phaethon	96/09/18	16.1	0.66	1.24	54.0
	96/10/29	16.4	0.83	1.71	22.8
3691 1982 FT	96/01/22	16.7	0.85	1.75	18.0
3752 Camillo	95/07/22	17.8	0.93	1.76	26.3
4055 Magellan	95/05/05	18.0	1.29	2.00	25.8
4197 1982 TA	96/09/18	15.4	0.57	1.49	25.4

Table 2.1 - *Continued*

Object	Date(s) Observed	V (mag)	Δ (AU)	r (AU)	ϕ (deg)
5131 1990 BG	95/01/01	17.8	1.55	2.32	18.8
5143 Heracles	96/10/29	15.9	0.95	1.92	9.7
6053 1993 BW3	96/01/23	15.7	0.62	1.59	10.2
6491 1991 OA	95/07/22	17.8	0.26	1.17	48.9
6569 1993 MO	95/04/26	17.9	0.65	1.44	37.8
7088 Ishtar	95/04/25	18.8	0.99	1.96	10.1
7336 1989 RS1	96/09/18	16.2	0.19	1.20	9.8
7358 1995 YA3	96/01/23	16.9	1.19	2.16	6.2
7822 1991 CS	96/08/19	15.2	0.12	1.05	66.9
7889 1994 LX	95/06/27	18.0	1.10	1.41	45.9
7977 1977 QQ5	95/01/01	19.3	1.62	2.25	22.9
1990 VB	95/01/01	17.9	0.97	1.93	7.2
1991 BB	96/01/22	17.8	0.76	1.49	36.2
1992 QN	96/09/18	17.4	0.58	1.58	8.5
1996 AE2	96/01/23	16.7	0.16	1.13	20.5
1996 FQ3	96/04/17	17.7	0.11	1.09	34.5
	96/04/18	17.8	0.11	1.09	35.6
1996 JG	96/05/19	14.3	0.06	1.07	9.3

NOTE— Visual magnitude, geocentric and heliocentric distance, and phase angle are given as listed in ephemeris at time of observation.

the exposure or reset to the parallactic angle at the start of every exposure to minimize the effects of differential refraction.

Observations of solar analog stars (see Section 2.4) were made at least once and usually several times per night, at airmasses as near as possible to those at which the targets were observed. One or two spectroscopic standards were also

Table 2.2: Observational Circumstances for non-NEOs

Object	Date(s) Observed	V (mag)	Δ (AU)	r (AU)	ϕ (deg)
8 Flora	95/03/22	10.6	1.95	2.07	28.5
225 Henrietta	95/09/21	12.4	1.73	2.72	4.0
631 Philippina	95/03/22	12.9	1.79	2.57	16.8
692 Hippodamia	96/04/18	14.1	2.22	3.05	12.3
1279 Uganda	97/01/31	17.7	2.44	2.82	19.9
1653 Yakhontovia	97/01/31	15.4	1.94	1.78	30.4
2078 Nanking	96/10/29	15.4	1.37	1.73	35.1
2906 Caltech	96/04/18	15.5	2.65	3.47	11.0
2938 Hopi	95/04/26	18.0	3.50	4.17	11.3
6144 1994 EQ3	95/05/05	17.3	2.90	3.41	15.9
1996 PW	96/08/19	17.2	1.53	2.54	2.6

NOTE— Visual magnitude, geocentric and heliocentric distance, and phase angle are given as listed in ephemeris at time of observation.

usually observed.

2.3 Data Reduction

The spectra were extracted and calibrated using a combination of standard IRAF procedures and custom-written routines in IDL. The sequence of processing steps is described in this section; the potential for random or systematic errors introduced by these reduction steps is described in detail in Section 2.5.

About 10 bias frames were obtained at the beginning or end of the night of observation. Unfortunately, because of problems with the DIS blue CCD chip (described in Section 2.5.1 below) bias frame subtraction was usually not performed.

Spectroscopic flat fields were taken using "white-light" quartz lamps

illuminating the closed mirror cover. This results in a non-uniform and instrument-rotation dependent illumination pattern which does not match sky illumination very well. Typically, a stack of $\sim$ 10 flat field images was combined into an outlier-resistant average flat field. This flat was then divided by a median-filtered copy to remove the spectral continuum and lower-order illumination variations across the field. This normalized flat was then divided into each object exposure to remove pixel-to-pixel variations in efficiency.

Spectra were extracted using a custom-written routine in IDL. Night sky lines were fit in regions surrounding the object aperture, and were subtracted out. Cosmic ray removal was performed manually on the extracted spectra.

Wavelength dispersion solutions were accomplished by taking reference He-Ne-Ar arcs at the beginning or end of the night. Formal fits using the arcs were generally good to a small fraction of an Ångstrom. However, the central wavelength settings of the gratings changed throughout the night (especially in the early runs of this survey), often from one exposure to the next, and sometimes by more than a hundred Ångstroms. Wavelength zeropoints were determined for each individual spectrum by examination of night-sky line positions (in longer exposures, for fainter objects) or by locating strong solar or telluric atmospheric absorption features (in brighter objects). Wavelengths in calibrated spectra are believed to be accurate to within $\sim$2–3 Å.

Atmospheric extinction was corrected by division with the Kitt Peak standard extinction table provided in IRAF. Object spectra were divided by spectra of solar analog stars to remove the solar spectrum and instrumental efficiency variations, and the red and blue portions of the spectrum were knitted together to produce the final reflectance spectrum.

2.4 Solar Analog Stars

A key step in the reduction of reflectance spectra is the ratioing of the object spectrum by the reflected solar spectrum in order to reveal the underlying solid-state absorption characteristics of the object being studied. The Sun is far too bright to directly observe with the same instrument as other solar system objects. Since there are no perfect artificial sources of the solar spectrum, and since there

Table 2.3: Solar Analog Star Properties

Name	Type	V (mag)	$U-B$ (mag)	$B-V$ (mag)
Hyades 64	G2V	8.12	0.20	0.69
16 Cygni B	G1.5V	5.96	0.19	0.64
HD 120528	G5V	8.56	0.21	0.66
HD 144873	G5	8.5	... [a]	... [a]
HD 191854	G5	7.46	0.22	0.69

[a] Colors not available for HD 144873.

are no perfectly colorless reflectors of sunlight available in the night sky, one must use sun-like stars as surrogate providers for the solar spectrum. It is also highly desirable that the surrogate solar spectrum pass through the same atmospheric and optical path, is observed with the same detectors, and is calibrated in the same manner as the solar system objects being studied, since this congruency greatly reduces the possibility that instrumental or data reduction artifacts are introduced into the final reflectance spectrum.

The Sun is a fundamental calibrator in many astronomical photometric and abundance studies, which makes true solar analogs scientifically valuable for the reasons listed above. Such stars have been sought for decades (cf. Hardorp 1978). A list of five solar analog stars was drawn from catalogs prepared by Hardorp. Some characteristics of these stars (drawn from SIMBAD) are presented in Table 2.3.

2.5 Possible Sources of Systematic and Random Error

Each of the data reduction steps described in Section 2.3 can introduce error. The magnitude of these potential errors and their characteristics are described below.

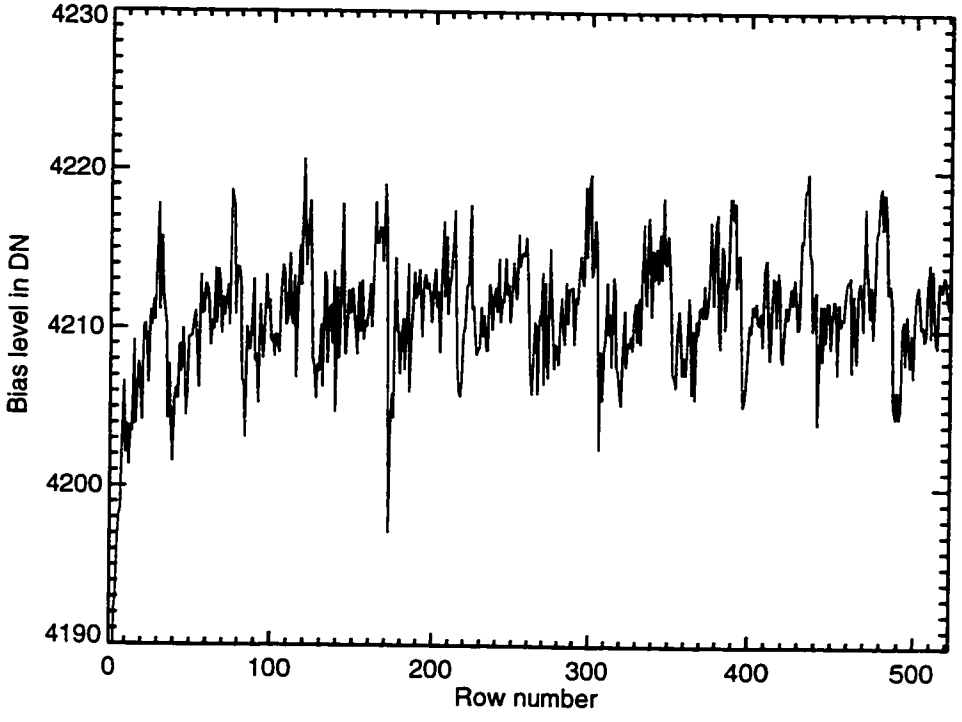


Figure 2.1: Average profile for central 10 columns in the DIS "blue" CCD chip bias frame. A non-stationary "ripple" presumably due to 60-Hz electrical interference is visible.

2.5.1 Bias Subtraction

During most of the course of the survey, the DIS blue CCD chip bias suffered from a non-stationary noise pattern apparently due to 60-Hz electrical interference (Figure 2.1). While much of this pattern could be removed by fitting and dividing out the "ripple," the subtraction of a bias frame still usually increased the noise level in the final product. Therefore, most spectra were not bias-subtracted. Those that were subject to bias subtraction were examined carefully for any artifacts which may have been introduced; none were found.

2.5.2 Flat Fielding

The DIS red CCD chip is cosmetically unclean. Various spots and hairlike features are evident in flat fields and in images with high sky backgrounds (Figure 2.2). As described in Section 2.5, the illumination pattern in flat fields does not match

that of the night sky. This causes slight differences in the positions of the "hairs" and "spots" between flat fields and night sky images or spectra. When dividing by these imperfect flat fields, residuals appear at the locations of these defects. An examination of spectra for flat-field residuals indicates that such residuals appear most strongly near 7800 Å, slightly redward of the strong telluric oxygen absorption, and less conspicuously near 6100 Å, 6500 Å, and 9200 Å. When present, these artifacts appear as spikes or dips of $\sim$ 20 DN amplitude over scales of around 50 - 100 Å. Such artifacts are generally systematic in effect: a given flat-field defect will repeatedly produce either a positive spike or a negative dip. These regions must be treated with care in further analyses.

In addition to these defects, the DIS red CCD chip also suffers from fringing in the near-infrared, which is apparent in spectroscopic flat field images and in the spectra of some of the objects observed. (Fringes are visible in Figure 2.2 as alternating dark and bright horizontal bands in the upper half of the image). Since atmospheric night-sky emission lines (particularly the near-infrared OH^- lines) cause stronger fringing than the "white" light in which flats are obtained, flat-fielding when strong fringing is present can introduce additional noise (cf. Tyson 1989). This artifact can be seen in the reflectance spectrum of 1627 Ivar as another "ripple" of $\sim$ 5 DN amplitude through the wavelength range $\sim$ 7600 - 9600 Å (Figure 2.3). Fringe artifacts can be thought of as a source of pseudo-random noise, as negative dips alternate with positive "bumps" around a common mean level. Since the scale of these fringes ($\sim$ 90 Å) is much smaller than most solid-state absorption features (which are typically hundreds of Ångstroms wide), they are an unsightly but relatively insignificant problem in this application.

2.5.3 Poor Subtraction of Night Sky Lines

In wavelength regimes or in exposures where the contrast between object and night sky spectra was very low, night sky line removal was problematic and apt to result in residual artifacts. This problem is particularly troublesome in the near-infrared redward of 9000 Å, where night-sky OH^- emission is bright, the reflected solar flux is low, instrumental efficiencies drop rapidly, and many asteroids display absorption features (further reducing the available flux). Frustratingly, these effects often conspire to introduce a thicket of powerful noise spikes in the



Figure 2.2: Flat-field image for the DIS "red" CCD chip. In addition to the "hairs" and "spots" visible in this image, fringing is apparent in the upper half, corresponding to the longer-wavelength end of the "red" spectra.

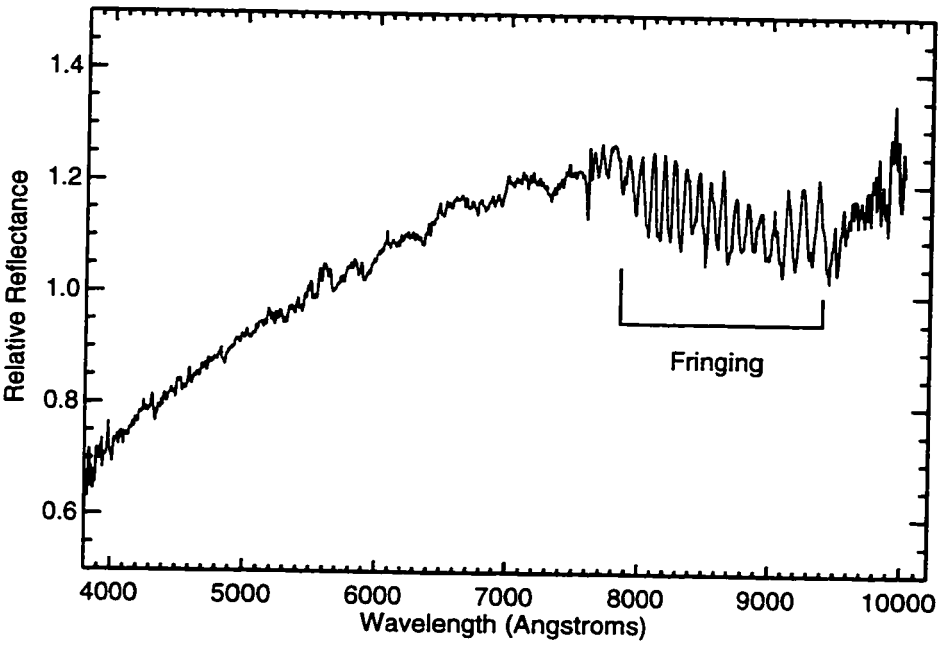


Figure 2.3: Fringing is apparent in extracted spectra as a $\sim$ 5% amplitude "ripple" in the 7600 – 9400 Åregion.

mineralogically important 9000 – 10,000 Å region. In general, over-subtraction of night sky lines is as likely as under-subtraction, leading to an increase in random noise levels in those spectral regions affected rather than a systematic shift in continuum level.

2.5.4 Poor Choice of Solar Analog?

The ultimate step in the production of a reflectance spectrum is the division by the solar (or as is usually the case for astronomical observations, the solar analog) spectrum. The choice of a solar analog star for use as a reference is thus critical for the quality and reliability of the product. Several solar analog stars from Table 2.3 were observed on the same night as a check on the similarity of their spectra. Figure 2.4 displays ratios of solar analog stars with one another. The spectra used in these ratios were all obtained on the same night, but at significantly different airmasses (in this respect, the figure may also be used to judge the accuracy of the atmospheric extinction correction). Apart from artifacts introduced by strong or changing telluric absorptions, the spectral ratios are flat

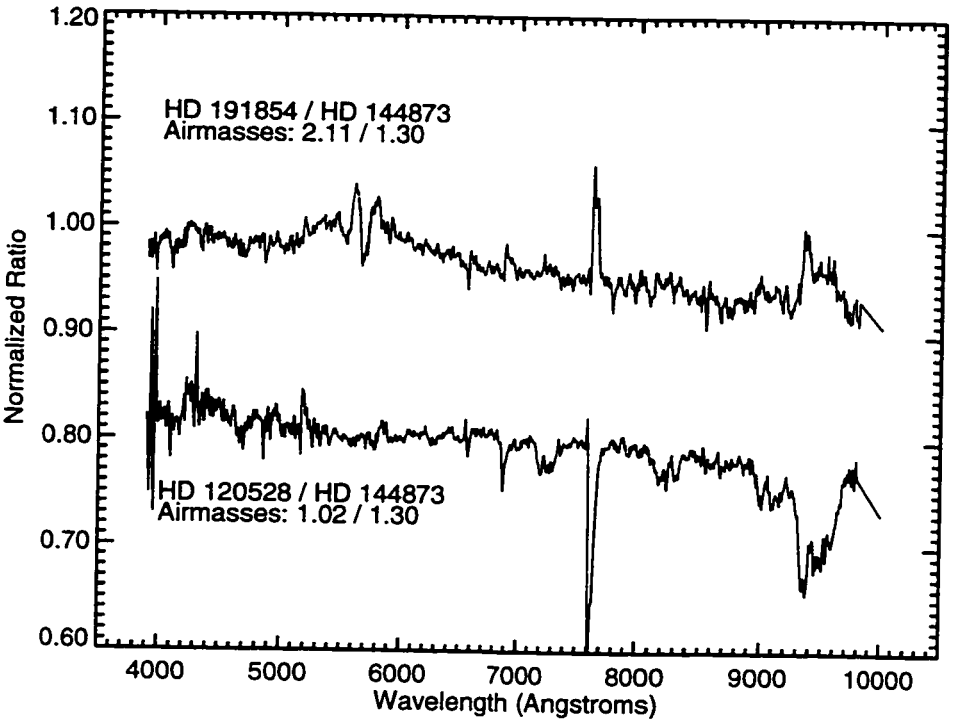


Figure 2.4: Ratios of solar analog stars observed on 05/22/96. The spectra have been offset vertically by 0.2 units for clarity.

to within 10% over the wavelength range 3800 Å to 10200 Å.

2.5.5 Misregistration of Object and Solar Analog Spectra

A misregistration between the solar analog star spectrum and the object spectrum may result in strong spurious features. There are several different manifestations of these artifacts: a misregistration of the relatively narrow solar or telluric atmospheric absorption features will cause correspondingly sharp dips and/or spikes to appear, while a misregistration of larger scale trends in instrumental efficiency can cause systematic bends in reflectance spectra. These latter artifacts will be especially apparent where the efficiency changes rapidly. These two effects are illustrated in Figure 2.5. A poor removal of absorption features may also be due to a change in the instrumental point spread function between the object spectrum and the solar analog star spectrum; at some level, this effect is inevitable.

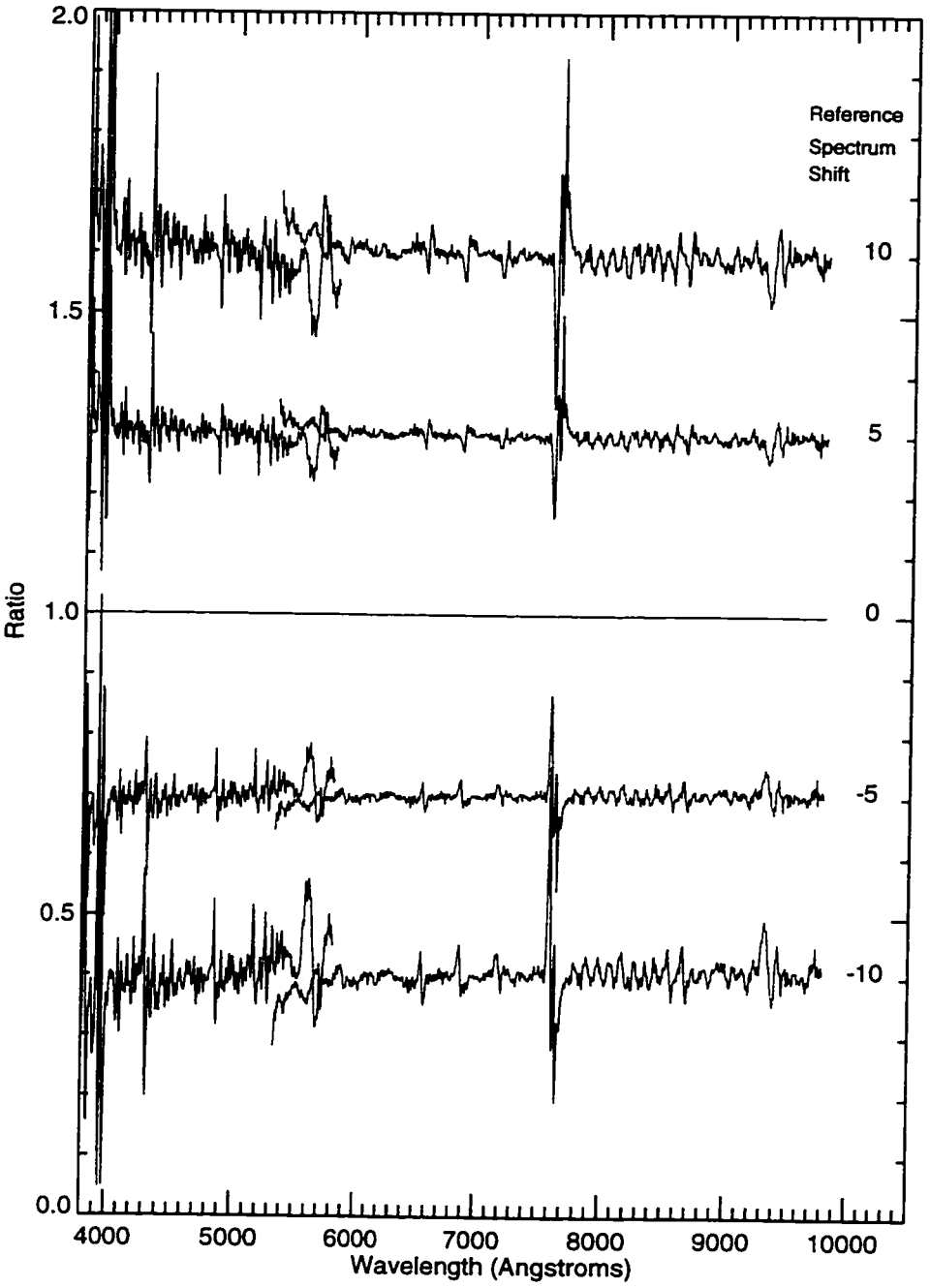


Figure 2.5: Ratios of artificially mis-registered solar analog spectra, illustrating the possible systematic shifts in continuum levels and the increase in apparently "random" noise due to poorly canceled stellar and telluric absorption lines. Ratios have been offset by 0.3 units for clarity.

If these spurious features can not be removed even through reprocessing the spectrum, then regions of the reflectance spectrum around the strong telluric oxygen absorption band and regions at both the far red and far blue ends of the spectrum are suspect and should be avoided in further analyses. Those spectra in which this effect was thought to have been important have had the appropriate ends truncated. Most other analyses are inherently resistant to possible errors near the telluric oxygen absorptions.

2.5.6 Improper Correction for Telluric Absorption

Since the solar analog star spectra were obtained at airmasses similar to the object spectra, differences in atmospheric extinction between the two were minimized, and the effects of any improper correction for extinction are likely to be small. However, certain telluric absorption features are particularly prominent and heavily dependent on airmass. For example, a change in the abundance of water vapor in the atmosphere may cause a faulty correction in the reflectance spectrum near 9600 Å, or a poor airmass match may cause strong residuals near the telluric oxygen absorption. Effects such as these are expected to be most important for observations at higher airmasses. Therefore, most observations were limited to airmasses less than 2.0, and almost all below 3.0.

2.6 Other Factors Which May Affect Spectra

2.6.1 Differential Refraction

Filippenko (1982) strongly cautioned astronomers about the phenomenon of atmospheric differential refraction. Unless observers plan for this effect, significant systematic errors in measured continuum intensities may arise, especially when observing at high airmasses. To avoid these errors in this work, the spectrograph slit was either maintained at the parallactic angle during the exposures, or reset to the parallactic angle at the beginning of each exposure. Also, observations were limited to relatively low airmasses (as detailed above).

2.6.2 Phase Reddening

Adams and Filice (1967) first showed that powdered rock samples can vary in color with phase angle and particle size. Since then it has been found that main-belt asteroids, the Moon, and powdered samples of meteorites also appear redder at higher phase angles (Gradie & Veverka 1986), over the phase angle ranges at which these objects are likely to be observed. Relatively little new work has been done on this subject. Luu & Jewitt (1990) estimate the effect for S-type asteroids as a linear increase in continuum slope by $\sim$0.16% $\pm$ 0.13% / 10^3 Å per degree, and $\sim$0.15% $\pm$ 0.17% / 10^3 Å per degree for C-type asteroids. Over the $\sim$6,000 Å range typically observed in this study, this effect could amount to a systematic reddening of about 1% per degree. Studies of the Moon and rocks powders indicate that phase reddening is not a monotonic function of phase angle, however, so that observations at very high phase angles would not necessarily be reddened as severely as a simplistic application of the Luu & Jewitt (1990) estimate would suggest (Adams & Filice 1967). Since NEOs are often observed at far greater phase angles than main-belt asteroids, the reader is cautioned that a simple comparison between NEO and main-belt object continuum slopes may be compromised to some degree by the effects of phase reddening.

Chapter 3

Summary of Results

Speak to the Earth, and it shall teach thee.

— Bible, Job 12:8

The reflectance spectra of the 34 NEOs which were observed during the course of this survey are presented in Figures 3.1 - 3.6. Reflectance spectra of the other objects, including the non-NEO cometary candidates, are presented in Figures 3.7 - 3.9. All spectra have been normalized to a reflectance of unity at 5500 Å by convention. The full-resolution, unsmoothed spectra are plotted as thin lines. The fainter objects show a considerable degree of scatter, especially at short or long wavelengths where the observed flux is very low. In some cases, the spectra have been truncated below $\sim$ 4000 Å or above $\sim$ 9500 Å to avoid these degraded data. To aid the eye, spectra were also binned in 250 Å intervals. The mean reflectances in these bins are overplotted as filled circles; the error bars correspond to plus and minus one standard deviation of the data in each bin. All of the automatic fits described in Section 5.2 were performed on the full-resolution, unsmoothed spectra. Some of the taxonomic classifications (Section 3.1) were performed by eye using the binned data.

Selected physical parameters for the NEOs are listed in Table 3.1, and non-NEOs in Table 3.2. Taxonomic classifications are described in Section 3.1. The Tisserand invariant T is a quasi-constant of motion for the restricted three-body problem of the Sun-Jupiter system (cf. Weissman et al. 1989), and is defined as

$$T = \frac{a_j}{a} + 2\sqrt{(\frac{a_j}{a})(1-e^2)}\cos i \qquad (3.1)$$

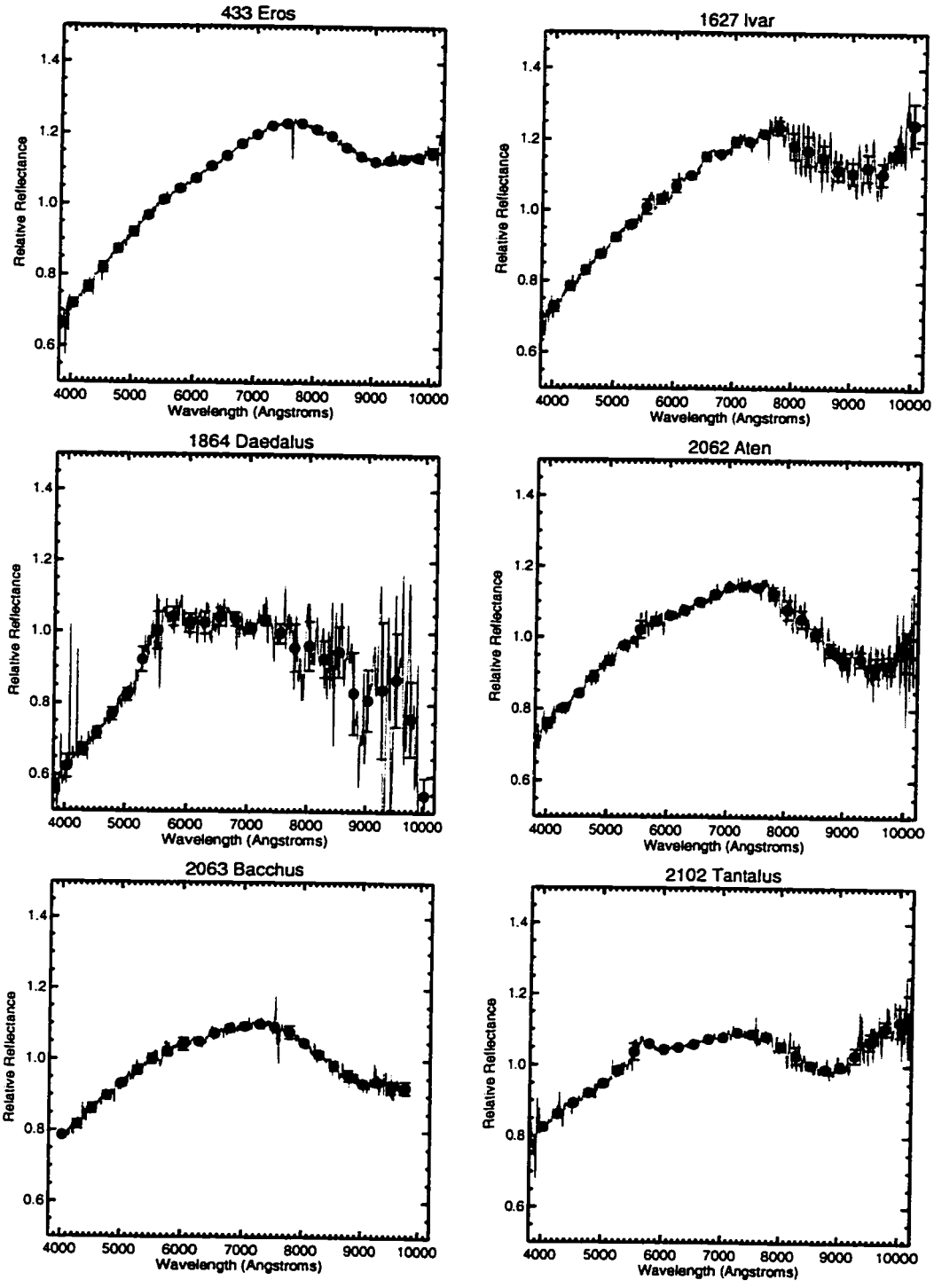


Figure 3.1: NEO reflectance spectra for 433 Eros, 1627 Ivar, 1864 Daedalus, 2062 Aten, 2063 Bacchus, and 2102 Tantalus.

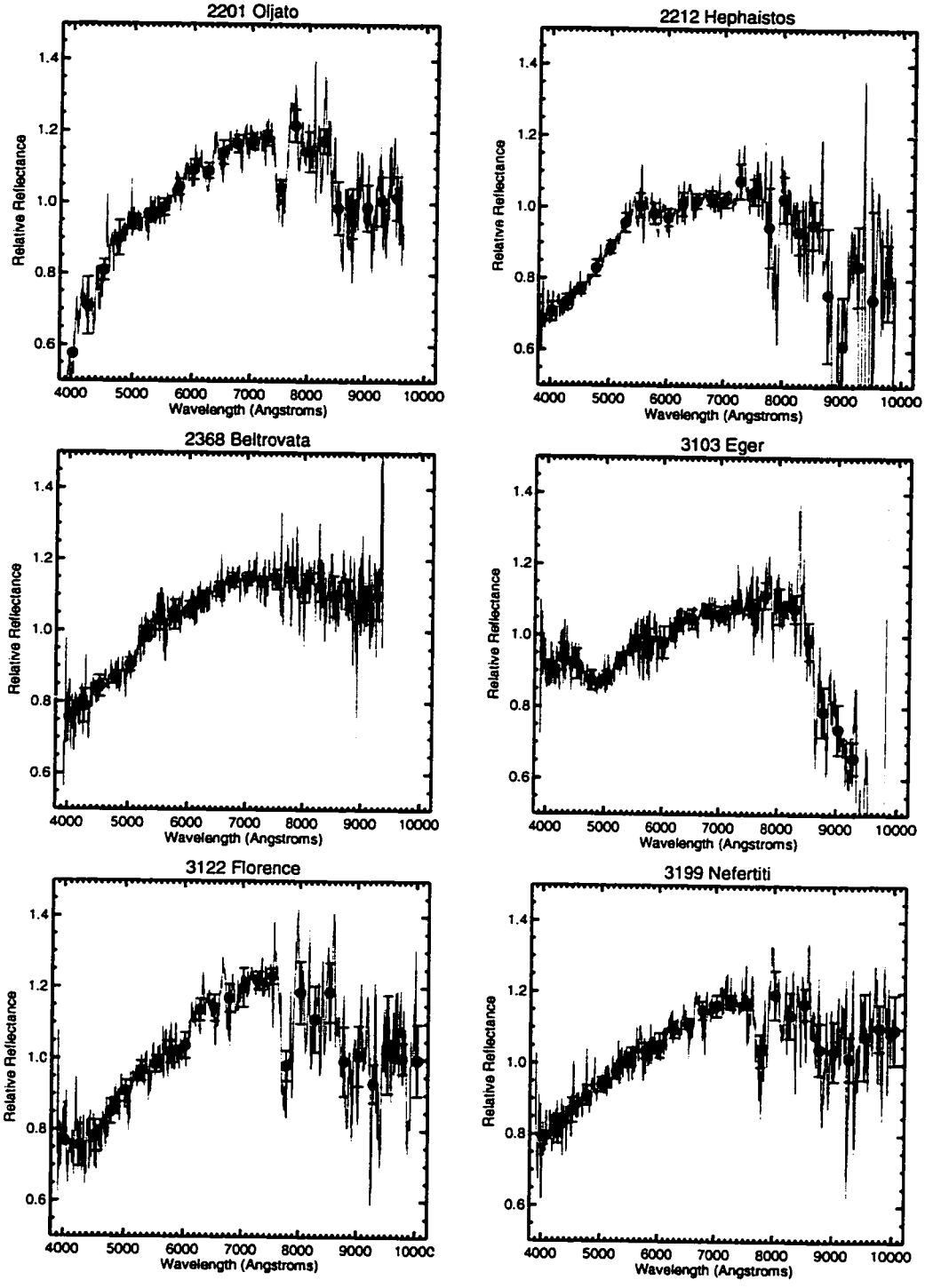


Figure 3.2: NEO reflectance spectra for 2201 Oljato, 2212 Hephaistos, 2368 Beltrovata, 3103 Eger, 3122 Florence, and 3199 Nefertiti.

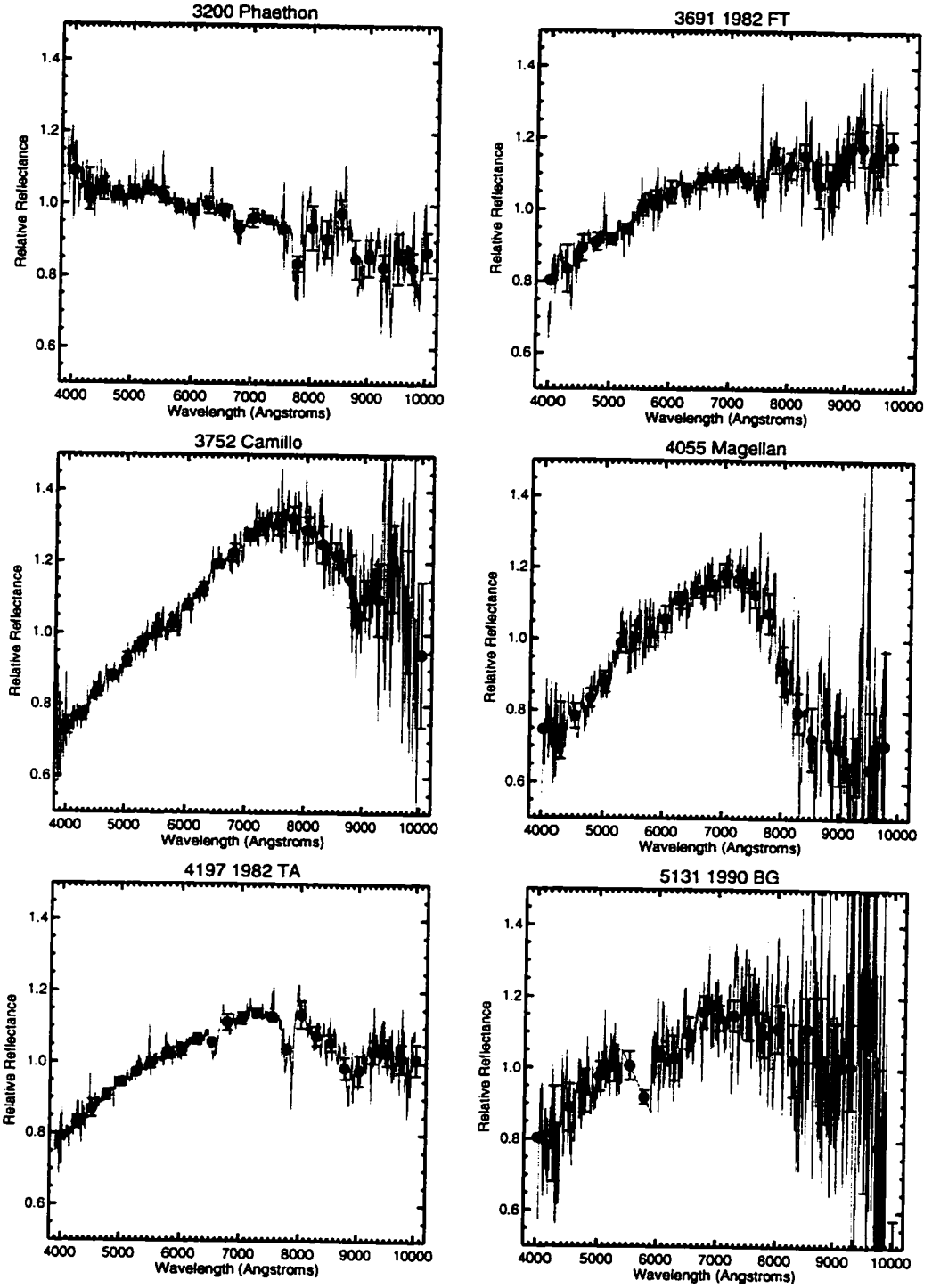


Figure 3.3: NEO reflectance spectra for 3200 Phaethon, 3691 1982 FT, 3752 Camillo, 4055 Magellan, 4197 1982 TA, and 5131 1990 BG.

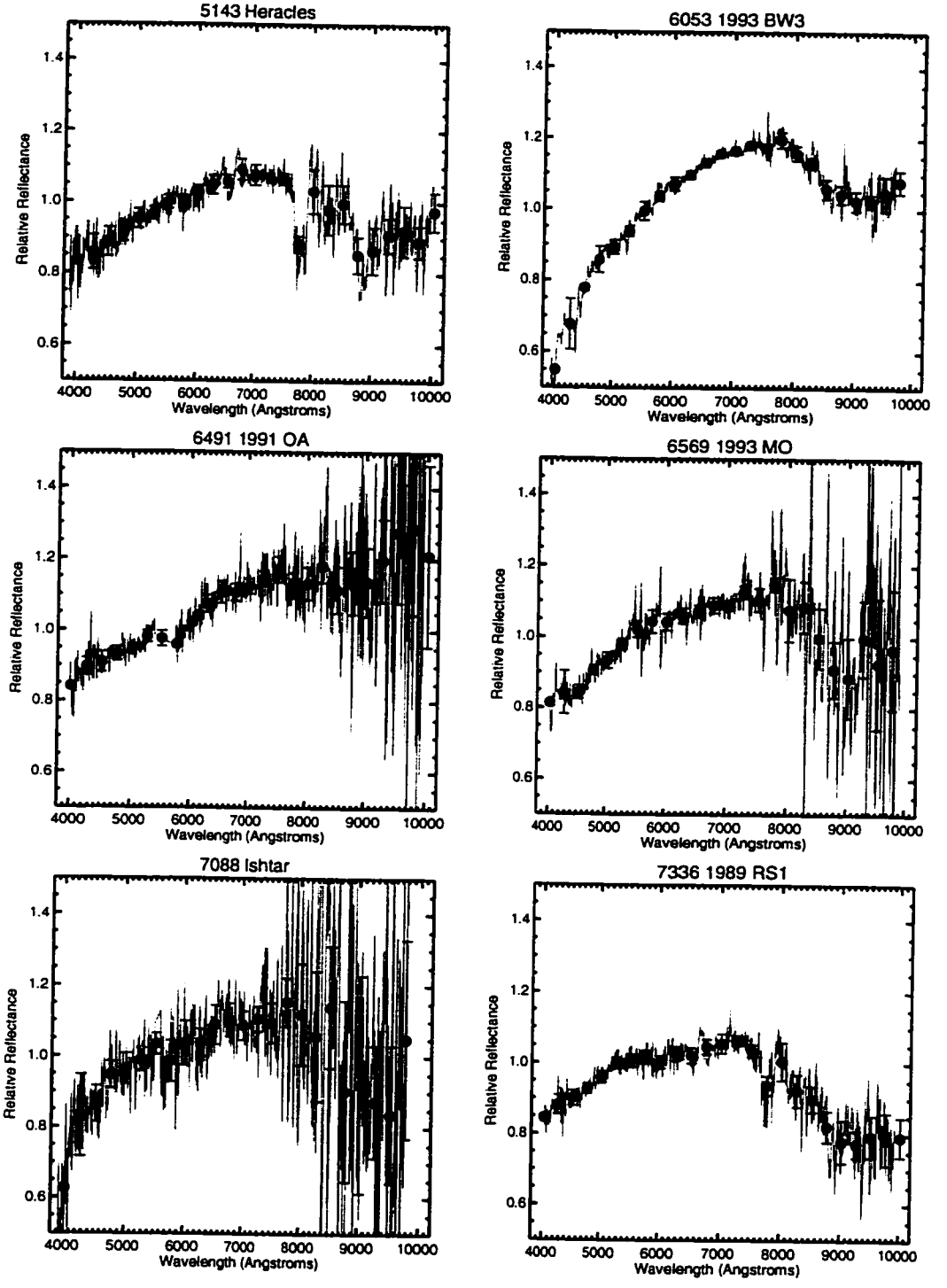


Figure 3.4: NEO reflectance spectra for 5143 Heracles, 6053 1993 BW3, 6491 1991 OA, 6569 1993 MO, 7088 Ishtar, and 7336 1989 RS1.

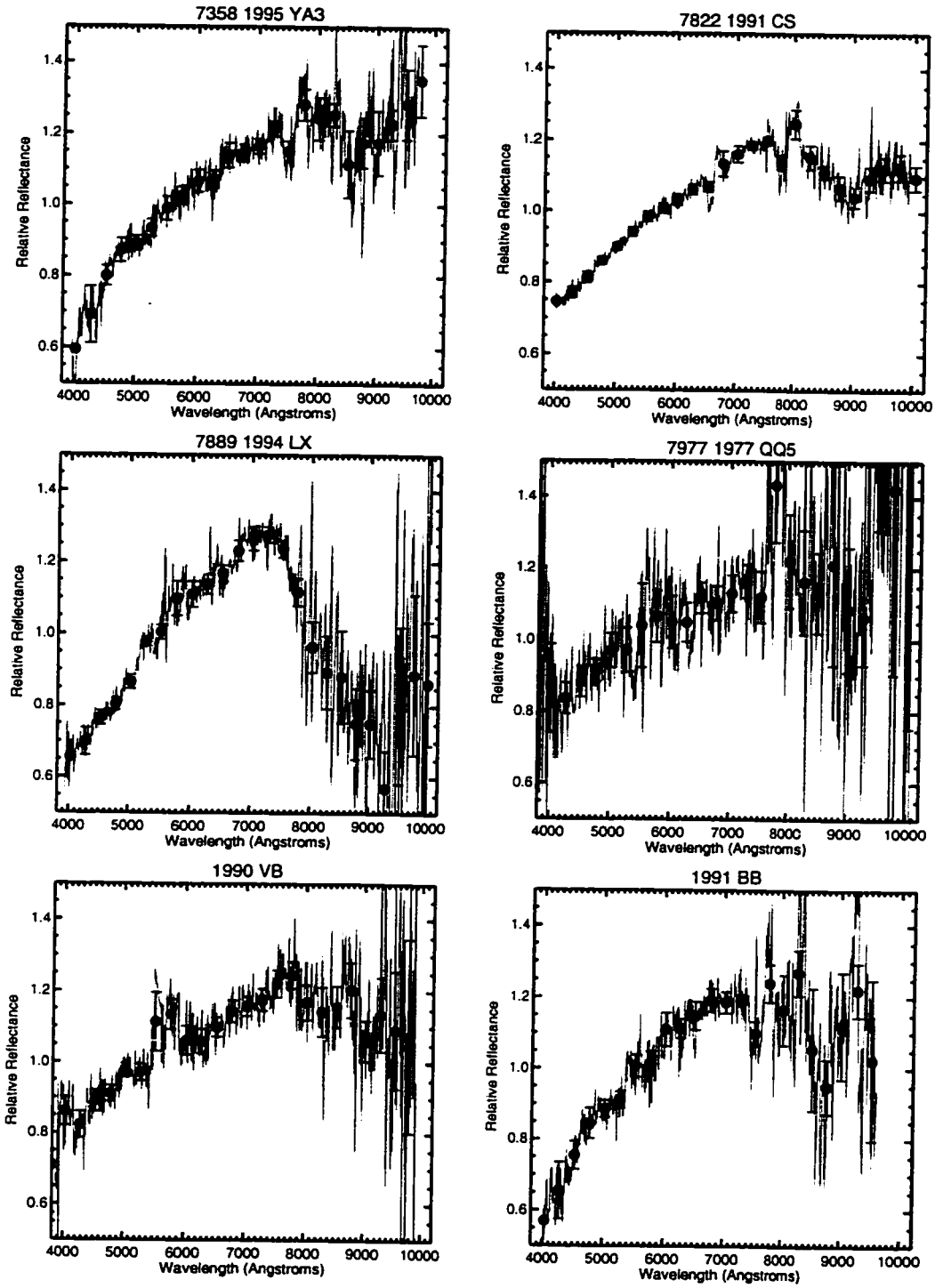


Figure 3.5: NEO reflectance spectra for 7358 1995 YA3, 7822 1991 CS, 7889 1994 LX, 7977 1977 QQ5, 1990 VB, and 1991 BB.

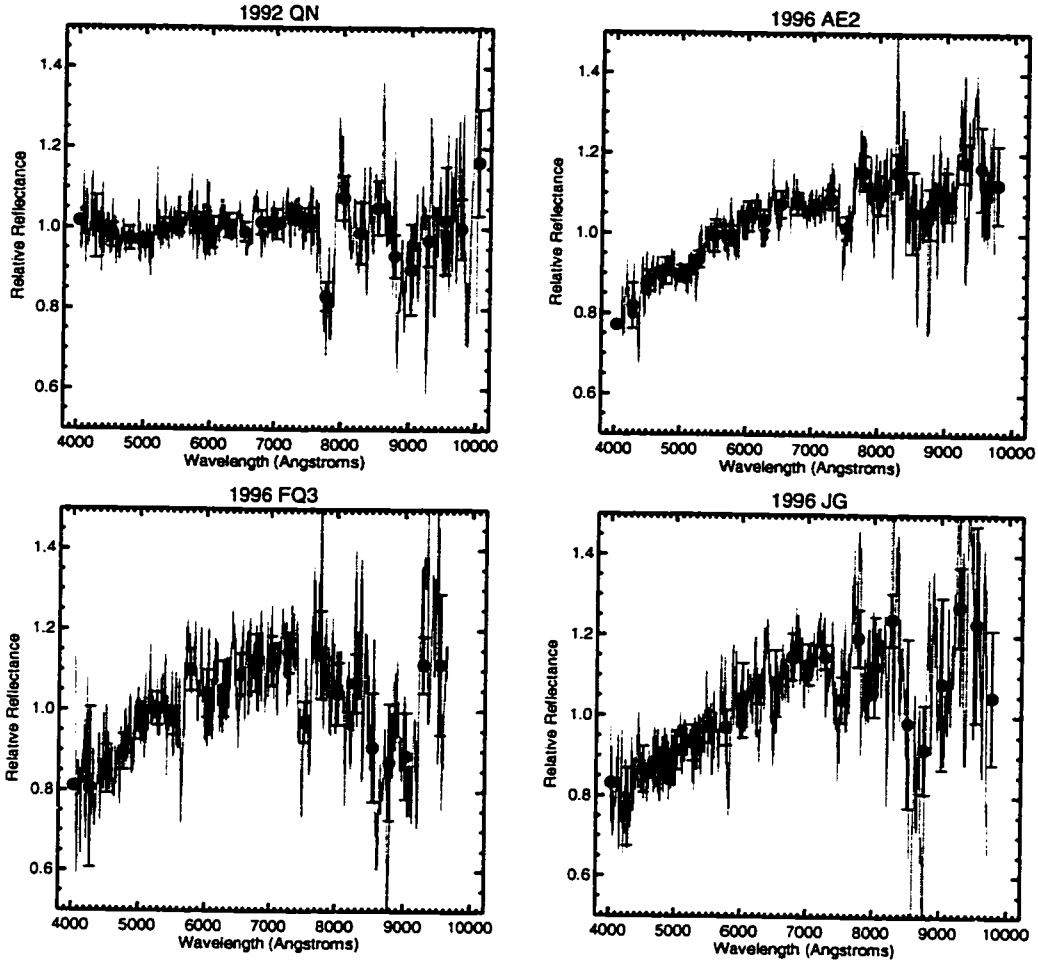


Figure 3.6: NEO reflectance spectra for 1992 QN, 1996 AE2, 1996 FQ3, and 1996 JG.

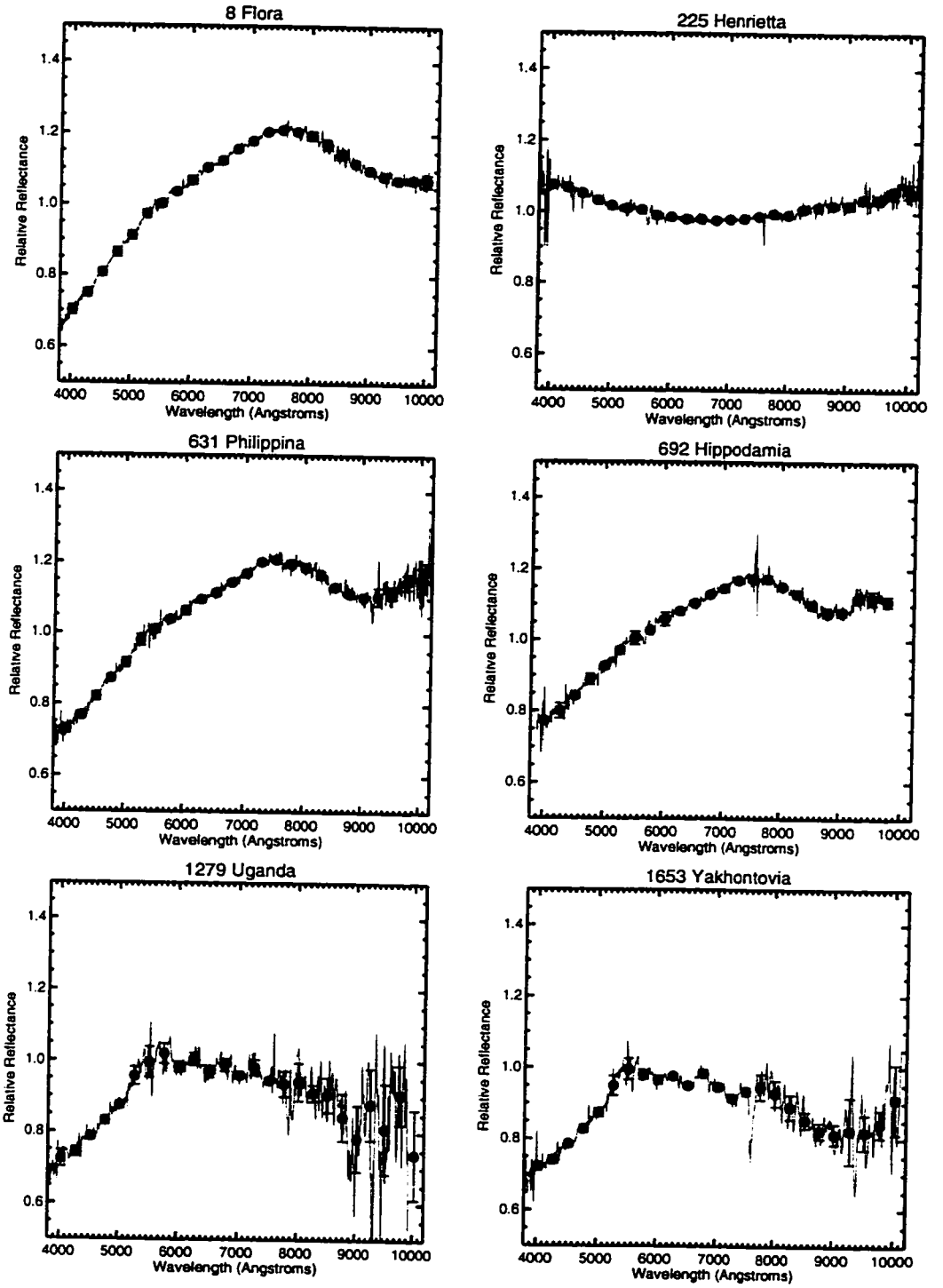


Figure 3.7: Main-belt asteroid reflectance spectra (for 8 Flora, 225 Henrietta, 631 Philippina, 692 Hippodamia, 1279 Uganda, and 1653 Yakhontovia).

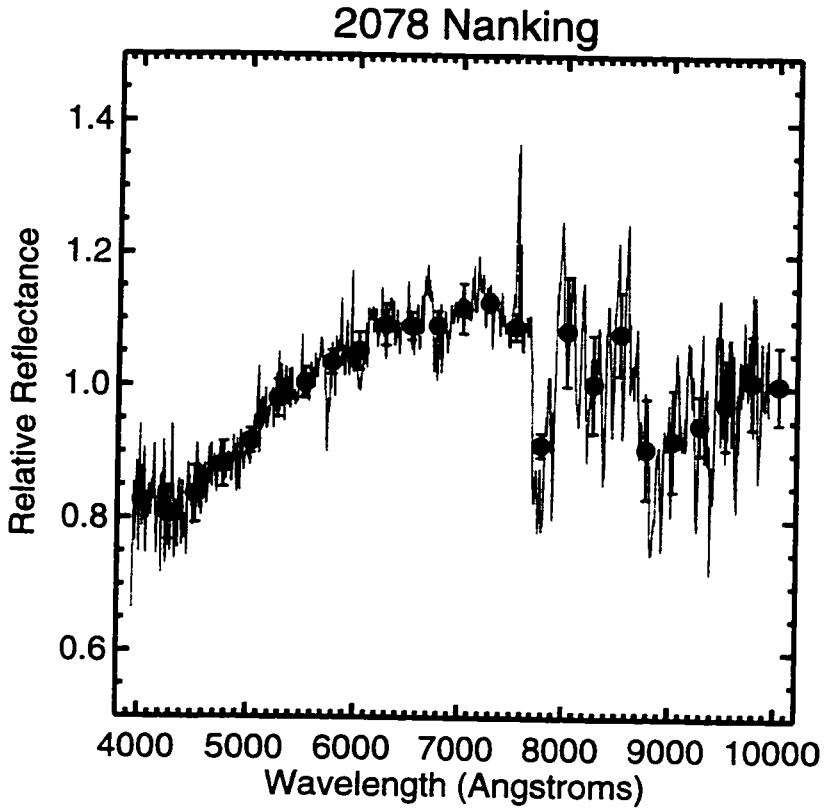


Figure 3.8: Reflectance spectrum of Mars-crossing asteroid 2078 Nanking

where a_j is the semimajor axis of Jupiter (≈ 5.2 AU), a is the semimajor axis of the object, e the object's orbital eccentricity, and i the inclination between the orbits of the object and Jupiter. The Tisserand invariant has the property that it remains nearly constant even if an object has close encounters with Jupiter. Close approaches to Jupiter can only occur if $T \leq 3$, and if stable resonances do not prevent them (as they do the Trojan asteroids). Almost all short-period comets have $T < 3$, and nearly all asteroids $T > 3$. Therefore, the dividing line $T = 3$ is often used to dynamically distinguish cometary-type orbits from asteroidal ones. For the purposes of this work, the ecliptic inclination of the object orbit was used instead of its relative inclination to Jupiter, as a simplifying assumption. This might cause an error of a few percent in T for the higher inclination objects.

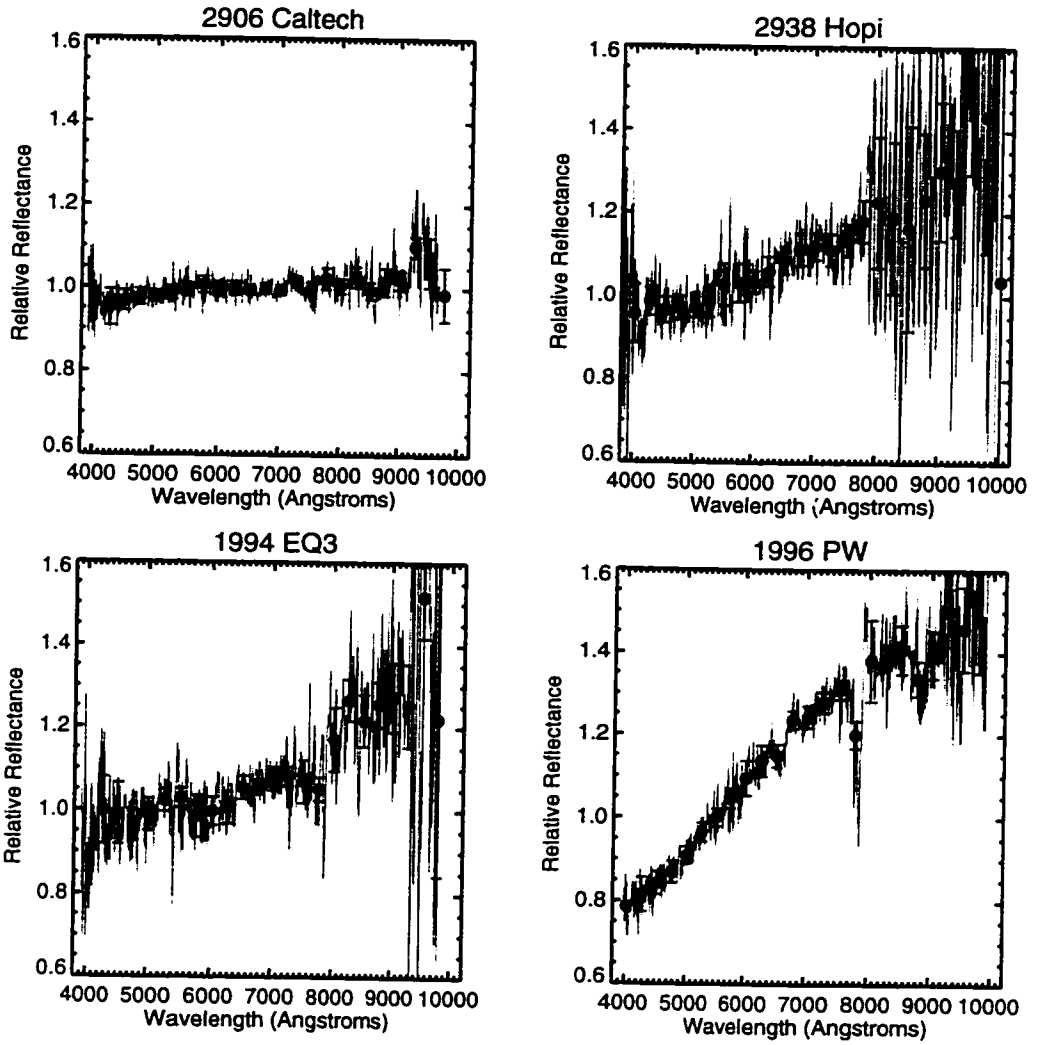


Figure 3.9: Non-NEO cometary candidate asteroid reflectance spectra (for 2906 Caltech, 2938 Hopi, 1994 EQ3, and 1996 PW).

3.1 Taxonomic Classification

A historical overview of the various asteroid taxonomic classification systems is given in Tholen & Barucci (1989). A brief summary of issues relating to these asteroid taxonomies is related in Section 1.3.

In order to taxonomically classify the objects observed during this survey according to the Tholen scheme, it would be necessary to calculate synthetic ECAS medium-band colors for them. Although early tests showed that this resulted in reasonable classifications (Hammergren 1996), and that different

Table 3.1: Physical parameters for the 34 NEOs observed

Object	a (AU)	e	i (deg)	T^a	H (mag)	D (km)	Taxonomic Type
433 Eros	1.458	0.223	10.83	4.58	11.16	19	S
1627 Ivar	1.863	0.397	8.44	3.88	13.20	7.4	S
1864 Daedalus	1.461	0.615	22.18	4.33	14.85	3.5	SQ
2062 Aten	0.966	0.183	18.93	6.18	16.80	1.3	Q
2063 Bacchus	1.078	0.349	9.43	5.67	17.10	1.1	Q
2102 Tantalus	1.290	0.299	64.01	4.45	16.20	1.9	SQ
2201 Oljato	2.173	0.713	2.52	3.30	15.25	1.90[b]	Q?
2212 Hephaistos	2.168	0.833	11.77	3.10	13.87	4.9	QR
2368 Beltrovata	2.104	0.414	5.25	3.63	15.21	2.9	SQ
3103 Eger	1.406	0.355	20.94	4.61	15.38	1.7	E?
3122 Florence	1.768	0.423	22.18	3.92	14.20	4.7	S
3199 Nefertiti	1.575	0.284	32.98	4.19	14.84	3.5	S
3200 Phaethon	1.271	0.890	22.10	4.51	14.60	5.20[b]	F
3691 1982 FT	1.774	0.284	20.38	3.98	14.90	[c]	XC
3752 Camillo	1.414	0.302	55.55	4.24	15.50	2.6	SA
4055 Magellan	1.820	0.326	23.24	3.89	14.80	3.3	V
4197 1982 TA	2.299	0.772	12.22	3.09	14.60	3.5	QS
5131 1990 BG	1.486	0.570	36.38	4.21	14.10	4.9	SQ
5143 Heracles	1.834	0.772	9.16	3.58	14.00	4.6	Q
6053 1993 BW3	2.146	0.529	21.60	3.44	15.10	2.8	QS
6491 1991 OA	2.508	0.587	5.52	3.19	18.50	0.6	SQ
6569 1993 MO	1.626	0.221	22.64	4.21	16.50	1.5	VQS
7088 Ishtar	1.981	0.390	8.29	3.75	16.70	1.5	SX?
7336 1989 RS1	2.305	0.482	7.18	3.41	18.70	0.5	RQ
7358 1995 YA3	2.198	0.502	4.66	3.49	14.40	4.3	S
7822 1991 CS	1.123	0.165	37.12	5.36	17.40	1.1	S
7889 1994 LX	1.261	0.346	36.90	4.86	15.30	2.6	V
7977 1977 QQ5	2.226	0.466	25.19	3.38	15.40	2.7	SX?

Table 3.1 – *Continued*

Object	a (AU)	e	i (deg)	T^a	H (mag)	D (km)	Taxonomic Type
1990 VB	2.442	0.528	14.57	3.26	15.90	2.1	S
1991 BB	1.186	0.272	38.48	5.11	16.04	1.8	QS
1992 QN	1.191	0.359	9.59	5.25	17.01	2.3	C
1996 AE2	1.368	0.257	37.35	4.59	19.49	[c]	X
1996 FQ3	2.031	0.471	1.07	3.66	21.09	0.2	Q
1996 JG	1.803	0.661	5.28	3.77	19.73	0.3	Q

[a] T = Tisserand Invariant.

[b] IRAS Diameter, listed in Bowell 1998.

[c] Diameters not computed for spectrally degenerate X-type objects.

NOTE— Diameters computed from H and average albedo for taxonomic type, except where indicated.

Table 3.2: Physical parameters for the non-NEOs observed

Object	a (AU)	e	i (deg)	T^a	H (mag)	D (km)	Taxonomic Type
8 Flora	2.201	0.156	5.89	3.64	6.49	141[b]	S
225 Henrietta	3.382	0.270	20.90	2.99	8.72	124[b]	F
631 Philippina	2.792	0.083	18.92	3.24	8.70	60.5[b]	S
692 Hippodamia	3.373	0.176	26.12	2.97	9.18	47.7[b]	S
1279 Uganda	2.371	0.208	5.73	3.51	12.51	19	SX?
1653 Yakhontovia	2.611	0.323	4.07	3.33	11.40	31	SX?
2078 Nanking	2.370	0.374	20.16	3.37	12.10	11	Q
2906 Caltech	3.161	0.116	30.69	2.98	10.00	62.0[b]	C
2938 Hopi	3.144	0.334	41.44	2.75	11.50	33	T
6144 1994 EQ3	4.691	0.367	5.73	2.87	11.50	33	T
1996 PW	329.968	0.992	29.82	1.73	14.01	10	D

[a] T = Tisserand Invariant.

[b] IRAS Diameter, listed in Bowell 1998.

NOTE— Diameters computed from H and average albedo for taxonomic type, except where indicated.

classes of objects separated fairly well in principal component space, it was felt that the loss of information incurred by using synthetic multicolor photometry rather than spectra was undesirable. Unfortunately, there are no formal taxonomies which use reflectance spectra as inputs.

Objects were instead classified by eye using human judgment, using mean reflectance spectra for various asteroid classes (Xu et al. 1994) as guides. This is obviously a subjective process, and one for which it is difficult if not impossible to estimate uncertainties. Also, as Binzel et al. (1998) note, smaller objects seem to possess spectral properties which span the differences between taxonomic classes, which makes unique classification very difficult. Taxonomic classes for the objects observed during this survey are presented in Tables 3.1 and 3.2.

Chapter 4

Discussion of Results for Individual Objects

In all matters of opinion, our adversaries are insane.
— Mark Twain

4.1 NEOs

4.1.1 Spectra similar to ordinary chondrite meteorites

Several of the objects which are apparently roughly similar to S-type asteroids have spectra which display stronger 1-μm absorption bands, and less reddened visual continua than typical S-type asteroids. These spectra are similar, and in some cases, for all intents identical to the spectra of ordinary chondrite meteorites. These objects include: 2062 Aten, 2063 Bacchus, 2102 Tantalus, 2201 Oljato, 2212 Hephaistos, 3199 Nefertiti, 4197 1982 TA, 5131 1990 BG, 5143 Heracles, 6053 1993 BW3, 6569 1993 MO, 7088 Ishtar, 7336 1989 RS1, 1996 FQ3, and 1996 JG. Hicks et al. (1998) also observed 2102 Tantalus, 2201 Oljato, and 6053 1993 BW3, all of which were in good agreement with observations in this work. Binzel et al. (1996) report an ordinary-chondrite-like spectrum for 2102 Tantalus, and a spectrum midway between an S-type asteroid and an ordinary chondrite for 2063 Bacchus. Lazzarin et al. (1997) also observed 2063 Bacchus, but their spectrum shows a featureless and slightly red-sloped continuum which they label as C-type. The reason for this inconsistency is unknown.

2212 Hephaistos and 2201 Oljato are traditionally known as strong cometary

candidates, owing mainly to their orbits (Weissman et al. 1989); 2212 Hephaistos has an orbit similar to that of Comet Encke and the δ Cancrids meteor stream, while that of 2201 Oljato is chaotic and possibly associated with several meteor streams (Weissman et al. 1989). Much of the excitement surrounding 2201 Oljato is due to an apparent excess of UV flux in photometry obtained by McFadden et al. (1984), and interpreted by McFadden et al. (1993) as a possible sign of fluorescent cometary emissions. Spectra from this work of 2212 Hephaistos and 2201 Oljato show fairly strong UV and 1-μm absorptions, reminiscent of absorptions seen in S- or Q-type asteroids. Since cometary nuclei are thought to be dark and neutral to red in color, with any absorption features muted by highly opaque surface materials (Luu 1993, 1994), these spectra argue strongly against a cometary origin for these objects.

4.1.2 Unusual spectra with UV- and 1-μm absorptions

Several of the objects observed in this survey display unusual spectra, with strong UV- and 1-μm absorption bands, but with relatively blue-sloped continua between 5500 Å and 7500 Å, where most asteroids and meteorites with silicate absorption features have significantly red-sloped continua. These objects are the NEOs 1864 Daedalus and 2212 Hephaistos, and the main-belt asteroids 1279 Uganda and 1653 Yakhontovia. The latter two were observed by Xu et al. (1995), whose spectra are similar to those of this study. 1864 Daedalus was observed by Lazzarin et al. (1997), who report an S-type spectrum for the object. I am unable to resolve this difference, although I note that their observations of 2063 Bacchus (see above) are also discrepant.

4.1.3 3103 Eger

The Apollo object 3103 Eger was reported by Wisniewski (1987) to have relatively neutral colors below about 7000 Å. Veeder et al. (1989) determined the radiometric albedo of 3103 Eger to be in the range 0.53 - 0.63. The high albedo and neutral colors were strongly suggestive of the E-type asteroids. Observations of this object by Gaffey et al. (1992) using continuously-variable filters (CVF) over the wavelength range 0.8 – 2.5μm seemed to confirm this classification. Gaffey et al. (1992) went on to conclude that 3103 Eger was the sole near-Earth parent

body for all enstatite achondrite (aubrite) meteorites. However, spectra obtained during the course of this survey display an unusually strong 1-μm absorption band, coupled with a relatively neutral continuum below about 8000 Å, except for a small apparent absorption feature near 5000 Å. The presence of such a strong feature near 1-μm – if real – is enough to invalidate the E-type classification, since the presence of ultramafic silicates is inconsistent with a putative enstatite composition. Ten separate spectra of 3103 Eger were obtained on 05/22/96. All of them show the identical shape. Three different solar analog stars were also observed on that night; ratios using these stars all show the same features and continuum slope. It is difficult to reconcile the spectra obtained in this work to those of Gaffey et al. (1992) unless one invokes an observational or data reduction error. This issue remains unresolved.

4.2 Mars Crosser – 2078 Nanking

The Mars crossing asteroid 2078 Nanking was observed by Xu et al. (1995), and found to have a spectrum similar to an H-chondrite. This work confirms those conclusions. It is noted that many NEOs also have spectra similar to the H- or L-chondrites, and that perhaps 2078 Nanking is evidence of a link from the main asteroid belt through a Mars crossing stage.

4.3 Non-NEO Cometary Candidates

Four non-NEO objects with unusual orbits were observed during the course of this survey. 2906 Caltech, 2938 Hopi, 1994 EQ3, and 1996 PW all have orbits which either cross or approach Jupiter's orbit, thus making them excellent candidates for extinct comets. All four of these objects have relatively featureless, red-sloped continua consistent with either the C, P or D spectral classes, and with known cometary nuclei (Luu 1993, 1994).

Chapter 5

Spectral Trends Among S-type Objects

Some circumstantial evidence is very strong, as when you find a trout in the milk.

— Henry David Thoreau

5.1 Space Weathering

Surface processes active on the Moon cause a lowering of albedo, a reddening of the spectral continuum, and a decrease in absorption band depth in regolith compared to lunar rocks (cf. Pieters et al. 1993). This process or collection of spectrally-altering processes is known as "space weathering," and has been proposed as one reason for the differences between asteroid and meteorite spectra, particularly for the spectral differences between S-type asteroids and the ordinary chondrite meteorites.

Pieters et al. (1993) found that lunar space weathering appears to be correlated to surface alterations, and that it is in only the finest fraction of lunar regolith that the red-sloped continuum is produced. Space weathering, at least on the Moon, appears to be directly related to regolith maturation. Galileo spacecraft images of main-belt S-type asteroid 951 Gaspra show color trends and differences in albedo and absorption band depth consistent with the nature (but not necessarily the magnitude) of lunar space weathering, demonstrating unambiguously the existence on S-type asteroids of some form of surface spectral alteration.

Since larger bodies will generally retain ejecta more effectively than smaller bodies, a dependence of regolith maturity on size may be expected to exist. Additionally, if small NEOs are produced from relatively recent catastrophic disruption events, they may be even less affected by potential space weathering. Previous investigations of small main-belt asteroids and NEOs have shown some interesting relationships between spectral slope, absorption band depth, and size. McFadden et al. (1984) found that small S-type NEOs displayed stronger absorption band depths than larger main-belt asteroids. In a study of small main-belt asteroids, Xu et al. (1995) also found this relation to be the case, but they interpret this effect as being due mainly to an increase in the diversity of spectral band strengths among the small S-type asteroids. Binzel et al. (1998) reiterate this supposition. In an investigation of 39 S-type asteroids ranging in size from ~30 – 300 km in diameter, Gaffey et al. (1993) too observed an increase in absorption band depth for smaller asteroids. They found no correlation between albedo and size, and a slight negative correlation between spectral slope and size (such that the smaller asteroids had relatively redder continua). Since this trend is opposite what might be expected from space weathering due to regolith maturity, they conclude that it is unlikely that the more steeply-sloped asteroids could be plausible ordinary chondrite parent bodies.

5.2 Modified Gaussian Model Fits

To study the relationships between spectral characteristics and object size, I first fit NEO spectra with a 5-parameter, two-component Modified Gaussian model, following those presented in Sunshine et al. (1990). A Modified Gaussian model is one in which an absorption band is fit by a Gaussian in inverse-energy – log reflectance space. In this work, a single Modified Gaussian was superimposed on a continuum which was linear in energy – log reflectance space. Figure 5.1 displays some example fits which result from the application of this model. In most cases, this 5-parameter fit does a wonderful job at fitting both the visual continuum and 1-μm band shape, returning reduced χ^2 values near 1.1-1.3. The fit does break down on objects with unusual spectra like 1864 Daedalus, or for obviously non-S-type objects. The model was also fit to spectra of S-type objects from Xu et al. (1995) for comparison. Those data include spectra of 5 NEOs;

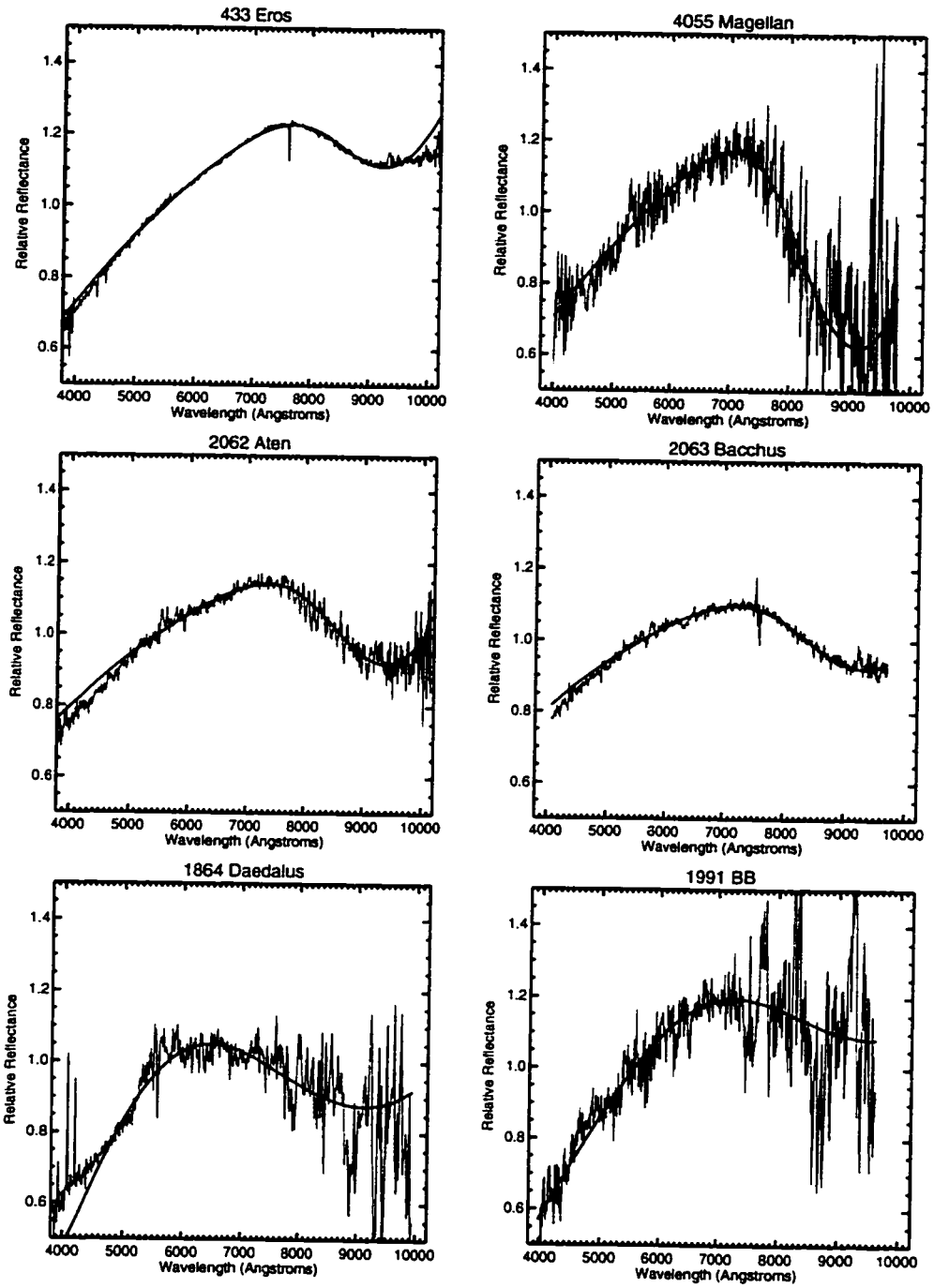


Figure 5.1: NEO spectra have been fit by a five-parameter, two-component Modified Gaussian model. S- and Q-type objects are well fit.

most others are of main-belt asteroids.

5.3 Spectral Trends with Diameter

The plot of 1-μm band strength versus size is shown in Figure 5.2. There is a slight trend towards deeper absorption bands for smaller objects. A formal fit to the equation

$$\text{BandStrength} = a_{str} + b_{str} \times \log_{10} D \tag{5.1}$$

yields $a_{str} = -0.230 \pm 0.001$ and $b_{str} = 0.0295 \pm 0.0007$, with a linear correlation coefficient r of 0.43. Most of the scatter around this fit is apparently intrinsic to the population of objects, and not due to observational errors.

A plot of 1-μm band center versus size is presented in Figure 5.3. While a slight trend towards bluer band centers with smaller diameters is apparent in a formal fit, the linear correlation coefficient r of 0.14 shows this trend to be relatively insignificant compared to the scatter in the data. The parameters of a formal log-linear fit to the equation

$$\text{BandCenter} = a_{ctr} + b_{ctr} \times \log_{10} D \tag{5.2}$$

are $a_{ctr} = 9243 \pm 4\mathring{A}$ and $b_{ctr} = 38 \pm 2$. If real, this trend would seem to imply that larger objects (or at least the spectrally important components of their surfaces) are relatively enriched in olivine or clinopyroxene. Hörz & Cintala (1997) have noted that such a trend towards olivine enrichment may be expected on asteroidal surfaces due to certain regolith processes, such as the selective crystallization of olivine relative to pyroxene in impact melts, and the production of much larger quantities of fine-grained olivine relative to pyroxene in comminution experiments.

Figure 5.4 presents a plot of "visual slope" versus size. The "visual slope" is a measure of the redness of a spectrum, and refers to the reflectivity gradient calculated in the 5000–7500 Å region, as introduced by Luu & Jewitt (1990). For objects larger than ~6 km, the trend towards redder visual slopes at larger diameters noted by Gaffey et al. (1993) is also observed, but quite strikingly, the reverse relationship is seen for smaller objects. A formal log-linear fit to those objects smaller than 6 km in diameter yields the relation:

$$\text{VisualSlope(D} < 6\text{km)(\%per10}^3\text{Å)} = (6.19 \pm 0.03) + (8.20 \pm 0.06) \times \log_{10} D(km) \tag{5.3}$$

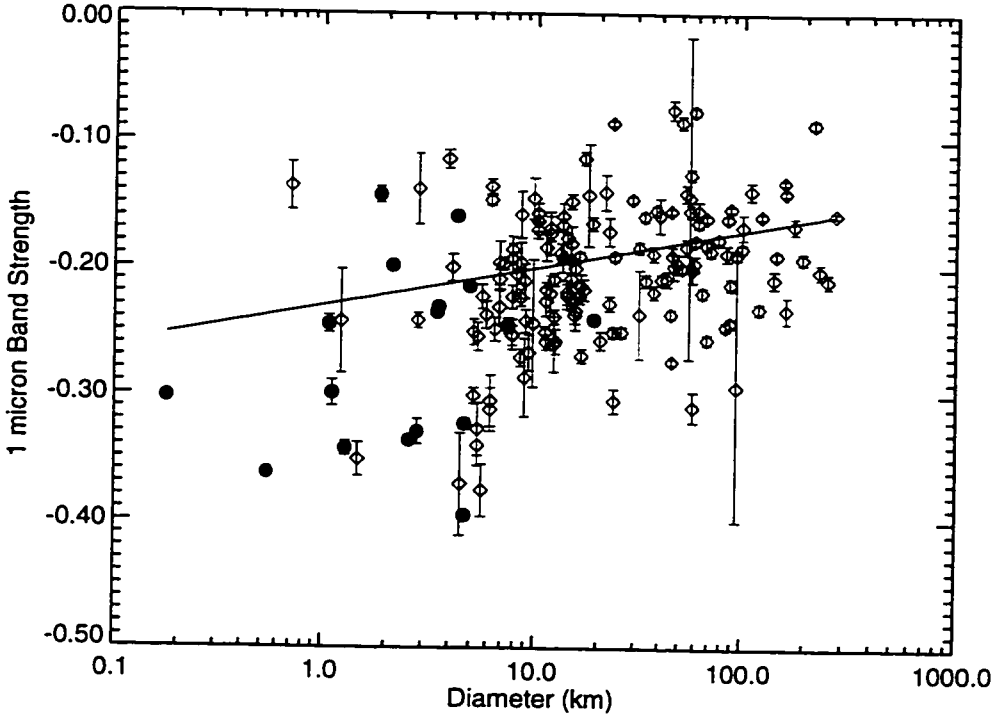


Figure 5.2: Plot of 1-μm band strength vs. size. NEOs observed in this survey are plotted as solid circles, and S-type asteroids from Xu et al. (1995) (mostly main-belt asteroids, but including five NEOs) are indicated by hollow diamonds. A log-linear fit to these data is shown.

For objects larger than 6 km, the fit is

$$\text{VisualSlope}(\text{D} \geq 6\text{km})(\%\text{per}10^3\text{Å}) = (12.16 \pm 0.02) + (-1.09 \pm 0.01) \times \log_{10} D(km) \quad (5.4)$$

This seems to indicate that the smaller S-type objects are systematically more spectrally similar to the ordinary chondrites than larger objects not only in terms of band depth but also visual continuum slope. It may be that there are two competing effects altering asteroid spectra, one which causes objects smaller than $\sim$6 km to redden with increasing size, and another which produces the opposite size dependence for larger objects.

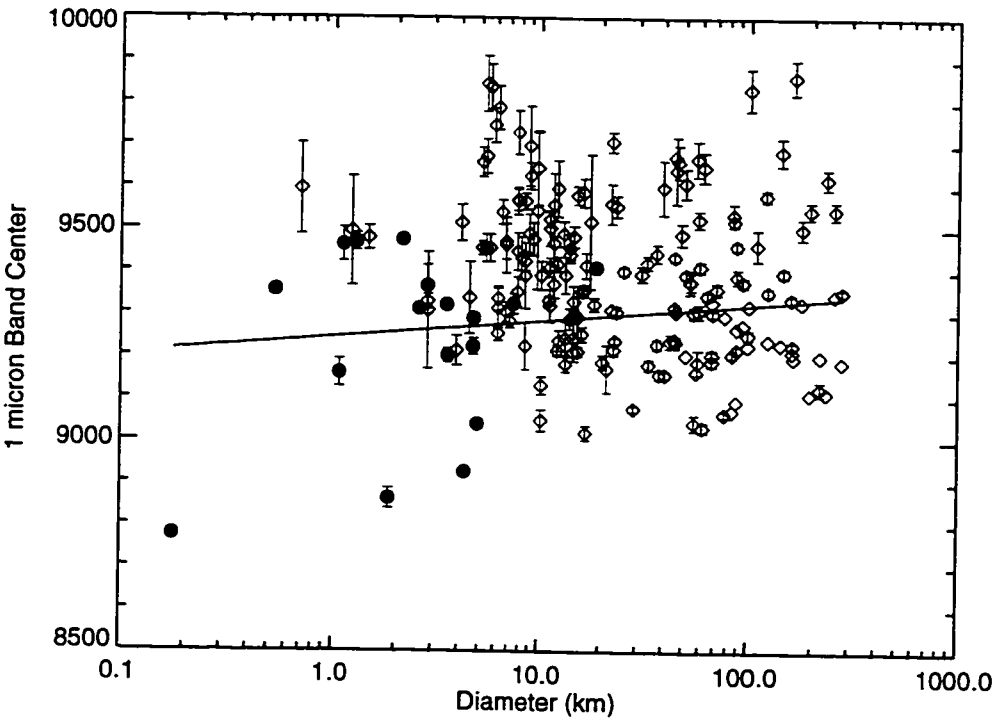


Figure 5.3: Plot of 1-μm band center vs. size. NEOs observed in this survey are plotted as solid circles, and S-type asteroids from Xu et al. (1995) (mostly main-belt asteroids, but including five NEOs) are indicated by hollow diamonds. A log-linear fit to these data is shown.

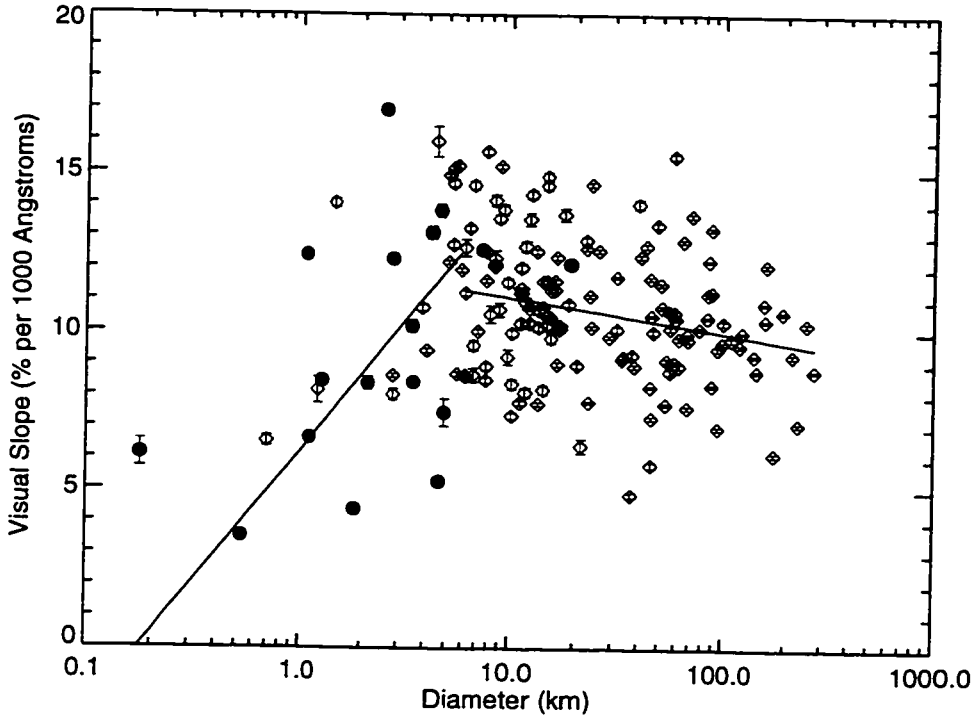


Figure 5.4: Plot of visual slope vs. size. Objects observed in this survey are plotted as solid circles, and S-type asteroids from Xu et al. (1995) are indicated by hollow diamonds. Separate fits to the <6 km and ≥6 km regimes are shown.

Chapter 6

Debiasing Asteroid Taxonomic Distributions

Truth is a good thing; but beware of barking too close to the heels of an error, lest you get your brains kicked out.

— Samuel Taylor Coleridge

6.1 Introduction

To some degree, the distribution of taxonomic types among the NEOs should resemble the distribution of types in its source population. In a simple universe, the NEO population would be identical to its progenitor population. There are several factors which complicate this prospect, however.

First, the NEOs may derive from several distinct sources. The two most commonly suggested sources are main-belt asteroids and extinct comets. While main-belt asteroids have been fairly well studied spectroscopically, only a handful of comet nuclei have been similarly observed (Luu & Jewitt 1990) . Extinct or dormant comets are even less well characterized, with only one definite example known (4015 Wilson-Harrington, e.g., Fernández et al. 1997). However, it is easier to state what a comet nucleus should *not* look like: comets contain copious amounts of organic material, which should render their surfaces dark and drastically reduce the contrast of any spectral absorption features (Luu 1993). There is no known process which would selectively remove this organic component (Rahe et al. 1994), so objects which have high albedos, or those which have strong spectral absorptions, can almost certainly be excluded as candidates for extinct

comets. We don't know very well what comet nuclei look like, but we do have a good idea of what they *don't* look like.

Second, the average size of NEOs which have been taxonomically classified is much smaller than similarly studied main-belt asteroids. If the taxonomic distributions are size-dependent, then it may be difficult to directly compare NEOs with such potential source populations. A possible taxonomic size dependence is amenable to observational investigation, however, so as long as this problem is acknowledged and taken into account, it may be possible to proceed with a comparison of taxonomic distributions.

Third, the mechanism which delivers progenitor objects into NEO orbits may have some compositional dependence. For example, stronger objects may better survive collisions which propel them into orbital resonances, leading to an over-representation of materially strong objects in the NEO population. The importance of material strength versus self-gravitation in regards to asteroid collisional survivability is a subject of some debate (see Love & Ahrens 1996, and references therein). However, most researchers believe that the transition from strength- to gravity-dominance in controlling asteroid structure occurs somewhere between diameters of 250 m up to a few km. Any possible strength-dependent delivery mechanism should thus be important only at smaller sizes, but since this size range includes the average size of NEOs, one must keep in mind this potential bias to the taxonomic distribution.

Alternatively, surface properties may affect the delivery of objects into resonances via the Yarkovsky effect (Farinella et al. 1998, Vokrouhlický & Farinella 1998). This effect, which is due to the offset between the direction of the absorption of sunlight and the reradiation of infrared radiation by a rotating body, is most significant for objects ranging between 0.1 and 100 m in diameter (Farinella et al. 1998). The strength of the effect is dependent on the insulating properties of the regolith. Additionally, those objects with longer collisional lifetimes will experience greater total orbital excursions due to the Yarkovsky effect than shorter-lived objects. This could cause taxonomic biases among the smallest solar system objects subject to this effect.

A fourth and perhaps most severe complication is that the observed populations of NEOs and main-belt asteroids are subject to strong observational

selection effects. Since most detection and physical observations are performed at visual or near-infrared wavelengths, the most prominent bias is that favoring objects with higher apparent visual brightnesses. The apparent brightness of any given object is dependent on several factors, including the geometric circumstances of its apparition (i.e., the Sun-object and Earth-object distances, and the Sun-object-Earth phase angle), and several intrinsic properties of the object (i.e., the object's diameter, slope parameter G, and geometric albedo in the appropriate wavelength band). The apparition geometry is in turn dependent on the orbital elements of the object, and the time at which the object is observed.

To provide a coherent sample for comparison purposes, the debiasing procedure (which will be described in the following sections) has also been performed on the observed taxonomic distributions of the main-belt asteroids.

6.2 Debiasing Methodology

The number of objects which have been taxonomically classified can be expressed as

$$N_{tax}(a, e, i, H, t) = N_{true}(a, e, i, H, T) P_{tax,total}(H) P_{det,total}(a, e, i, H, T), \quad (6.1)$$

where N_{true} represents the total number of objects, $P_{tax,total}$ the probability that an object which has already been detected has also been taxonomically classified (this may also be regarded, somewhat inaccurately, as the taxonomic completeness), and $P_{det,total}$ the detection completeness (or equivalently, the probability that an object has been detected in the first place). Before an object's taxonomic class is known, its apparent magnitude is typically calculated using the slope parameter appropriate for the S-type asteroids. Therefore, the selection of targets for taxonomical classification observations (which is almost always apparent magnitude-limited) is *not* dependent on taxonomy. Since various spectroscopic surveys have specifically targeted certain asteroid families, regions of the main belt, or orbital classes of objects (including NEOs), P_{tax} is somewhat dependent on the orbital elements of an object. P_{det} is dependent on properties which are themselves dependent on taxonomy, so P_{det} must be derived using models of the object population.

6.3 The Detection Simulation

The simulation I perform in this section follows the technique presented in Rabinowitz (1993), but extended to allow for differences in geometric albedo p, slope parameter G, and color for objects of different taxonomic types. Orbital element space was split up into 3024 cells: 24 bins in semimajor axis, 0.2 AU wide and ranging from 0.7 to 5.2 AU; 9 bins in eccentricity, 0.1 units wide and ranging from 0 to 0.9; and 14 bins in inclination, 5° wide and ranging from 0° to 70°. Test objects were created with orbital elements distributed uniformly within each bin; the remaining orbital elements Ω, ω, and M were chosen randomly from a uniform distribution between 0° and 360°. Object positions were then calculated analytically for a randomly chosen time of observation. If an object was found within 30° of the opposition point, it was scored as a "potential detection" and the circumstances of its apparition were recorded. The 30° limit was chosen as being representative of the observing practices of the major asteroid and NEO detection programs (Rabinowitz 1993, Helin & Dunbar 1990).

This process was continued within each orbital element bin until either 1000 objects were scored as "potential detections," or 1 million trials took place. In total, 243,918,704 objects were tested, with 2,954,807 scored as potential detections. The simple ratio of number "potentially detected" to number tested is a measure of the probability that an object with elements (a,e,i) is situated so that it is potentially detectable at any given time. This is displayed in Figure 6.1. In general, objects with lower inclinations and larger semimajor axes are more likely to be found close to the opposition point than objects with high inclinations and orbits closer to the Earth.

Another such run was performed, simulating the velocity-dependent detection criteria characteristic of Spacewatch NEO searches (Jedicke 1996). In this case, only those objects exceeding certain velocity thresholds were counted as potential detections. In this run, 1,987,727 "potential detections" were scored out of 1,717,326,208 trials. Figure 6.2 displays the orbital element dependence of the potential detection probability for the Spacewatch-type simulation. Following the procedure outlined in Rabinowitz (1993), each object potentially detected with elements (a,e,i) can be characterized by a magnitude offset $V^* - H$, which is dependent on the circumstances of the apparition. Using the magnitude law

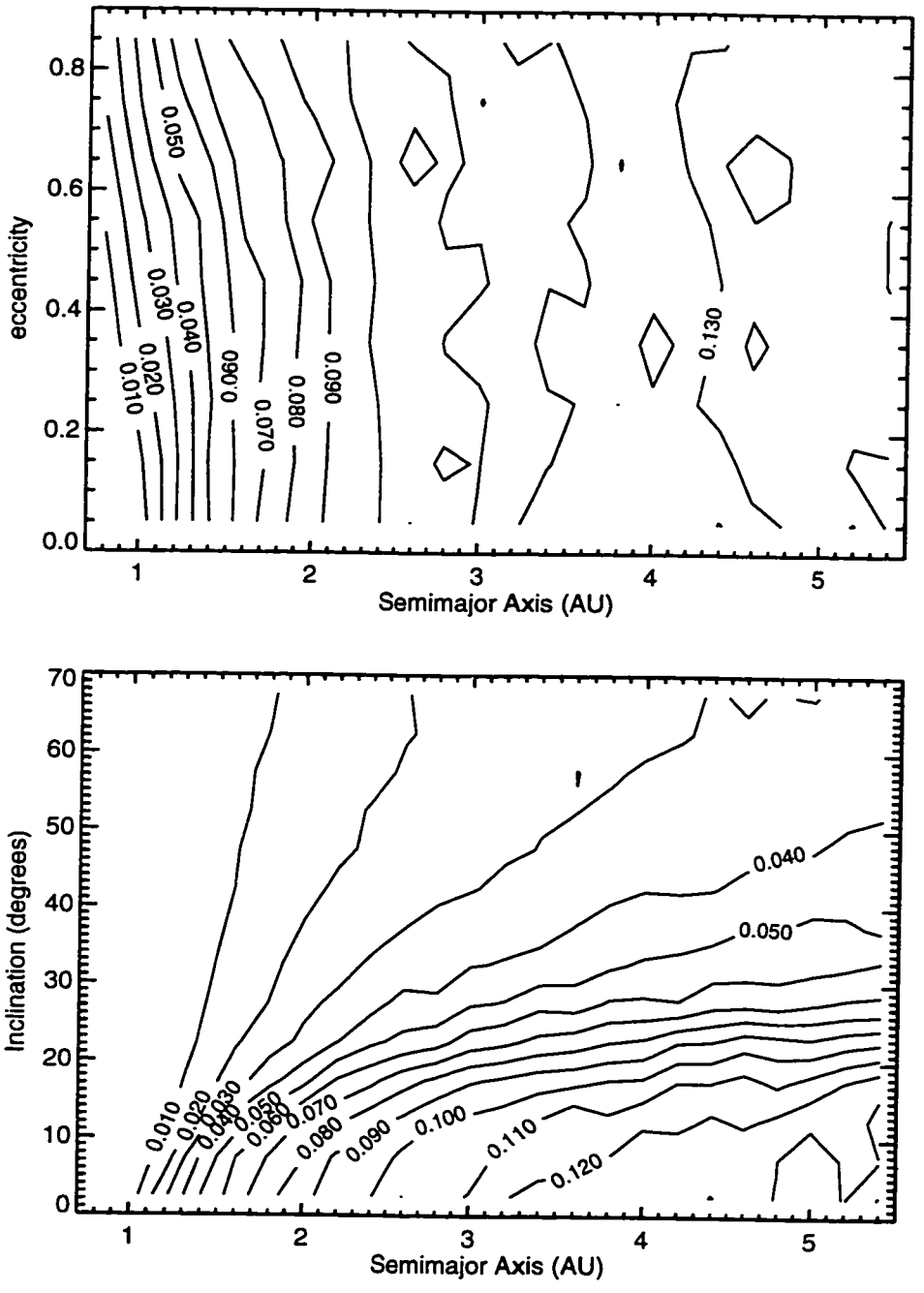


Figure 6.1: Probability that an object will be potentially observed at any given time, dependent on orbital semimajor axis, eccentricity, and inclination. The top e vs. a contour plot is for $i = 0°$, the bottom i vs. a contour plot is for $e = 0$.

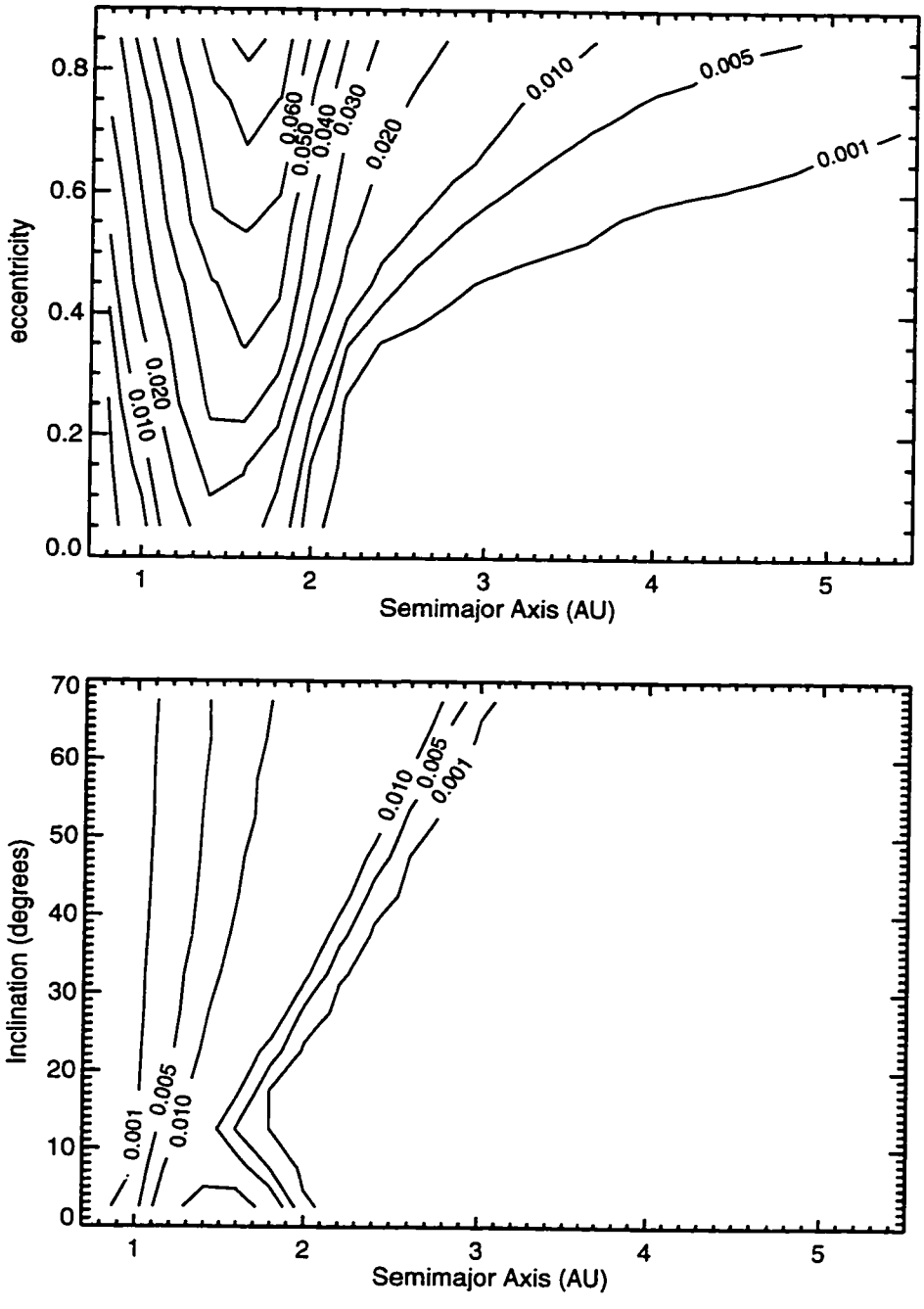


Figure 6.2: Fraction of objects potentially observed at any given time, dependent on orbital semimajor axis, eccentricity, and inclination, for a Spacewatch-type set of velocity-dependent detection criteria. The top e vs. a contour plot is for $i = 0°$, the bottom i vs. a contour plot is for $e = 0$.

Table 6.1: Average geometric albedo p, slope parameter G, and $B-V$ colors for 12 taxonomic classes.

Taxonomic Class	p	G	$B-V$ (mag)
A	0.262	0.350	1.042
B	0.071	0.100	0.672
C	0.052	0.100	0.702
D	0.043	0.100	0.751
E	0.422	0.350	0.706
F	0.042	0.100	0.632
G	0.074	0.100	0.744
M	0.150	0.190	0.703
P	0.043	0.100	0.700
Q	0.212	0.250	0.834
S	0.175	0.250	0.859
V	0.197	0.350	0.800

in Bowell et al. (1989), $V^* - H$ is calculated through

$$V^* - H = 5\log(r\Delta) + \delta_{trail} - 2.5\log[(1 - G(T))\Phi_1(\alpha) + G(T)\Phi_2(\alpha)], \quad (6.2)$$

where V^* is the apparent, trailed magnitude of the object, H is the absolute magnitude of the object, r and Δ the heliocentric and geocentric distances respectively, δ_{trail} is the trailing loss due to object motion during an exposure, G is the slope parameter, α the Sun-object-Earth phase angle, and Φ_1 and Φ_2 given by equation (A5) in Bowell et al. (1989). Since G is dependent on object surface characteristics (and taxonomic class T), so is the magnitude offset $V^* - H$. Table 6.1 lists average parameters for different taxonomic classes. It must be noted here that the X class represents those objects with E, M, or P-type spectra for which no albedo information is available, and since without albedo information it is impossible to derive sizes, X class objects are not explicitly handled as a unique group through the rest of these debiasing steps. Geometric albedo P and slope parameter G were derived from diameters and absolute magnitudes in Bowell (1998). $B - V$ colors were derived through synthetic photometry using

average spectra in Xu et al. (1995).

The trailing loss δ_{trail} is a function of exposure time, object apparent velocity, and atmospheric seeing, and becomes important when an object moves more than a seeing disk during an exposure. Trailing loss is calculated by the following fit to data presented in Figure 2 of Jedicke & Herron (1997):

$$\delta_{trail} = 1.25[\sqrt{(\xi - 0.035)^2 + 0.04262)} + (\xi - 0.035)], \tag{6.3}$$

and

$$\xi = \log(t_{exp}\omega/FWHM), \tag{6.4}$$

where t_{exp} is the exposure time, ω the apparent velocity, and $FWHM$ the full-width-half-maximum of the seeing disk, in the appropriate units. This expression has the advantage over the one presented in Jedicke & Herron (1997) in that this function is continuous rather than discrete.

The detection probability for an object of apparent magnitude V^* at a *single* apparition and in a single observation is

$$P_{det}(a, e, i, V^*, T) = V^*(a, e, i, T)\epsilon_{detect}(V^*)P_{pot}(a, e, i)(A_{det}/A_{survey}), \tag{6.5}$$

where ϵ_{det} is the threshold function for detection, $P_{pot}(a, e, i)$ is the probability that an object is positioned so as to be "potentially detectable," and A_{det}/A_{survey} is the ratio of the active detector area to the total survey area.

The threshold function was modeled in the form

$$\epsilon_{detect}(V^*) = 0.5(1 - \mathrm{erf}[(V^* - V_{50})/\sigma]), \tag{6.6}$$

where **erf** is the standard error function, V_{50} is the magnitude at 50% detection efficiency, and σ is a measure of the steepness of the dropoff in detection efficiency. For simulations of Spacewatch-type programs utilizing CCD's, V_{50} was set to 20.77, and σ to 0.230 to match the threshold function described in Jedicke & Herron (1997). For simulations of photographic search programs, V_{50} was set to 16.5, and σ to 0.7.

The individual detection probability for any object with elements (a,e,i) and absolute magnitude H is calculated as

$$P_{det}(a, e, i, H, T) = 1 - \prod_n (1 - P_{det}(a, e, i, V_n^*, T)). \tag{6.7}$$

It is now possible to estimate $P_{det,total}$ by calculating the probability that any object would be observed given the effective numbers of search exposures which have been made by the various asteroid and NEO search programs to date. At this point, it is necessary to derive the observed detection completeness as a function of absolute magnitude. This is the only part of the entire debiasing process which relies on model distributions of size, orbital elements, and taxonomic types. The (estimated) true size and orbital distributions for the main-belt asteroids were taken from Jedicke (1996). The main-belt was further broken down into three zones to allow the ratio between "bright" (i.e., S-type) asteroids and "dark" (C-type) asteroids to vary with heliocentric distance. The definitions of these zones, and the bright/dark ratios for the main belt were taken from Jedicke & Metcalfe (1998). The (estimated) true size and orbital distributions for the NEOs were taken from Rabinowitz (1993), with the numbers of "bright" and "dark" objects being set equal. I note here that it would be possible to rederive the estimated true distributions of the main-belt and NEO populations given the detection simulations which have been performed, but those results are extremely sensitive to the assumptions made about detector performance and other search criteria, and it was felt that such an undertaking was beyond the immediate scope of this work. The derivation of a taxonomic bias is less sensitive to such underlying assumptions, because it is the *ratio* of completenesses between different taxonomic classes that is important, and not the *absolute* level of completeness of any one of them.

The total detection probability may be expressed as

$$P_{det,total}(a, e, i, H, T) = 1 - \prod(1 - P_{det}(a, e, i, V^*, T))^{n_{exp}}, \qquad (6.8)$$

where n_{exp} is the effective number of exposures, which was estimated to be ~1,000 CCD and ~3,500 photographic exposures for main asteroid belt searches, and ~7,000 CCD and ~15,000 photographic exposures for the NEOs.

The taxonomic classification probability $P_{tax,total}$ is derived by calculating the fraction of known objects at a given magnitude which have been taxonomically classified. Since there may be some dependence of P_{tax} on orbital elements, it is necessary to split up the population of NEOs and main-belt asteroids into regions which have been similarly targeted. Table 6.2 lists 22 zones in which the selection effects are judged as most nearly identical. The boundaries of these zones are

Table 6.2: Orbital Element Zones

Zone	Description	Criteria
ATE	Atens	$a \leq 1.0$
APO	Apollos	$q \leq 1.0$
AMO	Amors	$q \leq 1.3$
HU	Hungaria Group	$1.78 \leq a \leq 2.00$, $e \leq 0.18$, $16° \leq i \leq 34°$
MC	Mars Crossers	$q \leq 1.666$
FL	Flora Family	$2.10 \leq a \leq 2.30$, $i \leq 11°$
PH	Phocaea Group	$2.25 \leq a \leq 2.50$, $e \geq 0.10$, $18° \leq i \leq 32°$
NY	Nysa Family	$2.41 \leq a \leq 2.50$, $0.12 \leq e \leq 0.21$, $1.5° \leq i \leq 4.3°$
VES	Vesta Family	$2.28 \leq q \leq 2.50$, $0.04 \leq e \leq 0.16$, $5° \leq i \leq 8°$
I	Main Belt	$2.30 < a \leq 2.50$, $i < 18°$
PAL	Pallas Zone	$2.50 < a < 2.82$, $33° \leq i \leq 38°$
IIa	Main Belt	$2.50 < a \leq 2.706$, $i < 33°$
IIb	Main Belt	$2.706 < a \leq 2.82$, $i < 33°$
KOR	Koronis Zone	$2.83 < a \leq 2.91$, $e \leq 0.11$, $i \leq 3.5°$
EOS	Eos Zone	$2.99 \leq a \leq 3.03$, $0.01 \leq e \leq 0.13$, $8° \leq i \leq 12°$
IIIa	Main Belt	$2.82 < a \leq 3.03$, $e \leq 0.35$, $i \leq 30°$
THE	Themis Zone	$3.08 \leq a \leq 3.24$, $0.09 \leq e \leq 0.22$, $i \leq 3°$
GR	Griqua Group	$3.10 \leq a \leq 3.27$, $e \geq 0.35$
IIIb	Main Belt	$3.03 < a \leq 3.27$, $e < 0.35$, $i \leq 30°$
CYB	Cybele Group	$3.27 < a \leq 3.70$, $e \leq 0.30$, $i \leq 25°$
HIL	Hilda Group	$3.70 < a \leq 4.20$, $e \leq 0.30$, $i \leq 20°$
T	Trojan Group	$5.05 \leq a \leq 5.40$

drawn from Zellner (1979) as augmented and altered by Gradie et al. (1989), and with the addition of a zone for the Vesta family of asteroids, which were targeted heavily by Xu et al. (1995), and the separation of the Amor-Aten-Apollo zone into its three component regions.

Figures 6.3 – 6.10 display the taxonomic probabilities for each of the 22 regions listed in Table 6.2. Taxonomic classifications for 1295 objects have been collected from the literature and produced for those objects observed in this

survey. The catalog of taxonomic classifications and selected other parameters are available from the author.

After the derivation of $P_{det,total}$ and $P_{tax,total}$, *correction factors* may be derived for every object such that any one taxonomically classified object corresponds to the true number of objects which exist in that taxonomic class, at that absolute magnitude, and in that orbital element zone. These correction factors are simply the reciprocals of $P_{det,total}$ and $P_{tax,total}$. A list of these corrections factors for 1295 taxonomically classified objects is available from the author by request. Systematic errors were estimated by varying the numbers of search exposures by factors of two, and recomputing the correction factors accordingly.

Once the true distribution of objects versus magnitude is known, it is then possible to derive the true size distribution. The absolute magnitude H of an object is dependent on its geometric albedo p and its diameter D through the following relation (Muinonen et al. 1995):

$$H(D,p) = 15.648 - 5\log D - 2.5\log p. \tag{6.9}$$

The results of these debiasing efforts are presented in Chapter 7.

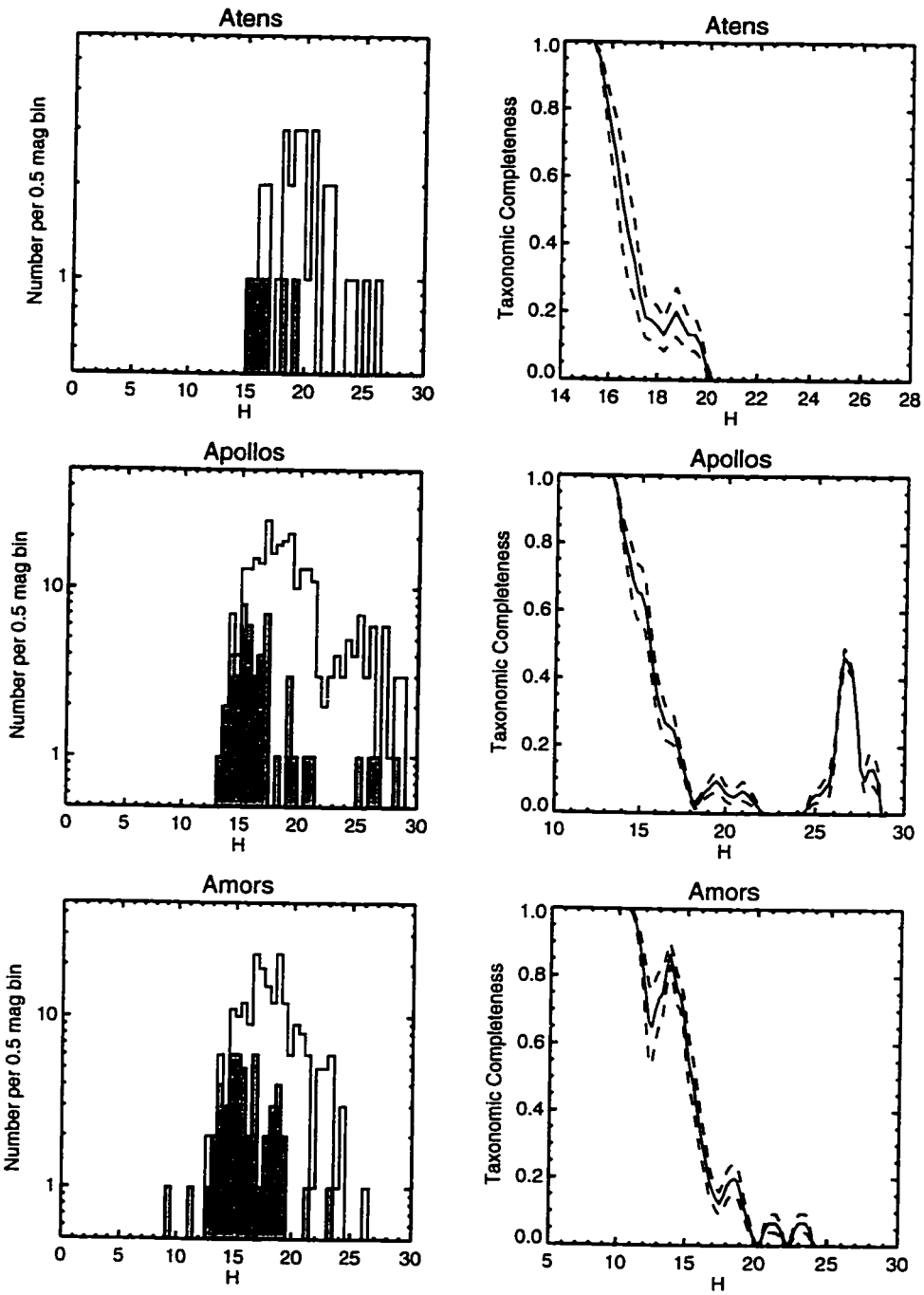


Figure 6.3: Taxonomic completenesses for the Atens, Apollos, and Amors. Figures on the left display histograms of the numbers of detected objects versus absolute magnitude H; the shaded areas represent those objects which have been taxonomically classified. Figures on the right present the derived taxonomic completeness functions versus absolute magnitude. Dashed lines represent 1 σ uncertainties.

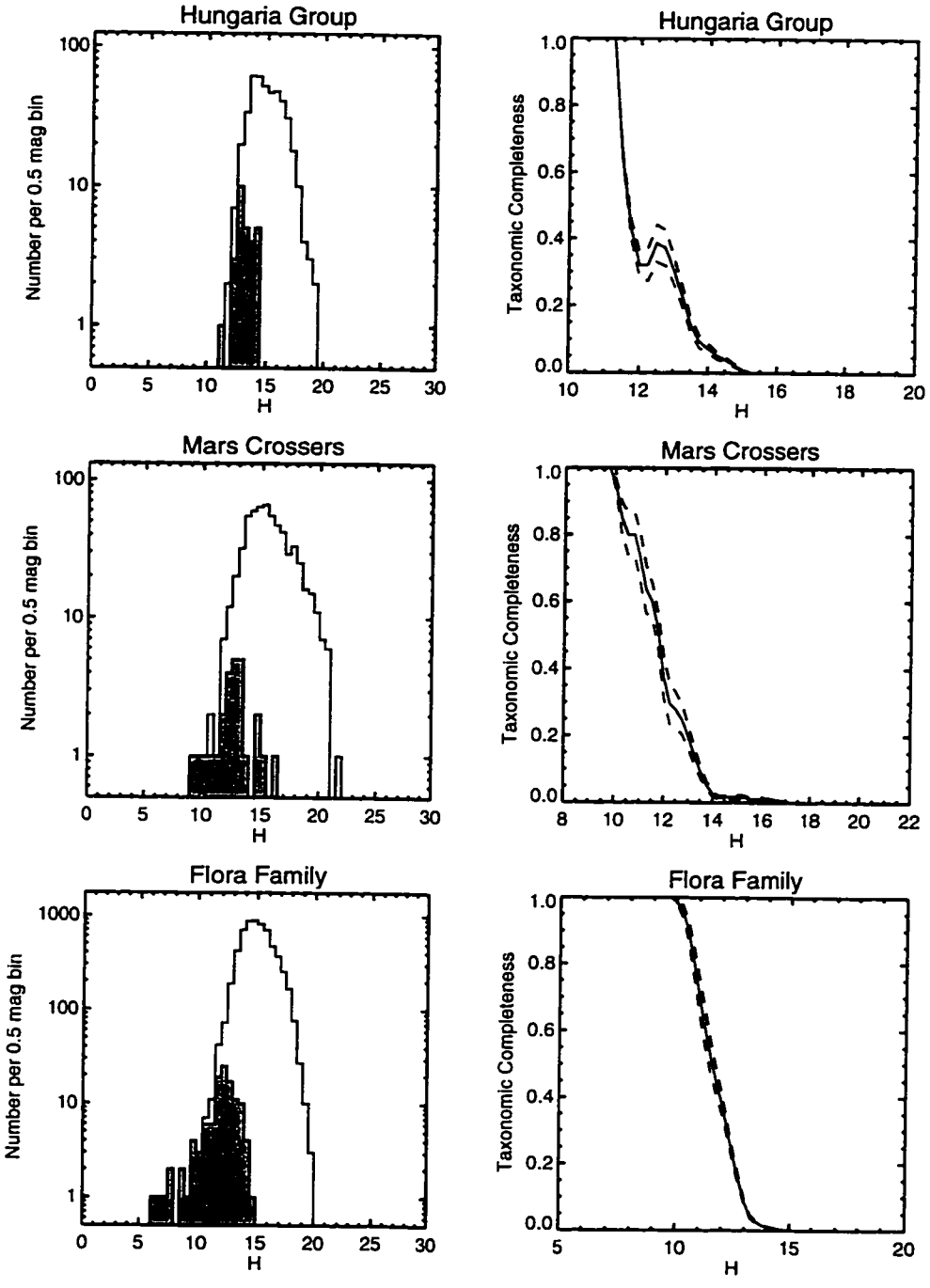


Figure 6.4: Taxonomic completenesses for the Hungaria Group, Mars Crossers, and Flora Family. Figures on the left display histograms of the numbers of detected objects versus absolute magnitude H; the shaded areas represent those objects which have been taxonomically classified. Figures on the right present the derived taxonomic completeness functions versus absolute magnitude. Dashed lines represent 1 σ uncertainties.

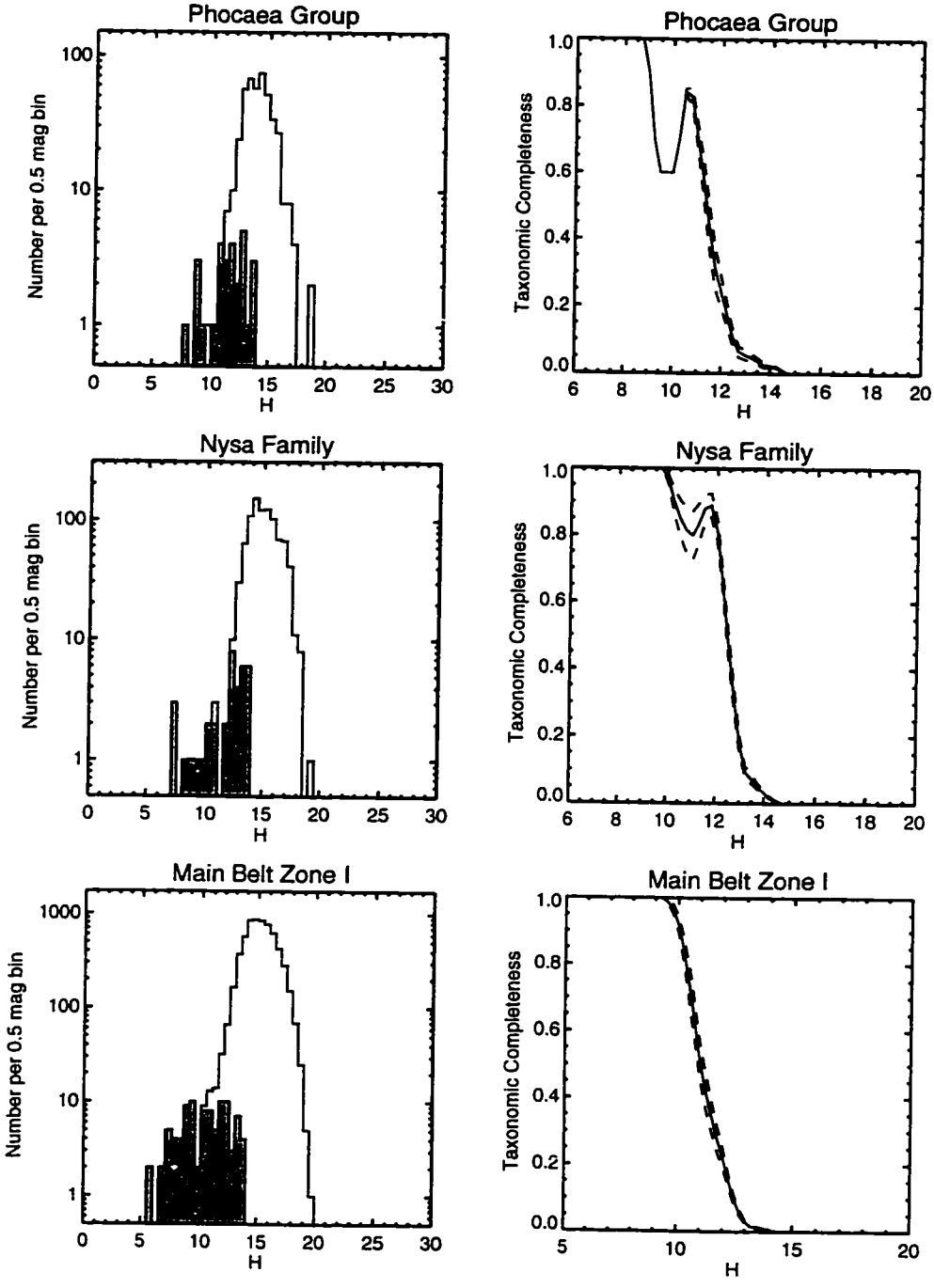


Figure 6.5: Taxonomic completenesses for the Phocaea Group, Nysa Family, and Main Belt I. Figures on the left display histograms of the numbers of detected objects versus absolute magnitude H; the shaded areas represent those objects which have been taxonomically classified. Figures on the right present the derived taxonomic completeness functions versus absolute magnitude. Dashed lines represent 1 σ uncertainties.

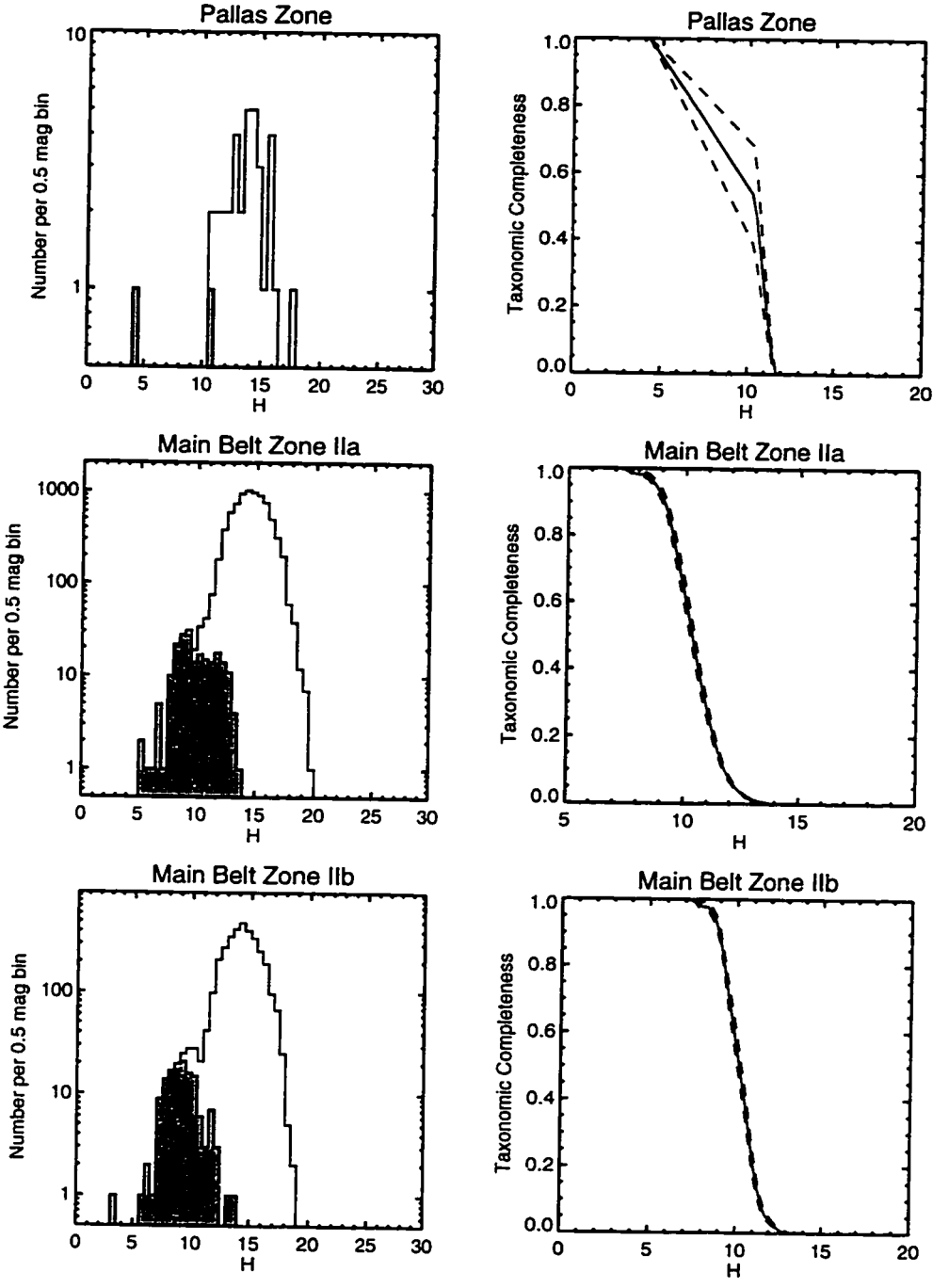


Figure 6.6: Taxonomic completenesses for the Pallas Zone, Main Belt IIa, and Main Belt IIb. Figures on the left display histograms of the numbers of detected objects versus absolute magnitude H; the shaded areas represent those objects which have been taxonomically classified. Figures on the right present the derived taxonomic completeness functions versus absolute magnitude. Dashed lines represent 1 σ uncertainties.

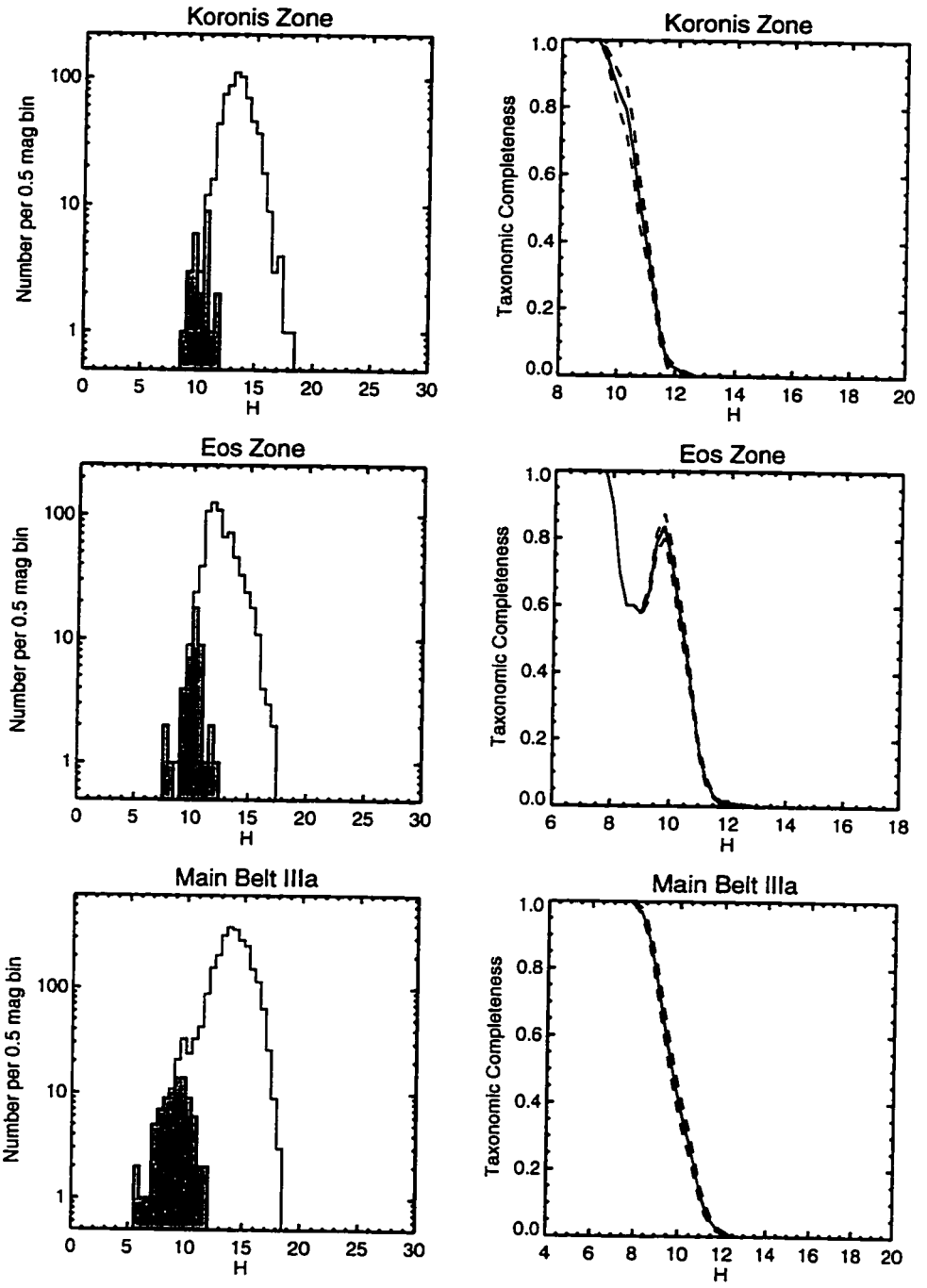


Figure 6.7: Taxonomic completenesses for the Koronis Zone, Eos Zone, and Main Belt IIIa. Figures on the left display histograms of the numbers of detected objects versus absolute magnitude H; the shaded areas represent those objects which have been taxonomically classified. Figures on the right present the derived taxonomic completeness functions versus absolute magnitude. Dashed lines represent 1 σ uncertainties.

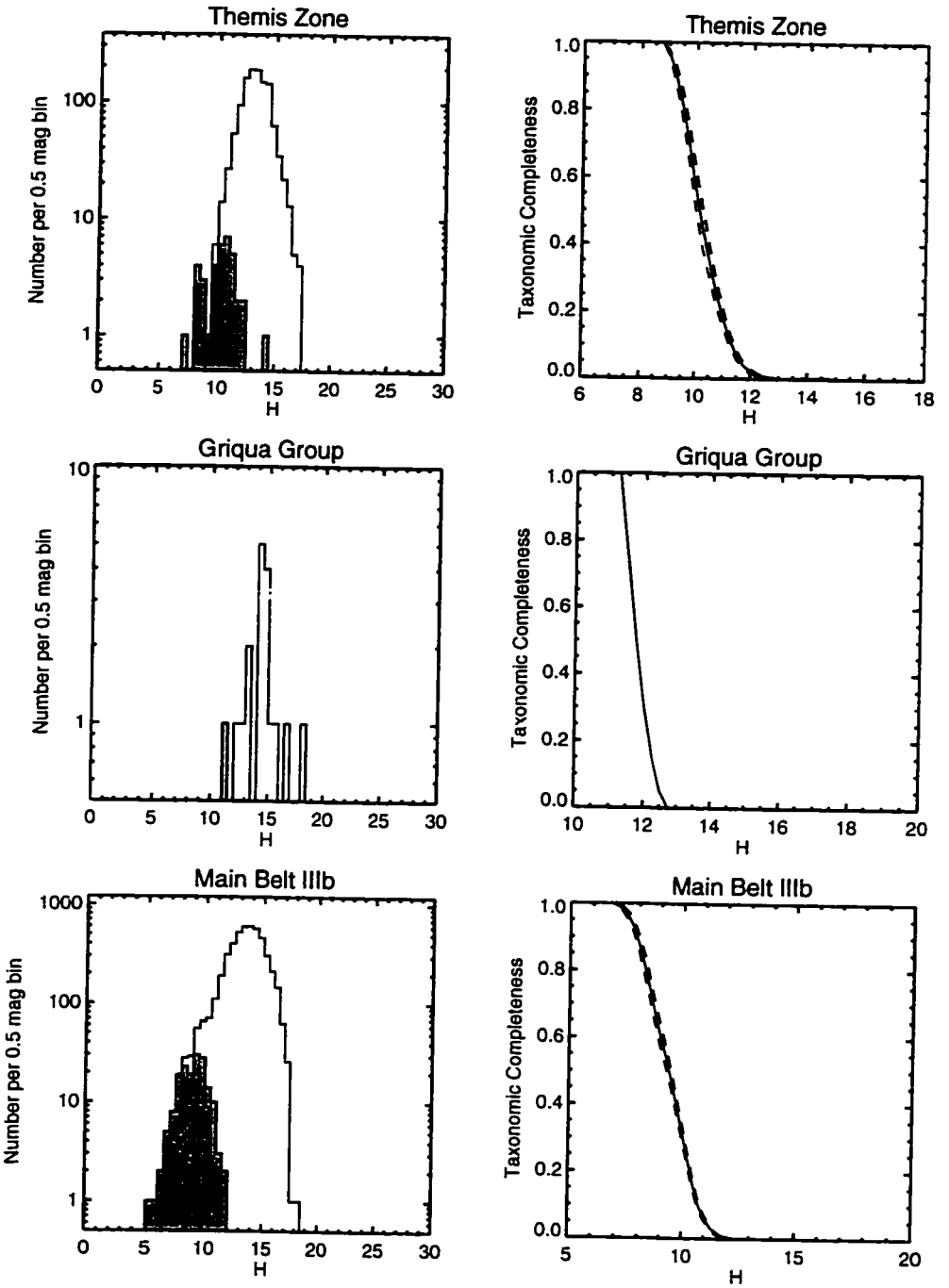


Figure 6.8: Taxonomic completenesses for the Themis Zone, Griqua Group, and Main Belt IIIb. Figures on the left display histograms of the numbers of detected objects versus absolute magnitude H; the shaded areas represent those objects which have been taxonomically classified. Figures on the right present the derived taxonomic completeness functions versus absolute magnitude. Dashed lines represent 1 σ uncertainties.

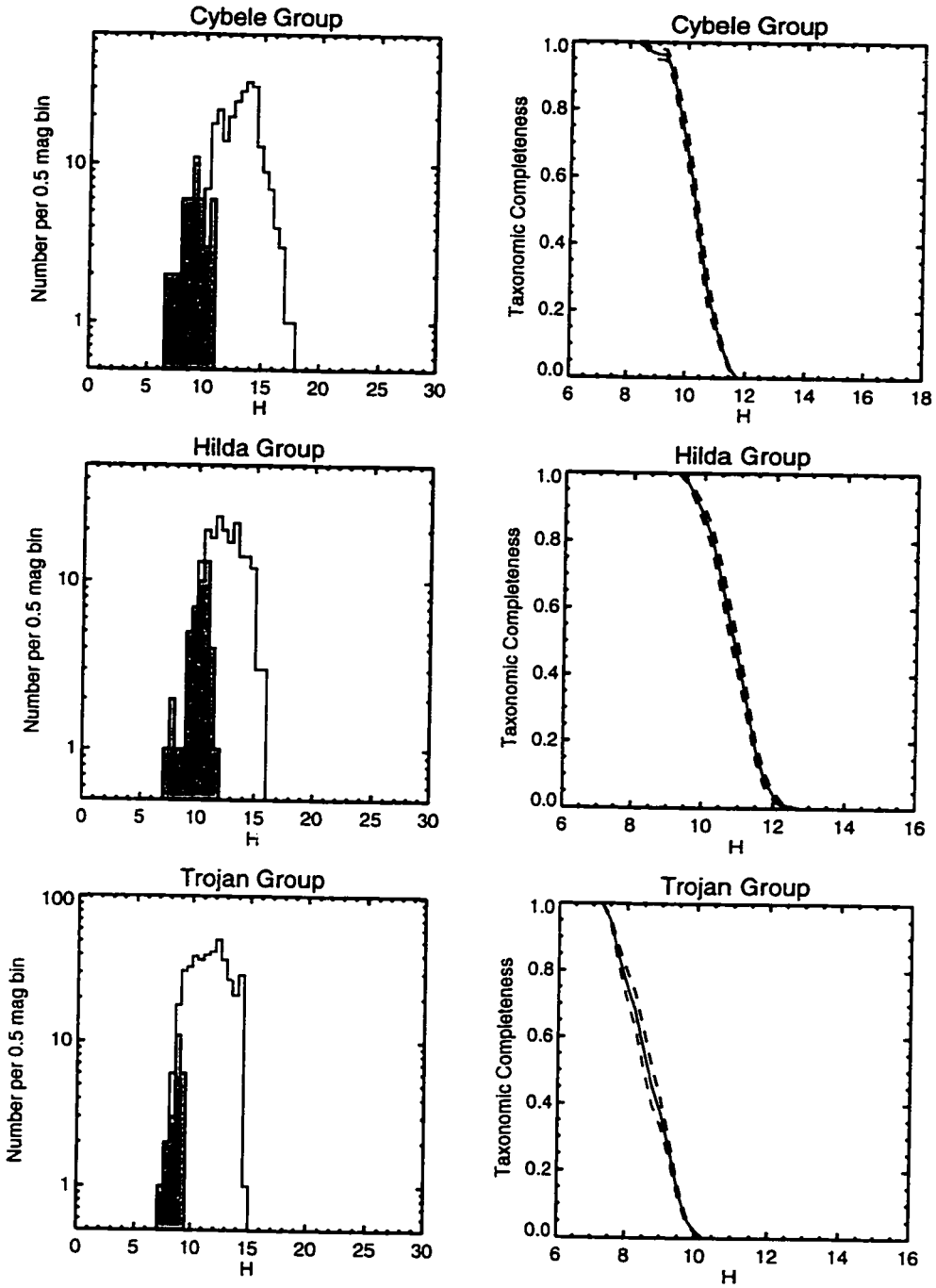


Figure 6.9: Taxonomic completenesses for the Cybele Group, Hilda Group, and Trojan Group. Figures on the left display histograms of the numbers of detected objects versus absolute magnitude H; the shaded areas represent those objects which have been taxonomically classified. Figures on the right present the derived taxonomic completeness functions versus absolute magnitude. Dashed lines represent 1 σ uncertainties.

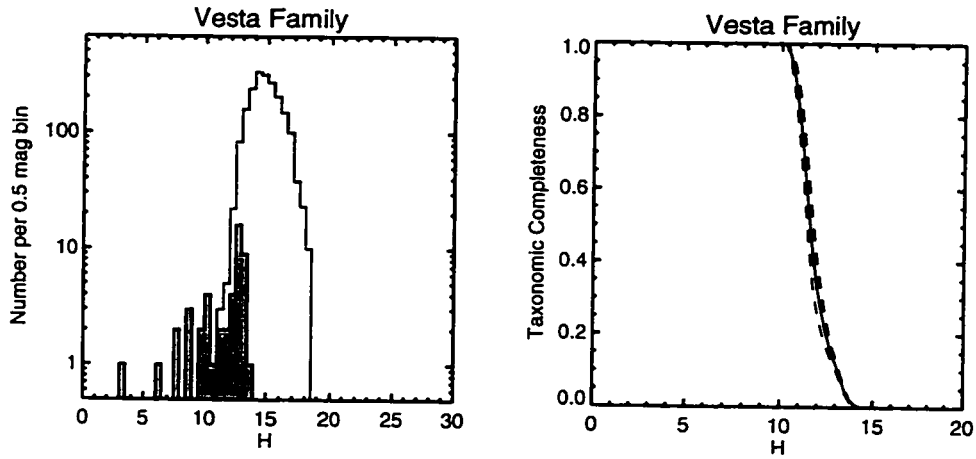


Figure 6.10: Taxonomic completenesses for the Vesta Family. The figure on the left displays a histogram of the numbers of detected objects versus absolute magnitude H; the shaded areas represent those objects which have been taxonomically classified. The figure on the right presents the derived taxonomic completeness function versus absolute magnitude. Dashed lines represent 1 σ uncertainties.

Chapter 7

Discussion of Debiasing Results

I wish I were a little rock,
A-sitting on a hill,
A-doing nothing all day long,
But just a sitting still;
I wouldn't eat, I wouldn't sleep,
I wouldn't even wash –
I'd sit and sit a thousand years,
And rest myself, b'Gosh!

— Frederick Palmer Latimer

Figures 7.1 – 7.22 present the bias-corrected cumulative size distributions of objects in each of 22 orbital element zones. Slopes for three different size exponents are overplotted; this is the exponent b in the traditional expression

$$N(> D) = AD^{-b+1}. \tag{7.1}$$

An exponent $b = 3.5$ corresponds to the expected value for a collisionally evolved system (Dohnanyi 1971).

Figure 7.23 displays the bias-corrected cumulative size distributions for the combined NEO population (i.e., the Amors, Atens, and Apollos).

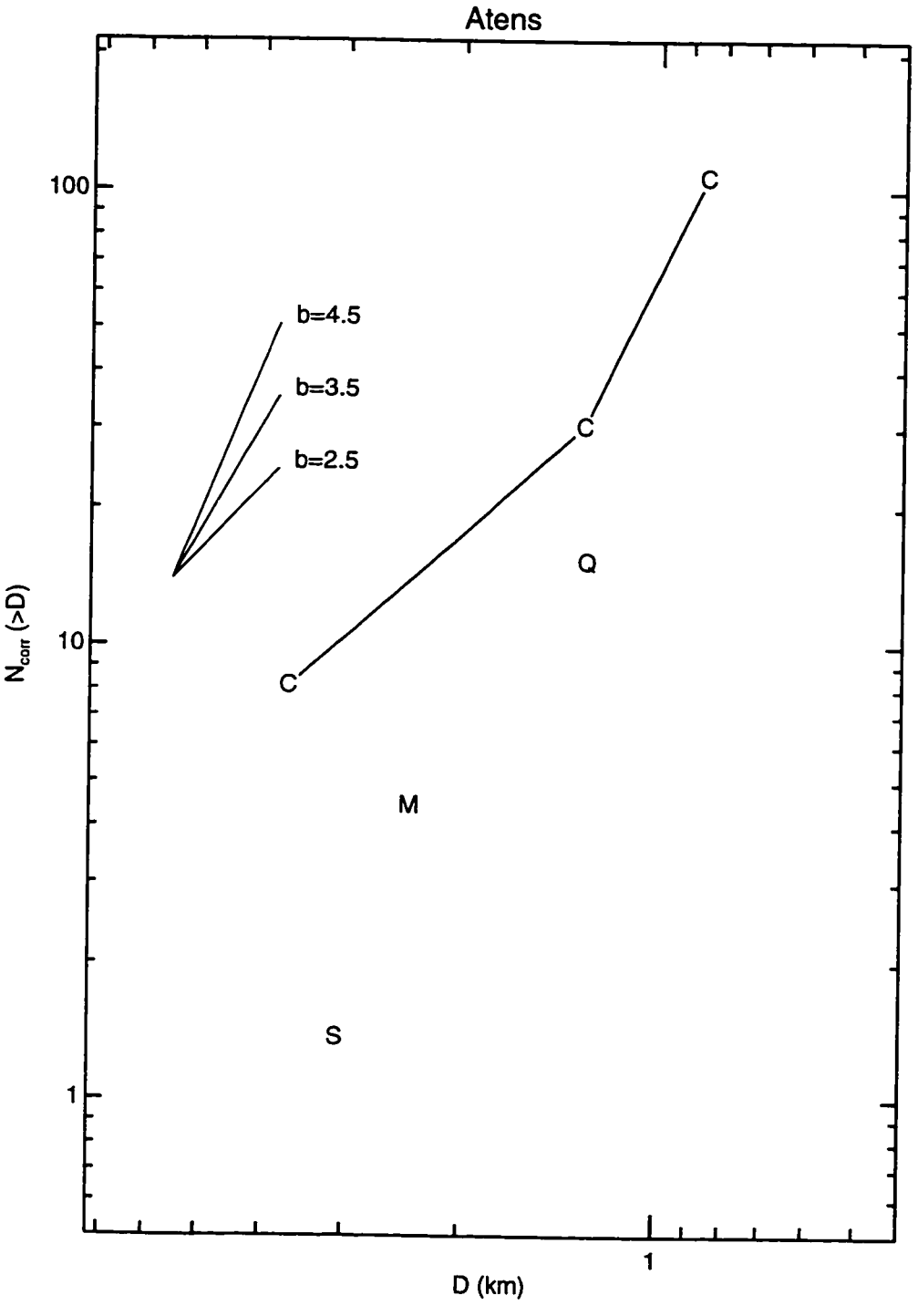


Figure 7.1: Cumulative Size-Frequency Diagram for the Atens.

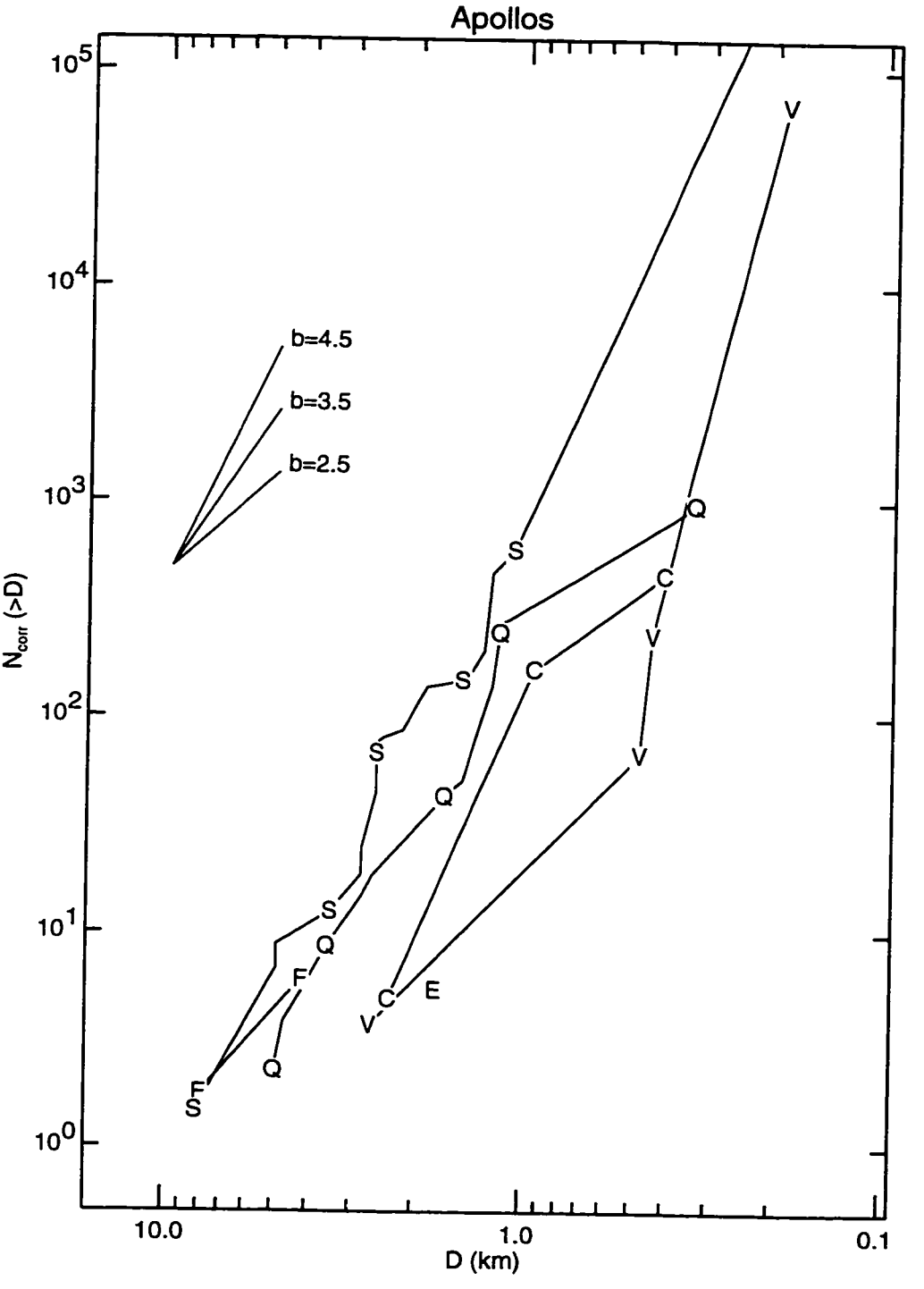


Figure 7.2: Cumulative Size-Frequency Diagram for the Apollos.

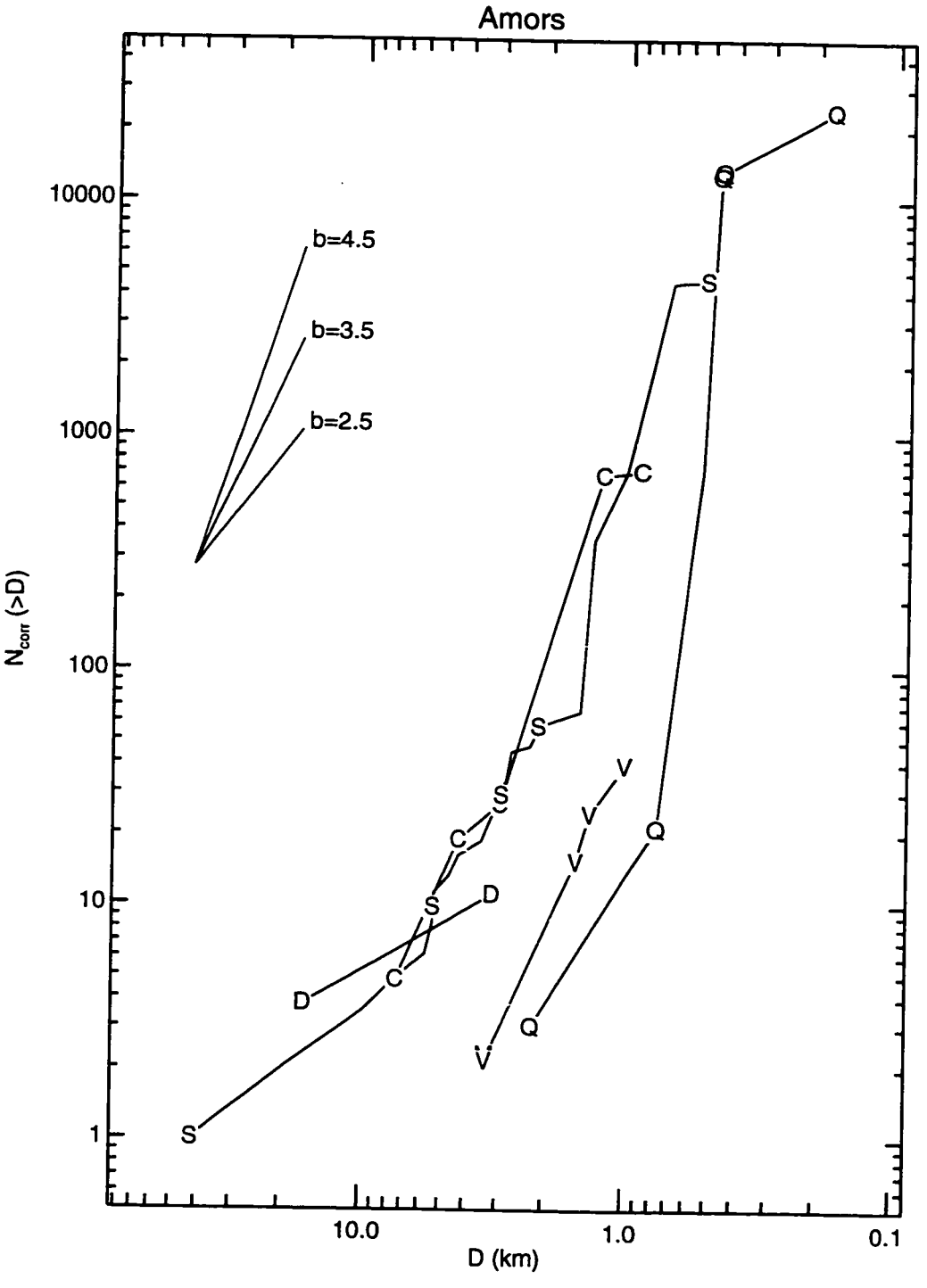


Figure 7.3: Cumulative Size-Frequency Diagram for the Amors.

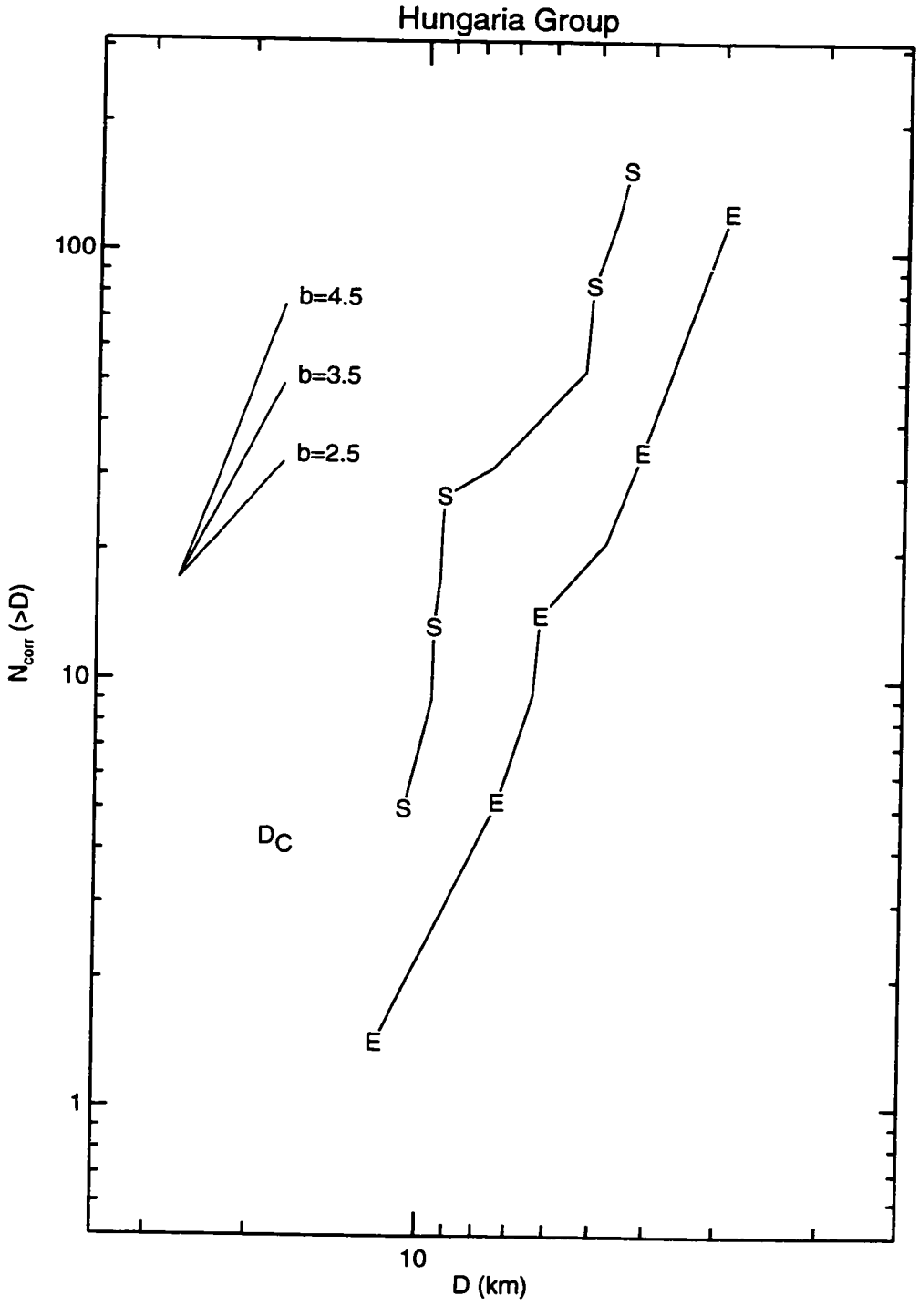


Figure 7.4: Cumulative Size-Frequency Diagram for the Hungaria Group.

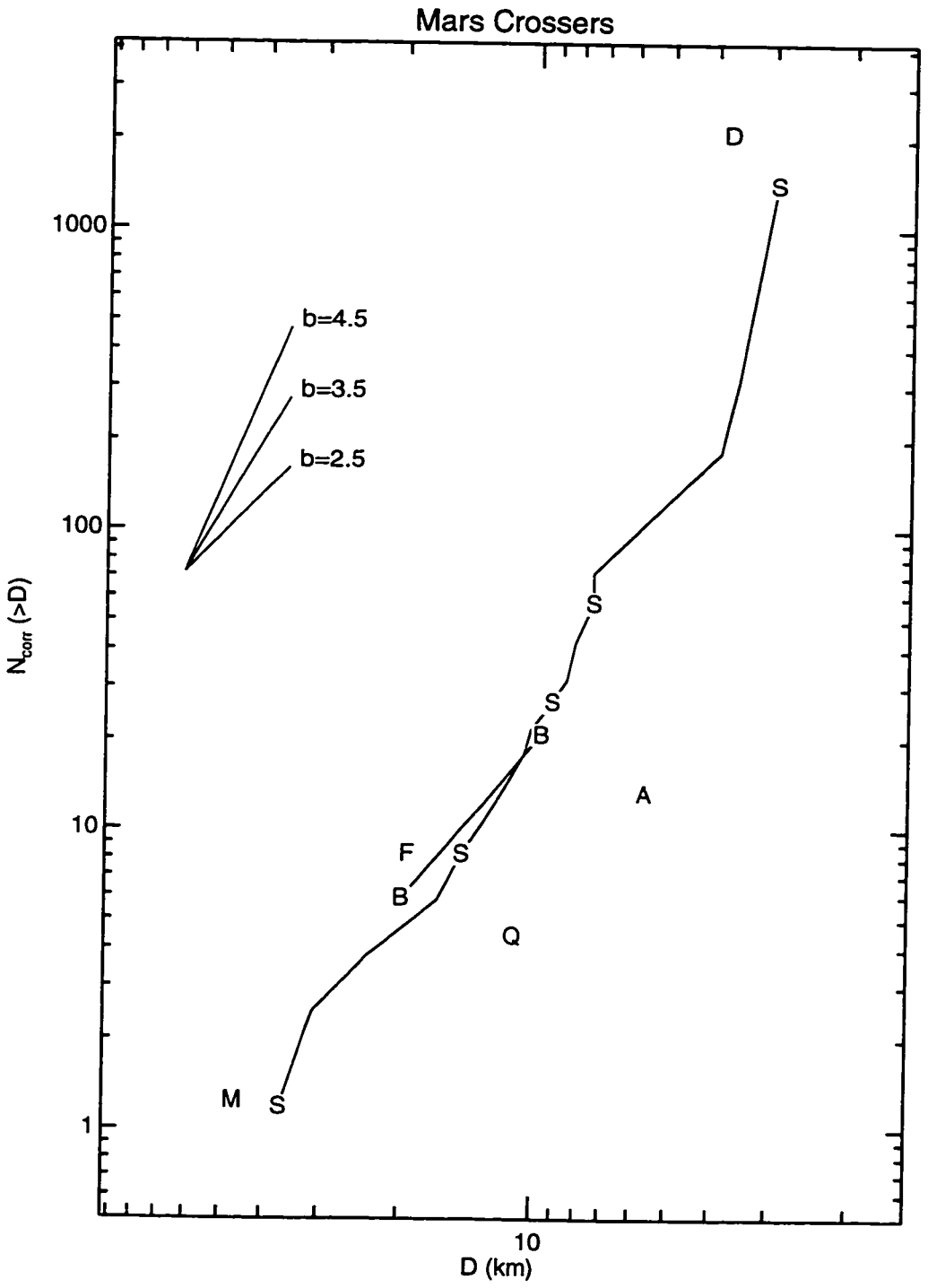


Figure 7.5: Cumulative Size-Frequency Diagram for the Mars Crossers.

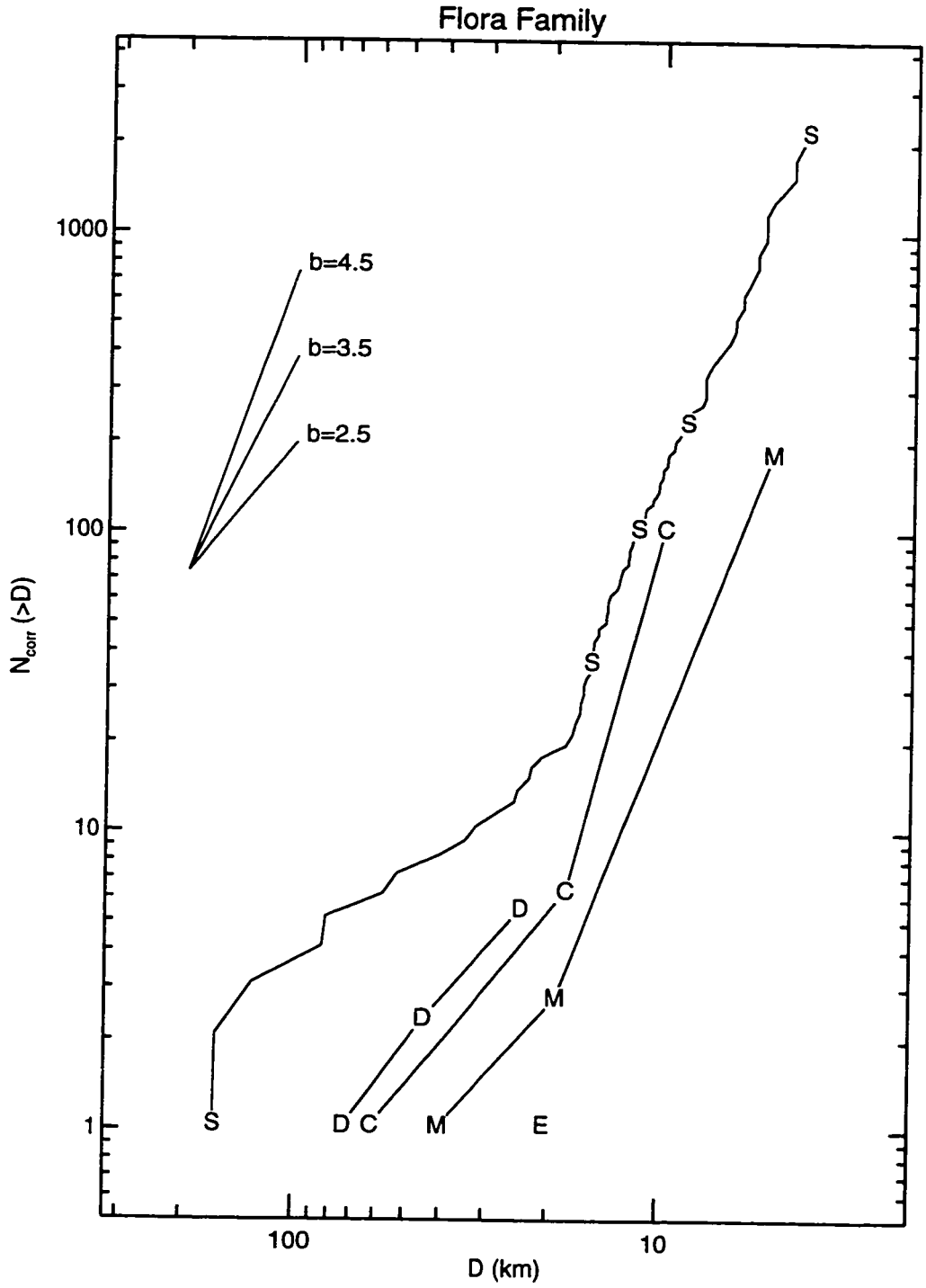


Figure 7.6: Cumulative Size-Frequency Diagram for the Flora Family.

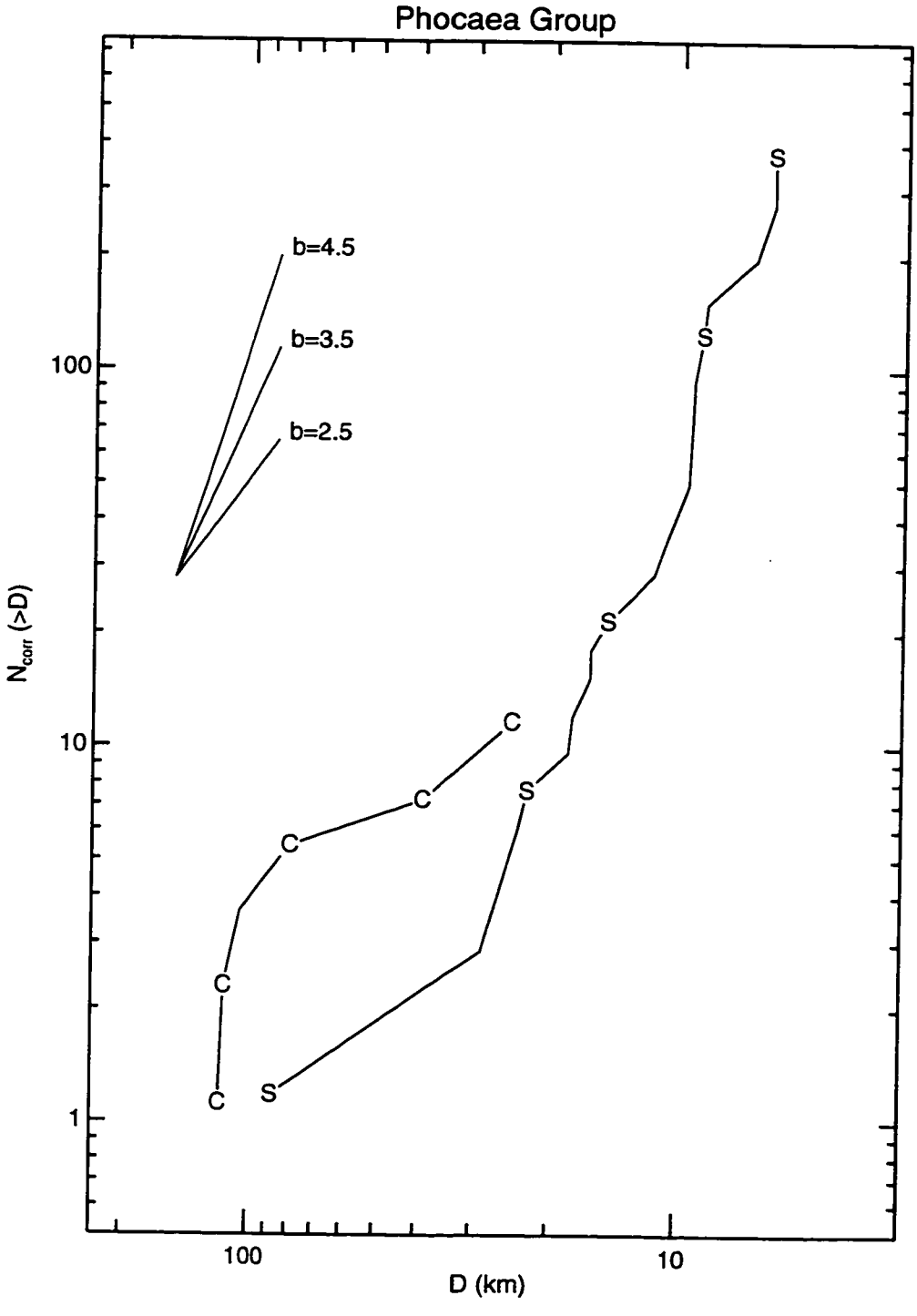


Figure 7.7: Cumulative Size-Frequency Diagram for the Phocaea Group.

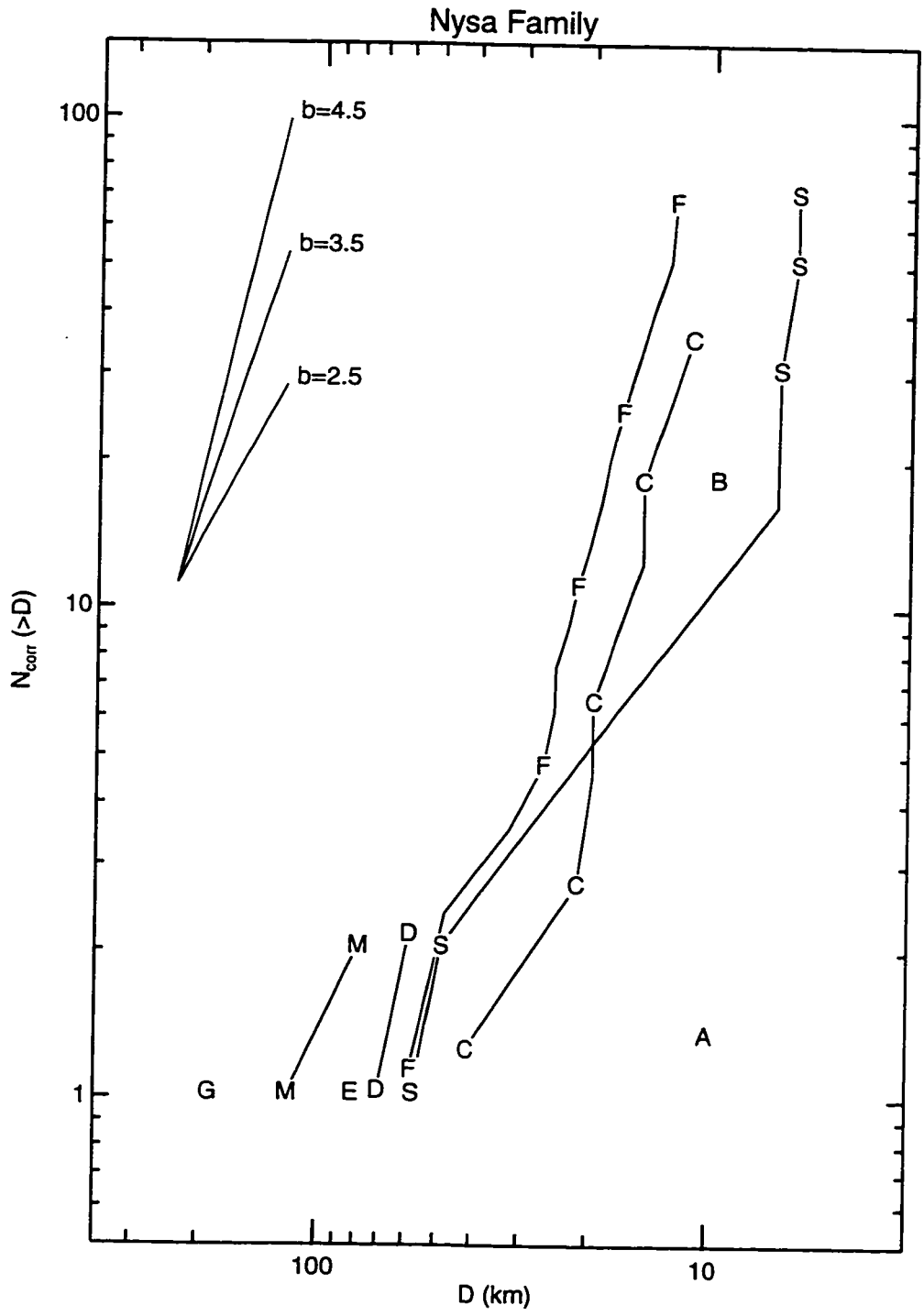


Figure 7.8: Cumulative Size-Frequency Diagram for the Nysa Family.

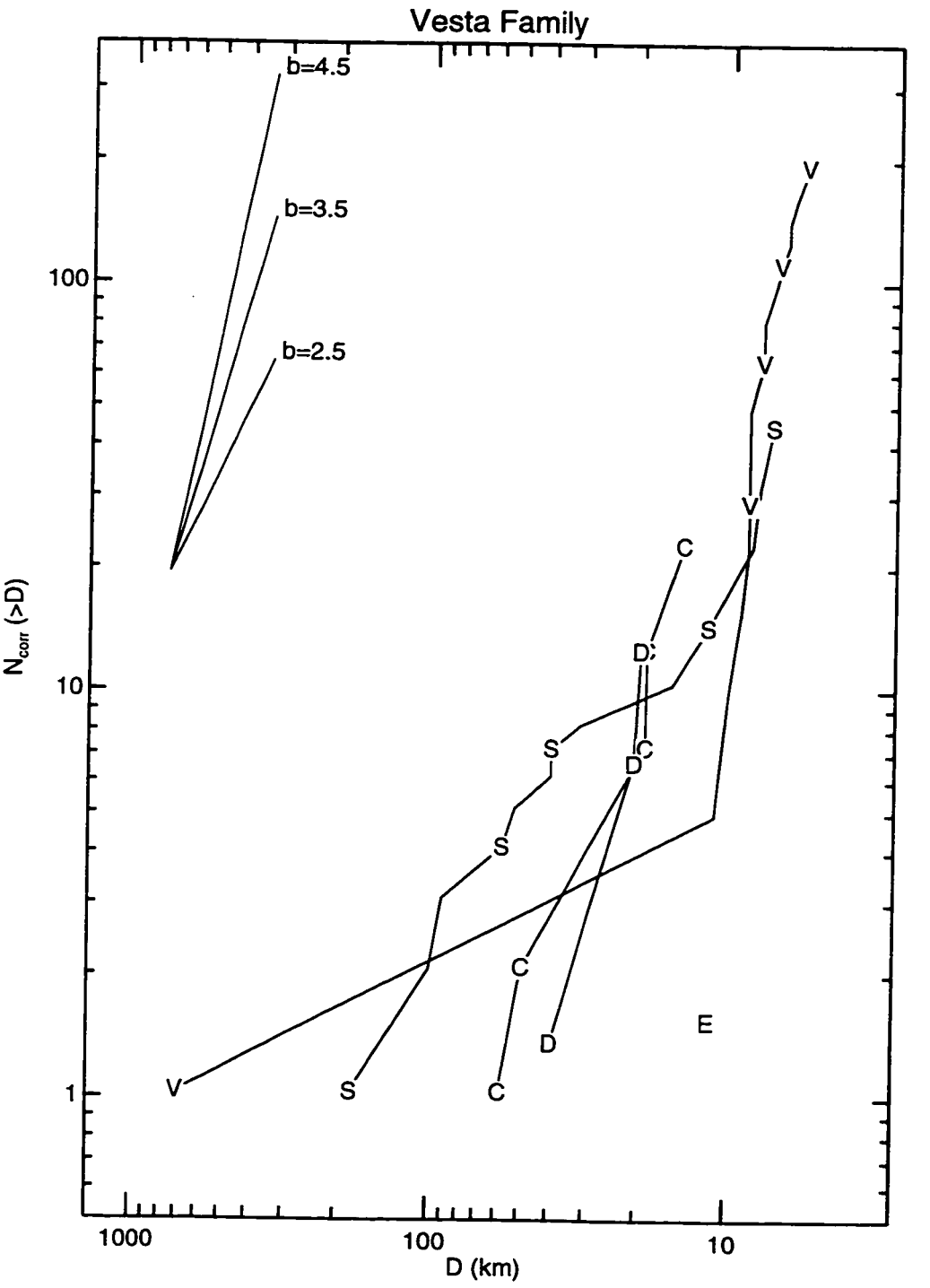


Figure 7.9: Cumulative Size-Frequency Diagram for the Vesta Family.

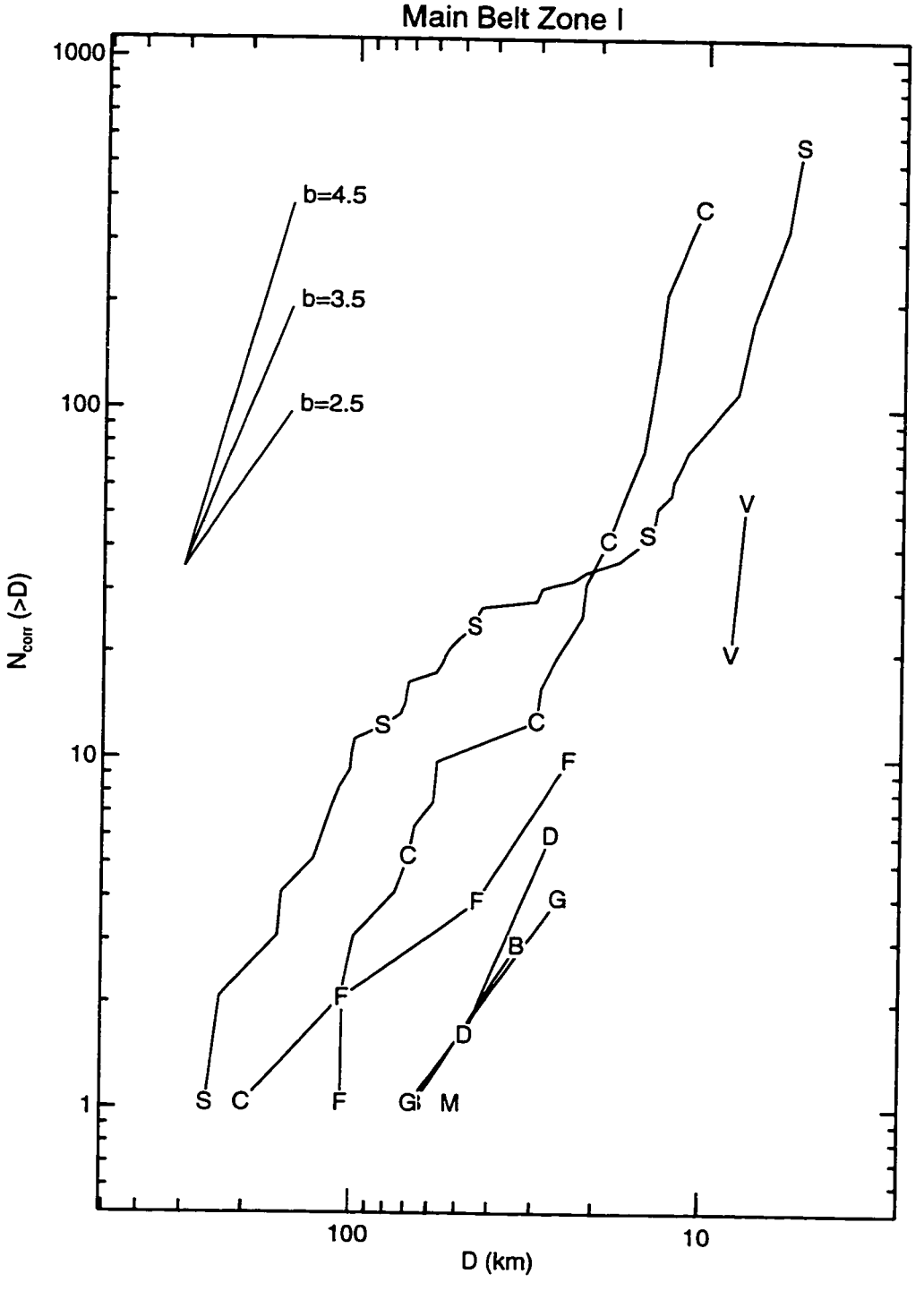


Figure 7.10: Cumulative Size-Frequency Diagram for Main Belt Zone I.

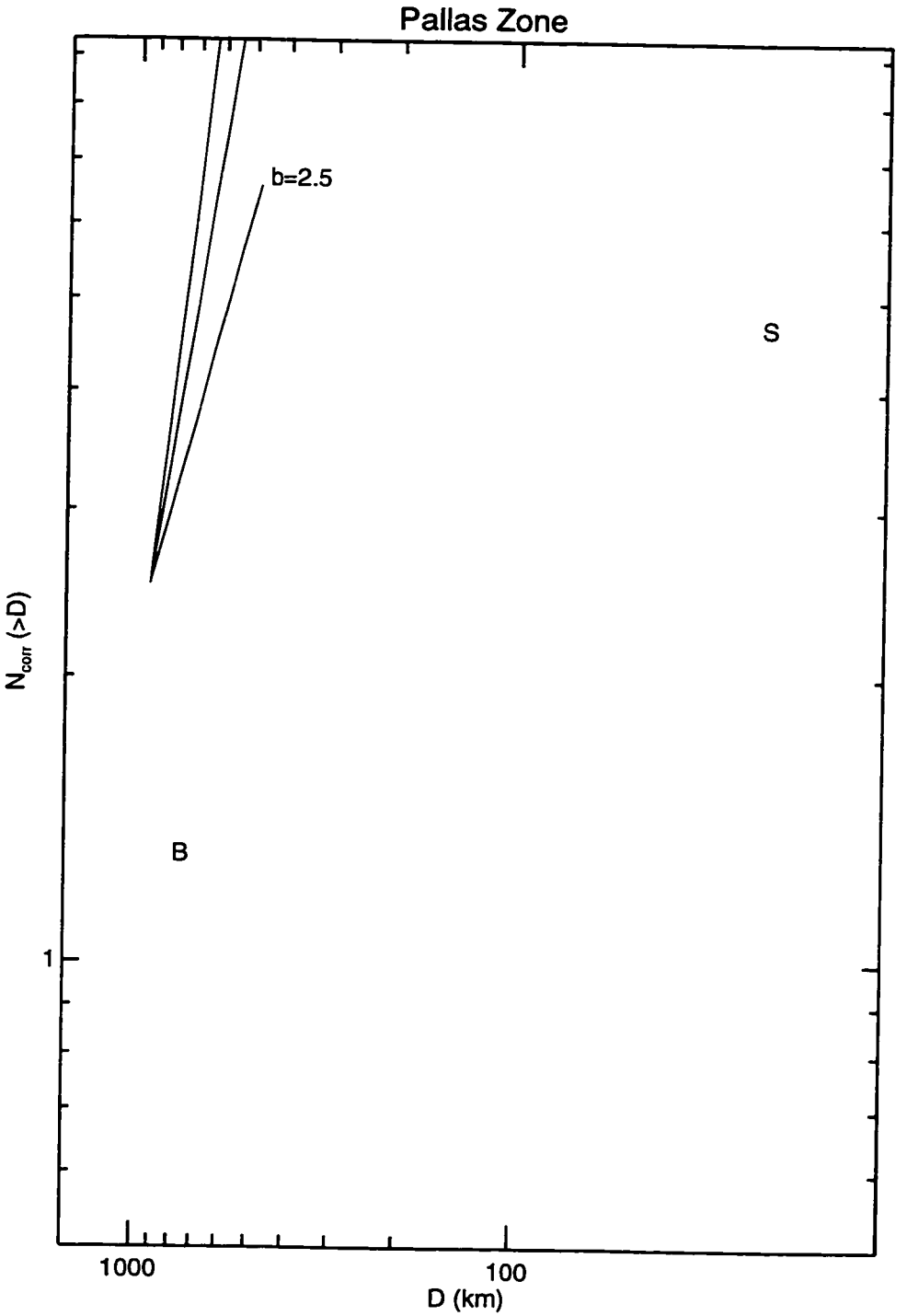


Figure 7.11: Cumulative Size-Frequency Diagram for the Pallas Zone.

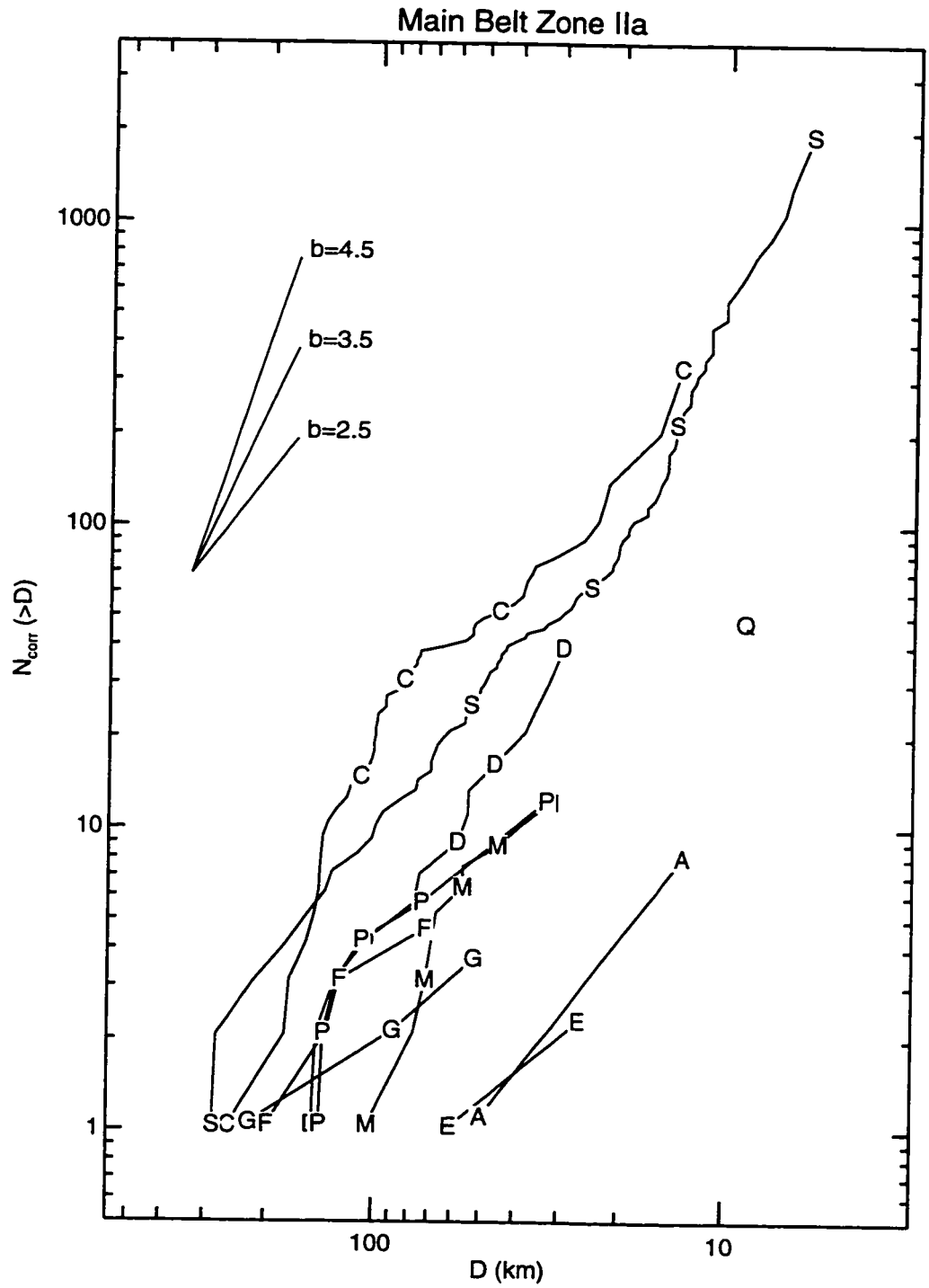


Figure 7.12: Cumulative Size-Frequency Diagram for Main Belt Zone IIa.

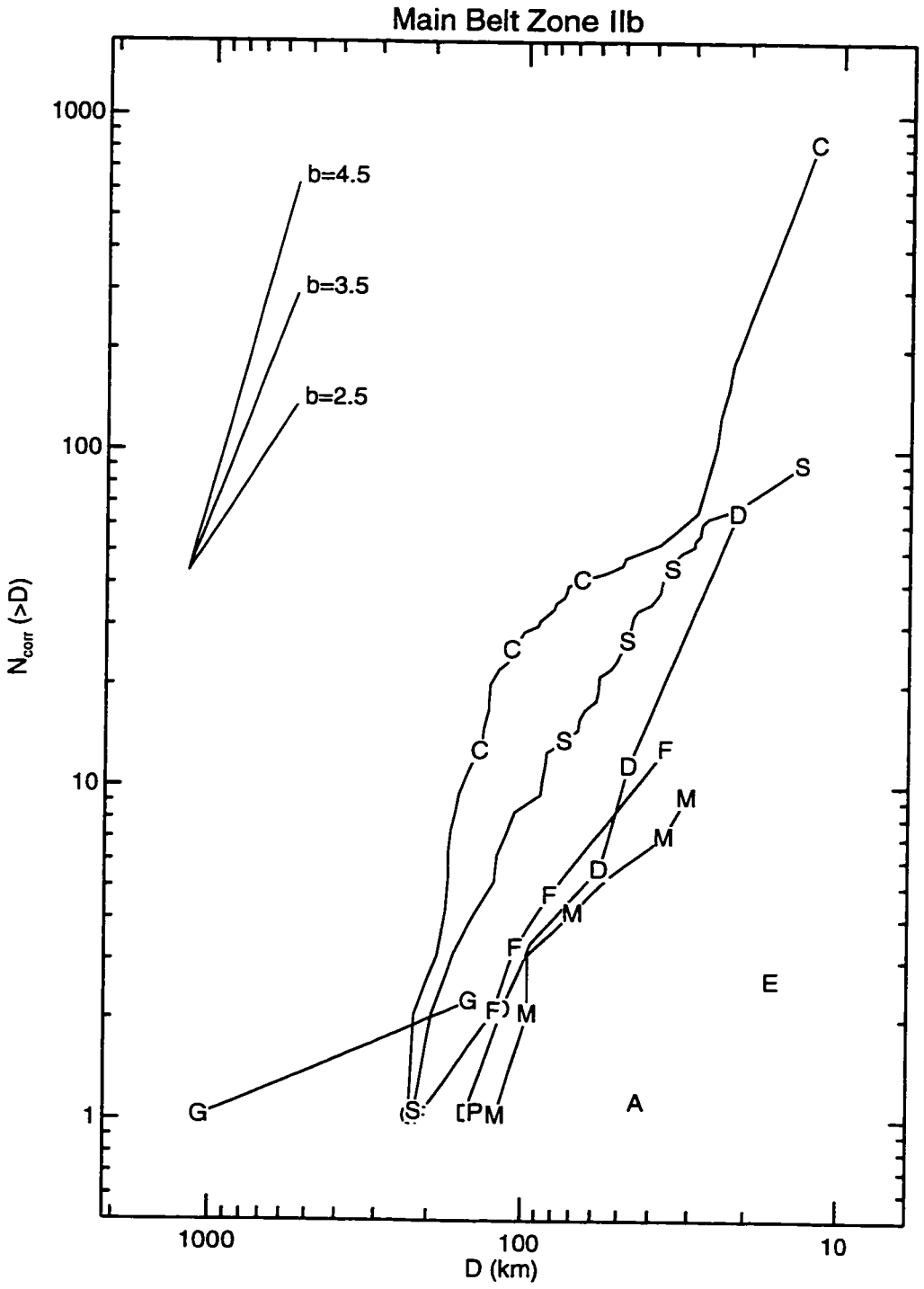


Figure 7.13: Cumulative Size-Frequency Diagram for Main Belt Zone IIb.

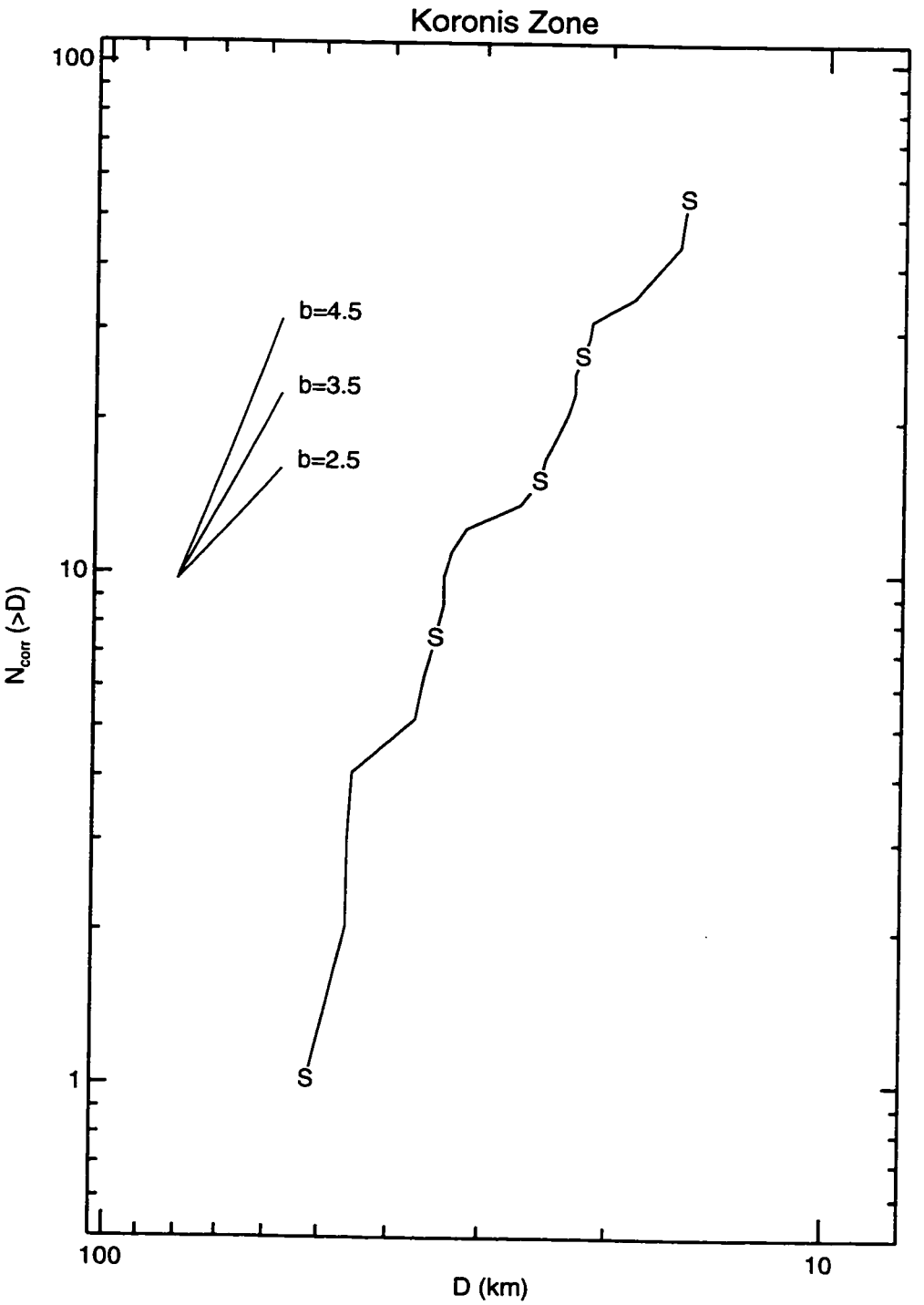


Figure 7.14: Cumulative Size-Frequency Diagram for the Koronis Zone.

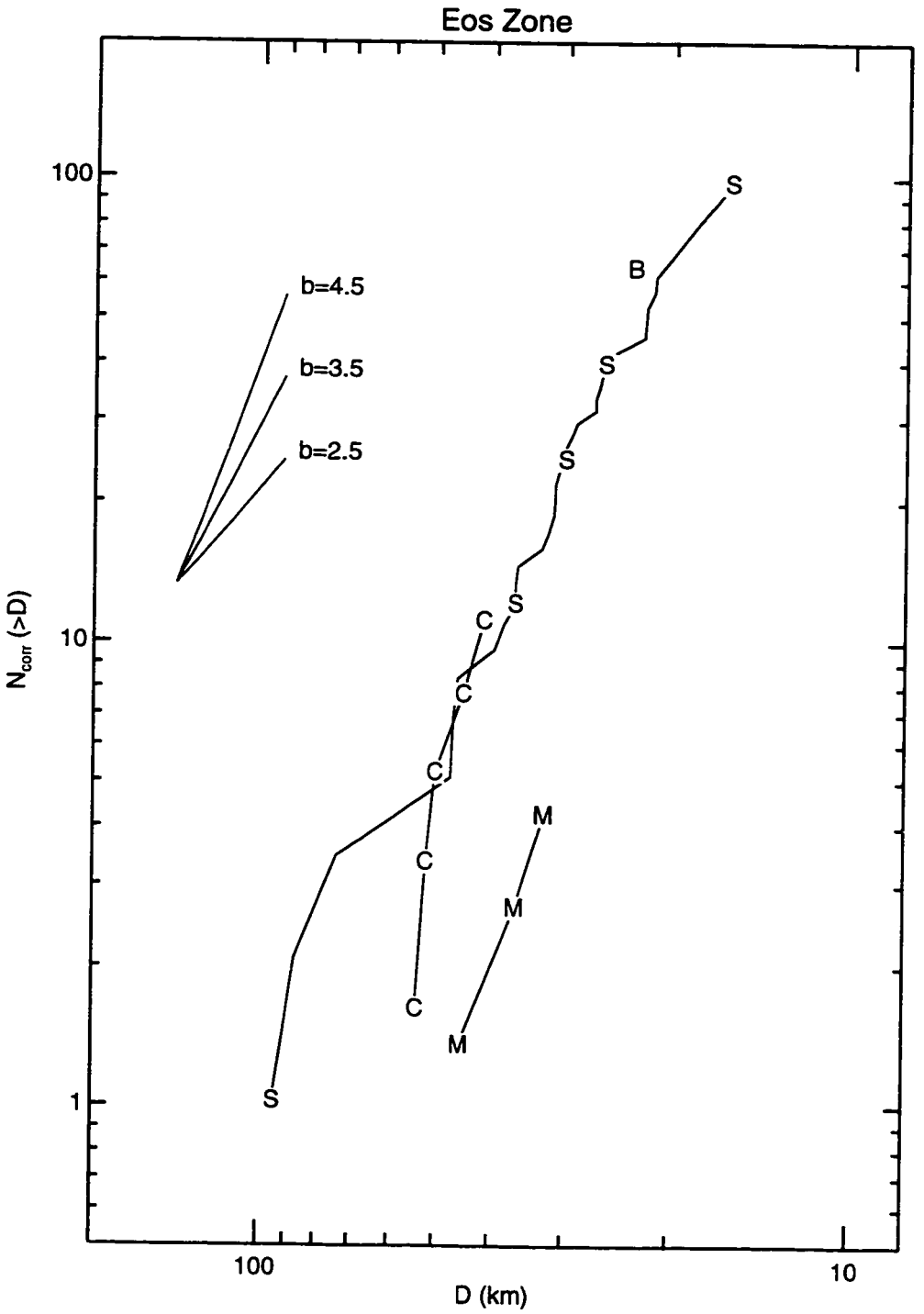


Figure 7.15: Cumulative Size-Frequency Diagram for the Eos Zone.

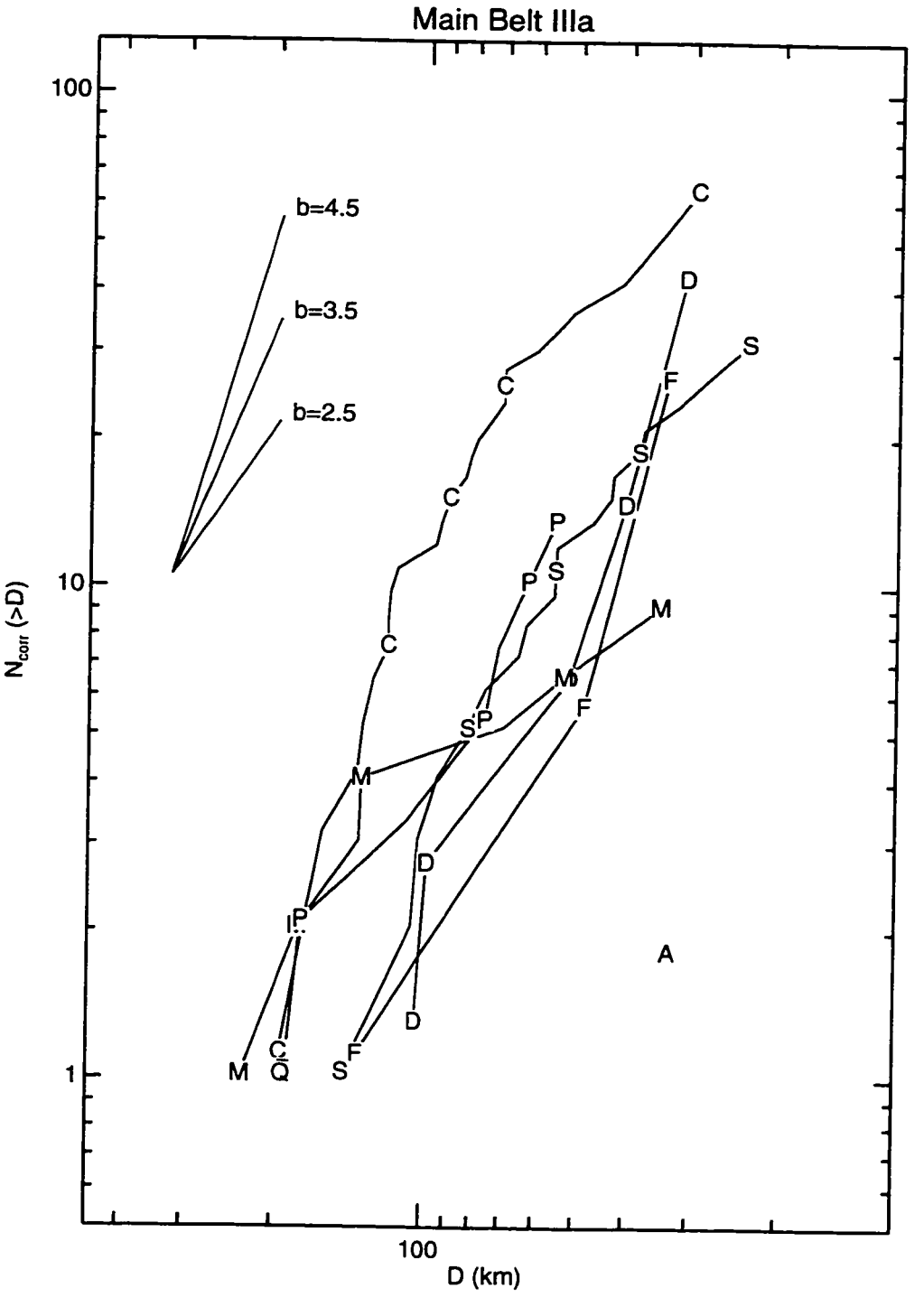


Figure 7.16: Cumulative Size-Frequency Diagram for Main Belt Zone IIIa.

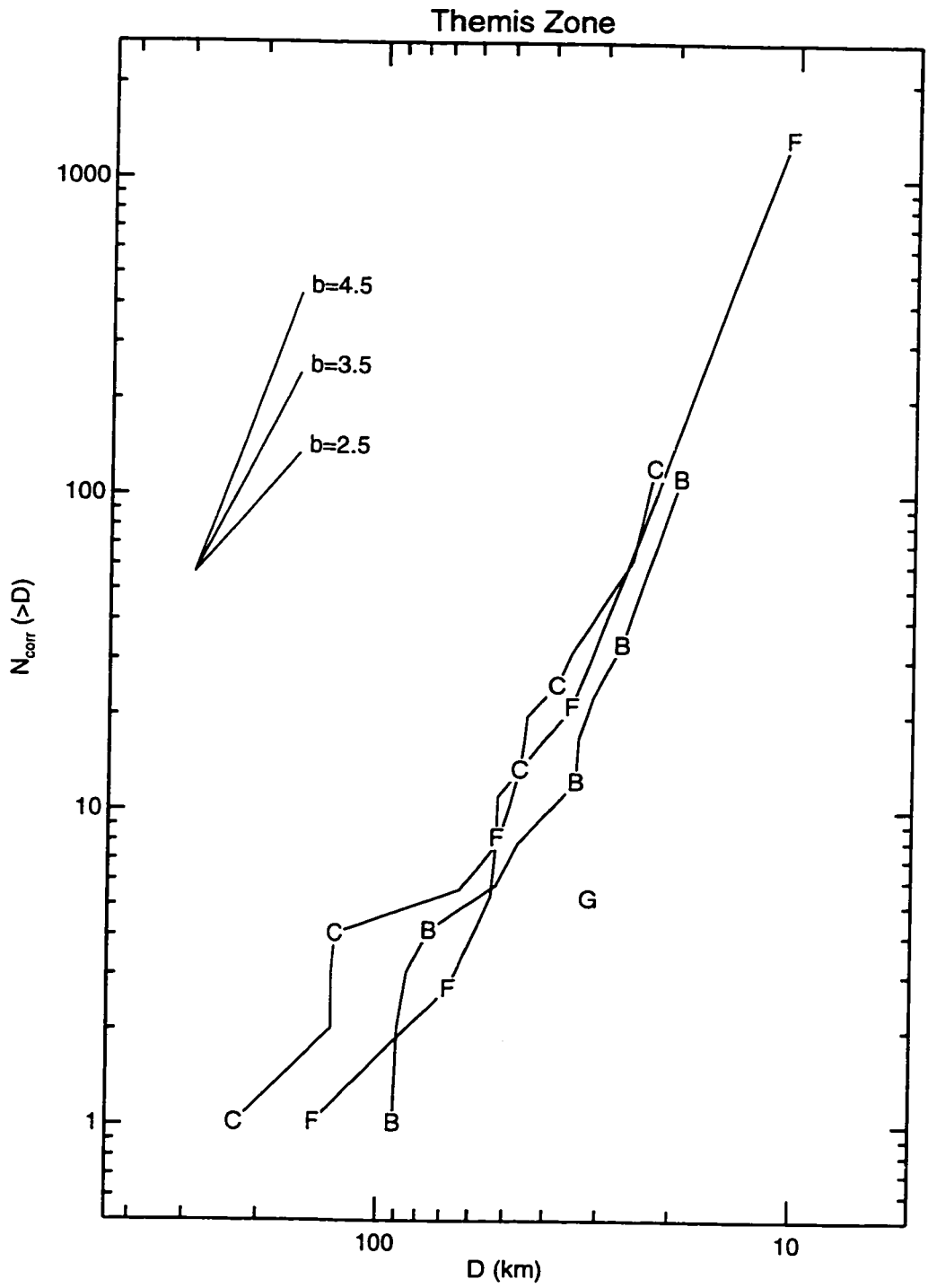


Figure 7.17: Cumulative Size-Frequency Diagram for the Themis Zone.

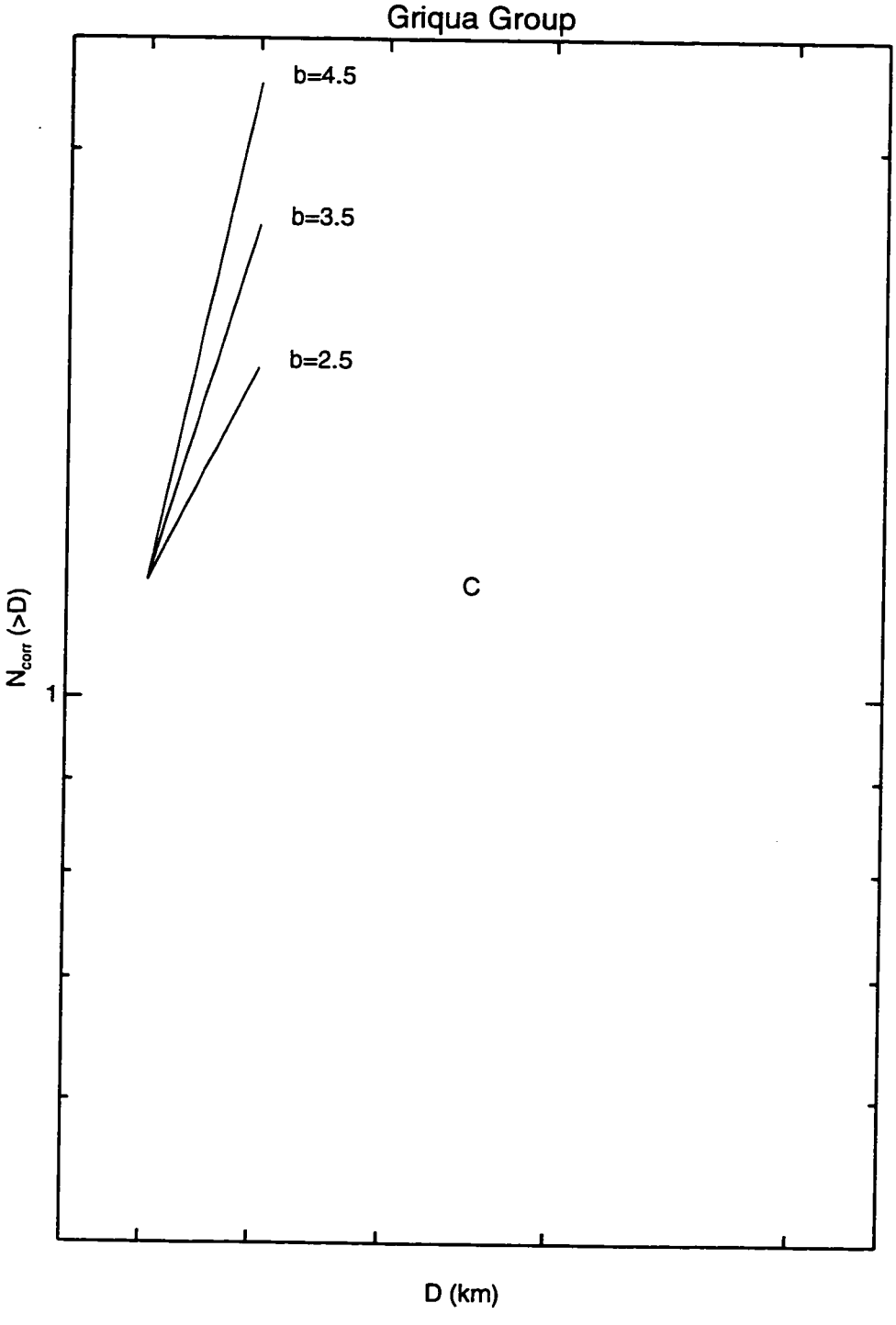


Figure 7.18: Cumulative Size-Frequency Diagram for the Griqua Group.

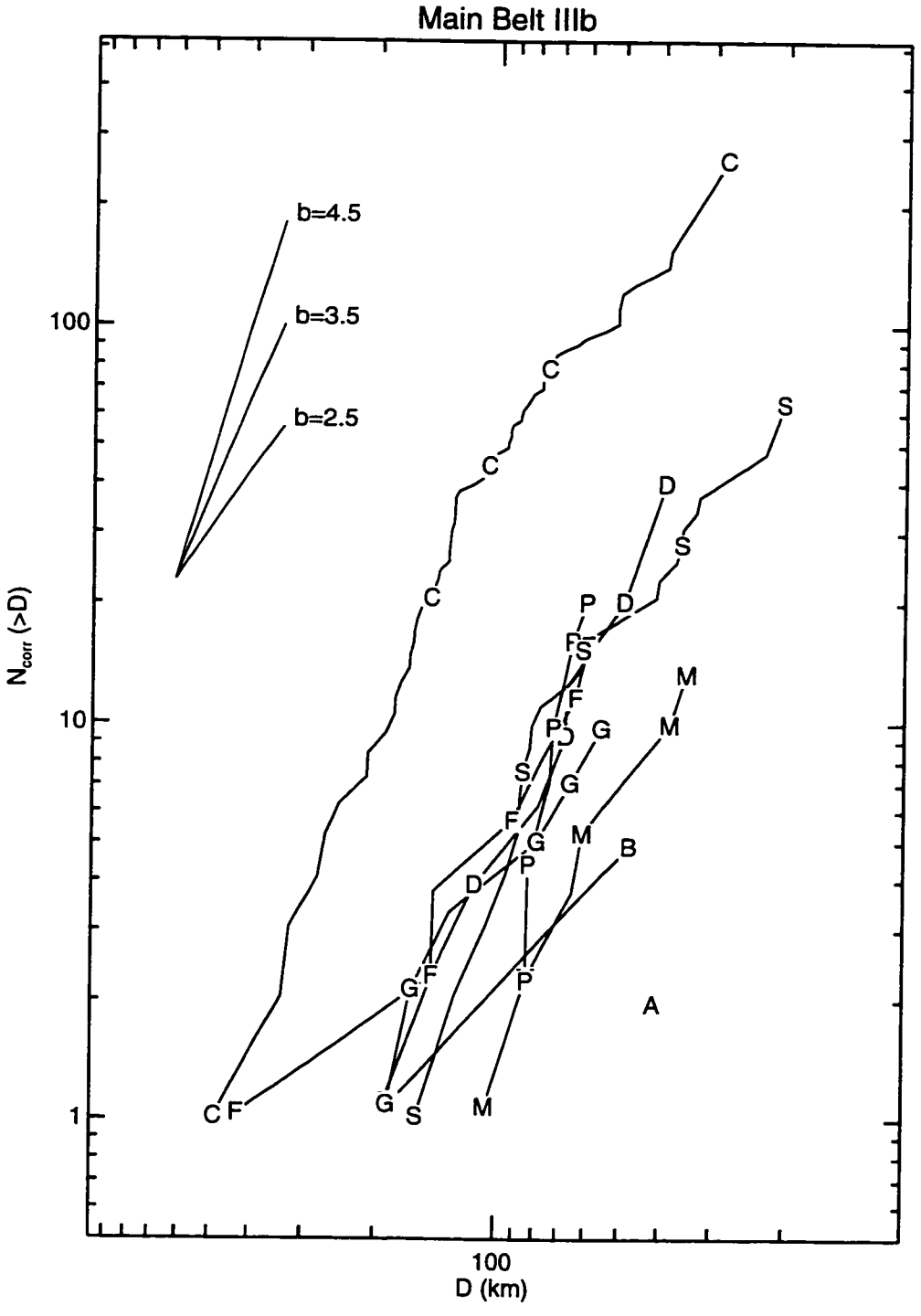


Figure 7.19: Cumulative Size-Frequency Diagram for Main Belt Zone IIIb.

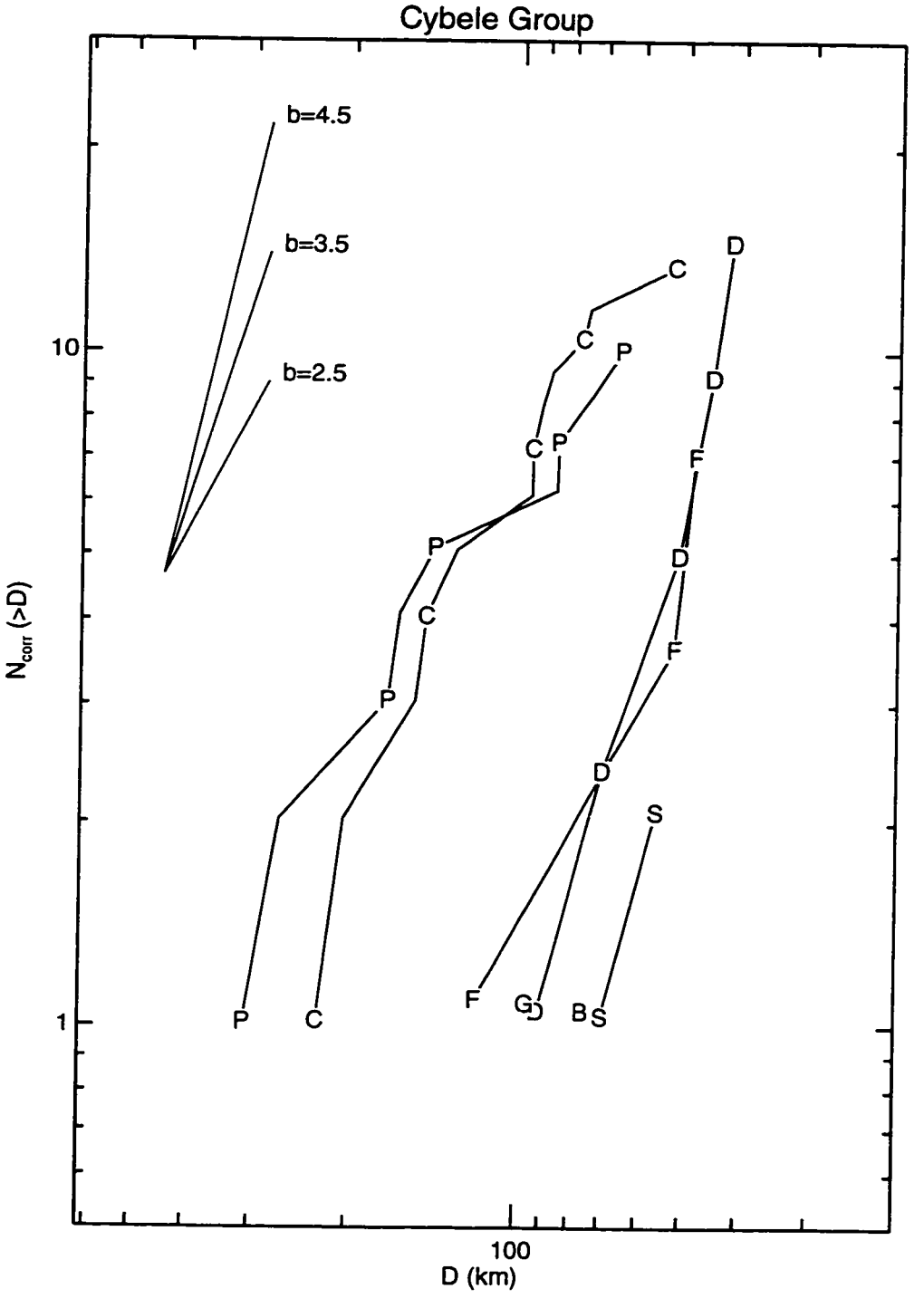


Figure 7.20: Cumulative Size-Frequency Diagram for the Cybele Group.

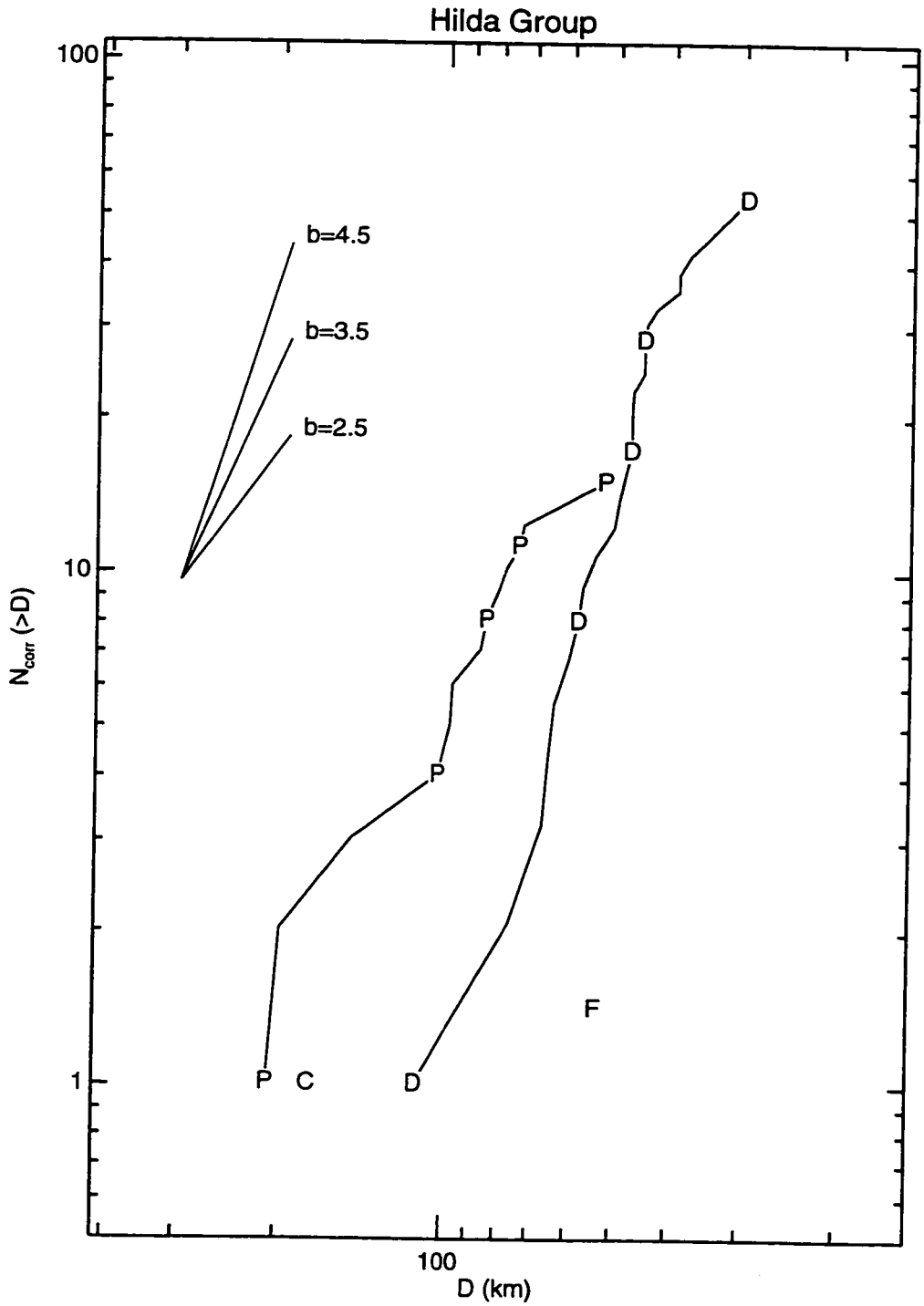


Figure 7.21: Cumulative Size-Frequency Diagram for the Hilda Group.

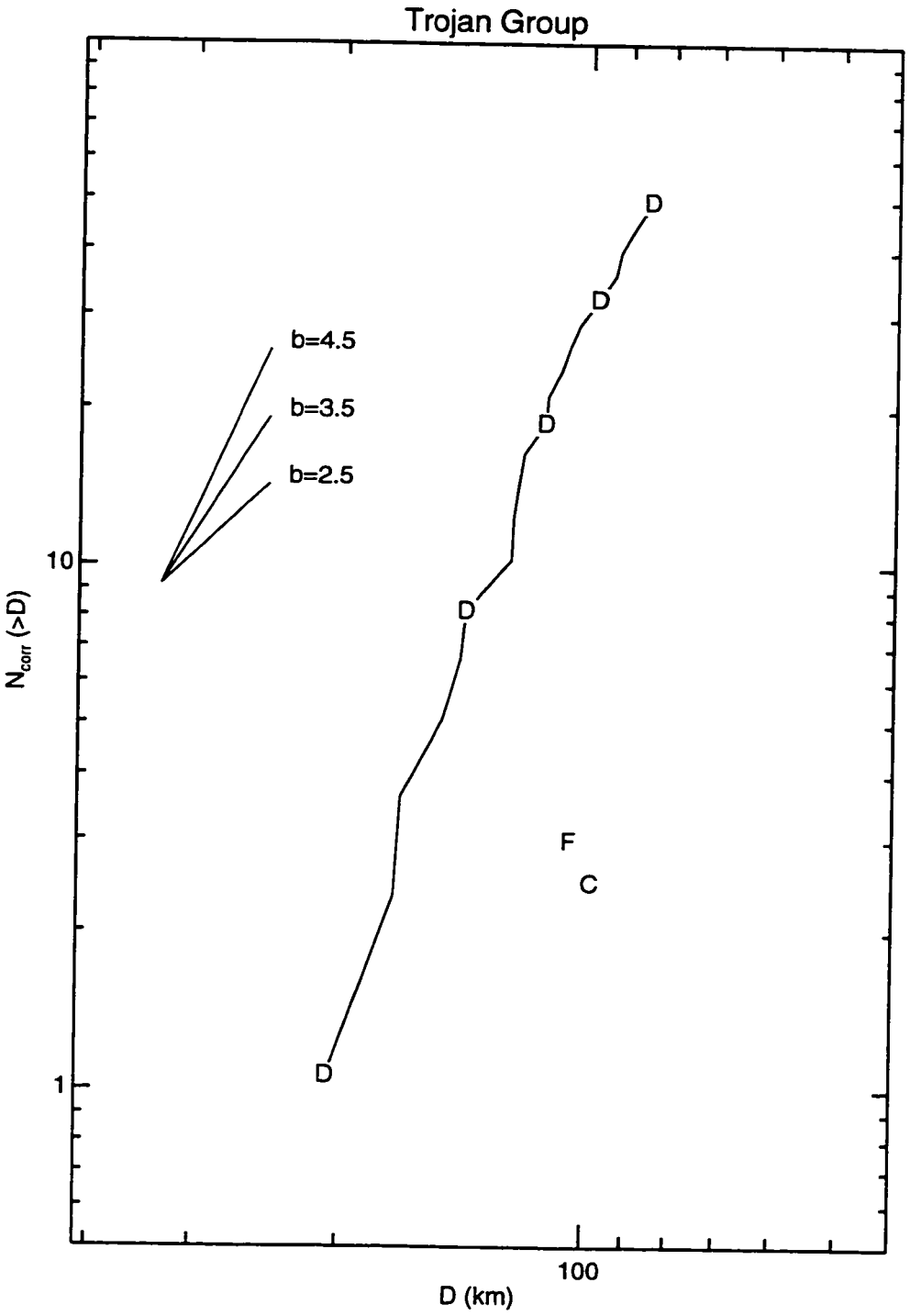


Figure 7.22: Cumulative Size-Frequency Diagram for the Trojan Group.

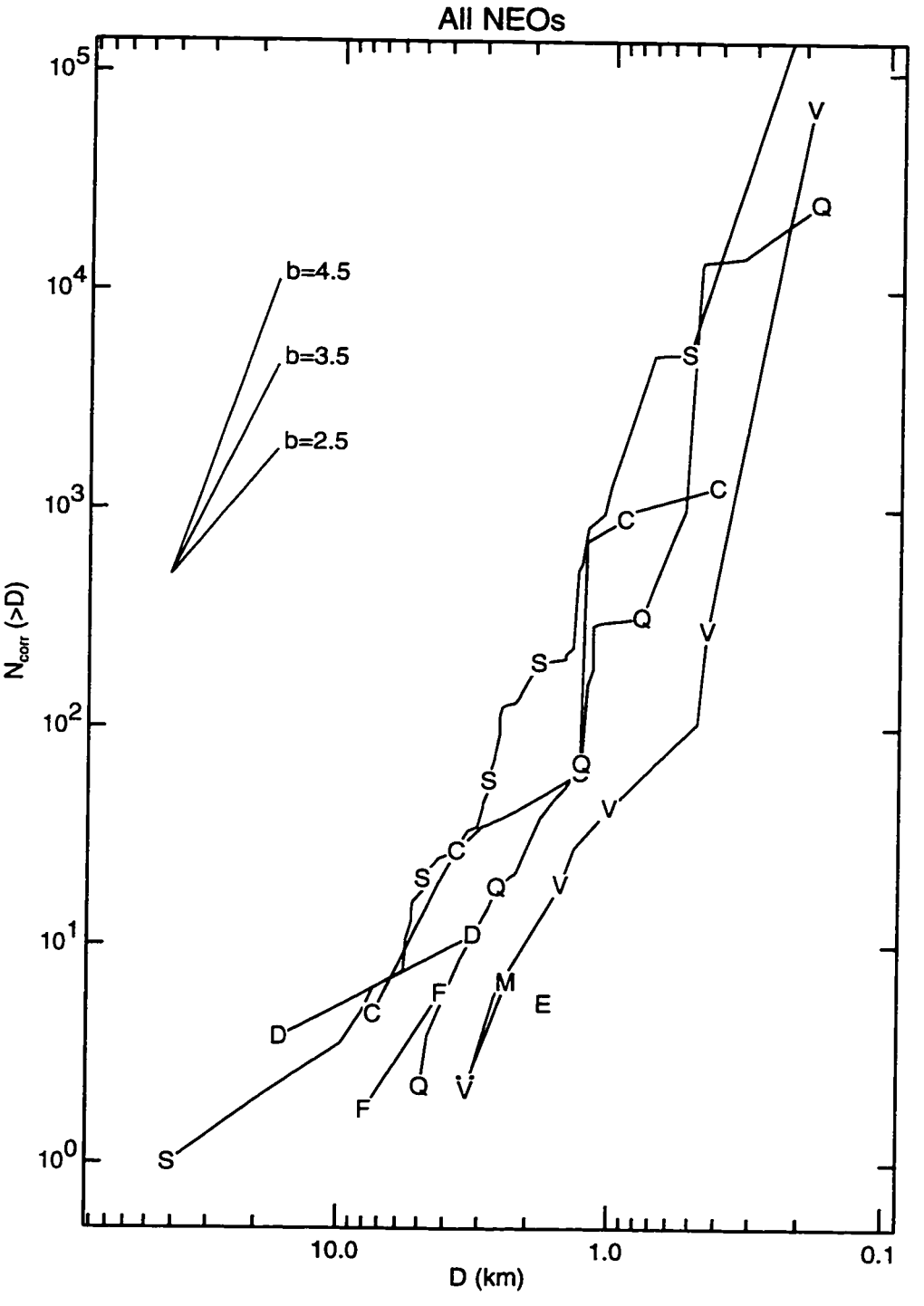


Figure 7.23: Cumulative Size-Frequency Diagram for the NEOs.

The overall slope of the size distribution in any given zone is sensitive to but not necessarily identical to the input model size distribution; caution must be exercised in drawing conclusions based on the overall slope. Changes in slope are less likely to derive from errors in the input model, and changes in relative numbers between different taxonomic classes still less so. For example, the sharp increase in the numbers of V-type objects in the Vesta Family is almost certainly real, and likewise the predominance of the F-types in the Nysa Family.

These diagrams point out the fact that no single ratio between taxonomic types can adequately describe the fraction of taxonomic types among either the asteroids or the NEOs. There are strong variations with size; this is perhaps most noticeable in the Vesta Family, where the V-types rise quickly in number at sizes smaller than 20 km.

Table 7.1 lists the taxonomic fractions in the 22 orbital element zones. These results are in good agreement with debiased distributions for main-belt asteroids reported by Luu & Jewitt (1990). Table 7.2 lists similar values, but for the combined NEO population. In this case, however, the new bias-corrected data showing an *excess* of S-types among the NEOs disagree with the results of Luu & Jewitt (1990), who found an S:C ratio of 0.2±0.1:1. Their results rely on spectra of NEOs obtained over the limited wavelength range 4200 - 7200 Å, which end far blueward of the 1μm absorption feature which characterizes the S-type. Furthermore, their results hinge on the application of a strictly linear correction for phase reddening, for phase angles exceeding 60°. The taxonomic bias corrections of Luu & Jewitt (1989) can not be directly compared to the correction factors derived in this work, as their corrections do not take sample completeness into account, and so refer only to a bias against the discovery of new objects and not that in the existing, observed population.

These new results also indicate that the NEO population is even more rich in S-types than the 3:1 resonance main-belt asteroids. If the NEOs do derive in part from 3:1 resonance asteroids, then it may be necessary to include a secondary source which contains a higher fraction of S-type asteroids. The inner main belt may provide such a source, especially along the border of the ν_6 secular resonance. It is also worth noting that the S:C ratio among the NEOs is roughly comparable to that of the Mars Crossers (though the C-types among the Mars Crossers seem

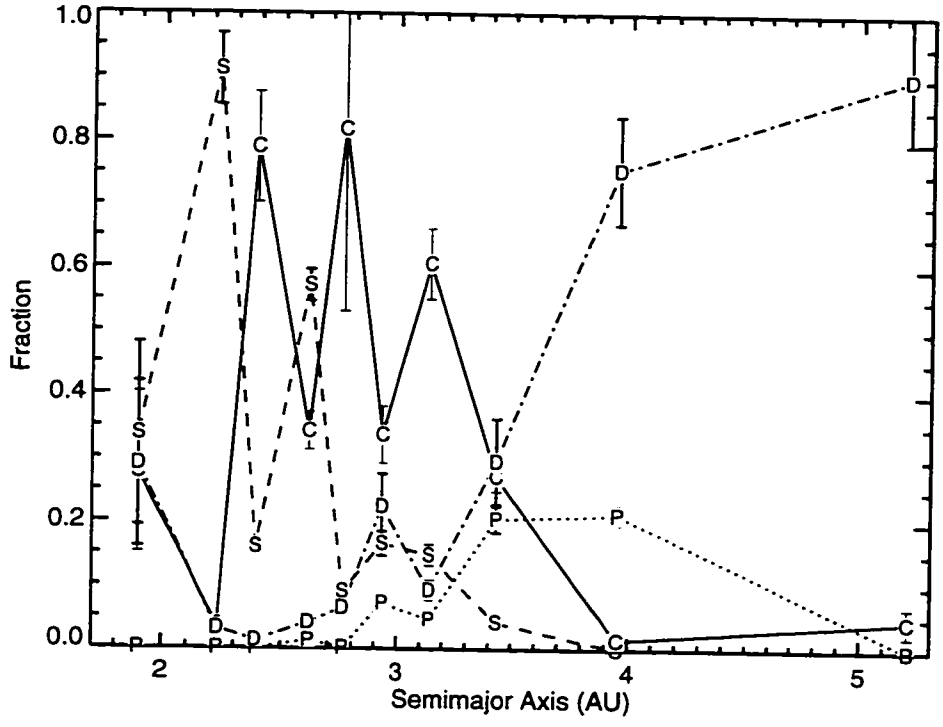


Figure 7.24: Taxonomic Gradient Across the Main Belt

to be of the B-subtype). These results hold true even if all the X-types among the NEOs are of the primitive P-type (Table 7.3).

Figure 7.23 presents the cumulative size-frequency diagram for the combined NEO population. It is immediately apparent that S-types dominate the numbers at nearly every magnitude. The anomalous D-type at a diameter of $\sim$13 km is 3552 Don Quixote, an Amor object with a Jupiter-crossing orbit, and thus a very good candidate for an extinct comet. It is also clear that the fraction of C-type or other "primitive" types do *not* increase relative at smaller sizes; if anything, there is a deficit of C-types near diameters of $\sim$2 km. This argues against there being a large population of extinct comets hidden among the $\sim$1 km diameter NEOs.

The well-known heliocentric variation in taxonomic types (cf. Gradie et al. 1989) is displayed using this new data in Figure 7.24.

Table 7.1: Bias-corrected Taxonomic Fractions

Atens				
Diam.	S	Q	V	A
1–10 km	0.03 ± 0.01	0.30 ± 0.12	0.00 ± 0.00	0.00 ± 0.00
10–100 km	...	...	...	...
100–1000 km	...	...	...	...
Diam.	E	M	C	F
1–10 km	0.00 ± 0.00	0.09 ± 0.04	0.59 ± 0.19	0.00 ± 0.00
10–100 km	...	...	...	...
100–1000 km	...	...	...	...
Diam.	B	G	P	D
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	...	...	...	...
100–1000 km	...	...	...	...
Apollos				
Diam.	S	Q	V	A
1–10 km	0.66 ± 0.12	0.31 ± 0.05	0.00 ± 0.00	0.00 ± 0.00
10–100 km	...	...	...	...
100–1000 km	...	...	...	...
Diam.	E	M	C	F
1–10 km	0.01 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
10–100 km	...	...	...	...
100–1000 km	...	...	...	...
Diam.	B	G	P	D
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	...	...	...	...
100–1000 km	...	...	...	...

Table 7.1 – *Continued*

Amors				
Diam.	S	Q	V	A
1–10 km	0.49 ± 0.15	0.00 ± 0.00	0.03 ± 0.01	0.00 ± 0.00
10–100 km	0.35 ± 0.04	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...
Diam.	E	M	C	F
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.48 ± 0.11	0.00 ± 0.00
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...
Diam.	B	G	P	D
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.65 ± 0.30
100–1000 km	...	...	...	...
Hungaria Group				
Diam.	S	Q	V	A
1–10 km	0.55 ± 0.06	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.34 ± 0.14	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...
Diam.	E	M	C	F
1–10 km	0.45 ± 0.08	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.10 ± 0.04	0.00 ± 0.00	0.28 ± 0.13	0.00 ± 0.00
100–1000 km	...	...	...	...
Diam.	B	G	P	D
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.29 ± 0.13
100–1000 km	...	...	...	...

Table 7.1 – *Continued*

Mars Crossers				
Diam.	S	Q	V	A
1–10 km	0.40 ± 0.11	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.52 ± 0.10	0.11 ± 0.05	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...
Diam.	E	M	C	F
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.00 ± 0.00	0.03 ± 0.01	0.00 ± 0.00	0.20 ± 0.07
100–1000 km	...	...	...	...
Diam.	B	G	P	D
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.60 ± 0.26
10–100 km	0.14 ± 0.06	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...
Flora Family				
Diam.	S	Q	V	A
1–10 km	0.88 ± 0.06	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.91 ± 0.06	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	1.00 ± 0.11	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Diam.	E	M	C	F
1–10 km	0.00 ± 0.00	0.08 ± 0.01	0.04 ± 0.01	0.00 ± 0.00
10–100 km	0.01 ± 0.00	0.02 ± 0.01	0.04 ± 0.01	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Diam.	B	G	P	D
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.03 ± 0.01
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Table 7.1 - *Continued*

Phocaea Group				
Diam.	S	Q	V	A
1–10 km	(1.00 ± 0.12)	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)
10–100 km	0.78 ± 0.12	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Diam.	E	M	C	F
1–10 km	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.22 ± 0.07	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	1.00 ± 0.22	0.00 ± 0.00
Diam.	B	G	P	D
1–10 km	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Nysa Family				
Diam.	S	Q	V	A
1–10 km	0.79 ± 0.12	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.02 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.01
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Diam.	E	M	C	F
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.01 ± 0.00	0.01 ± 0.00	0.32 ± 0.06	0.61 ± 0.08
100–1000 km	0.00 ± 0.00	0.50 ± 0.08	0.00 ± 0.00	0.00 ± 0.00
Diam.	B	G	P	D
1–10 km	0.21 ± 0.06	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.02 ± 0.00
100–1000 km	0.00 ± 0.00	0.50 ± 0.08	0.00 ± 0.00	0.00 ± 0.00

Table 7.1 - *Continued*

Main Belt Zone I				
Diam.	S	Q	V	A
1–10 km	0.90 ± 0.11	0.00 ± 0.00	0.10 ± 0.02	0.00 ± 0.00
10–100 km	0.14 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	0.71 ± 0.04	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Diam.	E	M	C	F
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.81 ± 0.09	0.02 ± 0.01
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.14 ± 0.02	0.14 ± 0.02

Diam.	B	G	P	D
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.01 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Pallas Zone				
Diam.	S	Q	V	A
1–10 km	...	...	...	...
10–100 km	1.00 ± 0.51	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Diam.	E	M	C	F
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Diam.	B	G	P	D
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	1.00 ± 0.42	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Table 7.1 - *Continued*

Main Belt Zone IIa				
Diam.	S	Q	V	A
1–10 km	(0.97 ± 0.12)	(0.03 ± 0.01)	(0.00 ± 0.00)	(0.00 ± 0.00)
10–100 km	0.59 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00
100–1000 km	0.22 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Diam.	E	M	C	F
1–10 km	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)
10–100 km	0.00 ± 0.00	0.01 ± 0.00	0.34 ± 0.03	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.02 ± 0.00	0.45 ± 0.03	0.07 ± 0.01
Diam.	B	G	P	D
1–10 km	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.04 ± 0.01
100–1000 km	0.00 ± 0.00	0.03 ± 0.01	0.10 ± 0.01	0.10 ± 0.01
Main Belt Zone IIb				
Diam.	S	Q	V	A
1–10 km	...	...	...	...
10–100 km	0.09 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	0.19 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Diam.	E	M	C	F
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.01 ± 0.00	0.83 ± 0.30	0.01 ± 0.00
100–1000 km	0.00 ± 0.00	0.02 ± 0.00	0.62 ± 0.03	0.07 ± 0.01
Diam.	B	G	P	D
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.07 ± 0.01
100–1000 km	0.00 ± 0.00	0.03 ± 0.01	0.02 ± 0.00	0.05 ± 0.01

Table 7.1 – *Continued*

Koronis Zone				
Diam.	S	Q	V	A
1–10 km	...	...	...	...
10–100 km	1.00 ± 0.11	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...
Diam.	E	M	C	F
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...
Diam.	B	G	P	D
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...
Eos Zone				
Diam.	S	Q	V	A
1–10 km	...	...	...	...
10–100 km	0.55 ± 0.06	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...
Diam.	E	M	C	F
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.02 ± 0.01	0.06 ± 0.01	0.00 ± 0.00
100–1000 km	...	...	...	..
Diam.	B	G	P	D
1–10 km	...	...	...	...
10–100 km	0.36 ± 0.12	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...

Table 7.1 - *Continued*

Main Belt IIIa				
Diam.	S	Q	V	A
1–10 km	...	...	...	...
10–100 km	0.17 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.01
100–1000 km	0.12 ± 0.01	0.04 ± 0.01	0.00 ± 0.00	0.00 ± 0.00

Diam.	E	M	C	F
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.03 ± 0.01	0.32 ± 0.05	0.15 ± 0.04
100–1000 km	0.00 ± 0.00	0.16 ± 0.01	0.44 ± 0.04	0.05 ± 0.01

Diam.	B	G	P	D
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.06 ± 0.02	0.25 ± 0.06
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.14 ± 0.02	0.05 ± 0.02

Themis Zone				
Diam.	S	Q	V	A
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Diam.	E	M	C	F
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.07 ± 0.01	0.85 ± 0.37
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.80 ± 0.05	0.20 ± 0.03

Diam.	B	G	P	D
1–10 km	...	...	...	...
10–100 km	0.07 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Table 7.1 – *Continued*

Griqua Group				
Diam.	S	Q	V	A
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...
Diam.	E	M	C	F
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	1.00 ± 0.34	0.00 ± 0.00
100–1000 km	...	...	...	...
Diam.	B	G	P	D
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	...	...	...	...
Main Belt IIIb				
Diam.	S	Q	V	A
1–10 km	...	...	...	...
10–100 km	0.17 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00
100–1000 km	0.05 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Diam.	E	M	C	F
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.03 ± 0.01	0.58 ± 0.06	0.03 ± 0.01
100–1000 km	0.00 ± 0.00	0.02 ± 0.00	0.75 ± 0.05	0.06 ± 0.01
Diam.	B	G	P	D
1–10 km	...	...	...	...
10–100 km	0.01 ± 0.00	0.02 ± 0.01	0.05 ± 0.01	0.10 ± 0.02
100–1000 km	0.02 ± 0.00	0.05 ± 0.01	0.00 ± 0.00	0.06 ± 0.02

Table 7.1 - *Continued*

Cybele Group				
Diam.	S	Q	V	A
1–10 km	...	...	...	...
10–100 km	0.05 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0 00 ± 0.00	0.00 ± 0.00
Diam.	E	M	C	F
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.22 ± 0.03	0.16 ± 0.05
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.45 ± 0.03	0.10 ± 0.03
Diam.	B	G	P	D
1–10 km	...	...	...	...
10–100 km	0.03 ± 0.01	0.03 ± 0.01	0.13 ± 0.03	0.38 ± 0.09
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.45 ± 0.03	0.00 ± 0.00
Hilda Group				
Diam.	S	Q	V	A
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Diam.	E	M	C	F
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.02 ± 0.01
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.17 ± 0.02	0.00 ± 0.00
Diam.	B	G	P	D
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.17 ± 0.02	0.81 ± 0.09
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.67 ± 0.04	0.17 ± 0.02

Table 7.1 - *Continued*

Trojan Group				
Diam.	S	Q	V	A
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Diam.	E	M	C	F
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.11 ± 0.05	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.09 ± 0.05
Diam.	B	G	P	D
1–10 km	...	...	...	...
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.89 ± 0.18
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.91 ± 0.12
Vesta Family				
Diam.	S	Q	V	A
1–10 km	0.14 ± 0.03	0.00 ± 0.00	0.86 ± 0.07	0.00 ± 0.00
10–100 km	0.23 ± 0.04	0.00 ± 0.00	0.15 ± 0.05	0.00 ± 0.00
100–1000 km	0.50 ± 0.09	0.00 ± 0.00	0.50 ± 0.09	0.00 ± 0.00
Diam.	E	M	C	F
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.03 ± 0.01	0.00 ± 0.00	0.39 ± 0.08	0.00 ± 0.00
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Diam.	B	G	P	D
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.21 ± 0.06
100–1000 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Table 7.2: Bias-corrected Taxonomic Fractions for the NEOs (excluding X-types)

Diam.	S	Q	V	A
0.1–1 km	(0.04 ± 0.01)	(0.25 ± 0.06)	(0.70 ± 0.30)	(0.00 ± 0.00)
1–10 km	0.54 ± 0.13	0.13 ± 0.02	0.02 ± 0.00	0.00 ± 0.00
10–100 km	0.35 ± 0.04	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Diam.	E	M	C	F
0.1–1 km	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.01 ± 0.00)	(0.00 ± 0.00)
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.30 ± 0.07	0.00 ± 0.00
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Diam.	B	G	P	D
0.1–1 km	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.65 ± 0.30

The cumulative size-frequency distribution for S-type NEOs shows a wave-like oscillation in number. This is displayed in Figure 7.25, where a power-law distribution with an exponent of $b = 3.7$ has been divided out to better show the variation in numbers with size. The error bars in diameter are derived from the estimated 1 σ uncertainties in magnitude for each object. The error bars in cumulative number are 1 σ random errors in corrected number only. These waves are qualitatively similar to those predicted by Campo Bagatin et al. (1994) for collisional systems with a small-size cutoff. In this theory, the wave pattern occurs because particles just larger than the cutoff are not destroyed by impacts with smaller particles, and are created by the disruption of even larger bodies faster than they are depleted. This pattern extends upwards in size to diameters of tens of kilometers.

While there are qualitative similarities to the apparent oscillations seen in this work, the waves observed here have a much shorter wavelength than the ones

Table 7.3: Bias-corrected Taxonomic Fractions for the NEOs (if all X-types are P-type)

Diam.	S	Q	V	A
0.1–1 km	(0.03 ± 0.01)	(0.21 ± 0.05)	(0.58 ± 0.25)	(0.00 ± 0.00)
1–10 km	0.52 ± 0.13	0.12 ± 0.02	0.02 ± 0.00	0.00 ± 0.00
10–100 km	0.35 ± 0.04	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Diam.	E	M	C	F
0.1–1 km	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.00 ± 0.00)
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.29 ± 0.07	0.00 ± 0.00
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Diam.	B	G	P	D
0.1–1 km	(0.00 ± 0.00)	(0.00 ± 0.00)	(0.17 ± 0.07)	(0.00 ± 0.00)
1–10 km	0.00 ± 0.00	0.00 ± 0.00	0.04 ± 0.01	0.00 ± 0.00
10–100 km	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.65 ± 0.30

in Campo Bagatin et al. (1994), where the peak-to-peak wavelength covers nearly two orders of magnitude in size (compared to a mere factor of $\sim$2.5 seen in this work). If indeed the oscillations are real, then it is possible they are due to some other mechanism, such as an influx into the NEOs of fragments of a preferred size range ($D \sim 2.5$ km), perhaps due to stochastic cratering and fragmentation events in the main belt (Durda & Dermott 1997).

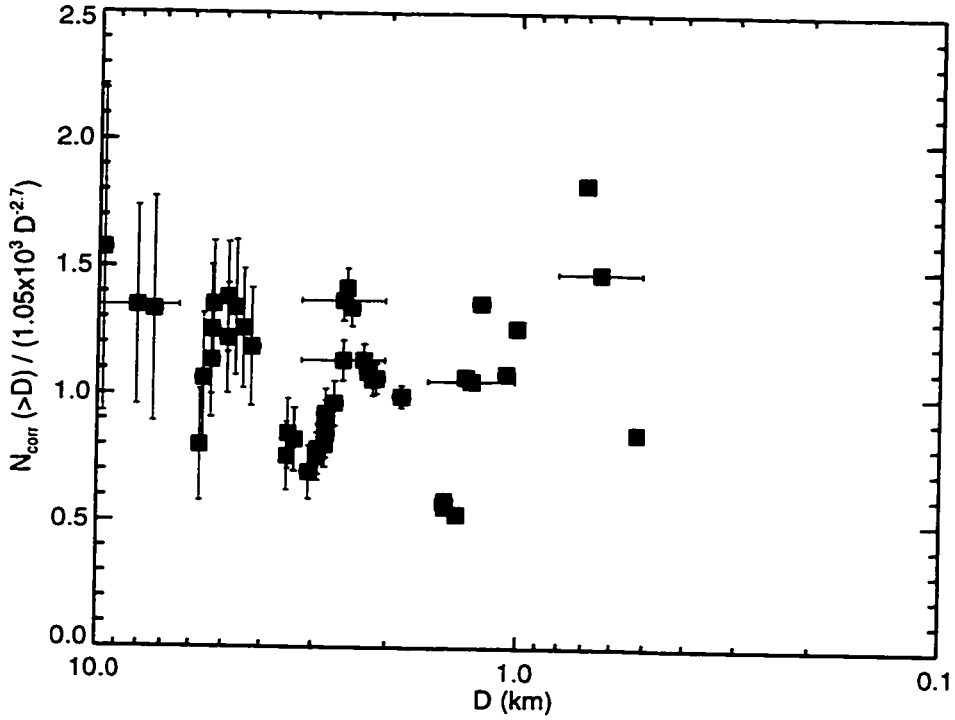


Figure 7.25: Wave-like Oscillation in the Size-Frequency Distribution of S-type NEOs.

Chapter 8

Conclusions

> *God keep me from ever completing anything. This whole book is but a draft - nay, but the draft of a draft. Oh Time, Strength, Cash, and Patience!*
>
> — Herman Melville

The observation and analysis of the reflectance spectra of a significant number of NEO and other objects, and their comparison with previously published NEO and main-belt asteroid spectra has allowed me to make several conclusions regarding individual objects and the general NEO population:

- The distribution of taxonomic types among the NEOs is similar to that of inner main-belt asteroids, and perhaps even richer in S-types than the 3:1 resonance asteroids. No evidence is found for any significant presence of extinct comets among the NEOs down to sizes of $\sim$1 km.
- Several of the NEOs displaying 1-μm absorption bands have spectra similar if not identical to ordinary chondrite meteorites.
- Two of the traditionally strongest candidates for extinct comets (2201 Oljato and 2212 Hephaistos) are seen to have strong 1-μm absorption bands, consistent with the presence of mafic silicates on their surfaces, and inconsistent with current ideas of cometary surfaces.
- Four objects in cometary-type orbits (2906 Caltech, 2938 Hopi, 1994 EQ3, and 1996 PW) are seen to have relatively featureless red-sloped spectra consistent with observed comet nuclei.

- 3103 Eger, classed as an E-type object and suggested by Gaffey et al (1992) as the parent body of the enstatite achondrite meteorites, is seen to have an unusual spectrum with a very strong 1-μm absorption band, completely inconsistent with an enstatite composition.

- The trend of increasing 1-μm absorption band depth with decreasing size is confirmed in this study. The positive correlation of visual slope with increasing size is also confirmed by this study for objects larger than 10 km; however, the opposite effect is seen for smaller objects. This may help reconcile the differences between the S-type asteroids and the ordinary chondrite meteorites.

References

Adams, J. B. 1975, in Infrared and Raman spectroscopy of lunar and terrestrial minerals, ed. C. Karr, Jr. (New York: Academic Press), 91

Adams, J. B. 1974, J. Geophys. Res., 79, 4829

Adams, J. B. & Filice, A. L. 1967, J. Geophys. Res., 72, 5705

Barucci, M. A., Capria, M. T., Coradini, A., & Fulchignoni, M. 1987, Icarus, 72, 304

Belton, M. J. S., Veverka, J., Thomas, P., Helfenstein, D., Simonelli, D., Chapman, C., Davies, M. E., Greeley, R., Greenberg, R., Head, J., Murchie, S., Klaasen, K., Johnson, T. V., McEwen, A., Morrison, D., Neukum, G., Fanale, F., Anger, C., Carr, M., & Pilcher, C. 1992, Science, 257, 1647

Binzel, R. P., Bus, S. J., & Burbine, T. H. 1998, Lunar and Planetary Science Conference, 29, 1222

Binzel, R. P. 1996, Private communication

Binzel, R. P., Bus, S. J., Burbine, T. H., & Sunshine, J. M. 1996, Science, 273, 946

Bowell, E. L. G. 1998, web page ftp://ftp.lowell.edu/pub/elgb/astorb.html

Burns, R. G. 1993, Mineralogical Applications of Crystal Field Theory (New York: Cambridge Univ. Press)

Campo Bagatin, A., Cellino, A., Davis, D. R., Farinella, P., & Paolicchi, P. 1994, Planet. Space Sci., 42, 1079

Charette, M. P., McCord, T. B., & Pieters, C. M. 1974, J. Geophys. Res., 79, 11

Chyba, C. F. 1990, Nature, 343, 129

Davenport, S. A., Wdowiak, T. J., Jones, D. D., & Wdowiak, P. 1990, in Global Catastrophes in Earth History: An Interdisciplinary Conference on Impacts, Volcanism, and Mass Mortality, eds. V. L. Sharpton & P. D. Ward, Geological Soc. of America Special Paper 247 (Boulder: Geological Soc. of America)

Dohnanyi, J. S. 1971, in Physical Studies of Minor Planets, ed. T. Gehrels (Washington, D. C.: N. A. S. A.), 263

Durda, D. D. & Dermott, S. F. 1997, Icarus, 130, 140

Farinella, P., Vokrouhlicky, D., & Hartmann, W. K. 1998, Icarus, 132, 378

Fernández, Y. R., McFadden, L. A., Lisse, C. M., Helin, E. F., & Chamberlin, A. B. 1997, Icarus, 128, 114

Filippenko, A. V. 1982, PASP, 94, 715

Gaffey, M. J., Burbine, T. H., & Binzel, R. P. 1993, Meteoritics, 28, 161

Gaffey, M. J., Reed, K. L., & Kelley, M. S. 1992, Icarus, 100, 95

Gradie, J. C., Chapman, C. R., & Tedesco, E. F. 1989, in Asteroids II, eds. R. P. Binzel, T. Gehrels, M. S. Matthews. (Tucson: University of Arizona), 316

Gradie, J. & Veverka, J. 1986, Icarus, 66, 455

Greenberg, R., Head, J., Murchie, S., Klaasen, K., Johnson, T. V., McEwen, A, Morrison, D., Neukum, G., Fanale, F., Anger, C., Carr, M., & Pilcher, C. 1992, Science, 257, 1647

Greenberg, R. & Nolan, M. C. 1989, in Asteroids II, eds. R. P. Binzel, T. Gehrels, M. S. Matthews (Tucson: University of Arizona), 778

Grieve, R. A. F. & Shoemaker, E. M. 1994, in Hazards Due to Comets and Asteroids, ed. T. Gehrels (Tucson: University of Arizona), 417.

Hardorp, J. 1978, A&A, 63, 383

Hammergren, M. 1996, AAS, 187, 4203

Helin, E. F. & Dunbar, R. S. 1990, Vistas in Astronomy, 33, 21

Hicks, M. D., Fink, U., & Grundy, W. M. 1998, Icarus, 133, 69

Hills, J. G. & Goda, M. P. 1993, AJ, 105, 1114

Hörz, F. & Cintala, M. 1997, Meteoritics and Planetary Science, 32, 179

Howell, E. S. 1994, Private communication.

Howell, E. S., Merenyi, E., & Lebofsky, L. A. 1994, J. Geophys. Res., 99, 110847

Hunt, G. R. & Salisbury, J. W. 1970, Modern Geology, 1, 283

Hunten, D. M. 1993, Science, 259, 915

Jedicke, R. 1996, AJ, 111, 970

Jedicke, R. & Heron, J. D. 1997, Icarus, 127, 494

Jedicke, R. & Metcalfe, T. S. 1998, Icarus, 131, 245

Kundic, T., Colley, W. N., Gott, J. R. III, Malhotra, S., Ue, Li-Pen, Rhoads, J. E., Stanek, K. Z., Turner, E. L, & Wambsganss, J. 1995, ApJ, 455, L5

Lazzarin, M., Di Martino, M., Barucci, M. A., Doressoundiram, A., & Florczak, M. 1997, A&A, 327, 388

Lipschutz, M. E., Gaffey, M. J., & Pellas, P. 1989, in Asteroids II, eds. R. P. Binzel, T. Gehrels, M. S. Matthews. (Tucson: University of Arizona), 740

Love, S. G. & Ahrens, T. J. 1996, Icarus, 124, 141

Luu, J. X. 1994, PASP, 106, 425

Luu, J. X. 1993, Icarus, 104, 138

Luu, J. X. & Jewitt, D. C. 1990, AJ, 100, 933

Luu, J. X. & Jewitt, D. C. 1989, AJ, 98, 1985

Marcus, J. N. & Olsen, M. A. 1991, in Comets in the Post-Halley Era, vol. 1, eds. R. L. Newburn, Jr., M. Neugebauer, & J. Rahe (Dordrecht: Kluwer)

McCord, T. B., Adams, J. B., & Johnson, T. V. 1970, Science, 168, 3938

McFadden, L. A., Cochran, A. L., Barker, E. S., Cruikshank, D. P., & Hartmann, W. K. 1993, J. Geophys. Res., 98, 3031

McFadden, L. A., Gaffey, M. J., & McCord, T. B. 1984, Icarus, 59, 25

Melosh, H. J. 1980, Annual Review of Earth and Planetary Sciences, 8, 65

Morrison, D. ed. 1992, The Spaceguard Survey: Report of the NASA International Near-Earth Object Detection Workshop (Pasadena: Jet Propulsion Laboratory)

Muinonen, K., Bowell, E., & Lumme, K. 1995, A&A, 293, 948

Pieters, C. M. 1993, J. Geophys. Res., 98, 20817

Pieters, C. M. & McFadden, L. A. 1994, Annual Review of Earth and Planetary Sciences, 22, 457

Rabinowitz, D. L. 1993, ApJ, 407, 412

Rahe, J., Vanysek, V., & Weissman, P. R. 1994, in Hazards Due to Comets and Asteroids, ed. T. Gehrels (Tucson: University of Arizona), 597

Ryan, E. V. & Melosh, H. J. 1998, Icarus, 133, 1

Shoemaker, E., Canavan, G., Darrah, J., Harris, A., Morrison, D., Mumma, M., & Rabinowitz, D. 1995, Report of Near-Earth Object Survey Working Group

Sunshine, J. M., Pieters, C. M., & Pratt, S. F. 1990, J. Geophys. Res., 95, 6955

Tedesco, E. F., Matson, D. L., & Veeder, G. J. 1989, in Asteroids II, eds. R. P. Binzel, T. Gehrels, M. S. Matthews (Tucson: University of Arizona), 290

Tholen, D. J. & Barucci, M. A. 1989, in Asteroids II, eds. R. P. Binzel, T. Gehrels, M. S. Matthews (Tucson: University of Arizona), 298

Tyson, J. A. 1989, in CCDs in Astronomy, ed. G. H. Jacoby, ASP Conference Series, 8, 1

Veeder, G. J., Tedesco, E. F., & Matson, D. L. 1989, in Asteroids II, eds. R. P. Binzel, T. Gehrels, M. S. Matthews (Tucson: University of Arizona), 282

Veverka, J., et al. 1996, Icarus, 120, 66

Vokrouhlický, D. & Farinella, P. 1998, A&A, 335, 351

Weissman, P. R., A'Hearn, M. F., McFadden, L. A., & Rickman, H. 1989, in Asteroids II, eds. R. P. Binzel, T. Gehrels, M. S. Matthews (Tucson: University of Arizona), 880

Wetherill, G. W. 1974, Annual Review of Earth and Planetary Sciences, 2, 303

Wisniewski, W. Z. 1987, Icarus, 70, 566

Xu, S., Binzel, R. P., Burbine, T. H., & Bus, S. J. 1995, Icarus, 115, 1

Zellner, B. 1979, in Asteroids, ed. T. Gehrels (Tucson: University of Arizona), 783

Zellner, B., Tholen, D. J., & Tedesco, E. F. 1985, Icarus, 61, 355

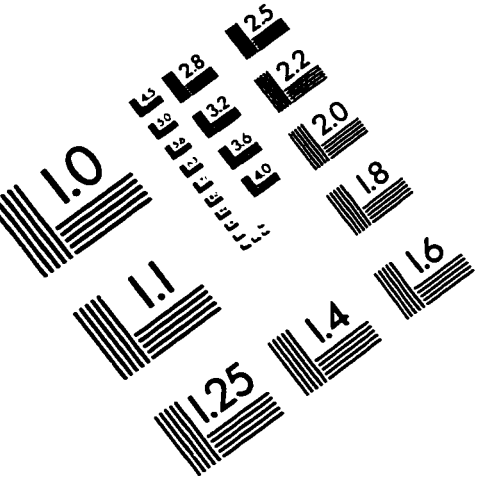
Curriculum Vitae

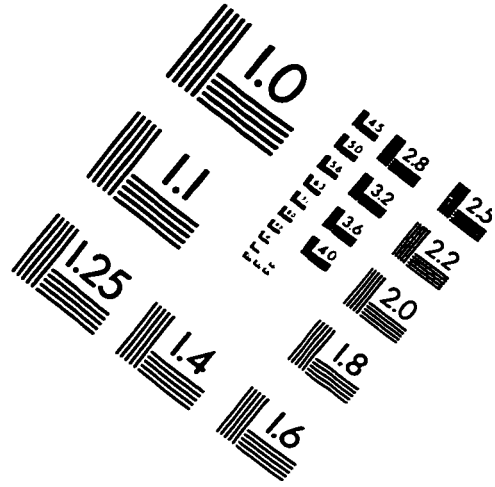
Mark Hammergren

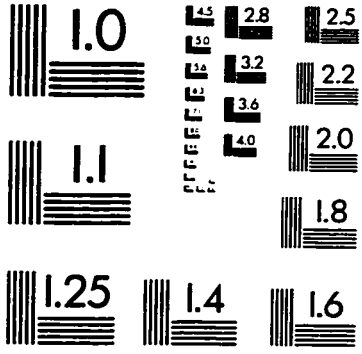
Mark Hammergren was born in the gritty lakeside town of Waukegan, Illinois, on May 14, 1964. He attended the University of Illinois at Urbana-Champaign, and earned two BS degrees (Physics and Astronomy) in 1986. In 1991, he went to graduate school at the University of Washington Department of Astronomy. On the way to his PhD (to be awarded 1998), he picked up an MS in Astronomy in 1992.

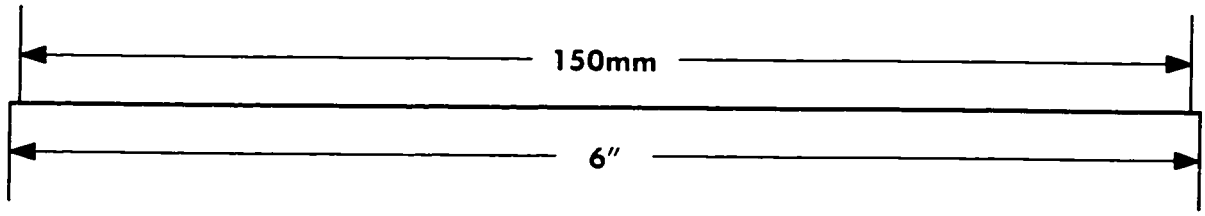
In between, he usually found things to occupy his time, only some of which involved beer.

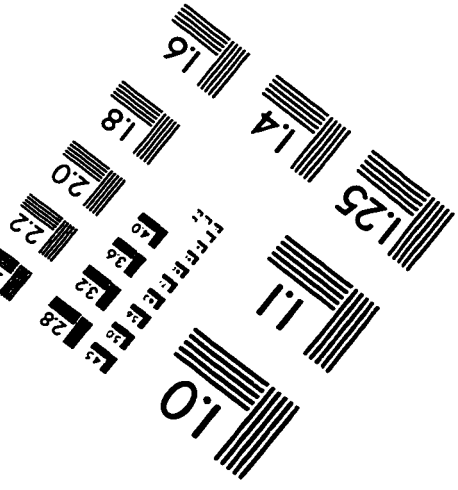
IMAGE EVALUATION TEST TARGET (QA-3)

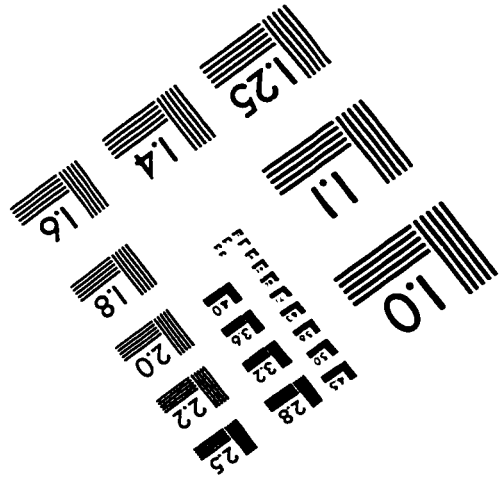












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved

99

07908

UMI

MICROFILMED 1998

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

NOTE TO USERS

The original manuscript received by UMI contains pages with indistinct and slanted print. Pages were microfilmed as received.

This reproduction is the best copy available

UMI

THEORY AND EVALUATION OF A NEW PHYSIOLOGIC SAMPLING PUMP

by

Cheryl K. Hart

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

1998

Approved by _____
Chairperson of Supervisory Committee

_____

_____

Program Authorized
to Offer Degree Department of Environmental Health

Date August 19, 1998

UMI Number: 9907908

Copyright 1998 by
Hart, Cheryl Kay

All rights reserved.

UMI Microform 9907908
Copyright 1998, by UMI Company. All rights reserved.

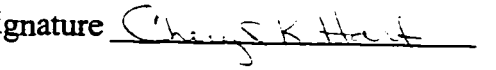
This microform edition is protected against unauthorized
copying under Title 17, United States Code.

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

© Copyright 1998
Cheryl K. Hart

Doctoral Dissertation

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date August 19, 1998

University of Washington

Abstract

Theory and Evaluation of a New Physiologic Sampling Pump

by Cheryl K. Hart

Chairperson of the Supervisory Committee: Associate Professor Michael G. Yost
Department of Environmental Health

Physiologic sampling pumps (PSPs) seek to change pump rate in proportion to the minute ventilation of the worker. They aim to account for (1) differences in minute ventilation between individuals and (2) correlation between air concentration of contaminants and minute ventilation. This dissertation presents PSP theory and evaluates a new PSP for use in the real world. Additionally, it develops an equation termed the "physiologic volume-weighted average" (P-VWA), which is comparable in significance to a time-weighted average (TWA).

A computer simulation evaluated the effect of correlation between air concentration and minute ventilation on inhaled dose estimates. The simulation outcome was a ratio of the inhaled dose estimates (D_{PSP}/D_{TSP}), termed the Exposure Ratio, which was strongly related to both the correlation between air concentration and minute ventilation, and the GSDs of their distributions.

The laboratory phase of this research evaluated a new heart rate-controlled PSP. Subjects bicycled on an ergometer to a set protocol while their minute ventilation and heart rates were recorded. Two methods were used to estimate minute ventilation from heart rate: (1) an individual calibration curve and (2) a predictive equation derived from a database of exercise test data. The PSP itself functioned according to design, but both estimation methods failed to achieve the a priori accuracy and precision goals.

A pilot field study was performed to validate that the PSP could be used in real-world conditions. The PSP performed very smoothly; it could easily be used by Industrial Hygienists in workplaces, given an acceptable predictive equation relating heart rate to minute ventilation.

This research also verified that the integrity of the charcoal tube sampling mechanism would not be compromised by breakthrough, even at high solvent levels.

This dissertation is significant because:

(1) it demonstrates that given a better predictive equation relating heart rate to minute ventilation, a PSP could be built that would fit all criteria for real-world application;

(2) it is the first time PSP theory has been developed and presented in the literature;

(3) it is the first analysis of the effect of correlation between air concentration and minute ventilation on inhaled dose estimates.

TABLE OF CONTENTS

List of Tables ix
List of Equations xii
Glossary xiv
List of Abbreviations xvi
Introduction 1
The Problem Statement 3
The Subproblems 3
The Hypotheses 4
The Delimitations 5
Review of Related Literature 5
Research by Satoh, Higashi, Sakurai, and Omae 8
Research by Levine 8
Research by Yost, Hart, and Lopez 10
Synopsis 13
Chapter 1: Physiologic Sampling Pump Theory 14
Inhaled Dose Estimates Using TSP versus PSP 14
PSP Measurements in Relation to TWAs 18
Criteria for an Ideal PSP 20
Heart Rate as a Predictor of Minute Ventilation 21
Physiology 21
Advantages and Disadvantages to Heart Rate 25
Significance 26
Chapter 2: Assessment of the Effect of a Correlation Between Air Concentration and Minute Ventilation on the Estimation of Inhaled Dose 29
Methods 32

Results 37
Discussion 60
Conclusions 64
Chapter 3: Laboratory Validation of a Physiologic Sampling Pump 66
Introduction 66
Methods 69
Pneumo Method Calibration 72
RIP Method Calibration 72
Direct Heart Rate Method 75
Indirect Heart Rate Method 77
Satoh's Heart Rate Method 84
Laboratory Simulation 85
RPD Analysis 89
Concordance Analysis 90
Dynamic Pump Response 92
Results 95
Human Subject Characterization 95
Direct Calibration Results 96
Indirect Calibration Results 98
Calibration Curve Comparison 101
PSCU Total Inhaled Volume Results 104
Mass Flowmeter Total Inhaled Volume Results 113
RPD Analysis Results 114
Concordance Analysis Results 118
Dynamic Pump Response Results 123
Discussion 128
Conclusions 141
Chapter 4: Pilot study of Charcoal Tube Breakthrough Characteristics When Using A Physiologic Sampling Pump 143

Introduction....143
Methods....143
Results....146
Discussion....147
Conclusions....148
Chapter 5: Field Testing of the Physiologic Sampling Pump....149
Introduction....149
Methods....150
Equipment Calibration....151
Experiment Preparations....151
Sample Collection....153
Charcoal Tube Analysis....153
Results....153
Subject #1....154
Subject #2....166
Subject #3....173
Discussion....181
Conclusions....185
Chapter 6: Summary and Conclusions....186
Bibliography....191
Appendix A: Yost/Hart/Lopez Manuscript....194
Appendix B: Computer Simulation Visual C++ Code....219
Appendix C: Phone Interview....225
Appendix D: Subject Data Collection and Processing Checklist....227
Appendix E: Summary of Notations Made on Checklist....231
Appendix F: Human Subject Consent Form....233
Appendix G: SUBJHX Excel Macro....237
Appendix H: Information Collected Via the Subject History Macro....243
Appendix I: RIPCAL.bat Code....247

Appendix J: ISO Excel Macro 248
Appendix K: RIPCAL Excel Macro 254
Appendix L: HRATE.bat Code 257
Appendix M: HRCAL Excel Macro 258
Appendix N: PREDHRLOGVI Excel Macro 261
Appendix O: S-Plus Code for Bootstrap Analysis 265
Appendix P: TXDAT.bat Code 267
Appendix Q: SENDDAT.bas Code 268
Appendix R: ANALOG0 PSCU Code 270
Appendix S: ANALOG2 PSCU Code 271
Appendix T: ANALOG6 PSCU Code 272
Appendix U: LABPSP PSCU Code 273
Appendix V: Primary Standard Calculations 277
Appendix W: VTEST15S PSCU Code 279
Appendix X: VTEST60S PSCU Code 281
Appendix Y: NIOSH Method 1500 283
Appendix Z: PUMPCAL PSCU Code 288
Appendix AA: FIELDPSP PSCU Code 290

LIST OF FIGURES

Number *Page*

Figure 1. Possible correlation mechanisms....16

Figure 2: Exposure Ratio versus GSD[P] when K= 0.2 and –0.2....38

Figure 3: Exposure Ratio versus GSD[P] when K= 0.4 and –0.4....39

Figure 4: Exposure Ratio versus GSD[P] when K= 0.6. 40

Figure 5: Exposure Ratio versus GSD[P] when K= 0.8. 41

Figure 6: A graph of all data points generated from positive K values. Exposure Ratio plotted versus GSD[P]....43

Figure 7: Exposure Ratio versus GSD[P] when GSD[C]= 1.2....45

Figure 8: Exposure Ratio versus GSD[P] when GSD[C]= 1.6....46

Figure 9: Exposure Ratio versus GSD[P] when GSD[C]= 2.0....47

Figure 10: Exposure Ratio versus GSD[P] when GSD[C]= 2.4....48

Figure 11: Exposure Ratio versus GSD[P] when GSD[C]= 2.8....49

Figure 12: Exposure Ratio versus GSD[P] when GSD[C]= 4.0....50

Figure 13: Exposure Ratio versus GSD[P] when GSD[C]= 6.0....51

Figure 14: Three-dimensional plot of the Exposure Ratio values for K= 0.2....56

Figure 15: Three-dimensional plot of the Exposure Ratio values for K= 0.4....57

Figure 16: Three-dimensional plot of the Exposure Ratio values for K= 0.6....58

Figure 17: Three-dimensional plot of the Exposure Ratio values for K= 0.8....59

Figure 18: Mouth unit attached to a subject's head. 71

Figure 19: Plot of the abdominal band signal versus the ribcage signal during the three isovolume maneuvers for Subject #6, Session #1, Calibration #1....74

Figure 20: Plot of the sum of the adjusted RIP band derivatives versus the minute ventilation. 75

Figure 21: Histogram of the correlation coefficients between delta heart rate and the logarithm of minute ventilation for all database subjects. 81

Figure 22: Diagram of raw subject data (consisting of the four recorded signals) being sent from the computer to the PSCU. 87
Figure 23: Diagram of the equipment set-up. 89
Figure 24: Plot of Session #1 versus Session #2 estimated minute ventilation values at the three calibration points (0, 40, and 80 Watts) using the Direct HR Method. 97
Figure 25: Histogram of delta heart rate coefficients. 98
Figure 26: Histogram of weight coefficients. 99
Figure 27: Histogram of intercept coefficients. 100
Figure 28: Plot of the estimates of the logarithm of minute ventilation based on each heart rate calibration method (Indirect HR Method and Direct HR Method) versus heart rate. 102
Figure 29: Plot of the difference of the logarithm of minute ventilation (Indirect HR Method - Direct HR Method) versus heart rate. 103
Figure 30: Difference in minute ventilation estimates (Indirect HR Method - Direct HR Method), stratified by subject's fitness 104
Figure 31: Plot of Pneumo Method versus Primary Standard PSCU results. 107
Figure 32: Plot of Direct HR Method versus Primary Standard PSCU results. 108
Figure 33: Plot of Indirect HR Method versus Primary Standard PSCU results. 109
Figure 34: Plot of RIP Method versus Primary Standard PSCU results. 110
Figure 35: Plot of Satoh's HR Method versus Primary Standard PSCU results. 111
Figure 36: Plot of Indirect HR Method versus Direct HR Method PSCU results. 112
Figure 37: RPD of the total inhaled volume calculated via the Pneumo Method versus the Primary Standard. 114
Figure 38: RPD of the total inhaled volume calculated via the RIP Method versus the Primary Standard. 115
Figure 39: RPD of the total inhaled volume calculated via the Direct HR Method versus the Primary Standard. 116

Figure 40: RPD of the total inhaled volume calculated via the Indirect HR Method versus the Primary Standard. 117
Figure 41: RPD of the total inhaled volume calculated via the Satoh's HR Method versus the Primary Standard. 118
Figure 42. Relationship of mass flowmeter estimates for all Methods to PSCU total inhaled volume estimates. 122
Figure 43: Original input voltage to PSCU (60 second steps). 123
Figure 44: Mass flowmeter voltages, measuring the flows generated by the pump, for pumps A, B, C, and D, using 15-second steps. 124
Figure 45: Comparison of mass flowmeter voltage in response to the same sized step using pump A. 125
Figure 46: Mass flowmeter voltage during step input to the sampling pump, under two conditions. 126
Figure 47: Same as Figure 46, except that the range is restricted to largest positive step and mass flowmeter response at 30 seconds into step shifted to overlap with original recorded values. 127
Figure 48. Timeline of minute ventilation estimates for Subject #11 using the Pneumo Method and Direct HR Method. 130
Figure 49. Timeline of minute ventilation estimates for Subject #18 using the Pneumo Method and Direct HR Method. 131
Figure 50. Time line of Pneumo Method, Direct HR Method, and Indirect HR Method, for Subject #11. 133
Figure 51. Time line of Pneumo Method, Direct HR Method, and Indirect HR Method, for Subject #30. 134
Figure 52. Timeline of minute ventilation estimates for Subject #11 using the Pneumo Method and the RIP Method. 137
Figure 53. Timeline of minute ventilation estimates for Subject #32 using the Pneumo Method and the RIP Method. 138

Figure 54: Example of how laboratory subject population may have differed from the true population. 141
Figure 55: Histogram of 15-second averages of heart rate for Subject #1. 156
Figure 56: Histogram of 15-second averages of estimated minute ventilation for Subject #1 157
Figure 57: Histogram of 15-second averages of toluene concentration for Subject #1 ... 158
Figure 58: Minute ventilation and toluene concentration over time. 159
Figure 59: Comparison of analysis results between the three charcoal tubes sampled using the TSP versus the three sampled via the PSP. 163
Figure 60: TWA versus P-VWA for Subject #1. Each data point represents a different chemical species. 165
Figure 61: Relationship between air concentration of toluene (ppm) and minute ventilation (LPM) for Subject #1. 166
Figure 62: Histogram of Subject #2's 15-second heart rate averages. 168
Figure 63: Histogram of Subject #2's 15-second estimated minute ventilation averages 169
Figure 64: Minute ventilation over time. 170
Figure 65: Comparison of analysis results between the charcoal tube sampled using the TSP versus the one sampled via the PSP. 172
Figure 66: Histogram of Subject #3's 15-second heart rate averages. 174
Figure 67: Histogram of Subject #3's 15-second estimated minute ventilation averages 175
Figure 68: Histogram of 30-second xylene concentration averages. 176
Figure 69: Minute ventilation and xylene concentration over time. 177
Figure 70: Comparison of analysis results between the charcoal tube sampled using the TSP versus the one sampled via the PSP. 179
Figure 71: Relationship between air concentration of xylene (ppm) and minute ventilation (LPM) for Subject #3. 180

LIST OF TABLES

Number *Page*

Table 1: Mean and range of the R values for each K value. 37

Table 2: Exposure Ratios when K= 0.2, for all combinations of GSD[C] and GSD[Vi]. 52

Table 3: Exposure Ratios when K= 0.4, for all combinations of GSD[C] and GSD[Vi]. 53

Table 4: Exposure Ratios when K= 0.6, for all combinations of GSD[C] and GSD[Vi]. 54

Table 5: Exposure Ratios when K= 0.8, for all combinations of GSD[C] and GSD[Vi]. 55

Table 6: Lowest GSD combination for each positive K value. 64

Table 7: Subject matrix showing the number of subjects in database for each matrix cell. 78

Table 8: Subject matrix showing the subjects excluded from the analysis for each matrix cell. 82

Table 9: Summary of male subjects' age, height, weight, and body mass index. 95

Table 10: Summary of female subjects' age, height, weight, and body mass index. 95

Table 11: First session steady-state heart rate and minute ventilation statistics for 0, 40, and 80 Watts. 98

Table 12: PSCU estimates of total inhaled volume (in liters) from the first sampling period for all 30 subjects. 105

Table 13: PSCU estimates of total inhaled volume (in liters) from the second sampling period for 16 of the subjects. 106

Table 14: Mass flowmeter estimates of total inhaled volume (in liters) from the first sampling period for all 30 subjects. 113

Table 15: Concordance component values for first session PSCU results. 119
Table 16: Comparison of concordance component values for first session PSCU results, of Direct HR Method to Primary Standard, when including all subjects versus only Caucasian subjects. 120
Table 17. Concordance component values for first session PSCU versus mass flowmeter results 121
Table 18: Four conditions tested in this experiment. 144
Table 19: Expected masses of toluene loaded onto charcoal tubes that were under and over the breakthrough point for each sampling condition. 146
Table 20: Recovered masses and apparent recovery efficiencies for the front sections of the charcoal tubes surrounding the breakthrough point for all four conditions. 147
Table 21: Field subject characteristics. 154
Table 22: Mass of analyte (in micrograms) on the charcoal tube samples collected on the TSP and PSP for Subject #1 160
Table 23: Sample times and volumes for each charcoal tube. 161
Table 24: Air concentration results (in milligrams per cubic meter) based on charcoal tube samples collected on the TSP and PSP for Subject #1. 161
Table 25: TWA and P-VWA results for Subject #1's work shift on 3/18/97. The toluene values are based on the first two charcoal tube results only. 164
Table 26: Mass of analyte (in micrograms) on the charcoal tube samples collected on the TSP and PSP for Subject #2 171
Table 27: Sample times and volumes for both charcoal tube. 171
Table 28: Air concentration results (in milligrams per cubic meter) based on charcoal tube samples collected on the TSP and PSP for Subject #2. 171
Table 29: Mass of analyte (in micrograms) on the charcoal tube samples collected on the TSP and PSP for Subject #3 178
Table 30: Sample times and volumes for both charcoal tubes 178

Table 31: Air concentration results (in milligrams per cubic meter) based on charcoal tube samples collected on the TSP and PSP for Subject #3. 178
Table 32: Summary of subjects' age, height, weight, and body mass index. 212

LIST OF EQUATIONS

Number ..... *Page*

Equation 1: Condition under which particle distributions through a filter would be equivalent. 7

Equation 2: Stoke's number. 7

Equation 3: Inhaled dose equation. 14

Equation 4: Inhaled dose equation when using a traditional sampling pump 15

Equation 5: Inhaled dose equation when using a physiologic sampling pump. 18

Equation 6: Time weighted average equation. 19

Equation 7: Physiologic volume weighted average equation. 19

Equation 8: Inhaled dose equation for samples collected on traditional sampling pumps, incorporating the TWA value. 20

Equation 9: Inhaled dose equation for samples collected on physiologic sampling pumps, incorporating the P-VWA value. 20

Equation 10: Equation for the geometric standard deviation of the product of two lognormal variables ($\dot{V}i$ and C). 31

Equation 11: Generation of a Data List that is correlated to the Random List via the correlation coefficient, K. 35

Equation 12: Equation for generating the list of 15-second averages of air concentration. 35

Equation 13: Equation for generating the list of 15-second averages of minute ventilation. 35

Equation 14: Calculation of the TWA for each iteration. 36

Equation 15: Calculation of the P-VWA for each iteration. 36

Equation 16: Predictive equation model for the direct heart rate calibration method. 76

Equation 17. Satoh *et al.*'s equation to relate heart rate to minute ventilation. 77

Equation 18: Predictive equation model for the indirect heart rate calibration method. 83

Equation 19. Satoh *et al.*'s equation to relate heart rate to minute ventilation. 85
Equation 20: Relative percent difference between two Methods.. 89
Equation 21: Definition of location shift in concordance analysis. 90
Equation 22: Definition of scale shift in concordance analysis. 90
Equation 23: Definition of concordance precision. 91
Equation 24: Definition of concordance accuracy. 91
Equation 25: Definition of concordance. 91
Equation 26: First order linear system differential equation in standard form. 92
Equation 27: Response of a first order linear system to a step input. 93
Equation 28: Predictive equation relating minute ventilation to heart rate derived from a large database of physiologic data. 101
Equation 29. Expected mass (in mg) collected on a charcoal tube. 145

GLOSSARY

Accuracy. A measure of degree of agreement with a standard or a true value.

Cardiotachometer. An instrument that measures the heart rate. Typically, this instrument sends out an electronic signal of 5 mV each time the heart beats, as measured on the subjects via three electrodes; the period between the heart beats can be converted into a heart rate (or beats per minute).

Concordance. A measure of degree of agreement between two methods.

Inhaled dose. The mass of a compound that is inhaled over a specified period of time. The uptake of the compound is related to the inhaled dose by the percent of the compound absorbed by the body (i.e. if 100% of the inhaled dose is absorbed then the uptake would equal the inhaled dose).

Macro. A small computer program that is written using a programming language called Visual Basic and can be executed in Excel.

Minute ventilation. The liters of air inspired per minute.

Permissible exposure limit. An occupational health standard enforced by either OSHA or a comparable state agency (i.e. WISHA in Washington State).

Physiologic sampling pump. A sampling pump controlled via a physiologic signal (i.e. from a pneumotachometer) such that it generates a pump flow that is proportional to minute ventilation.

Pneumotachometer. An instrument which measures inspiration and/or expiration rate. Often done by measuring the pressure difference across a wire mesh as air moves through the mesh in response to airflow changes.

Precision. A measure reflecting the scatter of data around a central point.

Relative percent difference. The experimental value minus the true value, divided by the true value, expressed as a percentage.

Respiratory inductive plethysmograph. An instrument that uses two (or more) bands, wrapped around the chest and the abdomen, to estimate minute ventilation from movements of the chest and abdominal cavities.

Threshold limit value. A guideline developed by the American Conference of Governmental Industrial Hygienists (ACGIH) to assist in control of health hazards.

Traditional sampling pump. A sampling pump commonly used by Industrial Hygienists; in this study, a Gilian GilAir 5 sampling pump.

LIST OF ABBREVIATIONS

PEL. Permissible exposure limit.

PSP. Physiologic sampling pump.

P-VWA. Physiologic volume-weighted average.

RIP. Respiratory inductive plethysmograph.

RPD. Relative percent difference.

TLV. Threshold limit value.

TSP. Traditional sampling pump.

TWA. Time-weighted average.

ACKNOWLEDGEMENTS

The author wishes to thank the following individuals: Michael Yost, her advisor, for his many hours spent counseling her along the way; Jacob Hildebrandt, her committee member and teacher, who helped train her how to think like a scientist; Michael Morgan, for being a model teacher and for always being there for her without question; Kathy Sietsma, for opening doors to get the database information as well as for the personal time she took out of her very busy life to assist me as much as possible; Ed Luschei, for assistance with the computer simulation coding; and Ted Hart, for his many hours spent carefully editing her work.

DEDICATION

The author wishes to dedicate this dissertation to her husband, Ted; her parents, Erich and Nancy Luschei; and her grandparents, Charles and June Strother. Her husband convinced her that she could accomplish a Ph.D., her father inspired her to get one, her mother gave her the tools to complete it, and her grandparents listened eagerly to the highlights as well as the periodic hurdles along the way. Without all of these people's support, this dissertation would not have been written.

INTRODUCTION

The ability to accurately quantify worker exposure to toxic compounds is important for industrial hygiene, epidemiological, and risk assessment purposes. The ultimate goal of exposure assessment as defined here, is to determine the mass of the compound of interest to which the body surface is exposed, called the barrier dose. Of the three main routes of entry for hazardous chemicals (inhalation, dermal, and ingestion), inhalation is considered to be of greatest significance.

In the past, area monitoring of air concentrations served as the measure of "exposure". With the development of personal sampling pumps, however, personal sampling has become the desired means of estimating an individual's exposure. Personal sampling pumps are small units that attach to the belt of an individual. Tubing connects these pumps to a collection device that is usually located on the lapel of the person being sampled. Thus, air from the actual breathing zone of the sampled individual is pulled through the collection device, and can later be analyzed for a variety of compounds.

While these devices yield a definite improvement over area sampling, they fail to account for a critical factor: variations in breathing rate. As a person works harder, his/her pulmonary ventilation, cardiac output, and metabolism all increase to meet the additional energy requirements.[1] When a person's pulmonary ventilation increases, the gas-exchanging surfaces in the lung are exposed to more of the hazardous contaminants in the air. Therefore, because these "traditional" sampling pumps (TSPs) have constant pump rates, our ability to accurately estimate inhaled dose is limited. A physiologic sampling pump, or PSP, eliminates this constraint; it is a sampling pump that has been modified such that the pump rate varies in proportion to pulmonary ventilation.

While a few physiologic sampling pumps have been reported in the literature, the PSP evaluated in this dissertation utilizes state-of-the-art technology (such as a sophisticated

digital controller) to uniquely combine several techniques which until now have only been attempted in a limited fashion. Additionally, this pump can be calibrated to a subject in the field in a short period of time, without the need for extensive training on the part of the operator. Perhaps most importantly, it is the only pump that can change its pump rate in proportion to pulmonary ventilation, by using heart rate as a surrogate measure.

This research goes beyond the mere evaluation of a new physiologic sampling pump by developing the theory on which it is based, testing that theory in a laboratory, and evaluating the pump in the field on real workers. No researcher to date has reported any of these listed activities.

In Chapter 1, physiologic sampling pump theory is presented by first deriving the equations necessary to relate the mass of a compound collected (via both traditional and physiologic sampling pumps) to inhaled dose. Second, a method for comparing physiologic sampler results to current regulatory and/or recommended standards is proposed. Third, the criteria for an ideal physiologic sampling pump are listed. Finally, the theory for using heart rate as a surrogate measure of minute ventilation is explored; it is important to understand the relationship between heart rate and minute ventilation in order to critically analyze the results of the experiments.

Chapter 2 addresses the issue of correlation between air concentration of a contaminant and the minute ventilation of a worker, and how this might affect exposure estimates. This issue is explored using a computer simulation and provides justification for the need of PSPs in some workplaces.

Chapter 3 presents the laboratory validation study. This phase involved developing calibration methods for using heart rate as a surrogate measure of pulmonary ventilation and then testing the PSP using data collected from 30 human subjects.

Chapter 4 examines the effect of variable pump flow on charcoal tube breakthrough characteristics. While seemingly a minor issue, this set of experiments provides necessary information for evaluating the potential use of the PSP under real world conditions. If the variable pump flow reduced the possible sampling time significantly, for example, then it would not be a practical device for Industrial Hygienists.

The results of the field test are discussed in Chapter 5. The primary purpose of this field test was to validate that the PSP fulfilled the criteria of an ideal PSP (presented in Chapter 1). A secondary purpose was to compare samples from a PSP with those of a TSP. Finally, this experiment investigated the existence of a correlation between air concentration and minute ventilation (the significance of which is described in Chapters 1 and 2).

Finally, Chapter 6 draws all of the results together to arrive at overall conclusions for the dissertation.

THE PROBLEM STATEMENT

This research proposes to validate the theory behind physiologic sampling pumps and evaluate a novel instrument, a heart rate-controlled physiologic sampling pump (with a sophisticated digital controller), for use in the field.

THE SUBPROBLEMS

I. Development of physiologic sampling pump theory, including: (1) comparison of inhaled dose estimates calculated from samples gathered on ventilation-controlled physiologic sampling pumps versus traditional sampling pumps, (2) an equation for use with the PSP, similar to the time-weighted average (TWA), to be referred to as the physiologic volume-weighted average (P-VWA), (3) criteria for an ideal PSP, and

(4) rationale for selecting heart rate as the surrogate measure of minute ventilation for use in the field.

II. Assessment of the effect of correlation between air concentration and minute ventilation on the estimation of inhaled dose via computer simulations.

III. Calibration method development for several input signals (to relate the input signal to minute ventilation) for the physiologic sampling pump: (1) heart rate (individually calibrated), (2) heart rate (using a predictive equation for use in the field), and (3) a respiratory inductive plethysmograph (for laboratory use only).

IV. Evaluation of input signals using human subject data collected in the laboratory, by comparing total inhaled volume estimates against the true total inhaled volume.

V. Exploration of charcoal tube breakthrough characteristics when sampling with the PSP.

VI. Comparison of exposure estimates based on samples collected via heart rate-controlled physiologic sampling pumps versus those collected by traditional sampling pumps, using workers in the field.

THE HYPOTHESES

I. A database composed of exercise test data from 181 subjects can be used to predict minute ventilation via heart rate with the following accuracy and precision goals: an average RPD less than 20% (a measure of accuracy) and a RPD standard deviation of less than 10% (a measure of precision) for leg work rates ranging from 0 to 120 Watts.

II. The degree of difference between inhaled dose estimates based on samples collected via traditional sampling pumps versus those collected by heart rate-controlled

sampling pumps is dependent on the magnitude of correlation between air concentration and minute ventilation.

III. Breakthrough of toluene through a charcoal tube is not likely to be a problem when using the PSP in the field.

IV. There is no significant difference between inhaled dose estimates based on samples collected on workers in the field via heart rate-controlled sampling pumps versus those collected by traditional sampling pumps.

THE DELIMITATIONS

I. The only chemical studied in the laboratory phases of the study will be deuterated toluene.

II. The only sampling media tested in conjunction with the physiologic sampling pumps will be charcoal tubes.

III. During the charcoal tube breakthrough experiment, only stationary air concentrations will be tested.

IV. The field study will be limited to University of Washington workers who are likely to be exposed to solvents (e.g. painters, laboratory personnel).

REVIEW OF RELATED LITERATURE

Three physiologic sampling pumps have already been reported in the literature. It is important to examine each of these research projects in order to demonstrate how the physiologic sampling pump evaluated in this dissertation research (PSP) is unique and builds upon this past work.

RESEARCH BY KUCHARSKI[2]

In 1980, Kucharski described a personal dust sampler that used heart rate to vary the pump rate in proportion to estimated minute ventilation. An electronic unit received an ECG signal from electrodes placed on a subject's chest. The electronic unit calculated the heart rate, which was then used to predict minute ventilation; unfortunately the author does not specify the equation, or equations, used to relate the heart rate to minute ventilation.

Kucharski tested his pump in the field by comparing the inhaled dose based on the amount of contaminant on respirator filters versus (1) area samples, (2) the personal dust sampler, and (3) a traditional sampling pump. The fact that the inhaled dose estimates based on the personal dust sampler were not significantly different from those based on the respirator filters led the author to conclude that the system worked as designed. The author does mention that the problem of size-selection of particles, which is dependent on a constant flow rate, was not addressed.

There are at least two significant problems with Kucharski's conclusion (that the PSP's flow rate was proportional to minute ventilation), stemming from the fact that he assumes that of the weight of the filtered particles on a respirator filter would be related to the weight of the particles on the personal dust sampler filter. The first problem is that the distribution of particle sizes pulled through the respirator filter would only be the same as the distribution of particle sizes pulled through the filter cartridge attached to the personal dust sampler under a single condition: if the ratio of the airflow rates was equal to the cubed ratio of the diameter of the filter inlets (see Equation 1). This condition is based on Equation 2 from Hinds[3] and the following assumptions: (1) that for the particle distribution to be the same, the Stoke's numbers would also have to be the same, (2) that the relaxation coefficient would be the same for both systems, and (3) that the air stream velocity is equal to the airflow rate divided by the area of the filter inlet. Kucharski gives us no information about the inlet diameters and does not discuss this issue.

$$\frac{Q_1}{Q_2} = \left(\frac{D_1}{D_2}\right)^3$$

Equation 1: Condition under which particle distributions through a filter would be equivalent, where Q= air flow rate and D= inlet diameter.

$$Stk = \frac{\tau \cdot U}{D}$$

Equation 2: Stoke's number, where τ= the relaxation coefficient, U= airflow velocity, and D= inlet diameter.

The second problem relates strictly to airflow rate: the airflow rate patterns through the two filters would be extremely different. The airflow rate through the PSP filter must be constant over some period of time (since the heart rate must be averaged over some period of time before an updated estimate of minute ventilation can be sent to the pump), while the airflow through the respirator would be constantly changing with each breath. Since the air stream velocity affects the sizes of particles entrained, the two filters would likely be exposed to very different sizes of particles even if Equation 1 were satisfied.

While the signal Kucharski used to control the personal dust sampler is similar to the proposed PSP, this research differs in the following ways:

- the theory behind the PSP will be validated;
- the PSP will sample for solvents (versus dust);
- the PSP will use a more sophisticated (non-linear) controlling device.

RESEARCH BY SATOH, HIGASHI, SAKURAI, AND OMAE[4]

For nearly a decade the physiologic sampling pump idea dropped out of the literature. It resurfaced in 1989 in a slightly different form; Satoh *et al.* designed a real-time instrument to record both heart rate and air concentrations simultaneously and continuously. They could then, theoretically, account for changes in minute ventilation post-sampling by weighting the instantaneous air concentrations by the corresponding minute ventilation (as estimated by pulse rate). The paper published in 1989, however, only addressed the use of heart rate as a surrogate measure of minute ventilation; the authors never published an article on the testing or use of the actual sampling device. Their device has several advantages over the physiologic sampling pump used in this research, all deriving from the detail of data one can acquire with real-time instruments. For example, one can determine the tasks that are contributing to the bulk of exposure and determine the relative contribution of both minute ventilation and air concentration. While real-time instruments can provide more detailed data, they restrict the number and types of compounds that one can measure (not to mention the fact that these instruments tend to be quite expensive). The proposed PSP fills a different niche than a device incorporating real-time instruments, just as TSPs fill a role today different from that of real-time instruments. Additionally, it should be noted that a heart rate monitor can be paired with any real time instrument in order to provide the same information gathered by the device built by Satoh *et al.*

RESEARCH BY LEVINE[5]

More recently, Levine developed a physiologic sampling pump that used thoracic impedance to estimate minute ventilation. Thoracic impedance is measured via two electrodes that are placed across the chest. A current is applied across the electrodes at a particular frequency, and as lung volume changes there are low frequency changes in the carrier signal related to changes in the electrical impedance of the body.

The major drawback of Levine's pump lies in its use of thoracic impedance. Levine identified two problems with his device: (1) false signal was introduced due to sudden motions by the subjects, and (2) field calibration would be difficult. The latter would pose a significance hindrance to use by Industrial Hygienists. In addition, there were four important research questions Levine did not address (and which have not been addressed in the literature):

How well does a calibration performed on subjects sitting on a bicycle apply to other body positions? Impedance is known to change with body position which, therefore, introduces error into minute ventilation estimates.[5] The degree of error introduced in this fashion is critical to the validation of the pump itself.

How long does a calibration remain stable during a real work shift? Levine did not (1) recalibrate his subjects post-sampling or (2) validate calibration on real workers in the field. This is a critical question since sweat or movement could affect either the electrode placement or the impedance itself.

Does normal body motion disrupt the signal? Levine noted that sudden motions by the subjects introduced false signal. What about motions performed by workers? It is important to be able to define which types of movement the signal remains stable for and which it does not. A physiologic sampling pump, after all, is designed to be worn by workers who are moving.

What happens to the relationship between inspired volume and estimated inspired volume above 100 Watts of work? Levine calibrated his subjects up to only 100 Watts. It is critical to know the relationship above 100 Watts, as workers undoubtedly work harder than 100 Watts for at least short periods of time, even if not for extended periods.

RESEARCH BY YOST, HART, AND LOPEZ

The physiologic sampling pump developed by Yost, Hart, and Lopez is evaluated in this dissertation. This PSP builds upon the earlier work of Kucharski[2] and Levine[5] by using elements from both to develop a novel instrument. While an extensive description of this device is included for reference (see Appendix A for an unpublished manuscript), a summary is provided below.

A commercial personal sampling pump (a Gilian "GilAir 5") was selected as the basis for this device. This pump was selected because it has the capability to give a wide dynamic range of flow rates, from approximately 75 ml per minute to a maximum of about 7.5 liters per minute. In addition, the analog control loop provides very stable and tight control of the pump flow over this range.

The pump was modified such that an external analog DC control signal (which ranged from 0 to 100 mV) can control the pump flow. The modification was designed such that it does not interfere with the pump's flow regulation, pressure compensation, or fault sensing circuits. Additionally, when an external signal is not provided to the pump, the pump is able to operate normally.

A second component of the physiologic sampler system is the Physiologic Sampler Control Unit (PSCU). The PSCU accepts a physiologic input signal and produces the analog output signal which controls the pump flow. The PSCU provides a number of system functions including:

- Conditioning the analog or digital physiologic input signals;
- Converting of the input signals to a calibrated measure of minute ventilation for the individual subject;
- Converting of the minute ventilation to a proportional calibrated flow rate for the pump;

- Data logging the pump flow rate over time to record the total sample volume;

- Sending an appropriate output signal to control the pump at specified time intervals.

These functions were implemented with a digital micro-controller and custom designed input-output hardware. The micro-controller used by the PSCU is a Tattletale model 4A (Onset Computer, Falmouth MA). The Tattletale 4A provides the following standard features: 8-channel, 12-bit analog-to-digital conversion, onboard clock/timer, 16 digital input/output lines, 2 channel serial interface (programmable UART), 32 K of non volatile RAM data storage, and a regulated 5 volt supply. Operating current is strongly dependent on the operating program activities; typical power requirements are 7 to 15 volts DC at a current of about 20 mA (although much lower power operation is possible).

Functions for this compact (2.25x 3.75 inch) micro-controller device are programmable in TxBASIC, a compiled BASIC-style language customized for the controller features. Control programs are written on a laptop or desktop PC, compiled, and downloaded into the Tattletale via a standard RS-232 interface connection. Functions in TxBASIC allow the user to provide input or control the operating program when the terminal is connected. When the terminal is disconnected, the program will operate unattended until the terminal connection is re-established. Data logging is done in a compact binary format, with time stamping of each record. The binary format is expanded to ASCII format for downloading. Retrieval of logged data is accomplished over the same RS-232 serial interface with a PC attached as a terminal.

In addition to the standard features of the Tattletale 4A, a number of auxiliary circuit functions were needed. The four main functions were: (1) input signal conditioning, (2) digital to analog conversion for the pump control signals, (3) output signal drivers, and

(4) an auxiliary power supply. These functions were developed on auxiliary "piggy-back" circuit boards that plugged directly onto a bus connector for the Tatteltale 4A.

The digital-to-analog converter for the PSCU uses an Analog Devices AD7225. This chip provides 4 independent analog output channels with 8-bit resolution, which allowed the PSCU to independently control up to 4 sampling pumps. Input lines are addressed sequentially via a 74HC595 serial shift register and the 4 outputs are updated all at once when latched. These outputs are buffered by a CA660 CMOS quad operational amplifier to provide low impedance signals required by the pumps.

The input signal conditioning board accepts signals from a pneumotachometer, a Respiratory Inductive Plethysmograph (RIP), or a cardiotachometer. The pneumotachometer and RIP signals are analog voltages, which are processed by a LM324 quad operational amplifier. The pneumotachometer signal resembled a positive rectified sinusoid. The DC offset of the pneumotachometer signal is level-shifted to +0.5 volts and the signal range is adjusted to span from 0 to +4.5 volts at the A/D converter. The RIP provides two bipolar sinusoidal analog signals, one signal for each band, corresponding to changes in the rib cage or abdominal cross-sectional area. The DC zero crossing of the RIP signals are level-shifted to an offset of +2.5 volts and the signal range is adjusted to span from 0 to +5 volts at the A/D converter.

The cardiotachometer signal is broadcast from a small, wireless transmitter unit (Polar Inc) strapped to the subject's chest with an elastic band. The heart rate input is a 15-millisecond digital pulse corresponding to the R-wave detected from the electrocardiogram. The receiver unit inside the PSCU can detect the 5-kHz transmitter signal up to 3 feet away. The heart beat input is processed by a 7473 flip flop operating in toggle mode to provide an alternating-stepped 5V transition for each heart beat. This arrangement eliminates the need for interrupts or high-speed digitization to reliably detect the heart beat. Software timing of the inter-beat interval using the internal clock provides a beat-by-beat estimate of heart rate.

SYNOPSIS

While each of these authors approached the idea of a PSP in a slightly different manner, the basic idea was the same; they attempted to account for the change in exposure due to change in minute ventilation. These authors all had some understanding of the theory behind physiologic sampling pumps; however, none of them developed this theory in detail or addressed the following issues:

1. How the sampling results could be integrated with current standards (P-VWA);
2. How potential correlation between minute ventilation and ambient air concentrations could affect the results.

The first issue is of critical importance for acceptance of any PSP into mainstream Industrial Hygiene use; the PSP results must be able to be evaluated in terms of current exposure standards and recommendations.

The second issue has not been addressed at all in the literature, and has the potential to be extremely important in any discussion about the necessity of using PSPs.

CHAPTER 1: PHYSIOLOGIC SAMPLING PUMP THEORY

The purpose of this chapter is: (1) to derive the inhaled dose equations for samples collected on both a traditional sampling pump (TSP) and a physiologic sampling pump (PSP), (2) to derive an equation that relates the sampling results from a PSP to the current regulatory framework, (3) to detail the characteristics of an ideal PSP, and (4) to describe why heart rate might function well as a surrogate for minute ventilation.

INHALED DOSE ESTIMATES USING TSP VERSUS PSP

In the work environment there are three variables that affect the magnitude of the inhaled dose: the air concentration, the minute ventilation of a worker, and time. Since air concentration and minute ventilation both fluctuate over time, these variables must be inside the integral in an inhaled dose equation. The inhaled dose D in a time interval T is mathematically defined as the following:

$$D = \int_0^T (C \cdot \dot{V}_i) dt$$

Equation 3: Inhaled dose equation, where C = air concentration, and $\dot{V}$ i = minute ventilation.

This inhaled dose equation is able to capture the instantaneous changes of ambient air concentration and minute ventilation. Unfortunately, TSPs capture only the changes in air concentration over time. Since minute ventilation is not inherently accounted for when using TSPs, it must be estimated and factored into the inhaled dose equation after the sampling is complete. Mathematically, this can be represented by removing the minute ventilation variable from the integral, thereby treating it as a constant (mean minute ventilation), as shown in Equation 4.

$$D_{TSP} = \bar{\dot{V}}_i \int_0^T (C)dt$$

Equation 4: Inhaled dose equation when using a traditional sampling pump.

There are three important weaknesses in our estimation of inhaled dose using this equation:

1) Minute ventilation is represented by a constant mean value. While there may be work situations for which this assumption is reasonably valid (e.g. sedentary workers), there are many jobs for which this assumption is not legitimate.

2) Mean minute ventilation must be estimated. The value that one assigns for minute ventilation must be estimated for many different populations and a variety of circumstances. The value assigned can dramatically change the estimates of inhaled dose, and because there are very few data on which to base this estimate, the accuracy of these exposure estimates is uncertain.

3) The equation does not account for any correlation between minute ventilation and air concentration. While there has been little work done regarding such a correlation, one could imagine three mechanisms, illustrated in Figure 1, that would lead to either a positive correlation, no correlation, or even negative correlation between air concentration and minute ventilation.

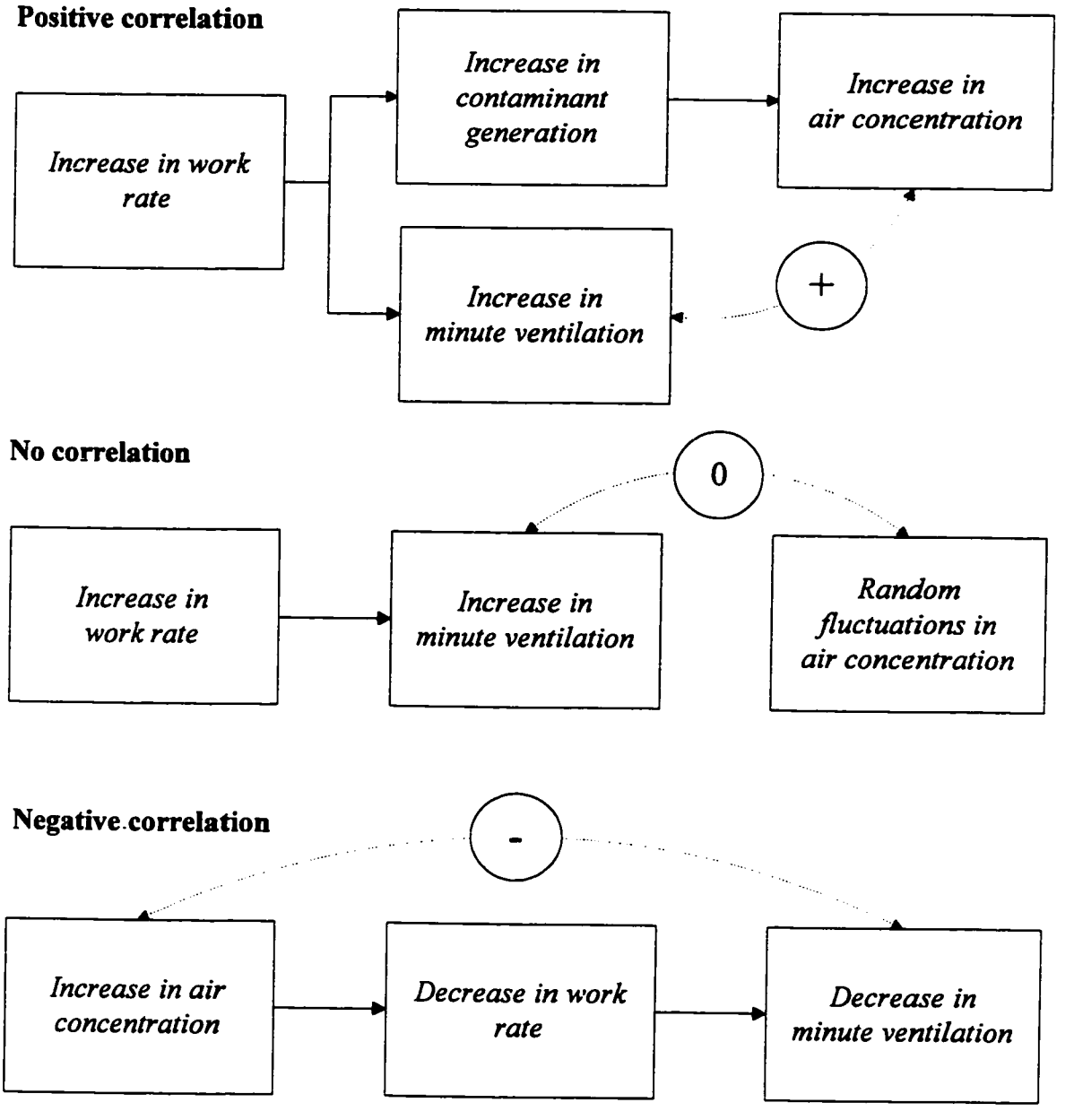


Figure 1. Possible correlation mechanisms.

A positive correlation between minute ventilation and air concentration could result from a situation where the contaminant generation is tied to what the worker is doing; for example, when an increase in shoveling dirt causes an increase in dust generation. This situation might exists for painters if the harder they work (and thereby increase their minute ventilation), the more contaminant they generate (and therefore increase the local air concentration).

If the air concentration of a contaminant is independent of a worker's work rate (and therefore their minute ventilation), then there would be no correlation between air concentration and minute ventilation.

A negative correlation could occur if increasing air concentrations tended to lower the work rate and therefore the minute ventilation. For example, if the air contaminant of interest had a negative side effect, (e.g. irritant) then workers might (consciously or otherwise) tailor their activity so as to work harder when the air concentration is lower. One could extend this hypothesis to day-to-day community exposures as well. For example, if people tend to exercise at times when the air pollution is lower, then there would be a negative correlation; on the other hand, some people might tend to exercise when air pollution is higher, possibly because of better weather conditions or because their schedules do not allow a choice. Assuming a single constant minute ventilation rate in either of these cases could significantly over- or under-estimate exposures.

Unfortunately the existence (or non-existence) of such a correlation has not been reported in the literature, and thus it is difficult to quantify the potential effect of such a correlation on actual inhaled doses. This issue, however, is critical to the decision regarding PSP development. If a correlation does not occur in the real world, or does not have a significant effect on actual inhaled doses, then TSPs do not need to be modified; the individual minute ventilation rates can be monitored independently of air sampling. If, on the other hand, it does prove to be a significant factor, then the investment in PSP

development is justifiable. This issue will become more clear with an understanding of the potential benefits of PSPs.

PSPs, as opposed to TSPs, can integrate minute ventilation into the inhaled dose estimate since pump rate changes in proportion with minute ventilation. Thus, the minute ventilation variable remains inside the integral in the inhaled dose equation for PSPs:

$$D_{PSP} = k\int_{0}^{T}(C \cdot \dot{Q})\,dt$$

Equation 5: Inhaled dose equation when using a physiologic sampling pump, where k = the proportionality constant between the pulmonary ventilation and the pump flow rate, and $\dot{Q}$ = the pump flow rate.

When minute ventilation is accounted for in this manner, all of the weaknesses previously mentioned with regard to TSPs no longer apply.

PSP MEASUREMENTS IN RELATION TO TWAs

Time-weighted average concentrations are the most common measure of exposure intensity. Their frequency of use has, to some degree, been driven by the use of TWAs in official and unofficial standards (PELs and TLVs, respectively). TWAs are calculated by averaging the ambient air concentration over time, as shown in Equation 6.

$$TWA = \frac{1}{T}\int_0^T C\,dt$$

Equation 6: Time weighted average equation.

The relationship of PSP data to TWAs is very important because collected samples must be evaluated as to whether the exposure levels are acceptable or unacceptable. Since TWAs are the most common exposure metrics used for this purpose (e.g. PELs and TLVs), it would be convenient to be able to calculate a similar exposure metric using PSP data. Based on an understanding of how TWAs are calculated using TSP data (Equation 6) and an understanding of the relationship of how PSPs differ from TSPs (Equations 4 and 5), one can generate an equivalent exposure metric called the Physiologic Volume Weighted Average, or P-VWA :

$$\text{P-VWA} = \frac{1}{V_{TOT}}\int_0^{V_{TOT}} C\,dV$$

Equation 7: Physiologic volume weighted average equation, where V = inhaled volume and V_{TOT} = total inhaled volume over the work shift.

A P-VWA applies to samples collected on a PSP and is analogous to the TWA except that the worker's minute ventilation is incorporated into the P-VWA.

The first logical question is, how will P-VWAs differ from TWAs? The answer is that they will not differ unless there is a correlation between air concentration and minute ventilation (see the Note at the end of the chapter for a proof). If there is a positive correlation then the P-VWA would be greater than the TWA, and vice versa; in either of these situations, the TWA could not accurately quantify the inhaled dose.

Even if there is no correlation, and the P-VWA equals the TWA, the P-VWA is still useful. To demonstrate this, the estimated inhaled dose equations can be re-written to incorporate TWAs and P-VWAs:

$$D_{TSP} = \bar{V}_i \cdot T \cdot TWA$$

Equation 8: Inhaled dose equation for samples collected on traditional sampling pumps, incorporating the TWA value.

$$D_{PSP} = V_{TOT} \cdot \text{P-VWA}$$

Equation 9: Inhaled dose equation for samples collected on physiologic sampling pumps, incorporating the P-VWA value.

If the TWA equals the P-VWA, then it appears that Equations 8 and 9 would be mathematically equivalent. The difference would arise from the fact that a PSP provides an estimate of the total inhaled volume for a particular worker. As noted earlier, however, if no correlation exists between minute ventilation and air concentration (and the TWA equals the P-VWA) then a PSP is unnecessary; air concentration can be obtained via a TSP, and minute ventilation can be obtained independently of any sampling device.

CRITERIA FOR AN IDEAL PSP

The criteria necessary for an ideal PSP are few but important if the technology is to be **actively used** to assess worker exposure in a real-world environment:

1. the pump must be easy to use (e.g. no additional training other than reading a simple manual);
2. the pump must be easily calibrated in the field (e.g. not technically challenging);
3. the time necessary to calibrate the worker must be minimal (e.g. less than 15 minutes, preferably less than 5);
4. it must have acceptable accuracy and precision (for this document "acceptable accuracy" was defined as having an average RPD less than 20% and "acceptable precision" was defined as having a RPD standard deviation of less than 10%).

HEART RATE AS A PREDICTOR OF MINUTE VENTILATION

Currently, inhaled volume may be estimated by attempting to determine the average work rate of a job. It is not measured directly because the equipment necessary to do so (a pneumotachometer or flowmeter mask) is a burden for both the worker and the industrial hygienist. The use of heart rate to predict pulmonary ventilation, while not a novel idea, has not been investigated thoroughly enough to be used by Industrial Hygienists in the field.

This section will first provide an overview of the physiological relationship between heart rate and minute ventilation, and then discuss why heart rate could theoretically function as a surrogate for minute ventilation.

PHYSIOLOGY

The respiratory system is composed of a series of branching airways that conduct air from outside the body to the gas-exchanging regions within the lungs, the alveoli. The volume of air inhaled and exhaled is called the tidal volume.

Ventilation is ultimately linked to the oxygen requirements and carbon dioxide production of the body. The lungs must maintain precise control of these two critical gases in order to allow the body to maintain homeostasis under widely varying conditions. As a consequence, minute ventilation is intricately linked to cellular respiration by the pulmonary and peripheral circulatory systems.[1]

When greater quantities of oxygen are utilized by mitochondria within muscle cells, for example, there is a greater production of carbon dioxide, the peripheral circulation dilates, blood flow increases due to an increase in stroke volume and heart rate, and minute ventilation increases due to an increase in tidal volume and breathing frequency.[1] Tidal volume is the primary means by which minute ventilation increases, especially at low exercise intensity.[1,6] As the tidal volume begins to exceed half of the vital capacity (the largest possible tidal volume), breathing frequency plays a larger part of the increase in minute ventilation.[1]

At the start of exercise, the profile for minute ventilation can be divided into three phases. The first phase is a rapid increase in minute ventilation that commences in the first 15-20 seconds according to Åstrand *et al.*[6], and in the first 5-15 seconds according to Miyamoto *et al.*[7] Åstrand *et al.*[6] credits the hypothalamic center with this phase, suggesting that the signals sent to the muscles to initiate exercise are paired with signals sent to the respiratory control centers. The second phase is exponential, whereby the degree of rise is related to the severity of the exercise. The carotid bodies, which are peripheral chemoreceptors, are thought to mediate this phase. The third phase is one of steady-state, whereby at exercise levels below the anaerobic threshold, the minute ventilation is linked closely to the carbon dioxide production.[6]

At the end of exercise, the above profile is reversed. The fast response phase leads to a sharp drop in minute ventilation, followed by a slower, exponential phase. Finally, minute ventilation plateaus at a level related to the new exercise level (or rest).[6]

The cardiovascular system is responsible for the transport of oxygen and carbon dioxide between the alveoli and the cells undergoing respiration. When cellular respiration increases (i.e. during an increase in work rate), the distribution of perfusion in the body changes and the cardiac output is increased. Cardiac output, which is defined as the flow of blood from the heart per unit time, is determined by the stroke volume and heart rate. Stroke volume (the volume of blood ejected per heart beat from either ventricle) increases immediately upon the start of exercise, the magnitude of which is dependent on several factors such as age, fitness and body size. After this initial increase in stroke volume, additional increases in cardiac output are attributed to increases in heart rate. [1]

Heart rate increases on the order of 10-35% of the total change within 10 seconds from the onset of exercise.[1,7] Like the phases of minute ventilation, an exponential phase follows this fast response. The fast response drop after the end of exercise is longer, and ranged from 10-30 seconds in studies conducted by Miyamoto *et al.*[7]

Just as minute ventilation is tightly coupled with oxygen demand and carbon dioxide elimination, so too is cardiac output. Since they are both linked to the same source (an increase in muscle activity), they must be fairly tightly coupled to each other. Since stroke volume is reasonably fixed, we will consider heart rate a surrogate for cardiac output. While both heart rate and minute ventilation respond to changes in muscular activity, there is a lag between the change in heart rate and change in minute ventilation. This lag is not simple to describe and depends on many factors, such as the rate of change of work rate and the pattern of change.

Bakker et al.[8], for example, examined the dynamics of minute ventilation and heart rate in response to sinusoidal (at seven different frequencies) and impulse workloads in four men. For all subjects and all conditions, the heart rate lag was **smaller** than the ventilation lag. Therefore, they concluded that the heart rate responds faster to changes in work load than does the minute ventilation. Other authors support this conclusion and

add that "the time constant of the heart rate change is shorter than that of the corresponding ventilation change."[6]

Despite the differences in transient responses between heart rate and minute ventilation, several researchers have found that the relationship between heart rate and minute ventilation remains tightly linked under both steady state[4] and non-steady state conditions.[9,10] Satoh *et al.*[4], for example, found that under steady state conditions, the average correlation coefficient between heart rate and minute ventilation was 0.97. Under non-steady state conditions, Treese *et al.*[9] found that the mean correlation coefficient between heart rate and minute ventilation was 0.9, and that the data was best-fit by using a linear curve after log-transforming the minute ventilation data.

Mermier *et al.*[10] also studied the relationship between heart rate and ventilation under non-steady state conditions; they examined the 15-second averages of each variable while subjects performed both progressive and non-progressive exercise tests. Like Treese *et al.*, they found that the within-subject relationship between these factors was curvilinear, and was best characterized by a linear regression of the logarithm of minute ventilation versus heart rate. One additional finding of interest was that the relationship between minute ventilation and heart rate remained constant, regardless of the pattern of workloads (e.g. progressive versus non-progressive). This finding is an important indicator that a heart rate to minute ventilation calibration curve derived via progressive exercise testing can be meaningfully applied to non-progressive activity (as would be expected in a work environment). In addition, they concluded that "during averaging times of 15 seconds, minute ventilation remains tightly coupled to heart rate."[10] This conclusion is important in terms of the PSP design and will be referred to in Chapter 3.

McCool and Paek[11] studied the within-subject minute ventilation to heart rate relationship at steady state for four different activities (cycling, arm cranking, lifting and pulling), and concluded that "the minute ventilation-heart rate relationship calibrated during one type of activity can be used to predict ventilation during another type of activity." Samet *et*

al.[12] concluded otherwise; they found that minute ventilation increased more steeply with upper body exercise as compared to lower body exercise. Note that this discrepancy is very important, because the ability to apply a calibration performed on an ergometer to a variety of activities that workers perform is directly related to the accuracy of using the direct or indirect calibration methods presented in Chapter 3.

McCool *et al.*[11] and Samet *et al.*[12] cannot both be right; other literature tends to support the conclusion, by Samet *et al.*[12], that the slope of the minute ventilation to heart rate relationship is steeper for upper body exercise. For example, Taguchi and Horvath[13] found that for a given heart rate, minute ventilation is higher for upper body exercise as compared to lower body exercise. They attributed the larger increase in ventilation to higher muscle spindle activity and stimulation of mechanoreceptors; this would stimulate the respiratory center, thereby inducing reflex ventilation. McCool *et al.*[11] suggest that the reason their results differed from Samet *et al.*[12] might have been due to the fact that their subjects had a much more restricted range of heart rates.

Advantages and Disadvantages to Heart Rate

There are both advantages and disadvantages of using heart rate as a surrogate measure of minute ventilation. The advantages include:

1. Heart rate is very easy to measure; three electrodes embedded in a band can easily detect the R-wave when wrapped around a subject's chest. Since there is one R-wave per heartbeat, one can time the period between R-waves and determine the frequency (or beats per minute);

2. Heart rate is tightly coupled to minute ventilation under both steady state and non-steady state condition, even when using a 15-second averaging period;

3. Large numbers of exercise tests have been performed which could lead to an extensive enough database to develop a reliable predictive equation between heart rate and ventilation;

4. Subjects are not likely to resist wearing a heart rate monitoring band.

The disadvantages are also important to emphasize:

1. Heart rate is not actually a measure of lung volume changes or airflow;

2. The relationship between heart rate and minute ventilation varies between subjects. Thus, this relationship must be predicted in some manner, either by use of a predictive equation or via individual calibrations;

3. The slope of the minute ventilation-heart rate relationship is steeper for arm exercise than for leg exercise; if the minute ventilation-heart rate calibration curve was developed using exercise tests which involve lower body exercise, then the minute ventilation could be underestimated when workers perform upper body activities. While this error may be relatively small at lower work rates, at higher work rates it could be significant;

4. Calibration curves derived from traditional exercise test data, relating heart rate to minute ventilation, may not be accurate in actual work situations.

SIGNIFICANCE

This research is significant for a variety of reasons. For example, by using P-VWAs (as opposed to TWAs) to characterize exposure intensity, this exposure metric becomes more "meaningful". It does so by better classifying subjects with regard to exposure, by being more closely related to inhaled dose, and by being flexible enough to be meaningful across a wide range of exposure-disease relationships. In addition, using P-VWAs in

conjunction with exposure time will yield more meaningful cumulative exposure metrics, for these same reasons.

By more accurately characterizing exposure, occupational epidemiologists would be able to improve their ability to identify exposure-response relationships, and risk assessors would be able to perform more accurate risk assessments, due to both the improved understanding of exposure-response relationships and the better summary of exposure in regard to the population on which the risk assessment is being performed. This, in turn, would lead to higher quality data that regulatory agencies could use as a basis for setting standards.

PSPs could have a significant impact on the field of Industrial Hygiene outside of their potential effect on exposure-disease relationships; they could refine task-based sampling by capturing the differences in work rates, as well as the ambient air concentrations, which are related to particular tasks. In addition, heart rates could be measured prior to sampling in order to stratify workers with regard to work rates.

Finally, PSPs allow two ways of identifying the most highly exposed population in a sampling campaign: by comparing either the P-VWAs or the cumulative inhaled doses. Furthermore, if cumulative inhaled doses are calculated, minute ventilation does not have to be factored in as a constant as it does when TSPs are used, and minute ventilation does not have to be roughly estimated for an individual or an entire group.

Note:

Proof that P-VWA = TWA when there is no correlation between concentration (C) and minute ventilation ($\dot{V}i$):

$$P-VWA=\frac{1}{V_{TOT}}\int_{0}^{V_{TOT}}CdV=\frac{1}{V_{TOT}}\int_{0}^{T}(C\cdot\dot{V}i)dt$$

When there is no correlation between C and $\dot{V}i$, then it is assumed that the average value of the product $C\cdot\dot{V}i$ will equal the product of the average values of C and $\dot{V}i$. Therefore, the integral can be split into two separate integrals, as follows:

$$\int_{0}^{T}(C\cdot\dot{V}i)dt=\frac{1}{T}\int_{0}^{T}(C)dt\cdot\int_{0}^{T}(\dot{V}i)dt$$

This calculation can be verified by the fact that both sides of this equation are equal to the product of the total time, the average air concentration and the average minute ventilation, given the assumption mentioned above. Therefore:

$$P-VWA=\frac{1}{T}\cdot\int_{0}^{T}(C)dt\cdot\frac{1}{V_{TOT}}\cdot\int_{0}^{T}(\dot{V}i)dt=TWA\cdot\frac{\bar{\dot{V}}\cdot T}{V_{TOT}}$$

or

$$P-VWA=TWA$$

CHAPTER 2: ASSESSMENT OF THE EFFECT OF A CORRELATION BETWEEN AIR CONCENTRATION AND MINUTE VENTILATION ON THE ESTIMATION OF INHALED DOSE

Assessing the effect of correlation between air concentration and minute ventilation on the estimation of inhaled dose is important in order to determine the value of using physiologic sampling pumps (PSPs) in the workplace.

Based on PSP theory (presented in Chapter 1), when there is no correlation then there is no additional value to using a PSP. Rather, it would be sufficient to estimate a worker's minute ventilation independently of the sampling pump.

When correlation does exist, however, then assessing the effect of correlation on the estimation of inhaled dose is not straightforward. To assist in the assessment, a computer simulation was performed. This simulation evaluated how changing the following three factors affected the ratio of the P-VWA to the TWA (hereafter referred to as the Exposure Ratio):

1. the correlation between the air concentration and minute ventilation;
2. the distribution characteristics of air concentration;
3. the distribution characteristics of minute ventilation.

The Exposure Ratio was selected as the method of expressing the difference between the P-VWA and the TWA because it can be easily interpreted. An Exposure Ratio of 1.2, for example, implies that the P-VWA is 20 percent greater than the TWA.

Recall that inhaled dose is calculated by multiplying either the P-VWA or the TWA by the estimated total inhaled volume (see Equations 8 and 9); therefore, a ratio of the

inhaled dose based on the P-VWA (Equation 9) to the inhaled dose based on the TWA (Equation 8) reduces to a ratio of the P-VWA to the TWA.

One interesting topic for debate is: At what Exposure Ratio value (hereafter referred to as the "Critical Exposure Ratio") would it be worth the extra effort and expense of using a PSP versus a TSP to quantify exposure?

For the purpose of discussing the results of the computer simulations, a single Critical Exposure Ratio of 1.2 was selected. It was felt that an inhaled dose estimate that was 20% higher than the dose estimated by using a TSP would be considered significant under most circumstances. Note that since the Critical Exposure Ratio for a given chemical would undoubtedly be related to its health risk, a single value would not realistically apply to all situations. For more hazardous chemicals, for example, the potential underestimation of exposure would entail more risk, so the Critical Exposure Ratio should be more conservative (i.e. lower).

Since there was no literature to assist in estimating the actual levels of correlation between air concentration and minute ventilation in work environments, the computer simulations were programmed to calculate the Exposure Ratio over the most probable range of correlation values. Since positive correlations are more likely (and probably stronger) than negative correlations, the range of correlation values tested was from –0.4 to +0.8.

A pilot computer simulation was performed that assumed that both air concentration and minute ventilation distributions were lognormally distributed variables. The Exposure Ratio was examined to determine how it was affected by changes in:

1. the geometric mean and geometric standard deviation of the air concentration distribution (GM[C] and GSD[C]);

2. the geometric mean and geometric standard deviation of the minute ventilation distribution (GM[Vi] and GSD[Vi]);

3. the correlation between air concentration and minute ventilation (R).

The pilot study revealed that the ratio was unaffected by both geometric means, but was significantly affected by both geometric standard deviations as well as the correlation between air concentration and minute ventilation.

The conclusion that the GSDs played such a significant role in the Exposure Ratio made designing the real computer simulation more difficult. This meant that three factors affected the outcome (Exposure Ratio). Since a four-dimensional relationship would be difficult to characterize, it was hoped that a surrogate measure could be found that would be a combination of all three factors. Then, perhaps, this surrogate measure could be related to the Exposure Ratio.

Based on the fact that the P-VWA is calculated by multiplying lognormal distributions (the minute ventilation and the air concentration) within an integral (see Equation 5), it was hypothesized that the following equation might be related to the Exposure Ratio; this equation calculates the GSD of the product of two lognormal variables:

$$GSD[P] = e^{\sqrt{(\ln GSD[\dot{V}i])^2 + (\ln GSD[C])^2 + 2 \cdot R \cdot \ln GSD[\dot{V}i] \cdot \ln GSD[C]}}$$

Equation 10: Equation for the geometric standard deviation of the product of two lognormal variables ($\dot{V}i$ and C), where R = the correlation coefficient between $\dot{V}i$ and C.

Equation 10 indicates that for a given value of R, there should be a unique relationship between GSD[P] and the Exposure Ratio.

A computer simulation was performed to examine this potential relationship between the GSD[P] and the Exposure Ratio. The methods used to generate the computer simulation will be presented in the methods section of this chapter. Then, the computer simulation results will be presented, followed by a discussion of their implications, significance, and conclusions.

METHODS

The Visual C++ software package (Microsoft, Redmond, WA) was used to perform the computer simulations. The design of the computer simulation program is described here in detail; the Visual C++ code itself is included in Appendix B.

The input parameters for each iteration of the simulation were the following:

- GM[C], the GM of the air concentration distribution
- GM[Vi], the GM of the minute ventilation distribution
- GSD[C], the GSD of the air concentration distribution
- GSD[Vi], the GSD of the minute ventilation distribution
- K, the target correlation factor between the air concentration and minute ventilation distributions

Since the GMs were found to not affect the output (the Exposure Ratio), the program was designed to vary GSD[C], GSD[Vi], and K only. For each of these parameters, the starting value, the step interval, and the number of steps were selected before the computer simulation was initiated.

The K values, as described in the Introduction, were set to range from –0.4 to +0.8 at step intervals of 0.2.

The GSD[C] values were set to range from 1.2 to 6.0 at step intervals of 0.4. GSD[C] values reported in the literature generally range from 1.5 to 3.0.[14,15,16] The GSD[C] range, in this simulation, extended higher than 3.0 due to the fact that the literature values are based on 8-hour averages, not 15-second averages. Since the effect of lengthening averaging time is to smooth the variance (and, therefore, lower the GSD), it is probable that the GSD[C] values based on 15-second averages can extend higher than 3.0.

The GSD[Vi] values also were set to range from 1.2 to 6.0 at step intervals of 0.4. Unfortunately, there is no published literature on which to base these numbers. An unpublished study by Peregrin Spielholz was examined to assist in the range selection.[17] In this study, the 15-second averages of heart rate for 5 drywallers were recorded during one workshift. The 15-second averages of minute ventilation were then estimated using the direct heart rate calibration curves for all 30 laboratory subjects (as will be discussed in Chapter 3). The GSD of each set of drywaller data was calculated 30 times (once per direct calibration curve from the laboratory phase). This method was used to ensure that the GSD was not underestimated.

The GSD of the drywallers' minute ventilation estimates ranged from 1.2 to 2.1. The GSD[Vi] range was set as high as 6.0 (much higher than the Spielholz[17] data) because the Spielholtz study was very restricted (5 subjects, each performing only 2 tasks), and the true GSD[Vi] values in work environments may be much higher.

The goal of the simulation was to estimate the average Exposure Ratio for each combination of the three input parameters (K, GSD[C], and GSD[Vi]); each specific combination of these parameters will be considered a "Condition".

In order to accomplish this, the program was designed with a series of four nested loops, using the following hierarchy:

- For each K value from –0.4 to +0.8 (step interval= 0.2);
    - For each GSD[C] value from 1.2 to 6.0 (step interval= 0.4);
        - For each GSD[Vi] value from 1.2 to 6.0 (step interval= 0.4);
            - Calculate the average Exposure Ratio (described below).

Calculating the average Exposure Ratio for each Condition consisted of averaging the individual Exposure Ratios calculated over 4990 iterations of the following steps:

- Generate new air concentration and minute ventilation data sets using the current Condition's GSD and GM values for those variables (described below);
- Calculate the actual correlation coefficient between air concentration and minute ventilation (R);
- Calculate the P-VWA and TWA based on the data sets;
- Calculate the Exposure Ratio.

For each iteration within each Condition, 15-second air concentration and minute ventilation values for a simulated work shift were generated using the current Condition's GSD and GM values for those variables. The goal was to generate a list of 1920 values, representing 15-second averages over an 8-hour workday, for each variable. These variables were designed to be correlated by a value close to the target correlation coefficient (K) for the current Condition.

To accomplish this, two lists were generated, each consisting of 1920 randomly selected numbers from a Normal Distribution [0,1]. One list was termed the Random List and the other was called the Error List. A third list, called the Data List, was calculated by Equation 11. This equation generated a list that was correlated to the Random List, with a correlation coefficient of K.

$$\text{Data List} = K \cdot \text{Random List} + (1 - K) \cdot \text{Error List}$$

Equation 11: Generation of a Data List that is correlated to the Random List via the correlation coefficient, K.

A list for the 15-second air concentration values was generated by using the GM[C] and GSD[C], and the Data List, as shown in Equation 12.

$$\text{C List} = \text{Exp}(\ln GM[C] + \ln GSD[C] \cdot \text{Data List})$$

Equation 12: Equation for generating the list of 15-second averages of air concentration.

A list for the 15-second minute ventilation values was generated by using the GM[Vi] and GSD[Vi], and the Random List, as shown in Equation 13.

$$\dot{V}i \text{ List} = \text{Exp}(\ln GM[\dot{V}i] + \ln GSD[\dot{V}i] \cdot \text{Random List})$$

Equation 13: Equation for generating the list of 15-second averages of minute ventilation.

Note that although the Data List and the Random List were correlated by a factor of K, these two lists were exponentiated in order to create the C List and the $\dot{V}i$ List. As a result, the C List and the $\dot{V}i$ List were correlated by a factor, R, that (1) differed slightly from K, and (2) depended on the two GSDs. Because there was a unique R for each

Condition, a table was created to display the average and range of R values for each K value. This table was designed to serve as a reference when examining the graphs of the simulation results (which are identified with the more simple K values).

The TWA for each iteration within each Condition was calculated by summing the list of air concentration values (the C List), and dividing it by the number of samples, as per Equation 14.

$$TWA = \frac{\sum C\ List}{1920}$$

Equation 14: Calculation of the TWA for each iteration.

The P-VWA for each iteration within each Condition was calculated by taking the dot product of the $\dot{V}i$ List and the C List, and dividing it by the sum of the $\dot{V}i$ List, as per Equation 15.

$$P\text{-}VWA = \frac{\sum(\dot{V}i\ List \cdot C\ List)}{\sum \dot{V}i\ List}$$

Equation 15: Calculation of the P-VWA for each iteration.

The Exposure Ratio was then calculated by dividing the P-VWA by the TWA.

The computer stored the values of K, R, GSD[Vi], GSD[C], and the average Exposure Ratio, for every Condition. These values were then written to a file and imported into Excel for analysis.

RESULTS

Table 1 displays the mean and range of the R values (the actual correlation between the generated air concentration and minute ventilation data) for each K value (the target correlation).

Table 1: Mean and range of the R values for each K value.

K	Mean (R)	Range (of R)
0.2	0.14	0.09-0.24
0.4	0.40	0.27-0.55
0.6	0.68	0.42-0.83
0.8	0.86	0.51-0.97

Figures 2 through 5 show the Exposure Ratio versus the GSD[P] for all the K values tested. The results for K= +0.2 and –0.2 have been displayed on the same graph (Figure 2), as have the results for K= +0.4 and –0.4 (Figure 3). Figures 4 and 5 show the results for K= +0.6 and +0.8, respectively.

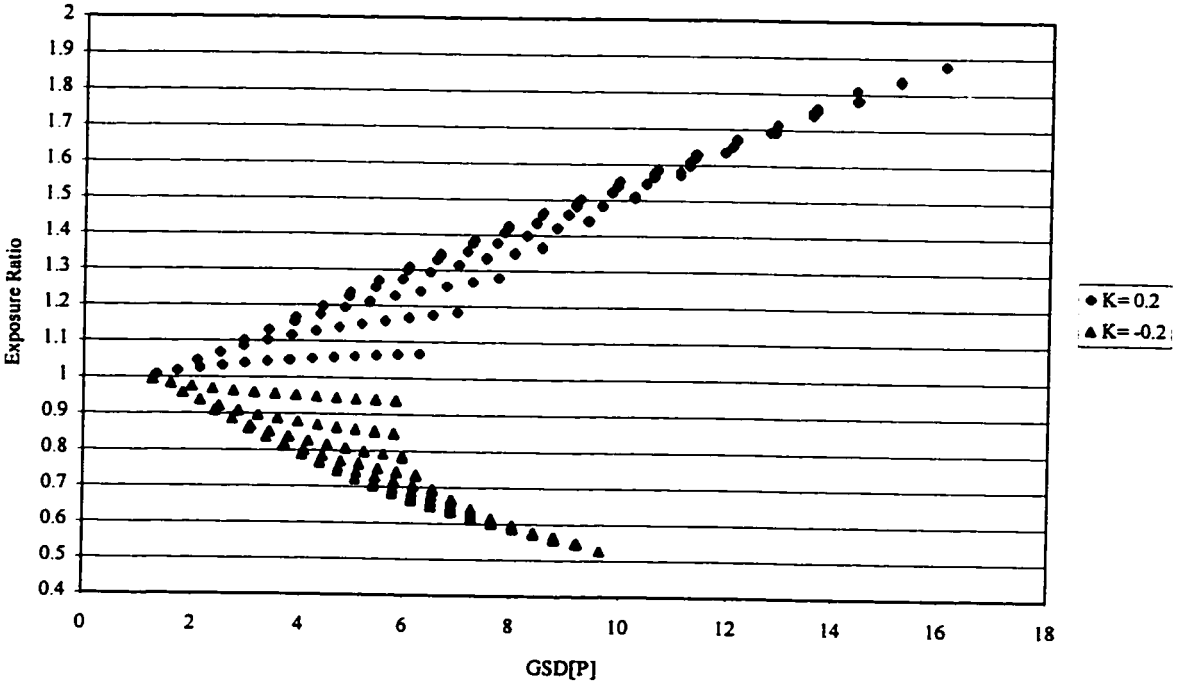


Figure 2: Exposure Ratio versus GSD[P] when K= 0.2 and –0.2.

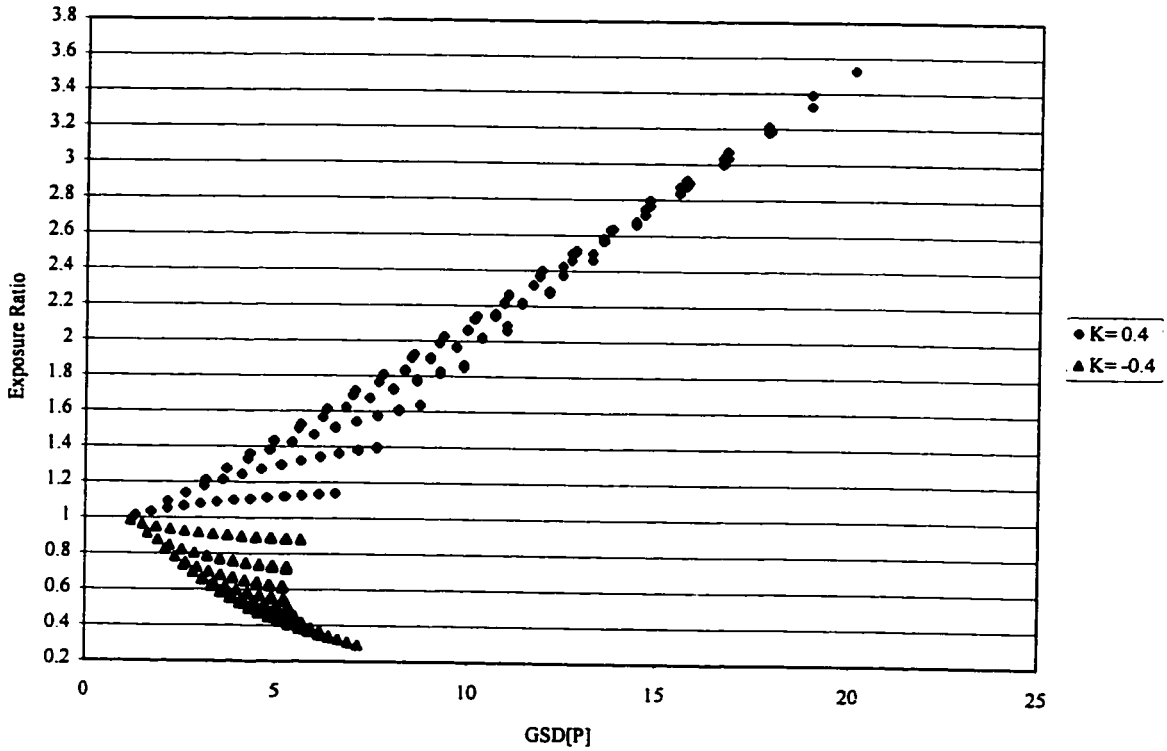


Figure 3: Exposure Ratio versus GSD[P] when K= 0.4 and –0.4

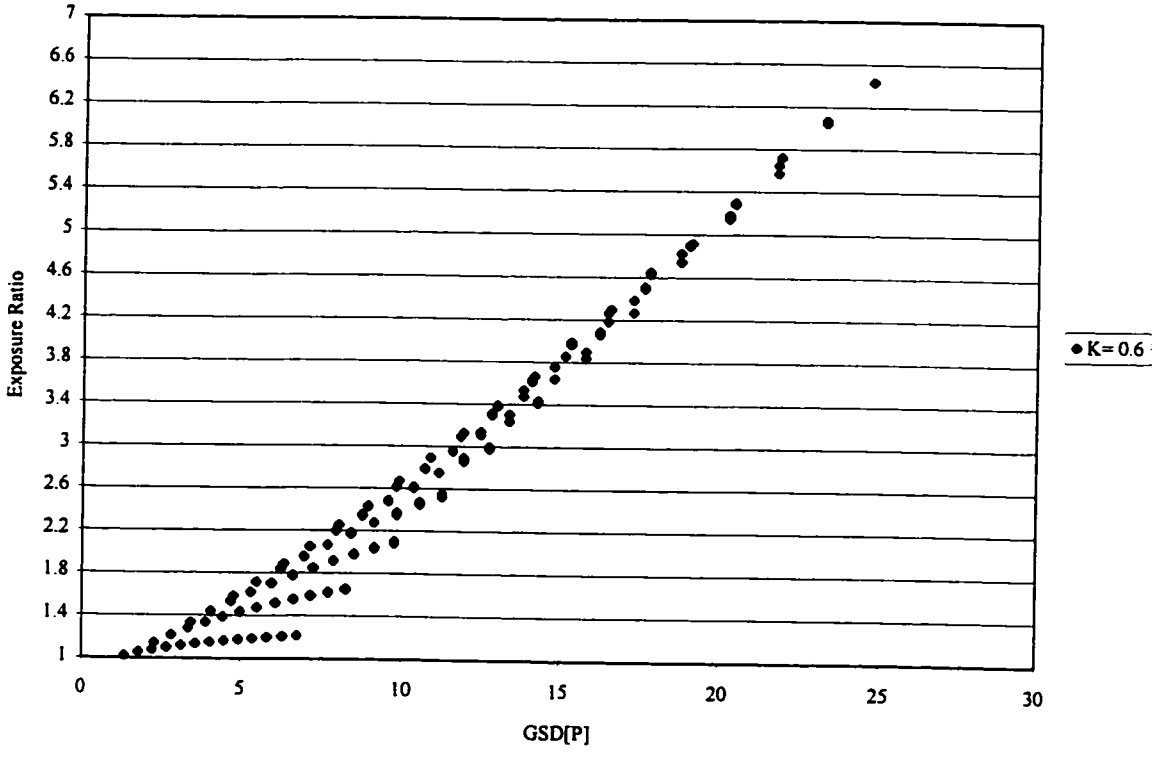


Figure 4: Exposure Ratio versus GSD[P] when K= 0.6.

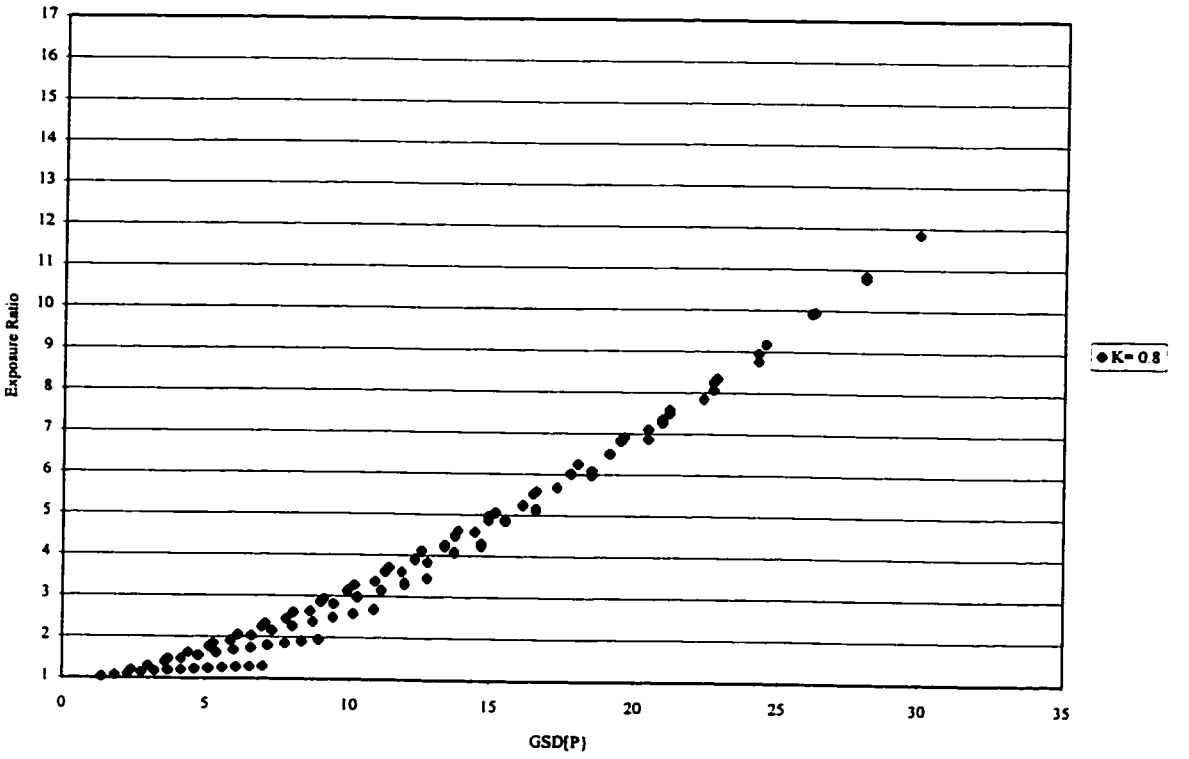


Figure 5: Exposure Ratio versus GSD[P] when K= 0.8.

The relationship between positive and negative correlation values, shown in Figures 2 and 3, reflect the nature of Equation 10. Within the square root of Equation 10, there are three summed components: (1) the square of the GSD[C], (2) the square of the GSD[Vi], and (3) the product of the GSD[C], the GSD[Vi], and R. When the correlation is positive, then this third component will increase the value under the square root. When the correlation is negative, however, this component will be subtracted from the squares of the GSDs; the effect is to compress the data along both axes.

Two interesting features of the data for positive correlation values in Figures 2 through 5 are (1) behavior of the Exposure Ratio at low GSD[P] values and (2) the change in the shape of the curve as the correlation becomes larger.

The behavior of the Exposure Ratio at low GSD[P] values is worth describing further. When either GSD[C] or GSD[Vi] was low (specifically 1.2, 1.6, and 2.0), the relationship between the Exposure Ratio and the GSD[P] appeared be more shallow than when GSD[C] and GSD[Vi] were both higher than 2.0.

As K increased, the data became steeper and the overall curve moved from quasi-linear to exponential. The change in the shape of the data as K is increased can be seen clearly in Figure 6; this graph shows all the data along with the best-fit power curves. Note that because of the behavior of the Exposure Ratio at low GSD[P] values, these best-fit power curves are not very useful (particularly at low GSD[P] values); they were included to provide a visual comparison only.

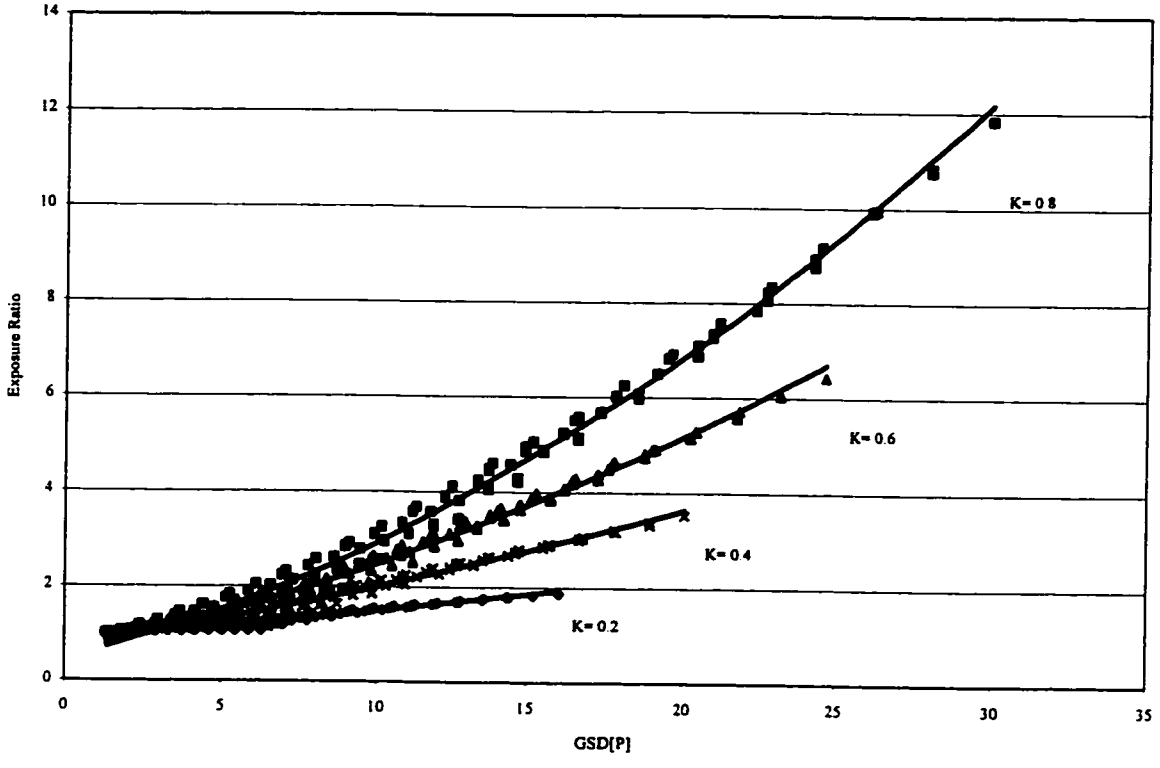


Figure 6: A graph of all data points generated from positive K values. Exposure Ratio plotted versus GSD[P]. Polynomial curves (order=2) fitted to each set of data (grouped by K value).

Figures 7 through 13 display the data in a way that is a little easier to interpret. Essentially, two of the three variables were "fixed". First, the GSD[C] values were fixed (at 1.2, 1.6, 2.0, 2.4, 2.8, 4.0 and 6.0) by plotting each value on a separate graph. Second, the K values were fixed by plotting data for each K value with a separate curve. The data points along each of these curves, therefore, represent the only changing variable: the GSD[Vi] values.

The graphs are presented in order of increasing fixed GSD[C] values. A bold line has been added (to the first five graphs) which represents a Critical Exposure Ratio of 1.2.

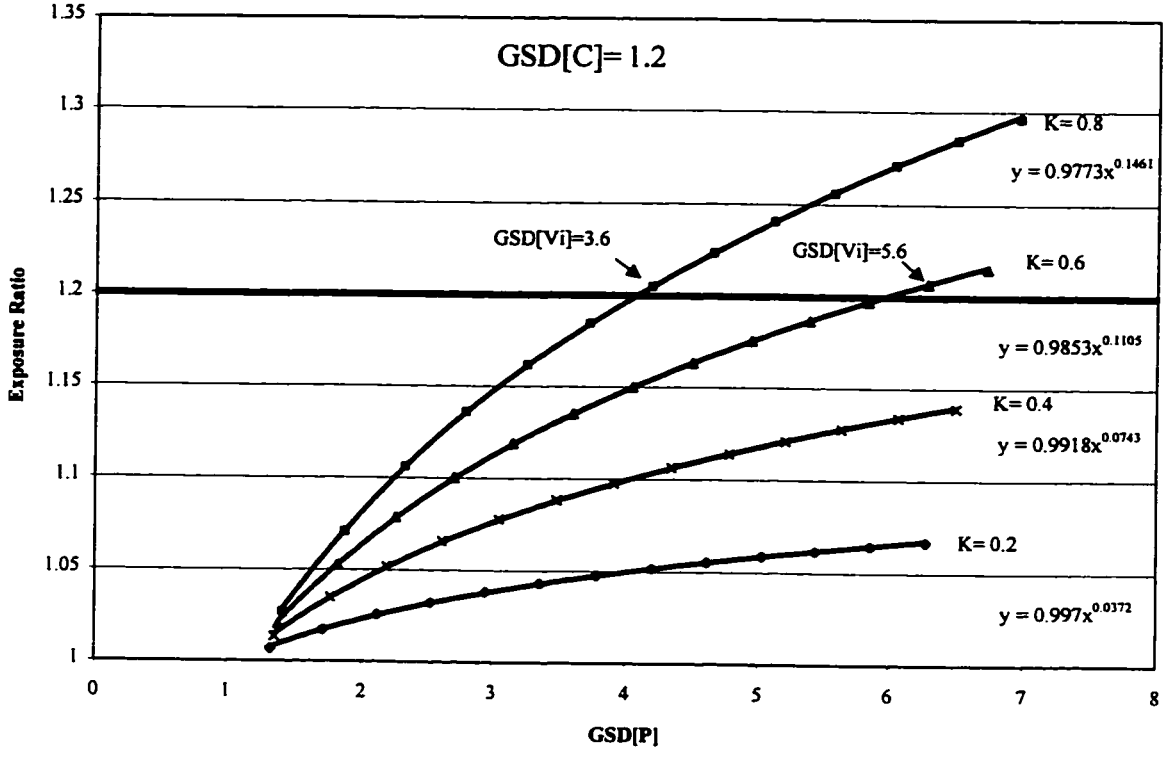


Figure 7: Exposure Ratio versus GSD[P] when GSD[C]= 1.2. Bold line represents Critical Exposure Ratio value.

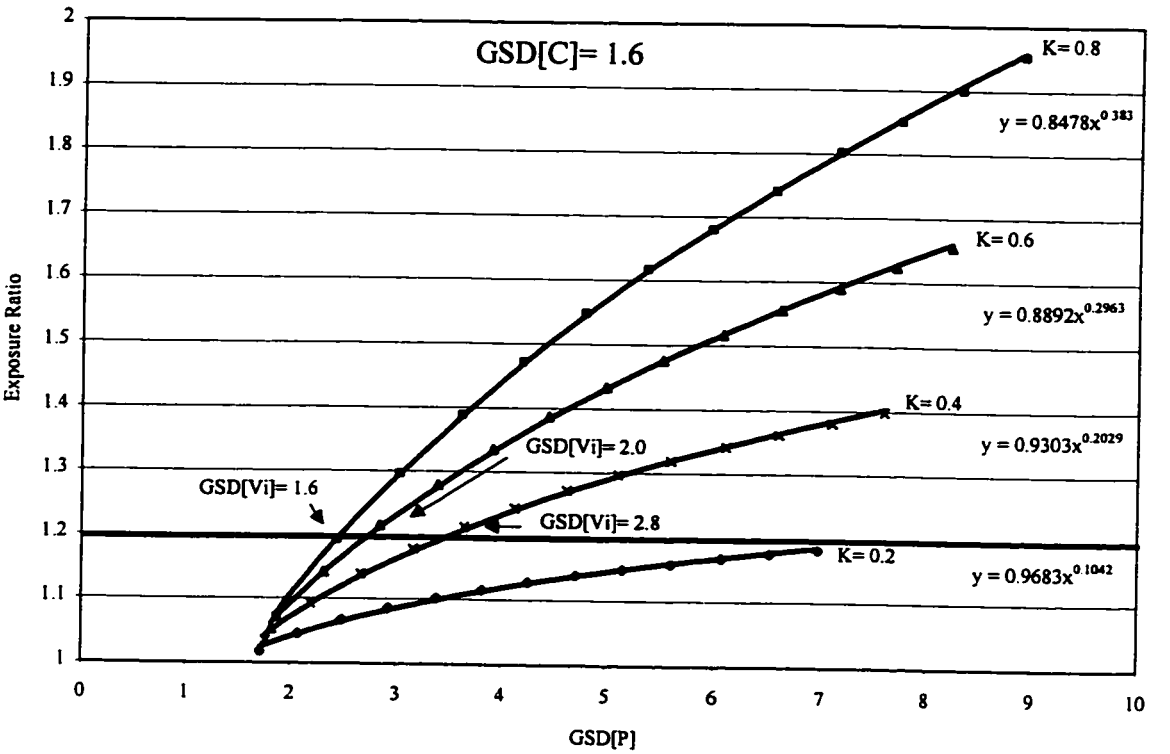


Figure 8: Exposure Ratio versus GSD[P] when GSD[C]= 1.6. Bold line represents the Critical Exposure Ratio.

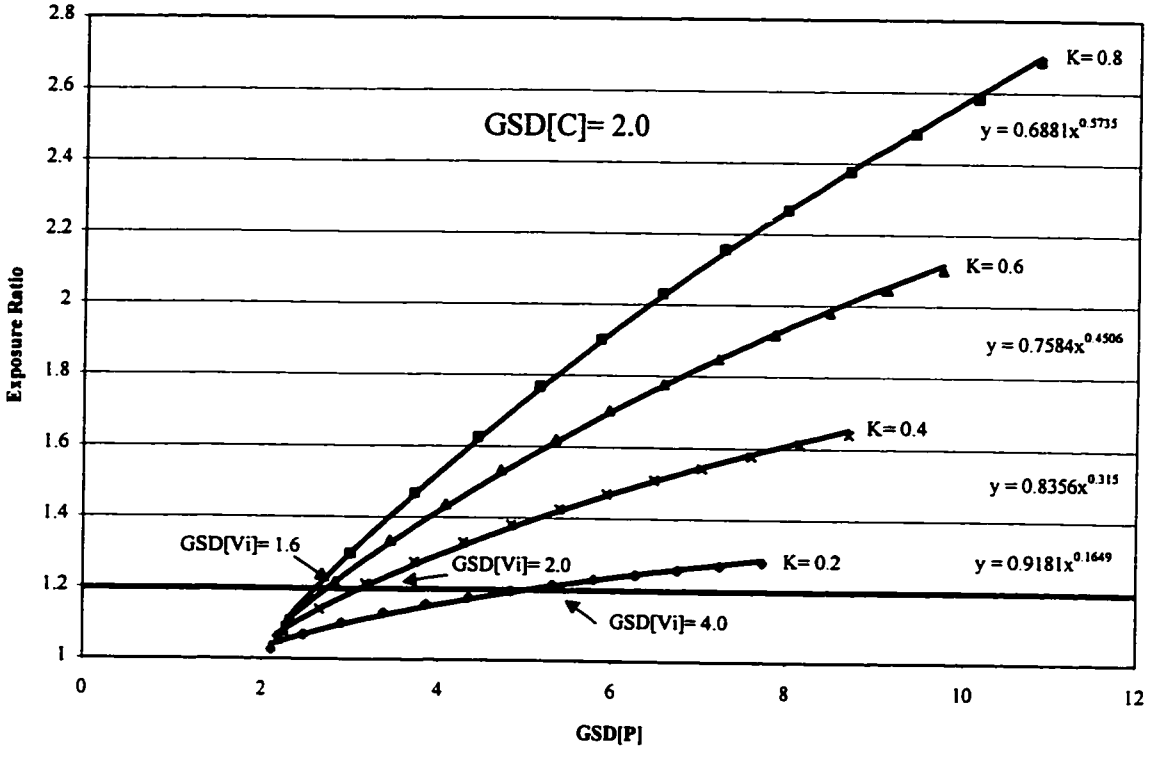


Figure 9: Exposure Ratio versus GSD[P] when GSD[C]= 2.0. Bold line represents the Critical Exposure Ratio.

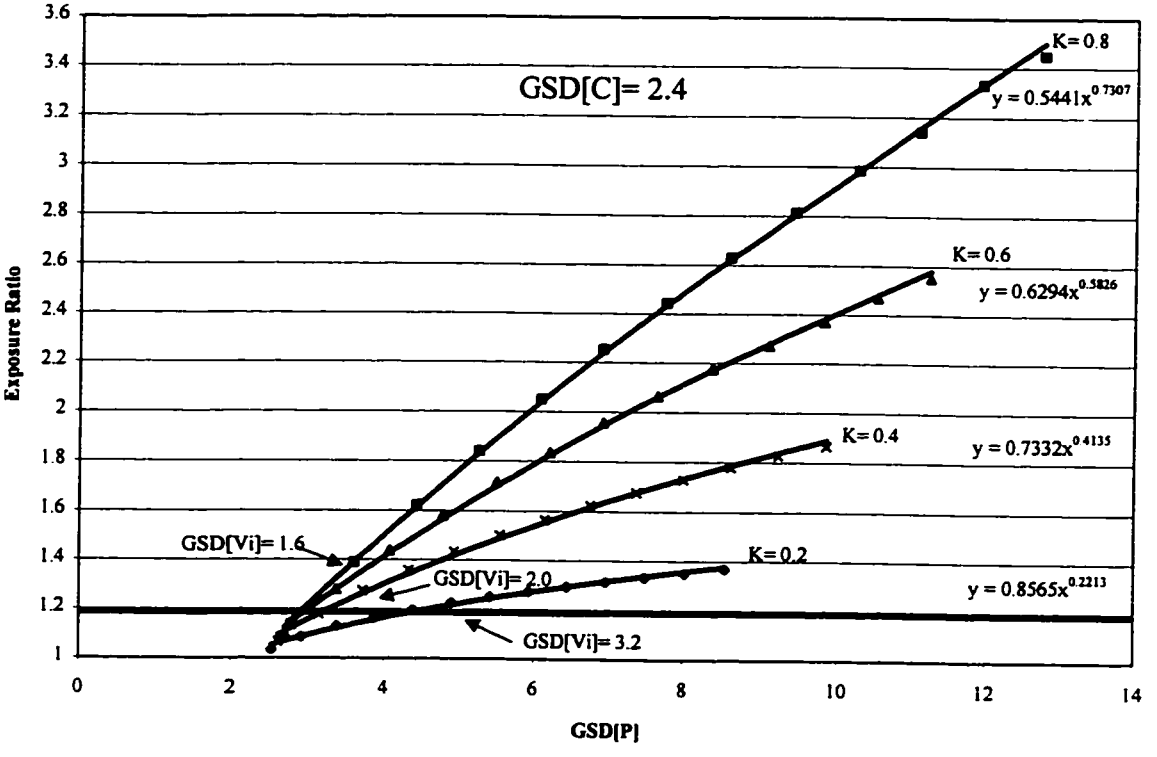


Figure 10: Exposure Ratio versus GSD[P] when GSD[C]= 2.4. Bold line represents the Critical Exposure Ratio.

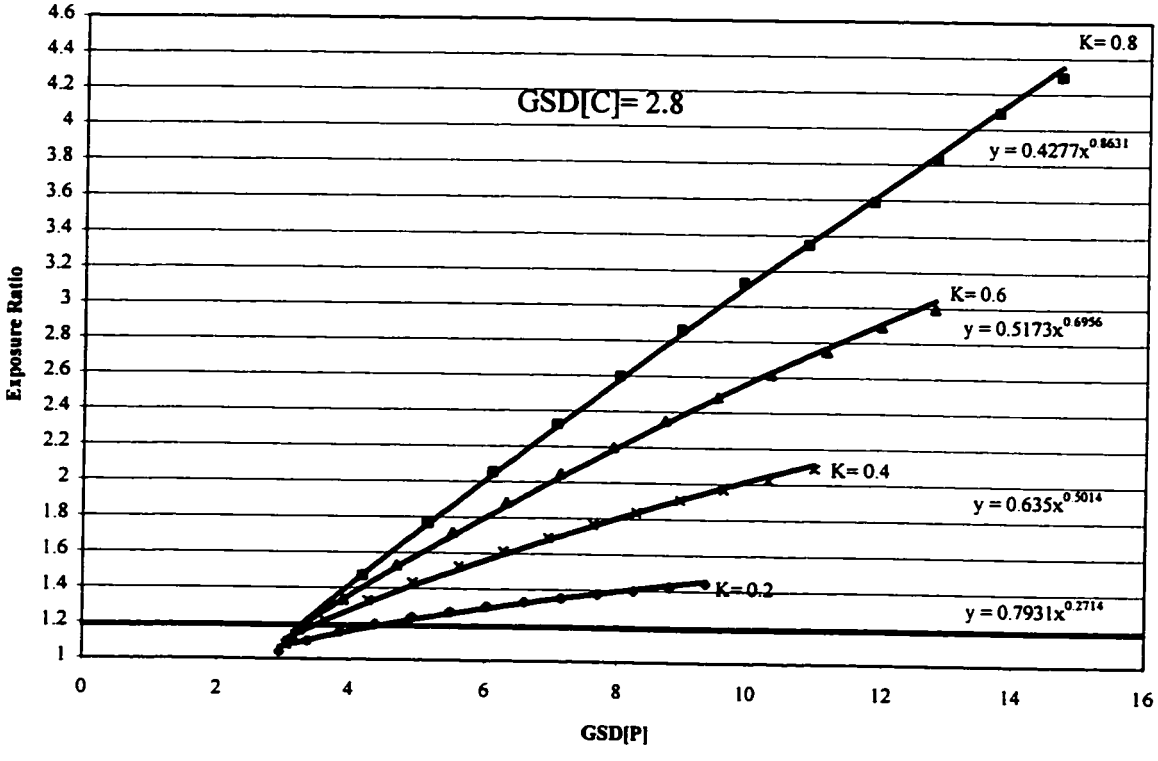


Figure 11: Exposure Ratio versus GSD[P] when GSD[C]= 2.8. Bold line represents the Critical Exposure Ratio.

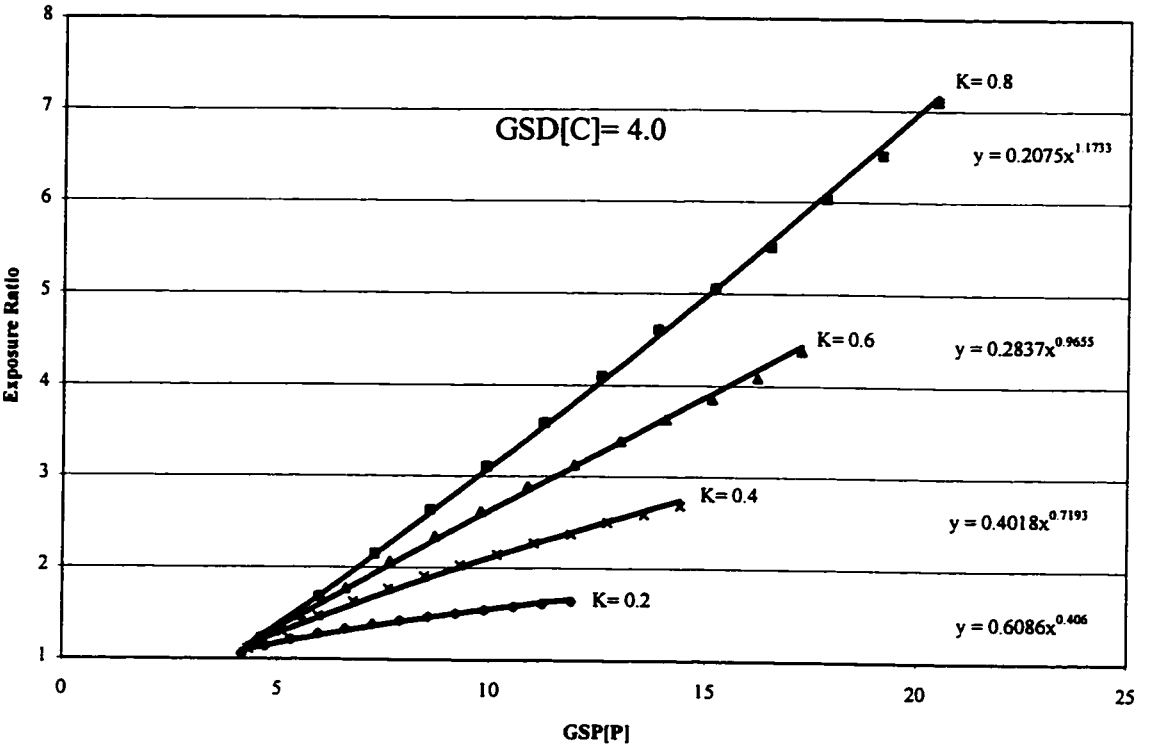


Figure 12: Exposure Ratio versus GSD[P] when GSD[C]= 4.0.

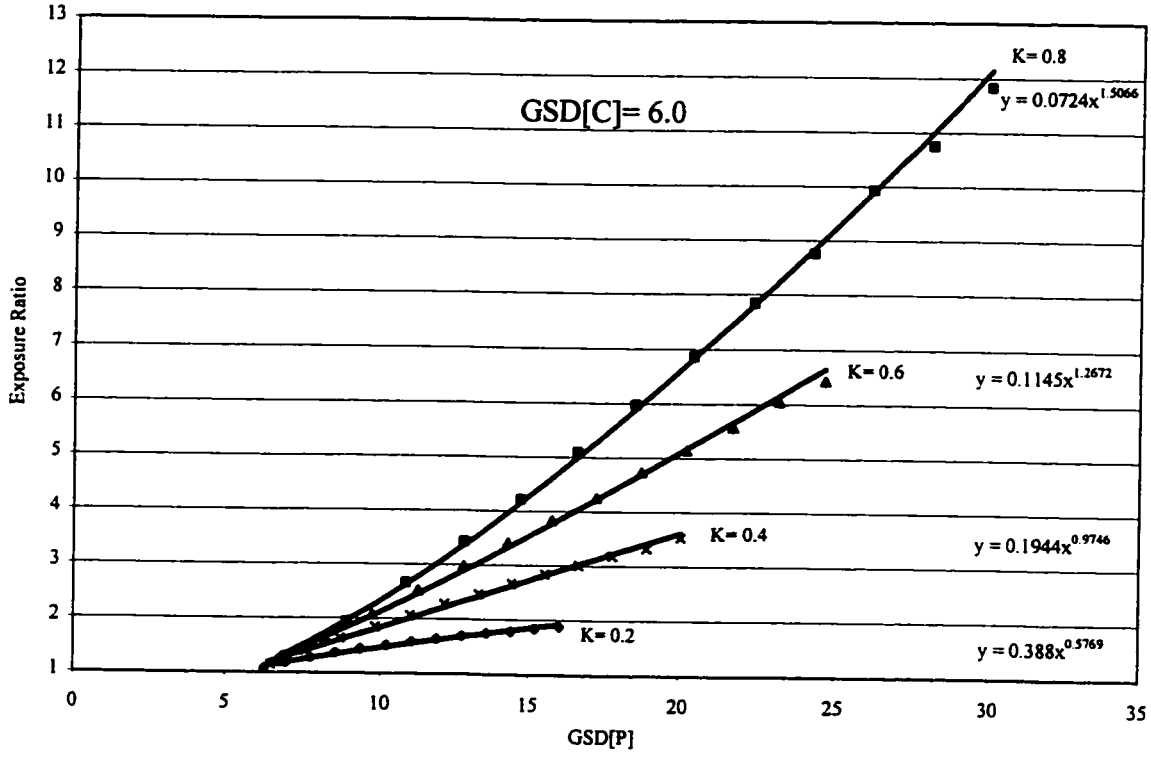


Figure 13: Exposure Ratio versus GSD[P] when GSD[C]= 6.0.

The simulation data is also presented in tabular format (see Tables 2 through 5); there is one table for each positive K value. These tables maximize the usage of this information for decision making by allowing a user to extract the exact Exposure Ratio values for any given set of GSD[C], GSD[Vi], and K. In addition, to assist in interpretation, all Exposure Ratios greater than 1.2 were shaded, and lines demarcating Exposure Ratios of 1.2, 2.0, 3.0, 4.0, 6.0 and 8.0 were added.

Table 2: Exposure Ratios when K= 0.2, for all combinations of GSD[C] and GSD[Vi].

	GSD[C]					**K= 0.2**							
GSD[Vi]	**1.20**	**1.60**	**2.00**	**2.40**	**2.80**	**3.20**	**3.60**	**4.00**	**4.40**	**4.80**	**5.20**	**5.60**	**6.00**
1.20	1.01	1.02	1.03	1.03	1.04	1.04	1.05	1.05	1.06	1.06	1.06	1.06	1.07
1.60	1.02	1.05	1.07	1.09	1.10	1.12	1.13	1.14	1.15	1.16	1.17	1.18	1.18
2.00	1.03	1.07	1.10	1.13	1.15	1.17	1.19						
2.40	1.03	1.09	1.13	1.17									
2.80	1.04	1.10	1.15										
3.20	1.04	1.12	1.17										
3.60	1.05	1.13	1.19										
4.00	1.05	1.14											
4.40	1.06	1.15											
4.80	1.06	1.16											
5.20	1.06	1.17											
5.60	1.06	1.17											
6.00	1.07	1.18											

>1.2

Table 3: Exposure Ratios when K= 0.4, for all combinations of GSD[C] and GSD[Vi].

	GSD[C]				K= 0.4								
GSD[Vi]	**1.20**	**1.60**	**2.00**	**2.40**	**2.80**	**3.20**	**3.60**	**4.00**	**4.40**	**4.80**	**5.20**	**5.60**	**6.00**
1.20	1.01	1.03	1.05	1.07	1.08	1.09	1.10	1.11	1.11	1.12	1.13	1.13	1.14
1.60	1.03	1.09	1.14	1.18									
2.00	1.05	1.14											
2.40	1.07	1.18											
2.80	1.08												
3.20	1.09												
3.60	1.10												
4.00	1.11												
4.40	1.11												
4.80	1.12												
5.20	1.13												
5.60	1.13												
6.00	1.14												

>1.2

>2.0

>3.0

Table 4: Exposure Ratios when K= 0.6, for all combinations of GSD[C] and GSD[Vi].

	GSD[C]					K= 0.6							
GSD[Vi]	**1.20**	**1.60**	**2.00**	**2.40**	**2.80**	**3.20**	**3.60**	**4.00**	**4.40**	**4.80**	**5.20**	**5.60**	**6.00**	
1.20	1.02	1.05	1.08	1.10	1.12	1.14	1.15	1.16	1.18	1.19				>1.2
1.60	1.05	1.14												
2.00	1.08													>2.0
2.40	1.10													
2.80	1.12													
3.20	1.14													>3.0
3.60	1.15													
4.00	1.16													>4.0
4.40	1.17													
4.80	1.19													
5.20														
5.60														>6.0
6.00														

Table 5: Exposure Ratios when K= 0.8, for all combinations of GSD[C] and GSD[Vi].

GSD[C] K= 0.8													
GSD[Vi]	**1.20**	**1.60**	**2.00**	**2.40**	**2.80**	**3.20**	**3.60**	**4.00**	**4.40**	**4.80**	**5.20**	**5.60**	**6.00**
1.20	1.03	1.07	1.11	1.14	1.16	1.18							
1.60	1.07	1.19											
2.00	1.11												
2.40	1.14												
2.80	1.16												
3.20	1.18												
3.60													
4.00													
4.40													
4.80													
5.20													
5.60													
6.00													

>1.2

>2.0

>3.0

>4.0

>6.0

>8.0

The data in Tables 2 through 5 are plotted, respectively, in Figures 14 through 17. The plots are three-dimensional and provide a feel for how the GSDs affect the Exposure Ratio. Each band on the graphed surface represents a range of Exposure Ratio values which correspond to the labels on the Y-axis. Note that the Y-axis in each graph has a different scale.

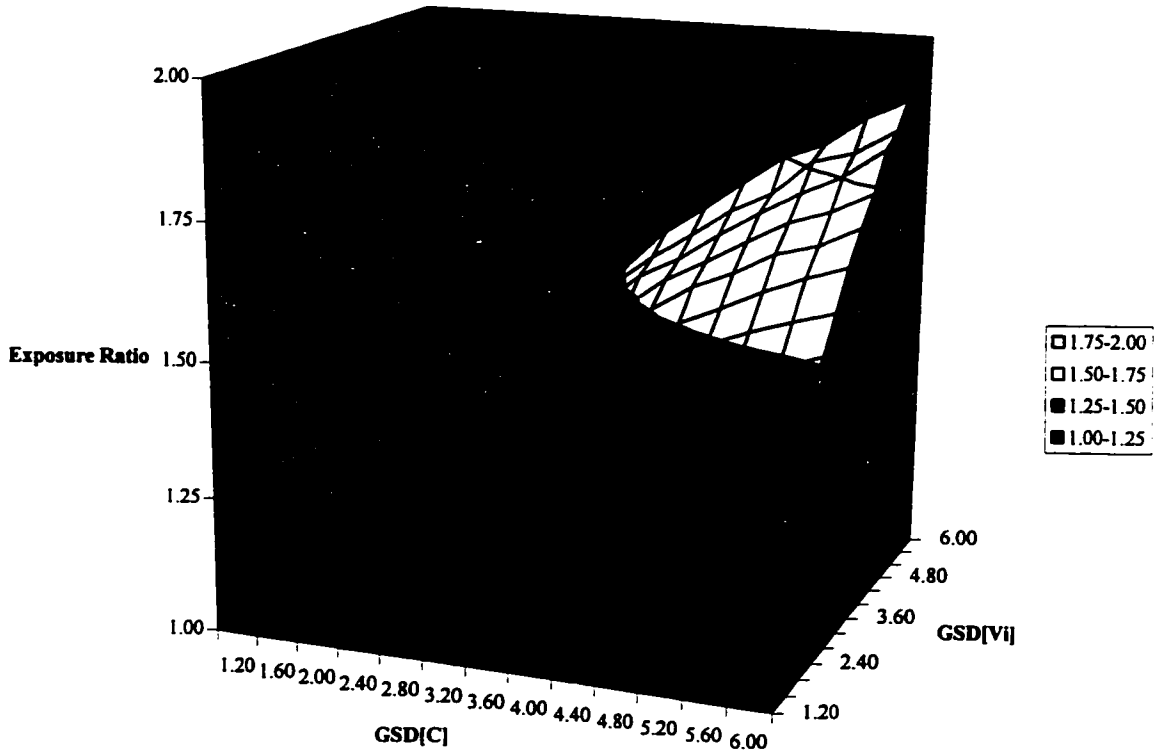


Figure 14: Three-dimensional plot of the Exposure Ratio values for K= 0.2.

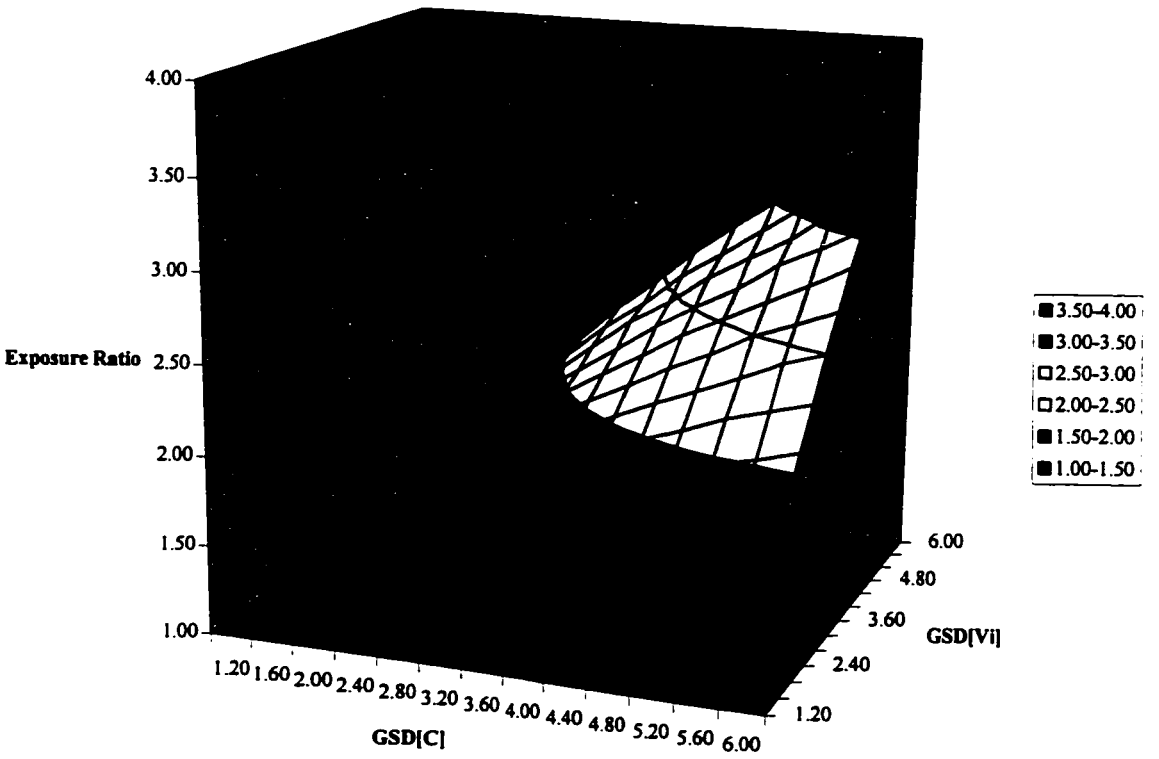


Figure 15: Three-dimensional plot of the Exposure Ratio values for K= 0.4.

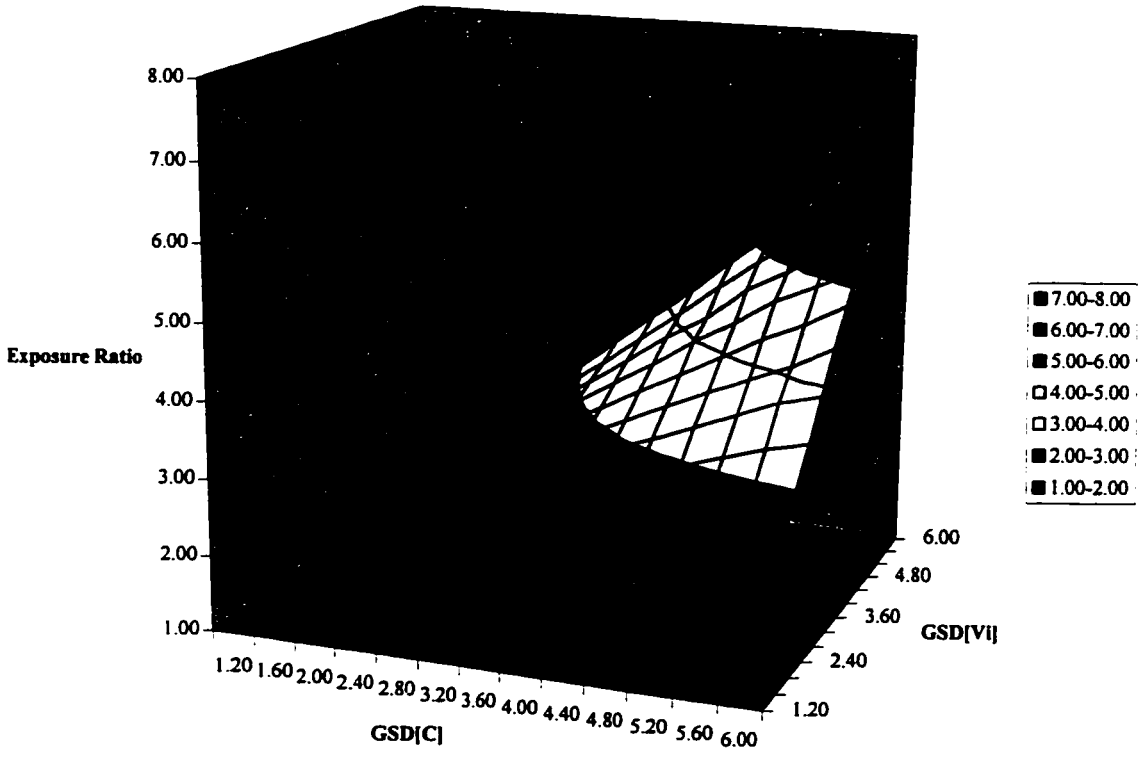


Figure 16: Three-dimensional plot of the Exposure Ratio values for K= 0.6.

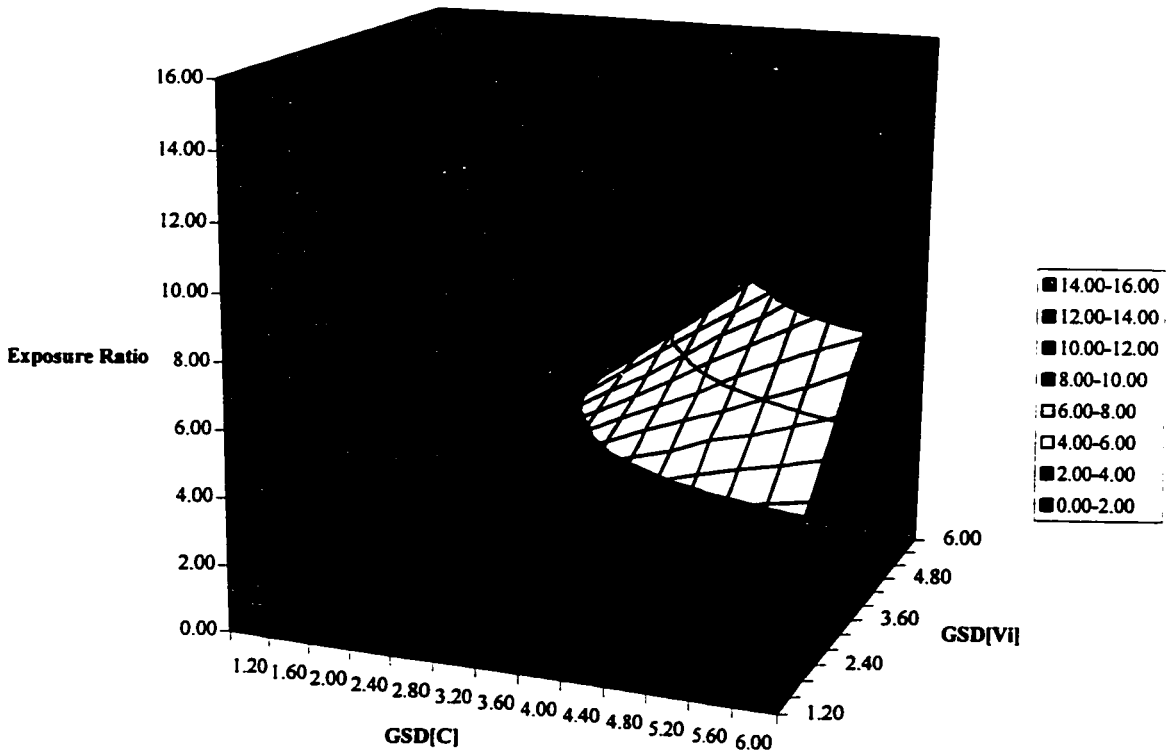


Figure 17: Three-dimensional plot of the Exposure Ratio values for K= 0.8.

DISCUSSION

The existence of a relationship between the Exposure Ratio and GSD[P] has been confirmed by this computer simulation. The relationship is complex, but can be summarized over the tested range of GSD and K values as follows:

- When K equals –0.4 (or –0.2), the Exposure Ratios tend to mirror the data when K is +0.4 (or +0.2) with a center around the Exposure Ratio value of 1, but with the range compressed.
- When K is positive, the larger the K value, the greater the Exposure Ratio for a given GSD[P].
- When K is greater than +0.4, the rise in the Exposure Ratio for a given combination of GSD[C] and GSD[Vi] is much more dramatic.
- For any given K, the data for all combinations of GSD[C] and GSD[Vi] values can be best fit with a second order polynomial curve; however, this curve does not fit the data well at the low GSD[P] values.
- When the data is restricted to a given K and a given GSD[C] (or GSD[Vi], see "Note", below), the data is best fit with a power curve. The higher the K value, the higher the exponential component of the best fit power curve.

Note: since GSD[C] and GSD[Vi] are treated identically in Equation 10 and the ranges tested are the same, Figures 7 through 13 can be inverted to "fix" a GSD[Vi] value while varying the GSD[C] values. To do this, simply swap GSD[C] for GSD[Vi] notations, and vice versa, wherever they appear in the Figure.

Tables 2 through 5 display the simulation data in a manner which is the most useful for decision making, particularly as the Conditions which exist in workplaces (specifically R, GSD[Vi] and GSD[C]) become better defined.

Instructions for using these tables are as follows:

1. Before examining the tables, try to decide on the following:

   a) What Critical Exposure Ratio value will be used for decision making?

   b) What is a good estimate range for GSD[C] based on available information?

   c) Can the possible range for the GSD[Vi] value be restricted based on known information (e.g. GSD probably greater than 2 but less than 4)?

   d) Can the possible range of correlation (K) be restricted based on known information (e.g. K greater than 0.4 is highly unlikely)?

2. For the tables corresponding to the included K values (from Step #1d), eliminate the cells in the tables that fall outside of the GSD[C] and GSD[Vi] ranges selected in Step #1b and 1c.

3. Compare the remaining cell values (after Step #2) to the chosen Critical Exposure Ratio (from Step #1a).

   a) If all the data points fall above the Critical Exposure Ratio, then using a PSP would be wise in order to prevent underestimation of exposure via TSPs.

   b) If all the data points fall below the Critical Exposure Ratio, then the costs of using a PSP would not justify the benefits of more accurately estimating exposure versus a TSP.

c) If some data points fall above the Critical Exposure Ratio and other fall below it, then one's best judgement must be used.

The results of this computer simulation demonstrate that the correlation between air concentration and minute ventilation (even at K= 0.2) could result in significant under-estimation of a worker's exposure if a TSP is used for sample collection. In addition to the potential implications regarding an individual worker's exposure, there are population implications as well: ignoring correlation could lead to misclassification of exposure. This means that workers who are exposed at a high level could be placed in a lower exposure group. When epidemiological studies are performed based on exposure data and worker health outcomes, this misclassification of exposure could lower the ability to identify exposure-response relationships.

Take, for example, the following hypothetical situation: Workers A and B are performing different tasks, but have the same average air concentration and minute ventilation over their 8-hour shifts. Worker A's minute ventilation is correlated to the air concentration in his/her breathing zone by a correlation coefficient of 0.4, while Worker B's minute ventilation is not correlated to his/her local air concentration. Based on the computer simulation, Worker A's actual exposure may be three times Worker B's exposure. If TSPs are used to estimate the workers exposure values, the exposure estimates for both workers would be the same. Biological monitoring samples, on the other hand, would reflect the difference in actual exposure. If one were trying to correlate the TSP-based exposure values to the biological monitoring results, the additional error caused by this misclassification of exposure would make the task much more difficult (particularly in terms of the power of any statistical tests).

Droz *et al.* (1991)[18] found a significant relationship between air monitoring and biological monitoring results, but point out that there was significant scatter, or variability, as well. For example, they looked at the agreement between air monitoring and biological monitoring results for two chemicals: mercury and perchloroethylene. The

workers measured for perchloroethylene exposure worked in the dry cleaning industry, while those measured for mercury exposure worked in manufacturing industries (specifically, luminous signs, batteries, and the production of acetaldehyde). Droz *et al.* categorized workers into two exposure groups (low and high) based on each method. There was excellent agreement between air monitoring and biological monitoring samples for perchloroethylene and poor agreement for mercury. The workers exposed to mercury would therefore have a greater chance at being misclassified with regard to exposure. While the authors list three factors that could explain why there is poor agreement for chemicals such as mercury, they do not take into account the potential of correlation between air concentration and minute ventilation. Correlation would be more likely in an industry where several tasks are performed; this could have contributed to the difference between the perchloroethylene and mercury results, since workers in manufacturing industries are more likely to have tasks requiring changes in physical workload.

When computer modeling is used to try to relate exposure data to biological monitoring data, the programmers try to account for the factors that contribute to variability. Droz *et al.* (1989)[19] performed such a computer model; to their credit, the model accounted for variation in air concentration and physical workload. However, when they applied this model they did not account for potential correlation between the physical workload and air concentration; if correlation exists in work environments, then the estimated variability of both exposure data and biological monitoring data would be underestimated.

The effect of ignoring possible correlation when using computer models is more obvious when a study by Thomas *et al.*[20] is examined. Their goal was very specific; they wanted to estimate the biological monitoring values (corresponding to BEIs) that would correlate to worker's exposure at the TLV. During the simulated work shifts, the air concentration was set to a constant value (the TLV). As they mention in the discussion, "if some model parameters have large intrapopulation variability, then deterministic models may produce

estimates of biological indicators that do not protect the majority of the worker population despite the protection afforded through the TLV."[20] If correlation exists, then it would be one such example of intrapopulation variability, and these researchers fall into the very trap that they suggest avoiding.

CONCLUSIONS

Correlation between air concentration and minute ventilation can dramatically affect a worker's inhaled dose. This dose is related to the magnitude of the correlation, as well as the geometric standard deviation of both the air concentration and minute ventilation. The computer simulation used in this experiment showed that these three factors can be expressed as the geometric standard deviation of the product, or GSD[P], which can then be related to the Exposure Ratio (P-VWA/TWA).

The relationship between the Exposure Ratio and GSD[P], regrettably, is not simple. In order to graph the relationship meaningfully, two of the three variables in the equation needed to be fixed. However, the general trend is evident even from the most basic of graphs: as any of the three variables becomes larger (K, GSD[C], and GSD[Vi]), the Exposure Ratio increases. In addition, at K values greater than 0.4, the Exposure Ratio becomes much more sensitive to changes in the GSDs.

The table below shows the lowest GSD combination for each positive K value for which the Exposure Ratio exceeded the Critical Exposure Ratio of 1.2:

Table 6: Lowest GSD combination for each positive K value.

K	GSD[C]	GSD[Vi]
0.2	2.8	2.4
0.4	2.0	2.0
0.6	2.0	1.6
0.8	1.6	1.6

These GSD combinations are well within possible limits in work environments. The implication is that if correlation between air concentration and minute ventilation exists, then it is very likely that we are underestimating exposure for that segment of the work force when traditional air sampling methods are used. The result of misclassifying exposure in this way has individual as well as population implications.

Finally, in the quest to understand the relationship between air monitoring and biological monitoring samples, the main focus has been to characterize the variability of air monitoring samples and biological kinetics (relating to the distribution and elimination of the chemical within the body). There has been no acknowledgement of the potential contribution that correlation between air concentration and minute ventilation could have on the variability of inhaled dose.

CHAPTER 3: LABORATORY VALIDATION OF A PHYSIOLOGIC SAMPLING PUMP

INTRODUCTION

The physiologic sampling pump (PSP) evaluated in this dissertation was developed by Yost, Hart, and Lopez and is described in detail in the manuscript contained in Appendix A. In brief, a modified GilAir 5 sampling pump is controlled by an analog signal from a PSCU (Physiologic Sampler Controller Unit), which is a small battery-powered datalogging computer (micro-controller). This micro-controller can be programmed in a simple Basic language to control the pump rate of the modified GilAir 5. The pump rate varies in proportion to an estimate of the minute ventilation of the person wearing the PSP. The PSCU in this laboratory experiment was designed to accept a variety of physiologic input signals from the following devices: (1) a pneumotachometer (providing a signal for true instantaneous inspiratory flow rate), (2) a cardiotachometer (providing heart rate data), and (3) a respiratory inductive plethysmograph (providing two signals which correspond to the cross-sectional areas of the chest and abdominal cavities). This design allowed for the evaluation of different indirect methods for estimating the minute ventilation and the comparison of these methods to actual minute ventilation under conditions of varying leg work rate.

Using these three input signals, five methods were used for estimating the minute ventilation (hereafter referred to as "Methods"). Each Method, along with the input signal from which it was derived and its purpose for being included in the study, is listed below:

Ventilation-based Methods:

1. Pneumotachometer Method. The PSP control signal was derived via the PSCU by integrating the pneumotachometer signal in 15-second batches. This method was designed to provide a reference for how well the PSCU processing performs compared to the raw signal (Primary Standard) and is abbreviated "Pneumo Method".

2. Respiratory Inductive Plethysmograph Method. The PSP control signal was derived from the respiratory inductive plethysmograph signals and a calibration curve relating these signals to minute ventilation. This method was designed to provide a comparison for heart rate-based estimates (although not deemed practical for use in the field at this time), and is abbreviated "RIP Method".

Heart rate-based Methods:

1. Direct Heart Rate Method. The PSP control signal was derived from the heart rate signal and an individually determined steady-state calibration curve of heart rate versus minute ventilation. This method was designed to provide a means of (A) assessing the potential of using heart rate as a surrogate for minute ventilation with this particular PSP design, and (B) comparison for the Indirect Heart Rate Method. This method is abbreviated "Direct HR Method".

2. Indirect Heart Rate Method. The PSP control signal was derived from the heart rate signal and an original predictive equation relating minute ventilation to heart rate. This method was the primary focus of this study, and is abbreviated "Indirect HR Method".

3. Satoh's Heart Rate Method. The PSP control signal was derived from the heart rate signal and a predictive equation relating minute ventilation to heart rate that was published by Satoh *et al.*[4] This method was designed to provide a comparison for the predictive equation used in the Indirect Heart Rate Method, and is abbreviated "Satoh's HR Method".

The results of all of these Methods were calculated by the PSCU (based on input signals) and provided estimates of minute ventilation. In order to evaluate these Methods meaningfully, the "true" (or reference) minute ventilation was necessary. For this purpose, a Primary Standard was also generated by processing (via positive half-wave rectification) and integrating the pneumotachometer signal; the resulting signal corresponded to inspired minute ventilation. The only difference between the Primary Standard and the Pneumo Method was the signal integration location (Dataq Software for the Primary Standard and the PSCU for the Pneumo Method). Differences between the Primary Standard and the Pneumo Method, therefore, could be specifically attributed to the PSCU's processing of the pneumotachometer signal.

One notable PSP design issue is the balance between the averaging time for the input signals and the update period for the output voltage to the pump. The shorter the averaging period, the better the PSP will be able to capture peak exposures; unfortunately, this must be balanced against the error introduced by "wind-up time". This particular error is inherent in a mechanical pump that changes flow rate because of inertia in the mechanical system; there must be a transient period during which the pump rate approaches the steady-state pump rate for a given input voltage. To minimize this error the update period between changes in flow rate should be much longer than the time constant for the pump to reach a new equilibrium condition. The more frequently the pump rate is altered, the more error will accumulate. A 15-second averaging period for the input signal was selected for two reasons: (1) it would be long enough to limit the error associated with the pump response transients just mentioned, and (2) research

performed by Mermier *et al.*[10] concluded that 15-second averages of heart rate were closely linked to minute ventilation (see Chapter 1).

The hypothesis underlying the research discussed in this chapter was that over a 40-minute exercise session, each Method could be used to predict minute ventilation with an average RPD of less than 20% percent and a RPD standard deviation of less than 10%, for leg work rates between 0 and 120 Watts.

A laboratory simulation was performed in order to collect human subject data to send to the PSP. The human subjects bicycled on an ergometer in order to simulate workers changing work rates on the job. Arm exercise, or combinations of arm and leg exercise, were not examined.

This chapter will first present the methodology for each Method's calibration equation development and the laboratory simulation. Second, it will detail how the RPDs were calculated, which are necessary for evaluating the hypotheses. Third, the method for performing concordance analysis will be described, and will assist in comparing each Method to the Primary Standard and showing specifically how they differ. Finally, conclusions will be drawn and the implications for the future of physiologic sampling pumps will be discussed.

METHODS

Potential subjects were recruited via fliers posted on the Environmental Health bulletin board and outside a busy café close to the laboratory. Interested parties were interviewed over the phone in a standard manner (see Appendix C). They were then scheduled for either one or two sessions and were sent a confirmation notice and a map.

Seven male and seven female subjects were scheduled for one testing session. A separate group of eight males and eight females were scheduled for two testing sessions, at least one week apart.

To track each subject session and ensure that each step was carried out, a checklist was used. The checklist included the following processes: preparation, subject handling, data collection, data processing, and sample collection. A blank checklist is included in Appendix D, and a summary of the notations made during each session is provided in Appendix E; these notations are important because they document when procedures were altered in any way.

When subjects arrived they were asked to read and sign the human subject consent form (see Appendix F). The project was also described to the subjects verbally, to provide them with a better feel for the purpose of their session. Subjects then filled out the subject history macro (SubjHx, Appendix G), which was reviewed by the author before proceeding further. The subject information collected via this macro is detailed in Appendix H.

Subjects were then asked to remove their shirts in order to attach the monitoring equipment. First, a band containing three electrodes, used to detect each heart beat, was strapped around their chests, just below the breast area. Water was applied to each electrode face to enhance conduction of the signal.

Second, each subject was fitted with a stress-test shirt. Stress-test shirts are tank-top shaped and made of an elastic mesh that stays tight against the body. The two RIP bands were placed over this shirt, one located around the chest and one located around the abdomen. The bands were attached to the shirt at four points (two in front, two in back) using straps made of webbing and velcro. These velcro straps prevented vertical shifting of the bands, without interfering with their expansion/contraction due to breathing movements.

Finally, the mouth unit was attached to the subject's head (see Figure 18). This mouth unit was composed of a PVC T-fitting, where one end was connected to a plastic mouthpiece and the opposite end to a tygon tube leading to a saliva-collection bottle.

Perpendicular, and pointing up when the subject had the mouthpiece in his/her mouth, was the pneumotachometer. Webbing and elastic were used to strap this mouth unit to the subject's head. One band wrapped around the back of the subject's neck and was secured to the top of the T-fitting on the opposite side of the stem from the subject. After this band was tightened, a second band, made of elastic and originating in front of the stem of the T-fitting, was placed over the top of the subject's head and attached behind the neck to the first strap with velcro. This point of attachment was reinforced with a simple barrette. Two or three more barrettes were used to secure the second band to the subject's hair.



Figure 18: Mouth unit attached to a subject's head.

Each Method used to estimate minute ventilation required a calibration equation. These calibration equations were necessary for the PSCU program to estimate minute

ventilation based on the instrument signals. The procedures for generating the calibration equations for each Method are presented below.

PNEUMO METHOD CALIBRATION

Prior to a subject's arrival, the pneumotachometer was heated for at least 30 minutes and then calibrated using a glass rotameter as a secondary standard (which was calibrated against a primary standard, a spirometer). A three-point calibration curve was generated, relating pneumotachometer voltage to airflow rate. During calibration, the air was set to flow in the direction the air flows during subject inspiration; the pneumotachometer was positioned such that this airflow direction produced a positive voltage in the pneumotachometer.

RIP METHOD CALIBRATION

The RIP calibration protocol involved two steps. For the first step, subjects were asked to perform three isovolume maneuvers while sitting on the ergometer. The maneuvers consisted of the subjects plugging their noses and holding their breath while moving their abdomens in and out several times. Each maneuver was performed at a different lung volume: for the first maneuver, the subjects breathed out briefly first (for a lung volume volume lower than FRC), for the second they simply plugged their noses and closed their mouths at a neutral state (to approximate FRC), and, finally, for the third, they inhaled briefly first (for a lung volume greater than FRC). The band signals during each maneuver were recorded using hardware and software from Codas; one file per isovolume maneuver was generated.

The isovolume maneuvers aid in determining the relationship between the band signals and the cross-sectional area of the torso. Theoretically, if the lung volume does not change, then any negative change in the cross-sectional area of the abdomen (caused by the subject moving the abdomen in) would cause a corresponding positive change in the cross-sectional area of the ribcage, and vice versa. Since the signals are related to the

cross-sectional area of the bands, a plot of the abdominal signal versus the ribcage signal during an isovolume maneuver should approximate a line (see Figure 19). The y-intercept should be related to the volume of air contained in the lungs during the procedure.

The second step of the calibration procedure required the subjects to sit on the ergometer at rest for two minutes, while the pneumotachometer and RIP band signals were recorded. This data could then be used to relate the derivative of the RIP volume signals to a known flow rate (provided by the pneumotachometer signal).

This protocol involving isovolume maneuvers and a two-minute period at rest was conducted twice: before and after the laboratory simulation.

The calibration curve was developed after the subject session was complete. First, the calibration files were processed using a batch file entitled RIP.bat (see Appendix I); this calculated the derivative of each band signal and stored it in the Codas file. Then a portion of the data from each isovolume maneuver and the period of breathing at rest was extracted and processed via the macros ISO and RIPCAL (Appendices J and K, respectively). The ISO macro opened all three isovolume maneuver files and calculated the line coefficients relating the abdominal signal to the ribcage signal for each. These three sets of data and their lines of best fit were plotted on a graph (see Figure 19). The most consistent isovolume data set was selected to relate the abdominal signal to the ribcage signal (for example, for the subject session shown in Figure 19, the middle isovolume maneuver was chosen to represent the band relationship). The slope and intercept for the selected data were then recorded for later use in the RIPCAL macro.

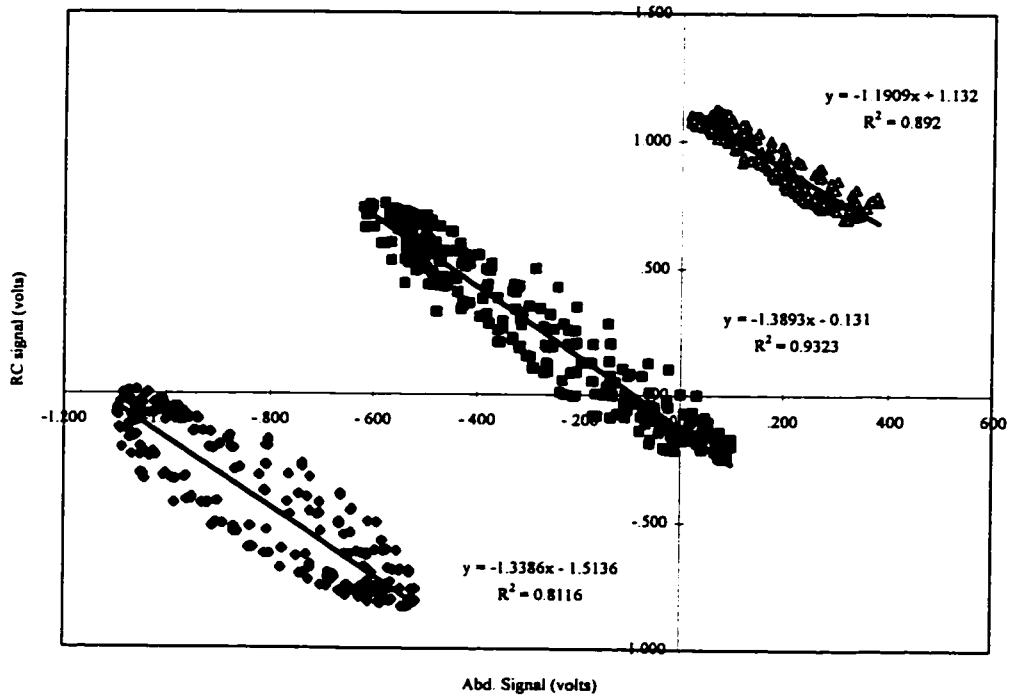


Figure 19: Plot of the abdominal band signal versus the ribcage signal during the three isovolume maneuvers for Subject #6, Session #1, Calibration #1.

The RIPCAL macro opened the data for breathing at rest, used the slope and intercept derived from the ISO macro to calculate the abdominal and ribcage calibration coefficients (abdominal coefficient= -slope/intercept, ribcage coefficient= 1/intercept), and multiplied the derivatives of the abdominal and ribcage signals by their respective calibration coefficients to get the adjusted derivative values. These two adjusted derivative values were then summed to get the final calibrated RIP values, which were plotted against the corresponding minute ventilation values in order to produce the final RIP calibration curve (see Figure 20).

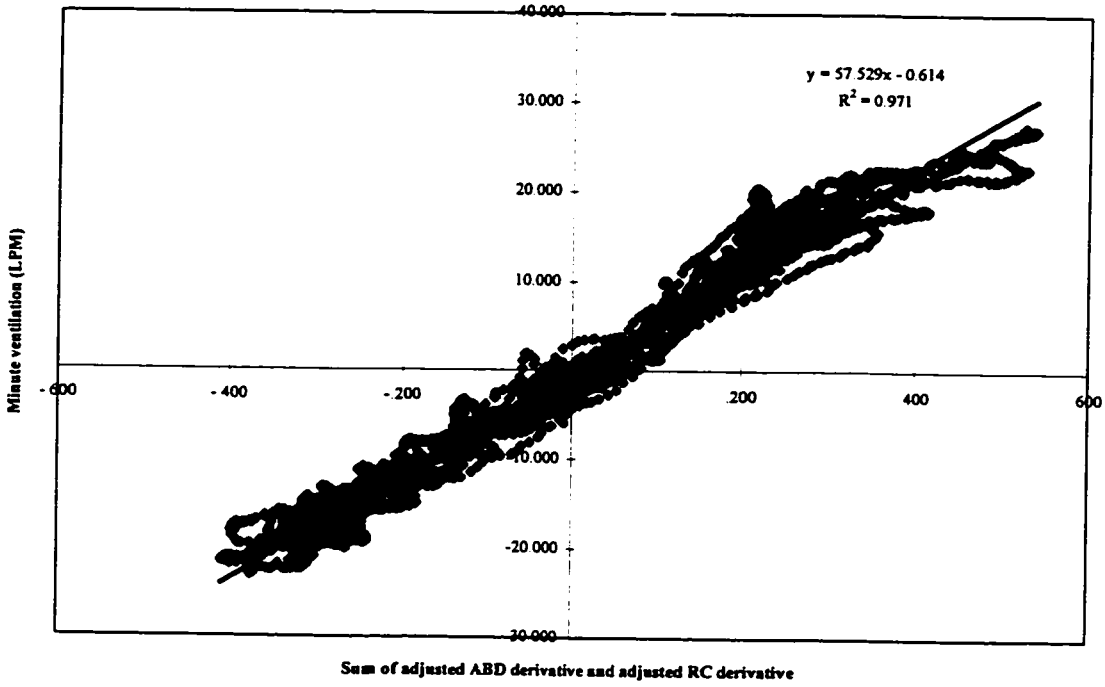


Figure 20: Plot of the sum of the adjusted RIP band derivatives versus the minute ventilation. The equation displayed on the figure is the RIP calibration curve for Subject #6, Session #1, Calibration #1.

The calibration information for each subject, consisting of the abdominal and ribcage coefficients as well as the slope and intercept of the calibration curve, was recorded in order to enter it into the PSCU program when the data collected from each subject were sent to the PSP.

DIRECT HEART RATE METHOD

In order to relate subjects' heart rates to their actual minute ventilation (for the Direct HR Method), a calibration curve linking these two variables was required. This curve was generated by recording the signals from the heart rate monitor and pneumotachometer

using software (Codas AT and Advanced Codas) and hardware (data acquisition card, Model DI-420 Parallel Port Module) produced by Dataq Instruments, for a total of 12 minutes: 2 minutes while sitting on a bicycle ergometer at rest, 5 minutes while biking at 40 Watts, and 5 minutes while biking at 80 Watts. After the subject session was complete, the Codas file was processed using a batch file (HRATE.bat, see Appendix L). This batch file calculated the moving averages of the heart rate and the positive half-wave rectified minute ventilation signal. One minute sections from each work rate period, corresponding to 45 to 105, 345 to 405, and 645 to 705 seconds from the start of the calibration file, were saved as a new file. An Excel macro named HRcal (see Appendix M) opened this file and generated a three-point calibration curve between the logarithm of the minute ventilation and the heart rate (see Equation 16). Note that this equation uses heart rate as opposed to delta heart rate, as per the Indirect HR Method which will be presented next; this format was chosen in order to avoid "clipping" the data due to potential overestimation of the true resting heart rate (this phenomenon will be addressed in the Discussion).

$$\text{Log}\,(\dot{V}\,i) = A \bullet HR + B$$

Equation 16: Predictive equation model for the direct heart rate calibration method. $\dot{V}\,i$ = minute ventilation (LPM) and HR = heart rate (BPM).

Another macro, PREDHRLOGVI (see Appendix N) used the resting heart rate and weight of the subject to predict the minute ventilation at each of the three steady-state heart rates using Equation 28 (see Results: Indirect HR Method). Then the actual data points, the measured calibration curve, and the predicted calibration curve were all plotted on a second graph, allowing visual comparison of the two calibration curves.

INDIRECT HEART RATE METHOD

The relationship between heart rate and minute ventilation, within subject, is very stable. This relationship can be fit with a straight line between heart rate and the logarithm of minute ventilation. Satoh *et al.*[4] developed an equation of this form using data from 34 healthy male subjects (see Equation 17).

$$\text{Log } \dot{V}i = (9.38 \bullet (HR - HR_0) + 4.22 \bullet H + 1.19 \bullet W + 2.22 \bullet A + HR_0) \times 10^{-3} - 0.0439$$

Equation 17. Satoh *et al.*'s[4] equation to relate heart rate (HR, in beats per minute) to minute ventilation ($\dot{V}i$, in liters per minute). HR_0 = resting heart rate (BPM), H = height (cm), W = weight (kg), A = age (yr).

The database compiled for our study was much larger, however, and consisted of 181 subjects. The data were accessed through the Gas Exchange Laboratory at the Harbor-UCLA Medical Center. Subject data were systematically abstracted from original records. Prior to data collection, a matrix was developed to guide the subject selection (see Table 7). Each cell was a unique combination of sex, age, health, and activity level (a surrogate measure of fitness). The goal for the data collection was to try to select five subjects per matrix cell.

Table 7 shows the actual number of subjects in the database for each matrix cell. In summary, the database contained data on 138 males and 43 females. 117 of these subjects were determined to have "Normal" health, 46 were "Mildly Diseased", and the remaining 18 were "Severely Diseased".

Table 7: Subject matrix showing the number of subjects in database for each matrix cell.

Health	Normal	Normal	Normal	Normal	Mild	Mild	Mild	Severe
Activity Exer.Freq.	Sedntry 0/wk	Avg. 1-3/wk	Active 4-5/wk	Athletic >5/wk	Sedntry 0/wk	Avg. 1-3/wk	Active 4-5/wk	Sedntry 0/wk
Males:								
Age <30	4	4	1	5	2	0	1	3
Age 30-39	3	10	3	3	4	1	4	0
Age 40-49	9	4	4	3	2	1	5	3
Age 50-59	4	8	7	2	6	2	3	5
Age ≥50	3	10	1	0	1	1	3	3
Females:								
Age <30	3	1	2	0	2	1	0	1
Age 30-39	2	2	2	2	1	1	0	0
Age 40-49	5	2	2	1	1	0	1	1
Age 50-59	2	1	1	0	1	1	0	2
Age ≥50	0	1	0	0	0	1	0	0

The subjects were drawn from three different sources: (1) asbestos-exposed individuals referred to the laboratory to take part in an Asbestos Case-Control study, (2) control subjects pulled from the local community for use in a variety of studies, including the Asbestos study, and (3) patients referred to the laboratory for clinical reasons. All of the controls were healthy; the remaining subjects were classified as healthy, mildly diseased, or severely diseased based on the physiologic data and the physician's written comments and conclusions. As a general rule, subjects were considered to be "mildly diseased" when they had some condition, such as asthma, diabetes or obesity, that was not severe enough to cause a major lifestyle change. When subjects had diseases that did seriously impede their health and lifestyle, such as pulmonary vascular disease or cardiovascular disease, they were considered to be "severely diseased".

All of the subjects performed an incremental work rate test as described in Principles of Exercise Testing and Interpretation.[1] The 30-second averages of heart rate, breathing

frequency, and minute ventilation were recorded over the span of the test (10-20 minutes). For all subjects, the following data were collected (except for weight, which was missing for one subject):

- sex
- age
- height
- weight
- maximum oxygen capacity
- predicted maximum oxygen capacity
- maximum heart rate
- maximum oxygen pulse
- maximum minute volume
- estimated overall health
- estimated fitness level

The following additional data were also collected when available:

- predicted maximum heart rate
- predicted maximum oxygen pulse
- blood pressure at rest
- blood pressure at maximum exercise
- arterial oxygen pressure at rest and at maximum exercise

- alveolar-arterial oxygen pressure difference at rest and at maximum exercise
- arterial end-tidal carbon dioxide pressure difference at rest and at maximum exercise
- the physiologic deadspace to tidal volume ratio
- whether the subject was a smoker at the time of the test
- smoking history
- whether the subject had respiratory disease and/or cardiovascular disease
- vital capacity
- inspiratory capacity
- forced expiratory volume in one second (FEV1)
- maximum ventilatory volume
- whether the subject was on any medication that would alter heart rate

Since the data was extracted from records, there was no control over the accuracy of the data. Instrument malfunction could have occurred, for example, during a subject's data collection. Therefore, the database was screened in order to eliminate potentially erroneous subject data. First, the subject data was visually assessed, with special attention given to the change in heart rate versus the logarithm minute ventilation. Second, the correlation between these two variables was calculated. Third, a histogram of all the subject's correlation coefficients was generated (see Figure 21).

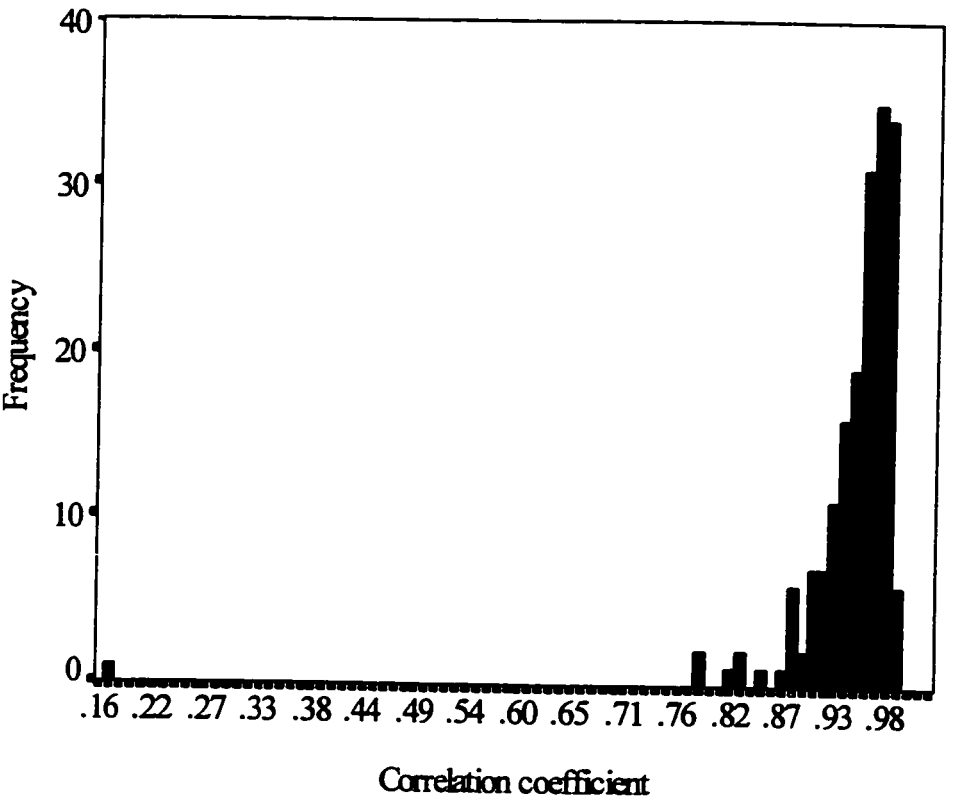


Figure 21: Histogram of the correlation coefficients between delta heart rate and the logarithm of minute ventilation for all database subjects.

The one-sided 95% lower confidence limit for the correlation coefficients was 0.84; six of the 181 subjects fell below this value and were excluded for that reason. The matrix cells representing each of the excluded subjects, as well as their subject identification number, are displayed in Table 8. As can be seen by this table, half of the excluded subjects fell into the category of normal health and no regular exercise.

Table 8: Subject matrix showing the subjects excluded from the analysis for each matrix cell.

Health	Normal	Normal	Normal	Normal	Mild	Mild	Mild	Severe
Activity Exer.Freq.	Sedntry 0/wk	Avg. 1-3/wk	Active 4-5/wk	Athletic >5/wk	Sedntry 0/wk	Avg. 1-3/wk	Active 4-5/wk	Sedntry 0/wk
Males:								
Age <30		#59						
Age 30-39								
Age 40-49								
Age 50-59								#124
Age ≥50	#109							
Females:								
Age <30								
Age 30-39	#177							
Age 40-49	#46							
Age 50-59						#185		
Age ≥50								

The predictive equation model was developed by performing a stepwise regression on the filtered database (of 175 remaining subjects) using SPSS®. The final equation included the minute ventilation as the independent variable and the change in heart rate (called delta heart rate) and weight as dependent variables (see Equation 18). Note that the following variables were tested and rejected for inclusion in this equation based on non-significance at the 0.05 level: sex, age, height, body mass index (Wt_kg/Ht_m^2), resting heart rate, health, and fitness.

$$\dot{V}i = e^{A \bullet (HR - HR0) + B \bullet W + C}$$

Equation 18: Predictive equation model for the indirect heart rate calibration method. $\dot{V}i$ = minute ventilation, HR = heart rate (beats per minute), HR_0 = resting heart rate, and W = weight in kilograms.

The coefficients for Equation 18 (A, B, and C) were determined by using a bootstrap statistical method as outlined by Efron[21], and LePage and Podgorski.[22] A bootstrap method was required in order to deal with the autocorrelation between minute ventilation and delta heart rate within-subject. Since the data were recorded as 30-second averages and each subject exercised for different periods of time, there were various numbers of pairs of delta heart rate and minute ventilation data for each subject. Inherently, these 30 second averages were autocorrelated, thus nullifying the assumption of independence necessary for the application of more traditional statistical methods.

The data was tabulated in the following format: each row contained one pair of minute ventilation and delta heart rate values for one subject for one 30-second period, as well as the subject's weight.

In order to calculate a linear regression without log-transforming the minute ventilation data, the statistical package S-Plus was used; this was necessary in order to preserve the variability structure of the underlying population. S-Plus has a log-link feature that fits a non-linear curve (in this case an exponential curve) to the data by minimizing the residual sums-of-squares.

The S-Plus code for the bootstrap is included in Appendix O; essentially, the following series of four steps was executed as many times as the user desired:

1) One row of data for each subject was randomly selected from all rows available for that subject.

2) The 175 rows selected in Step 1 (one for each subject) were extracted from the database and stored in a matrix.

3) A log-linked regression was performed on the extracted data.

4) The coefficients from the log-linked regression analysis were stored in another matrix.

A random distribution was used to ensure that a new set of rows of data was selected in Step 1 each time this step was executed. This bootstrap process produced a distribution of coefficients for each term in the predictive equation. The mean value of each distribution was then used in the final predictive equation.

SATOH'S HEART RATE METHOD

The following equation developed by Satoh *et al.*[4] was used to relate minute ventilation to heart rate. It incorporates the subject's resting heart rate, height, and weight in addition to heart rate.

$$\text{Log } \dot{V}i = (9.38 \bullet (HR - HR_0) + 4.22 \bullet H + 1.19 \bullet W + 2.22 \bullet A + HR_0) \times 10^{-3} - 0.0439$$

Equation 19. Satoh *et al.*'s [4] equation to relate heart rate (HR, in beats per minute) to minute ventilation ($\dot{V}i$, in liters per minute). HR_0 = resting heart rate (BPM), H = height (cm), W = weight (kg), A = age (yr).

LABORATORY SIMULATION

In order to compare the Methods listed in the introduction to this chapter, minute ventilation, heart rate, and RIP data for all 30 subjects were collected over a 40-minute sampling period that involved four work rates. The data were recorded to allow it to be replayed without requiring the subjects to return for additional sessions. The data were recorded with high fidelity; therefore, when it was replayed, the PSCU received the data as if the subject were actually hooked up to it directly. Note that if the subject had been hooked up to the PSP directly, the system would have worked the same.

The sampling period commenced immediately following the heart rate calibration data collection, after the subjects were given a water break. The mouth unit was re-attached, the nose clip was re-fastened, and the sampling period was started.

The 40-minute sampling period was broken into 8 five-minute sections, with the following work rates assigned to each section: rest, 40 Watts, rest, 80 Watts, 40 Watts, rest, 120 Watts, and rest. The cardiotachometer and pneumotachometer signals were recorded at 125 samples per second using the Dataq software and hardware. When biking, subjects were required to pedal at 50 rotations per minute; a metronome set to 100 beats per minute provided an auditory cadence, and the pedaling rate could be visually confirmed via a display on the ergometer itself. Subjects were allowed to stand during

the rest periods. In addition, they were allowed to read if they wished to and could do so while biking steadily at 50 rotations per minute. Music, however, was not allowed.

A batch file (TXDAT.bat, see Appendix P) was used to process the sampling period data. The processed file was then copied using a comma delimited format, opened in Word, and saved as a text file (after the Codas header was removed).

The subject data were sent to the PSCU via a Pentium computer, as diagrammed in Figure 22. The computer executed a QuickBASIC program called Senddat.bas (see Appendix Q); this program sent data from a comma delimited file to a digital-to-analog converter at a rate of 125 samples per second per signal.

For calibration purposes, prior to sending a subject's data, the Senddat.bas program was used to send zero volts to both the RIP and pneumotachometer channels of the PSCU. Three Txtools programs, Analog0 for the RIP abdominal signal (see Appendix R), Analog2 for the pneumotachometer signal (see Appendix S), and Analog6 for the RIP ribcage signal (Appendix T), were run on a laptop Pentium computer to measure the voltage offsets of the PSCU.

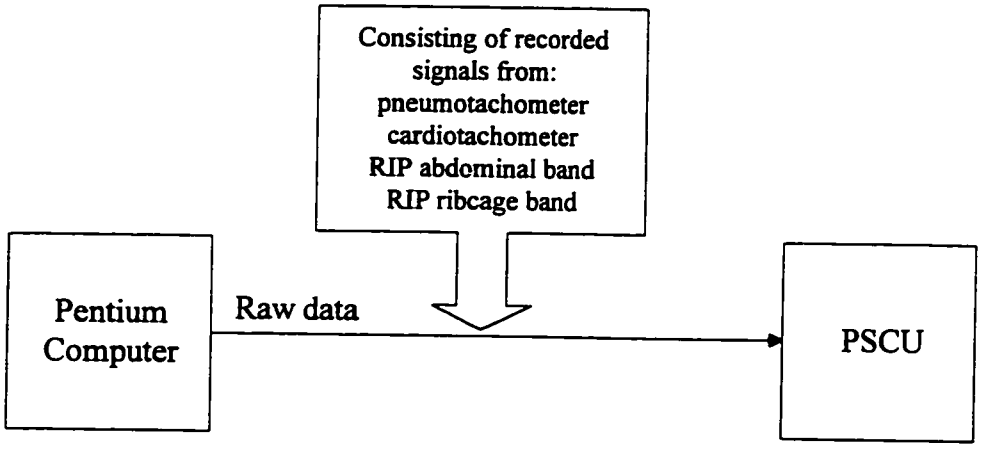


Figure 22: Diagram of raw subject data (consisting of the four recorded signals) being sent from the computer to the PSCU.

A TxBASIC program titled LABPSP (see Appendix U) was downloaded into the PSCU and started. This program prompted the user to input several values, including the PSCU offsets, the four pumps' calibration coefficients, the direct heart rate calibration coefficients, the isovolume slope, the RIP calibration coefficients, and the subject's age, height, weight and resting heart rate. The program's sampling protocol was initiated manually, to correspond with the start of the Pentium computer sending a subject's data file to the PSCU. The program then calculated and recorded the 15-second averages of minute ventilation estimates for each Method.

After each subject's 40-minute sampling period had been sent to the PSCU, the 15-second averages were downloaded from the PSCU to the laptop Pentium computer and saved as an Excel file. All five estimates were weighted by the averaging time and then summed over the sampling period to attain the total inhaled volume values.

The raw pneumotachometer data were processed in order to calculate the Primary Standard value. The processing of this signal involved summing the positive half-wave

rectified data over the sampling period and applying the pneumotachometer calibration curve. The spreadsheet that was used to calculate these values is included in Appendix V, along with the equations used for the calculations.

Mass flowmeters were used to validate that the volume of air pulled through charcoal tubes attached to the PSPs were proportional to the PSCU estimated total inhaled volumes. The equipment setup for this set of experiments is diagrammed in Figure 23. The raw subject signals were sent to the PSCU, as described previously. The PSCU was connected to four sampling pumps; these pumps were driven by the following Methods: (1) Pneumo Method, (2) Direct HR Method, (3) Indirect HR Method, and (4) RIP Method. Four GilAir 5 pumps were labeled with an identifier: A, B, C, and D. The four mass flowmeters were also individually labeled: 1, 2, 3 and 4. The mass flowmeters were calibrated against a primary standard, a Gilibrator®.

The mass flowmeters were attached upstream of the charcoal tube and allowed to warm up for at least 15 minutes prior to use. The pumps and mass flowmeters were randomly assigned each data set, in order to eliminate any pump or mass flowmeter bias. The output voltages of the four mass flowmeters were logged on a computer using the Dataq Instruments software and hardware. After sampling, the output voltages were integrated and multiplied by 40 (the inverse of the constant ratio of pump flow rate to minute ventilation) in order to verify that the pump flow rates were proportional to the total estimated minute ventilation values.

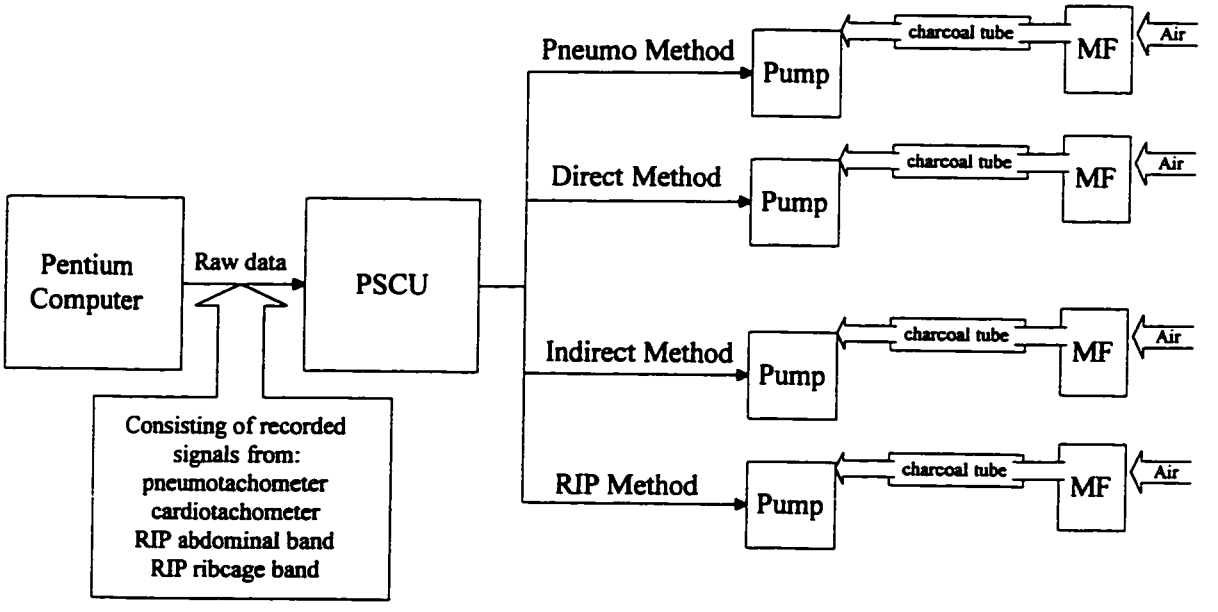


Figure 23: Diagram of the equipment set-up.

RPD ANALYSIS

The RPDs of total inhaled volume (over the 40-minute exercise period) between each Method and the Primary Standard was calculated using the following formula:

$$RPD = \frac{M - M_0}{M_0} \cdot 100$$

Equation 20: Relative percent difference between two Methods, where M_0 is the reference Method.

Histograms of the RPDs for each Method were plotted, and the mean and SD of each set of RPDs were calculated.

CONCORDANCE ANALYSIS

The PSCU results from all Methods (Pneumo Method, Direct HR Method, Indirect HR Method, RIP Method, and Satoh's HR Method) were compared to the Primary Standard using a method called concordance analysis as described by van Belle *et al.*[23] In addition, the Indirect HR Method was compared to the Direct HR Method in a similar manner. Concordance analysis is comprised of five components: Location Shift, Scale Shift, Concordance Precision, Concordance Accuracy, and overall Concordance. The reason this analysis was performed in addition to the RPD analysis was these components provided additional insight into how the Methods differed.

The concordance components are defined below in Equations 21 through 25.

$$\text{Location shift} = u = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1 \cdot \sigma_2}}$$

Equation 21: Definition of location shift in concordance analysis.

$$\text{Scale shift} = v = \frac{\sigma_1}{\sigma_2}$$

Equation 22: Definition of scale shift in concordance analysis.

$$\text{Concordance precision} = r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Equation 23: Definition of concordance precision, which is equal to Pearson's correlation coefficient.

$$\text{Concordance accuracy} = A = \frac{2}{v + \frac{1}{v} + u^2}$$

Equation 24: Definition of concordance accuracy.

$$\text{Concordance} = r \cdot A$$

Equation 25: Definition of concordance.

Location shift centers around zero; values less than or greater than zero signify that the comparison method underestimates or overestimates total inhaled volume compared to the standard method. The statistical significance of location shift was determined by performing a paired T-test between the two methods of interest. Scale shift is simply a ratio of the standard deviation of the comparison method to the standard deviation of the standard method. Values greater than one suggest that the variance of the comparison method is greater than the standard method. The statistical significance of scale shift was calculated by performing a linear regression through a scatterplot of data pairs for the

methods and testing the null hypothesis that the slope was not significantly different from one. The concordance precision is the Pearson correlation coefficient and the concordance accuracy is a combination of both the location shift and the scale shift (see Equation 24 for the exact relationship). Note that for both concordance accuracy and precision, higher values are considered better; this directionality may be anti-intuitive to some individuals. The term "concordance" is placed in from of these terms in order to remind the reader to recall the definitions and avoid confusion. The overall concordance is simply a product of the concordance precision and the concordance accuracy.[23]

Dynamic Pump Response

It is important to understand the behavior of both the pump and the mass flowmeter in order to assess the error associated with the pump's response to an input voltage. The inability of the pump to respond instantaneously to a new input voltage is characteristic of electro-mechanical systems in general. Usually these systems can be described in one of two ways; they are either a first order linear system or a second order linear system.

A first order linear system can be characterized by the differential equation shown below:

$$\tau \frac{dx}{dt} + x(t) = y(t)$$

Equation 26: First order linear system differential equation in standard form. Where y(t) is the input at time t, x(t) is the output at time t, and τ is the time constant of the system.[24]

The input in this equation is y(t). The steady state output is x(t). The third component in this equation, τ dx/dt, gives rise to the transient response, where τ is the time constant of the system. In the case of the sampling pump, the input is a voltage and the output is a flow rate. In the case of the mass flowmeter, it is just the reverse; the input is a flow rate and the output is a voltage. Naturally, there is a proportionality constant needed in order

to relate the voltage to the flow rate for these equations. If these instruments are hooked in series (and they are both first order linear systems), then the two first order linear equations representing each instrument will be related to each other by a linear equation.

The equation relating the response of a first order linear system to a rising step input is displayed in Equation 27.

$$x(t) = X_f + (X_i - X_f) \cdot e^{-t/\tau}$$

Equation 27: Response of a first order linear system to a step input. Where x(t) is the output response, X_f is the final value of x after the transients have died out, X_i is the initial value of x, just after the step was imposed, and τ is the time constant of the system.[24]

Graphically, the output of a first order system, in response to a step input, is an exponentially rising curve.

Second order linear systems are more complex than first order linear systems and have an additional parameter, the damping coefficient. While it is not critical to understand the mathematics behind second order linear systems for this dissertation, it is important to understand that they exhibit a different response to a step input that is distinct graphically from the response of a first order linear system.

This information was useful in setting up a series of experiments to determine: (1) whether the pump responded to the changes in voltage from the PSCU in a first order or second order fashion, (2) if the pumps all responded the same way, and (3) the approximate time constant of the system, if the pump responded in a first order linear fashion. These problems were complicated by the fact that the pump flows were measured using mass flowmeters, instruments that are also either a first or second order linear system.

In order to investigate the pump response, two Txtools programs were written to instruct the PSCU to generate of a series of voltage steps. The first program, called Vtest15s (see Appendix W), had voltage steps that were 15 seconds in length. The second program, Vtest60s (see Appendix X), was identical, except that the voltage steps were 60 seconds long.

After confirming that the predicted voltages corresponded to the actual measured voltages, each pump (A through D) was tested three times using Vtest15s. The mass flowmeter was attached upstream and its output voltage was recorded using Dataq software and hardware.

Finally, the Vtest60 program was sent to pumps "A" and "C" two times, hereafter referred to as "Run 1" and "Run 2". During Run 1, the mass flowmeter was located upstream of the pump. For Run 2, however, a valve was used to select whether the pump pulled air through the mass flowmeter or not. For each 60-second voltage step in the Vtest60 program, the air was pulled through the mass flowmeter for only the last 30 seconds to decouple the mass flowmeter response from the pump response.

The mass flowmeter output voltages for Runs 1 and 2 could then be overlaid; this allowed for a comparison of (a) the pump and mass flowmeter responses linked together (during Run 1) and (b) the mass flowmeter alone (during Run 2). Note that at 30 seconds into each step, the pump response is assumed to have stabilized, such that the mass flowmeter response would not be linked to the pump response.

The time constant for the mass flowmeter (whose published value was 0.08 seconds) and the time constant for the pump and mass flowmeter in series were calculated. The time constant for the pump could then be estimated, with the assumption that these two instruments in series behaved according to the law of superposition (meaning that the time constants would be additive).

RESULTS

HUMAN SUBJECT CHARACTERIZATION

Fifteen males and fifteen females participated in the study. Summary statistics for the age, height, weight, and BMI for males and females are provided in Tables 9 and 10, respectively.

Table 9: Summary of male subjects' age, height, weight, and body mass index.

MALES	Mean	Minimum	Maximum
Age (yr)	31.1	26	43
Height (cm)	182.5	165	193
Weight (kg)	84.9	69	130
BMI (kg/m^2)	25.5	21	40

Table 10: Summary of female subjects' age, height, weight, and body mass index.

FEMALES	Mean	Minimum	Maximum
Age (yr)	29.5	20	60
Height (cm)	166.0	155	173
Weight (kg)	62.5	52	86
BMI (kg/m^2)	22.6	19	32

Of the male subjects, seven classified themselves as above average in health, while the remaining eight decided they experienced normal health. One subject had mild exercise induced asthma; no other diseases were reported. None of the subjects reported taking any medication. Fourteen subjects were Caucasian and one subject was Asian (Subject #8). Only one male subject smoked (Subject #30), and he had a twelve-year history of smoking one pack per day. Subjects classified their fitness based on the number of times

per week they regularly exercised: sedentary (no exercise), average (exercise 1-3 times per week), active (exercise 4-5 times per week), and athletic (exercise 6-7 times per week). Two males were classified as sedentary, five as average, five as active, and three as athletic.

Of the female subjects, eight claimed they experienced above average health, and the remaining seven reported being of normal health. The only disease reported among the women were two cases of hypothyroidism. One subject smoked, Subject #32, although she only reported smoking two packs per month for approximately four years. Fourteen subjects were Caucasian and one was Filipino (Subject #32). In terms of their fitness, only one female was classified as sedentary, eight had average fitness, five were active, and one was athletic.

The two smokers were not obvious outliers with the exception that the Direct HR Method signficantly overpredicted the minute ventilation for the female smoker; this subject was also the only non-Caucasian female.

DIRECT CALIBRATION RESULTS

The average steady-state heart rate and minute ventilation values at 0 Watts, 40 Watts, and 80 Watts were significantly different between subjects (p-values of 0.004, 0.001, and 0.000 for heart rate and 0.021, 0.038, and 0.016 for minute ventilation).

The change in resting heart rate between Session #1 and Session #2 (within-subject) averaged 1.3%, but ranged from -22% to +32%.

When the steady-state minute ventilation values between Session #1 and Session #2 were compared (at 0, 40, and 80 Watts), it was found that the range of minute ventilation values was larger for Session #1 than for Session #2; this effect can be seen in Figure 24.

When the two sessions were compared overall (without separating the data by workrate), there was no significant difference found. However, Figure 24 shows that the minute

ventilation values tended to be more variable for Session #1 (evidenced by the greater range of values, and therefore lowering the slopes at each workrate).

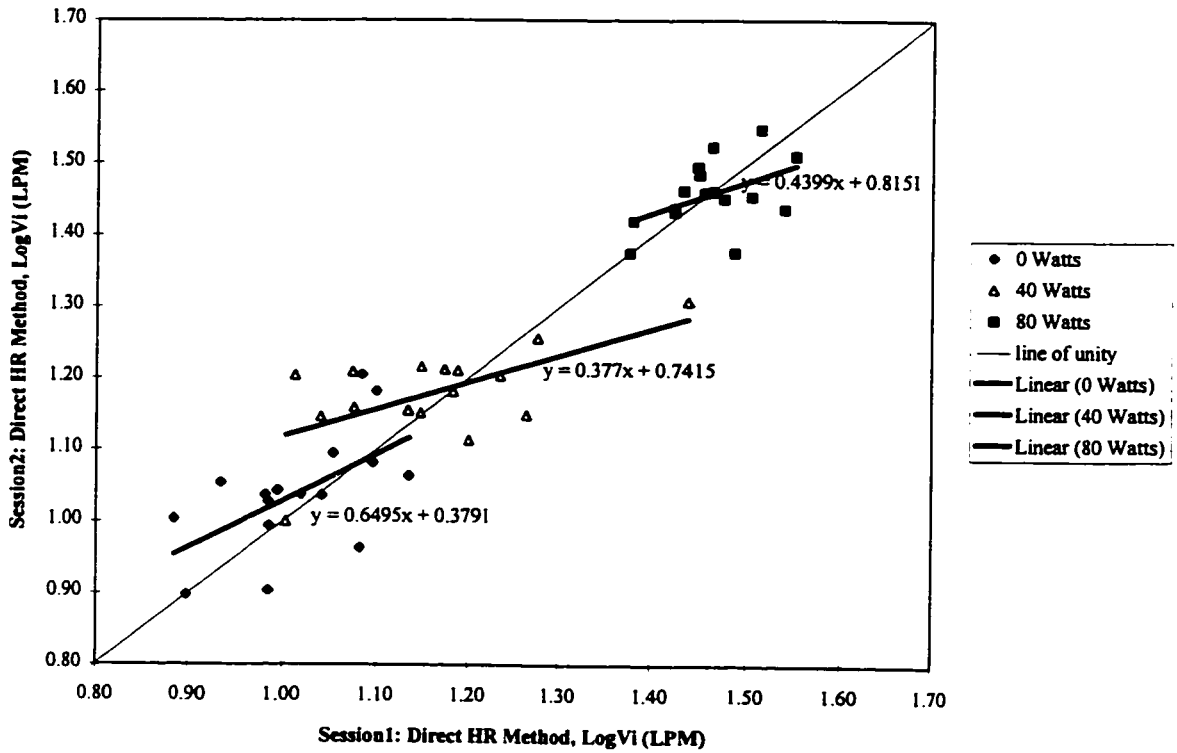


Figure 24: Plot of Session #1 versus Session #2 estimated minute ventilation values at the three calibration points (0, 40, and 80 Watts) using the Direct HR Method.

The average, minimum, and maximum steady-state values for heart rate and minute ventilation at all three work rates for the first session are displayed in Table 11. Second session data was not included in these statistics so as not to weight the averages twice for the sixteen subjects that performed a second session.

Table 11: First session steady-state heart rate and minute ventilation statistics for 0, 40, and 80 Watts.

	Work	N	Avg.	Min.	Max.
Heart Rate (BPM)	**0 Watts**	30	81.6	57.4	114.7
	40 Watts	30	92.0	60.7	137.0
	80 Watts	30	117.9	78.1	149.3
Minute Ventilation (LPM)	**0 Watts**	30	9.2	4.7	12.8
	40 Watts	30	16.8	12.5	30.8
	80 Watts	30	26.9	19.0	33.4

INDIRECT CALIBRATION RESULTS

The bootstrap process produced a population of coefficients. Histograms for the coefficients, overlaid with a normal curve, are displayed in Figures 25 through 27.

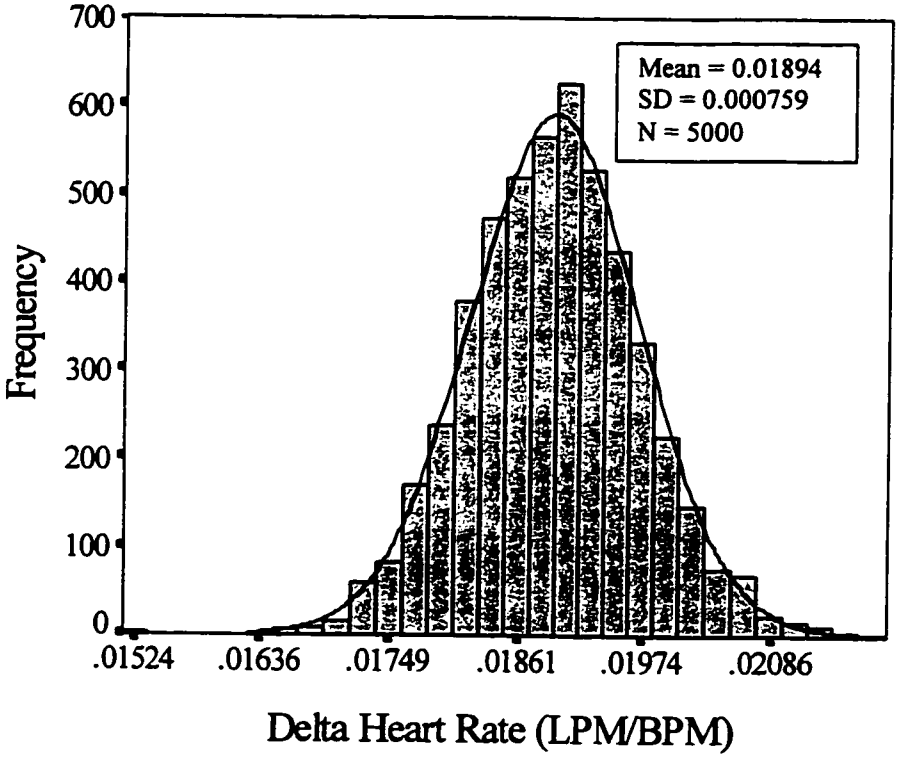


Figure 25: Histogram of delta heart rate coefficients.

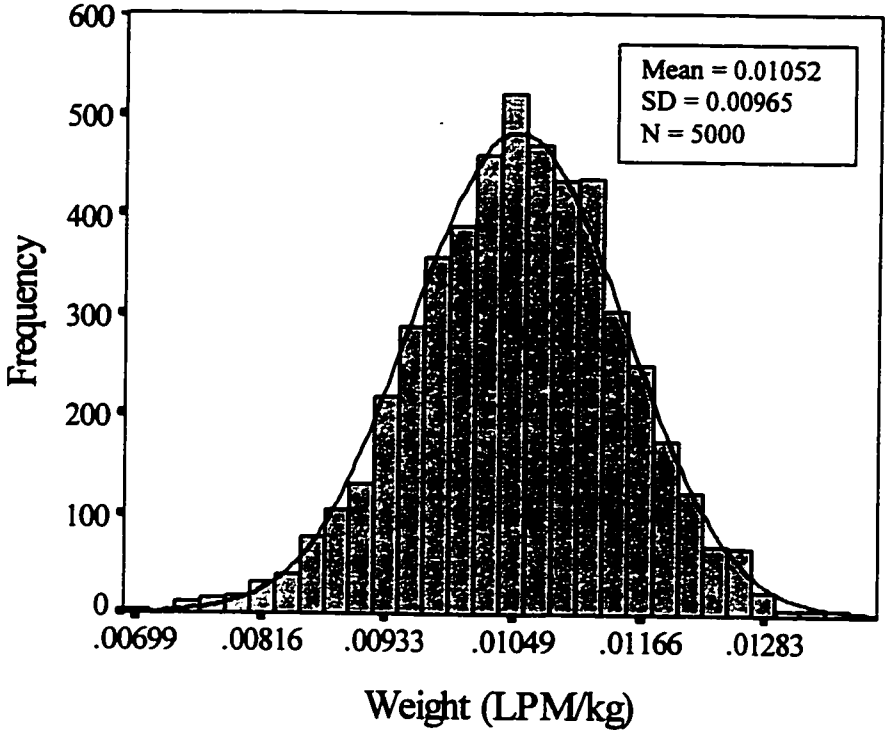


Figure 26: Histogram of weight coefficients.

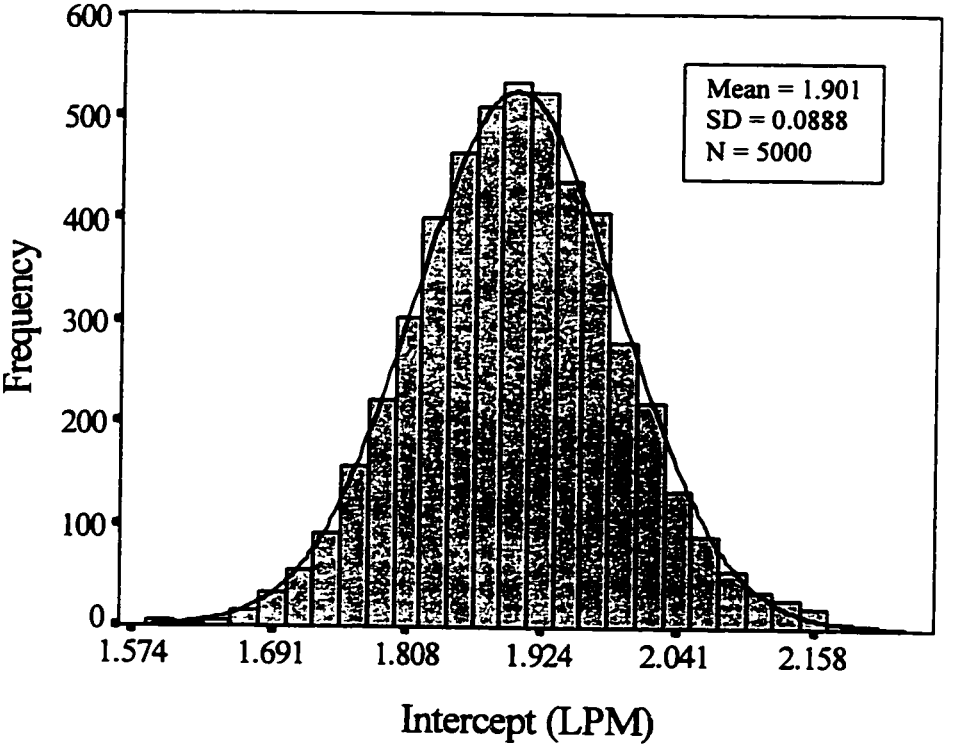


Figure 27: Histogram of intercept coefficients.

The final predictive equation (Equation 28) was derived by entering the means of each coefficient into Equation 18.

$$\dot{V}i = e^{0.01894 \bullet (HR-HR0) + 0.01052 \bullet W + 1.9008}$$

Equation 28: Predictive equation relating minute ventilation to heart rate derived from a large database of physiologic data. $\dot{V}i$ = minute ventilation (LPM), HR = heart rate (beats per minute), HR_0 = resting heart rate, and W = weight in kilograms.

At rest, when the delta heart rate value would be zero, the estimated minute ventilation would be dependent solely on weight of the subject.

CALIBRATION CURVE COMPARISON

For each subject session, the minute ventilation estimates for the steady state heart rate value corresponding to 0, 40 and 80 Watts of work were derived using both the Indirect HR Method and the Direct HR Method. Figure 28 shows a comparison of these values.

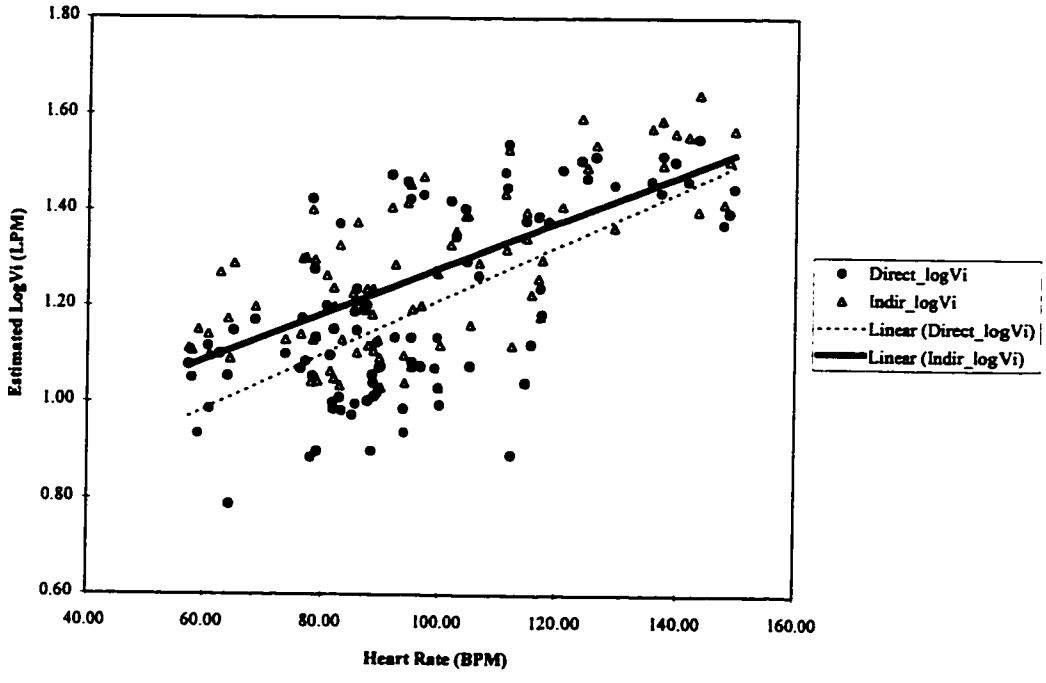


Figure 28: Plot of the estimates of the logarithm of minute ventilation based on each heart rate calibration method (Indirect HR Method and Direct HR Method) versus heart rate.

Figure 28 clearly shows that over the heart rate range corresponding to 0 to 80 Watts of work, the minute ventilation estimates based on the Indirect HR Method are higher than those derived using the Direct HR Method.

The difference between the two estimates versus heart rate is displayed in Figure 29. The Methods differ the most at lower work rates, and start to converge as work rate is increased.

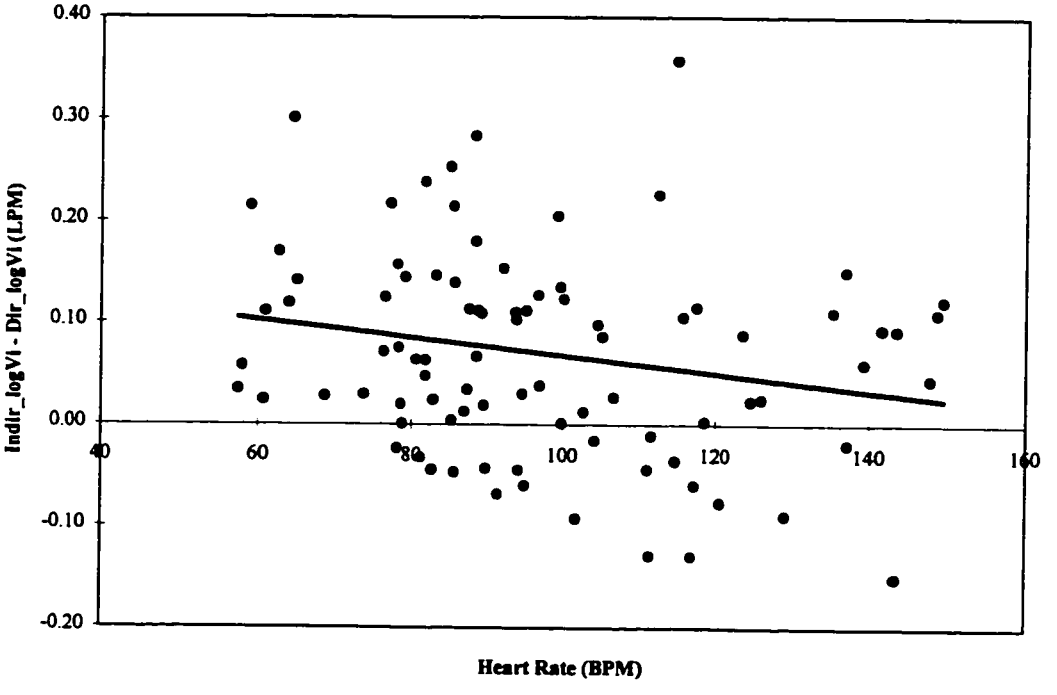


Figure 29: Plot of the difference of the logarithm of minute ventilation (Indirect HR Method - Direct HR Method) versus heart rate.

Several MANOVAs were performed to determine if sex, health status or fitness affected this difference between the Methods' estimates of minute ventilation. It was determined that only fitness had a significant effect. This effect is exceedingly apparent in Figure 30, where the difference between the minute ventilation estimates (Indirect HR Method - Direct HR Method) is stratified by fitness; in this case the fitness categories were collapsed into two categories: (1) exercise 0 to 3 times per week, and (2) exercise more than 3 times per week.

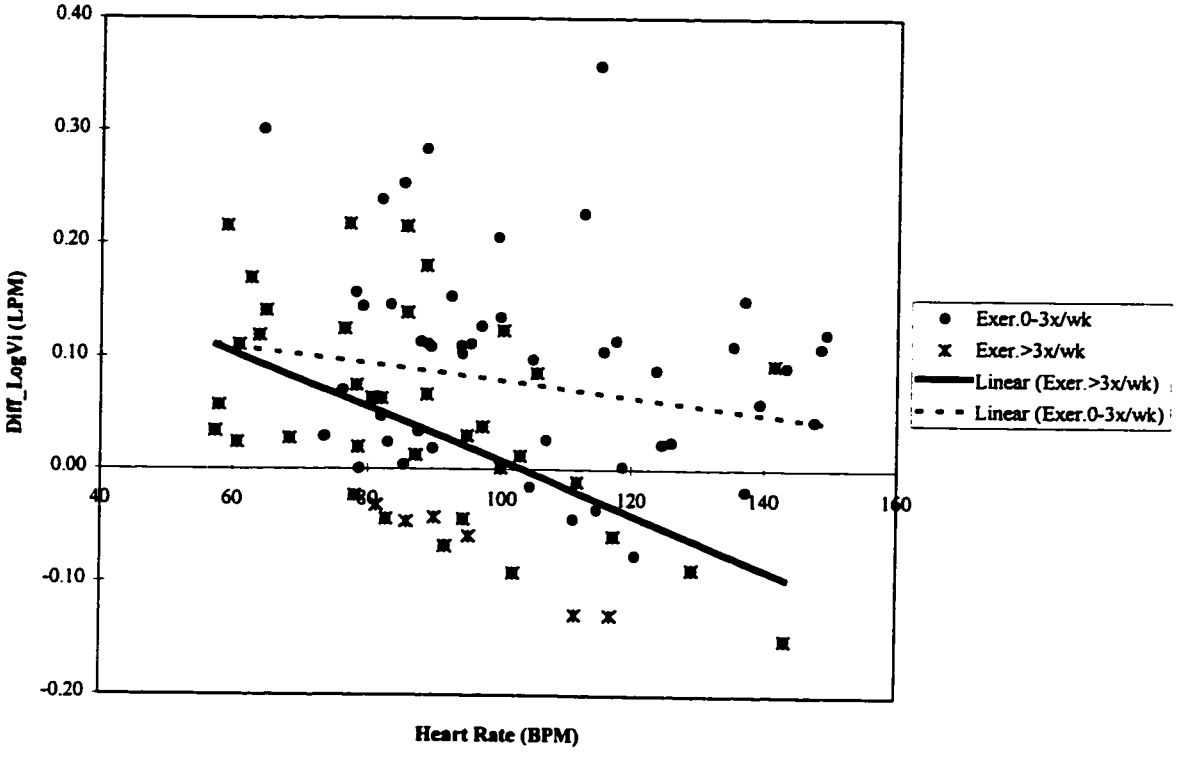


Figure 30: Difference in minute ventilation estimates (Indirect HR Method - Direct HR Method), stratified by subject's fitness (exercise 0-3 times per week or exercise more than 3 times per week).

Each Method was also examined separately to determine if there were any significant between subject effects for sex, health, or fitness; no significant effects were discovered.

PSCU Total Inhaled Volume Results

The PSCU estimates of total inhaled volume (liters of air inhaled over the 40 minute sampling period), from the first session for each subject, are summarized in Table 12. Results from the second session for the 16 subjects who performed a second testing session are summarized in Table 13. Note that these estimates are derived from the PSCU data files, not from the mass flowmeter data (with the exception of the Primary

Standard whose values were derived by integrating the pneumotachometer signal on a computer rather than via the PSCU).

Table 12: PSCU estimates of total inhaled volume (in liters) from the first sampling period for all 30 subjects.

Subj	Primary	Pneumo	Direct	Indirect	RIP	Satoh's
1	773	757	922	974	697	479
2	616	602	743	996	582	512
5	715	701	876	862	628	455
6	641	628	805	997	721	537
7	660	645	611	835	750	399
8	807	789	460	750	N/A	321
9	602	590	610	779	581	462
10	762	747	836	924	870	500
11	789	775	1023	1116	732	671
12	795	778	920	937	659	560
13	757	742	909	738	776	411
14	699	685	813	1028	730	588
15	624	613	865	698	N/A	382
16	665	653	648	726	629	422
17	936	918	886	1396	923	475
18	707	694	732	740	661	378
19	800	785	984	972	756	413
20	607	593	778	837	629	455
21	739	723	822	1051	354	586
22	670	654	746	749	707	415
23	598	585	691	1028	540	631
24	659	646	553	767	602	372
25	731	715	819	1123	804	606
26	756	745	845	873	705	404
27	709	694	768	743	638	410
28	680	672	715	906	725	503
29	706	694	908	1226	637	616
30	823	807	873	868	699	488
31	844	829	929	897	837	403
32	861	846	1362	984	678	498

Table 13: PSCU estimates of total inhaled volume (in liters) from the second sampling period for 16 of the subjects.

Subj	Primary	Pneumo	Direct	Indirect	RIP	Satoh's
1	791	777	656	853	685	411
10	675	661	800	965	646	521
11	876	861	1186	1062	836	631
13	823	808	793	722	810	399
14	798	785	836	958	683	545
16	777	764	887	898	603	540
17	834	820	978	1656	856	578
20	586	573	705	893	455	489
21	779	768	918	1051	N/A	587
22	709	695	767	859	604	485
25	734	718	883	1117	703	601
28	728	716	826	888	607	492
29	816	800	1096	1132	636	558
30	725	710	771	1043	1662	602
31	790	775	729	765	762	335

Figure 31 through 35 show graphs of the data in Table 12. The data from each Method is plotted against the Primary Standard values.

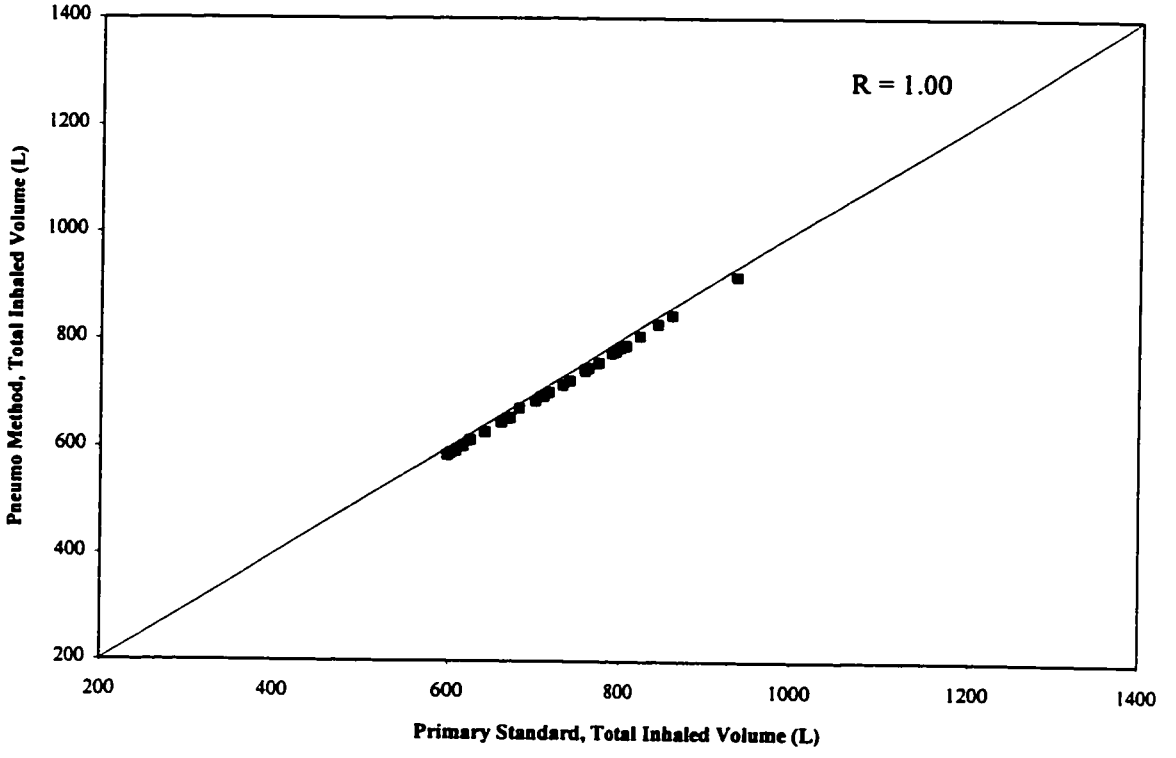


Figure 31: Plot of Pneumo Method versus Primary Standard PSCU results. The line is a line of unity for the Primary Standard.

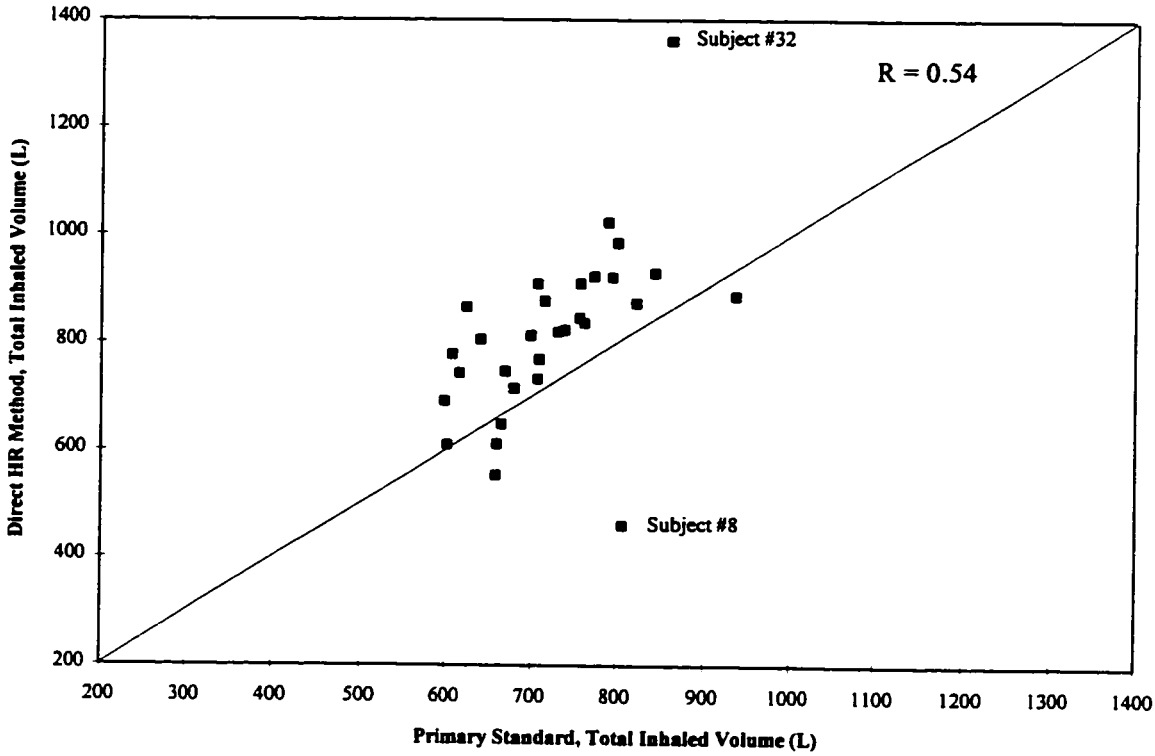


Figure 32: Plot of Direct HR Method versus Primary Standard PSCU results. The line is a line of unity for the Primary Standard.

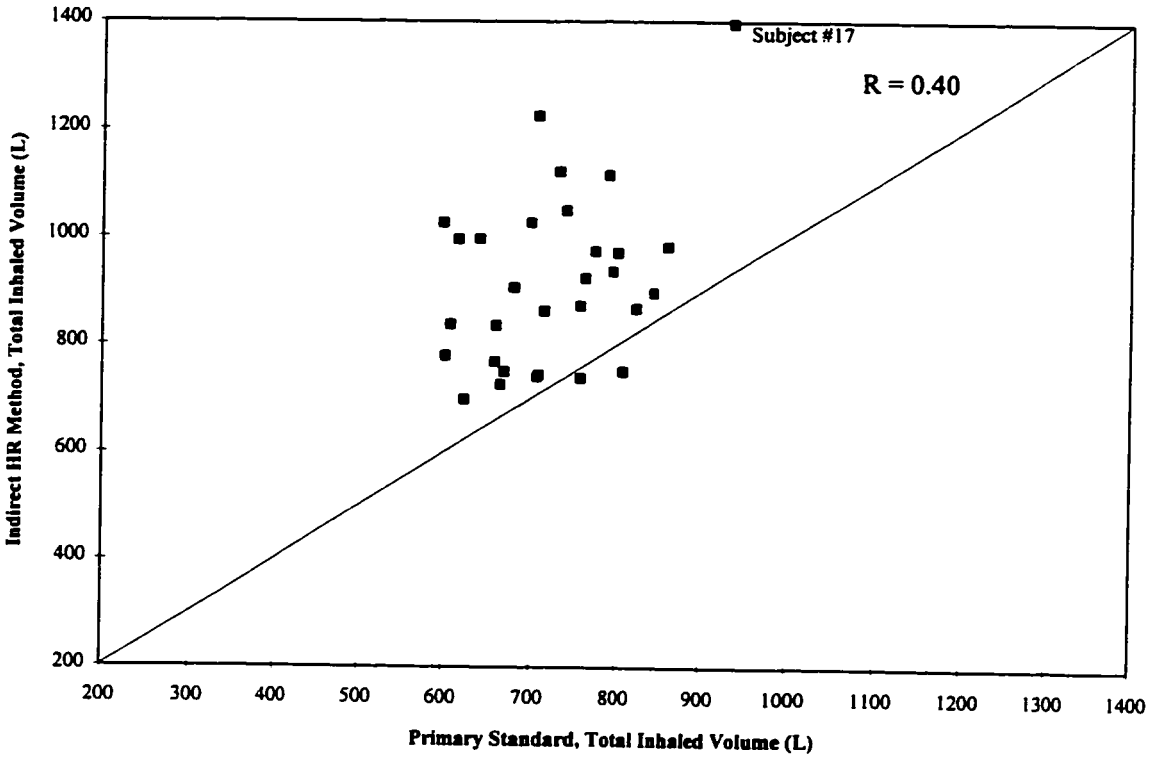


Figure 33: Plot of Indirect HR Method versus Primary Standard PSCU results. The line is a line of unity for the Primary Standard.

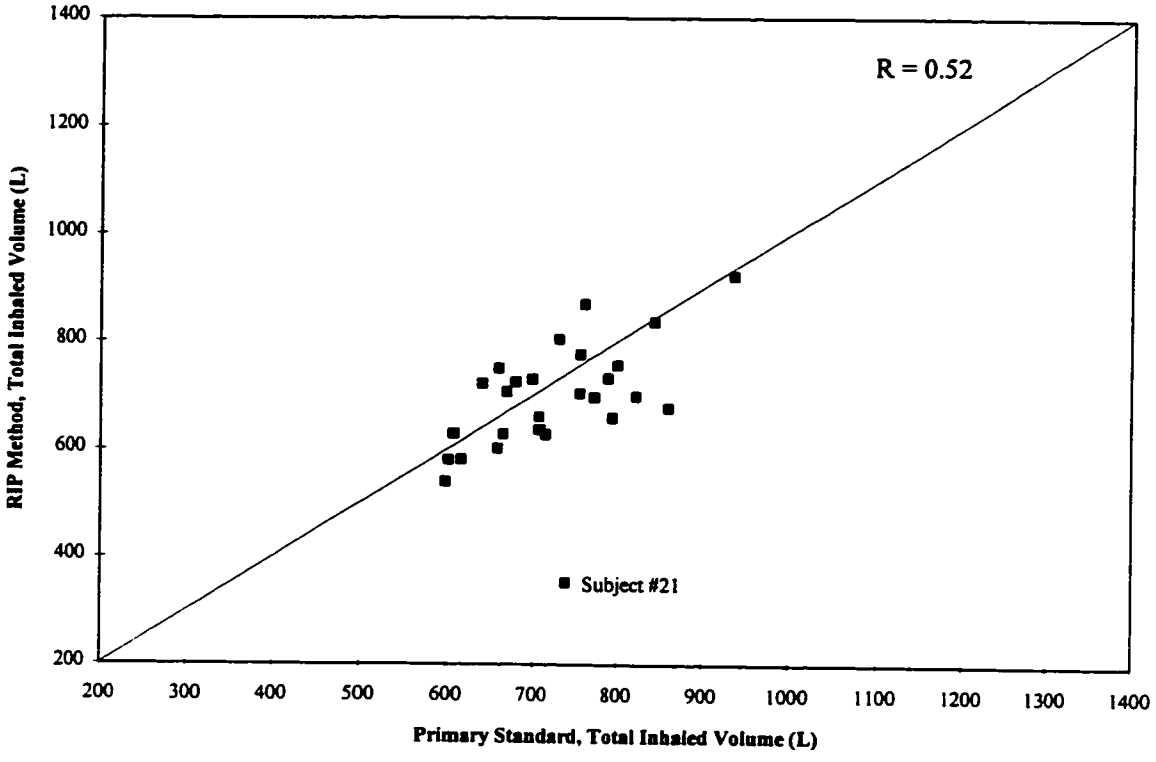


Figure 34: Plot of RIP Method versus Primary Standard PSCU results. The line is a line of unity for the Primary Standard.

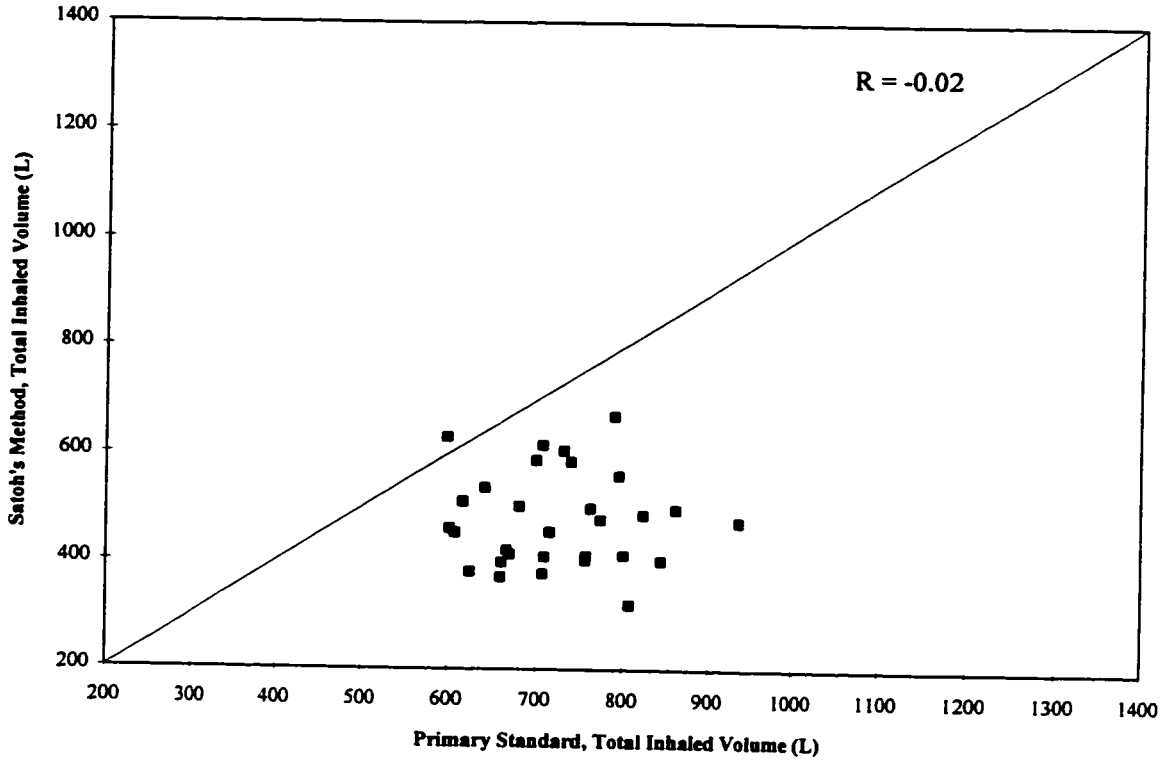


Figure 35: Plot of Satoh's HR Method versus Primary Standard PSCU results. The line is a line of unity for the Primary Standard.

Figure 36 graphs the Indirect HR Method results against the Direct HR Method results.

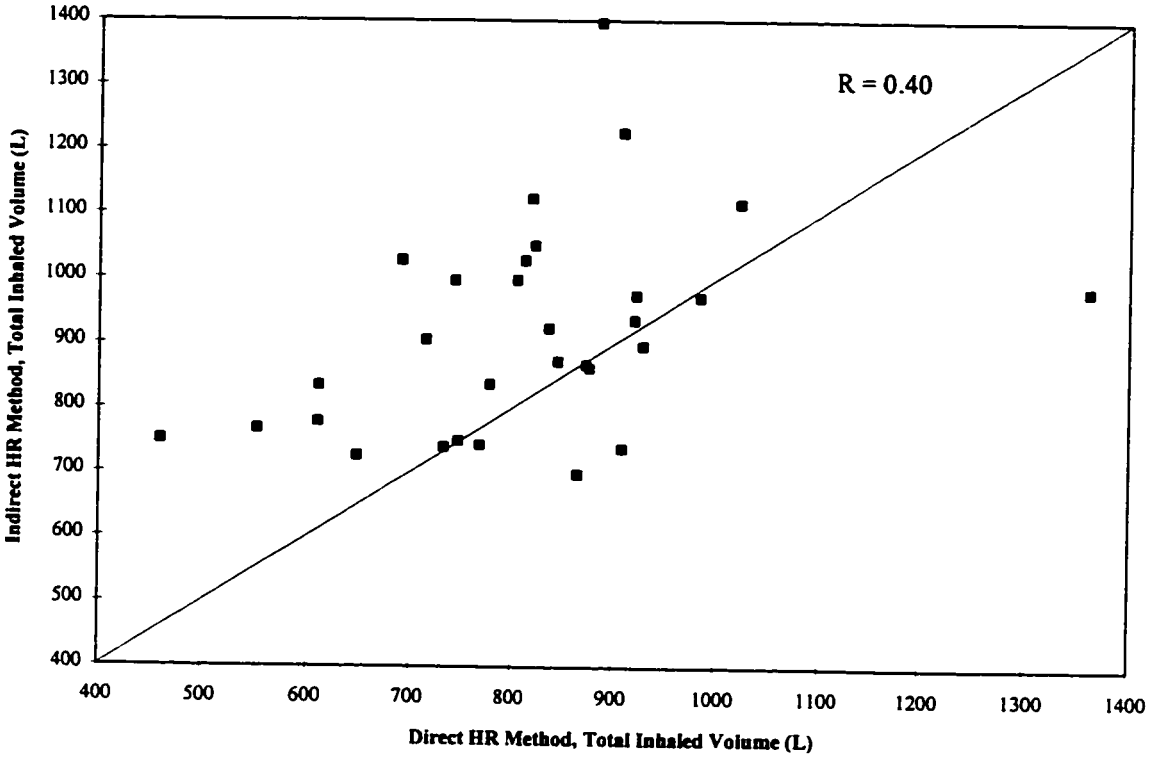


Figure 36: Plot of Indirect HR Method versus Direct HR Method PSCU results. The line is a line of unity for the Direct HR Method.

Mass Flowmeter Total Inhaled Volume Results

The mass flowmeter total inhaled volume (liters of air inhaled over the 40-minute period) results from the first session for each subject are summarized in Table 14.

Table 14: Mass flowmeter estimates of total inhaled volume (in liters) from the first sampling period for all 30 subjects.

Subj	Pneumo	Direct	Indirect	RIP
1	781.64	979.17	907.36	693.23
2	587.25	773.88	899.87	544.14
5	688.58	940.03	769.23	549.99
6	597.94	826.02	916.72	546.55
7	602.25	678.69	813.59	708.45
8	864.27	525.02	703.53	
9	564.52	686.92	705.04	580.79
10	774.87	882.75	866.13	851.59
11	740.19	1102.11	1033.97	700.82
12	780.36	937.32	848.21	541.56
13	753.36	1012.39	616.18	743.04
14	632.99	883.67	1026.76	741.93
15	644.94	924.21	577.87	
16	565.42	721.48	646.82	599.14
17	922.14	953.55	1323.18	907.58
18	715.00	764.93	722.63	592.18
19	783.20	1017.78	914.12	689.58
20	637.86	836.71	717.36	627.79
21	724.37	881.72	1030.03	341.57
22	601.98	830.11	719.28	759.03
23	619.21	766.93	1015.02	526.04
24	624.71	634.65	781.01	590.74
25	729.45	897.39	1104.50	727.97
26	737.50	895.14	738.57	696.40
27	723.44	824.51	695.43	566.52
28	681.31	757.97	823.06	664.65
29	663.05	958.61	1168.00	598.66
30	811.45	917.91	827.30	638.15
31	808.56	948.94	804.59	822.98
32	827.15	1397.00	915.69	637.75

RPD ANALYSIS RESULTS

Figures 37 through 41 display the RPDs for total inhaled volume between each Method and the Primary Standard.

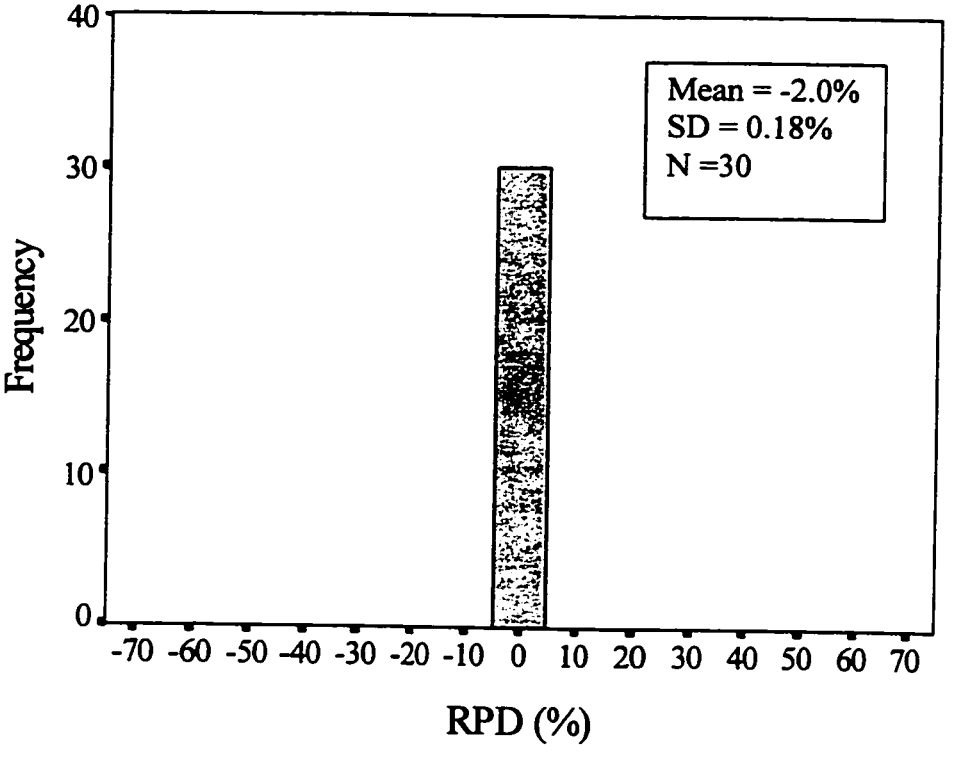


Figure 37: RPD of the total inhaled volume calculated via the Pneumo Method versus the Primary Standard.

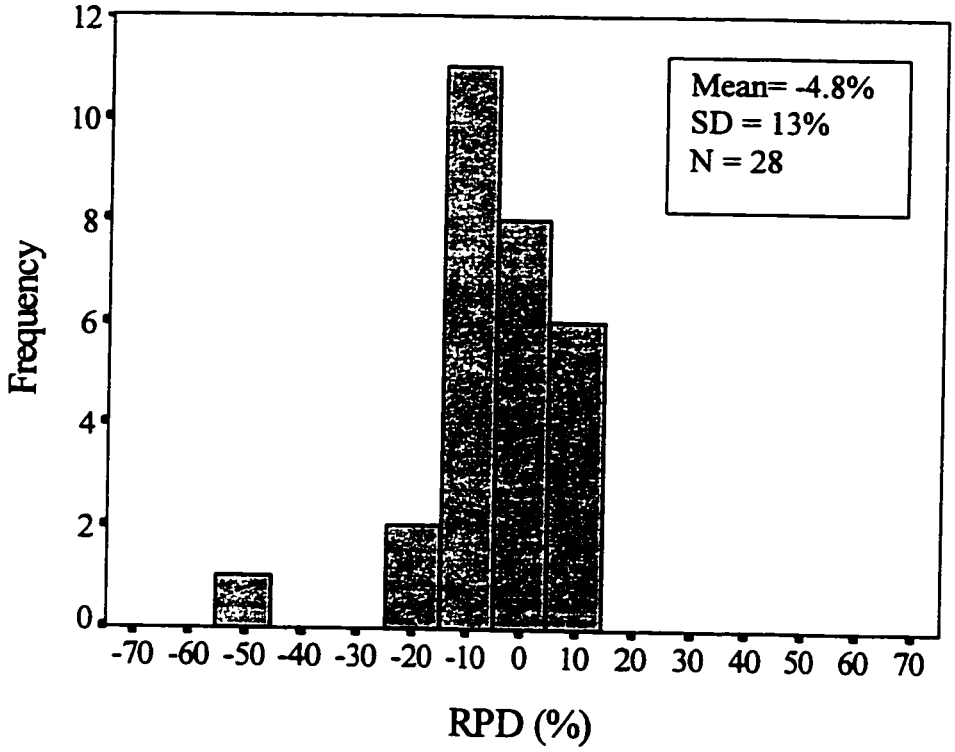


Figure 38: RPD of the total inhaled volume calculated via the RIP Method versus the Primary Standard.

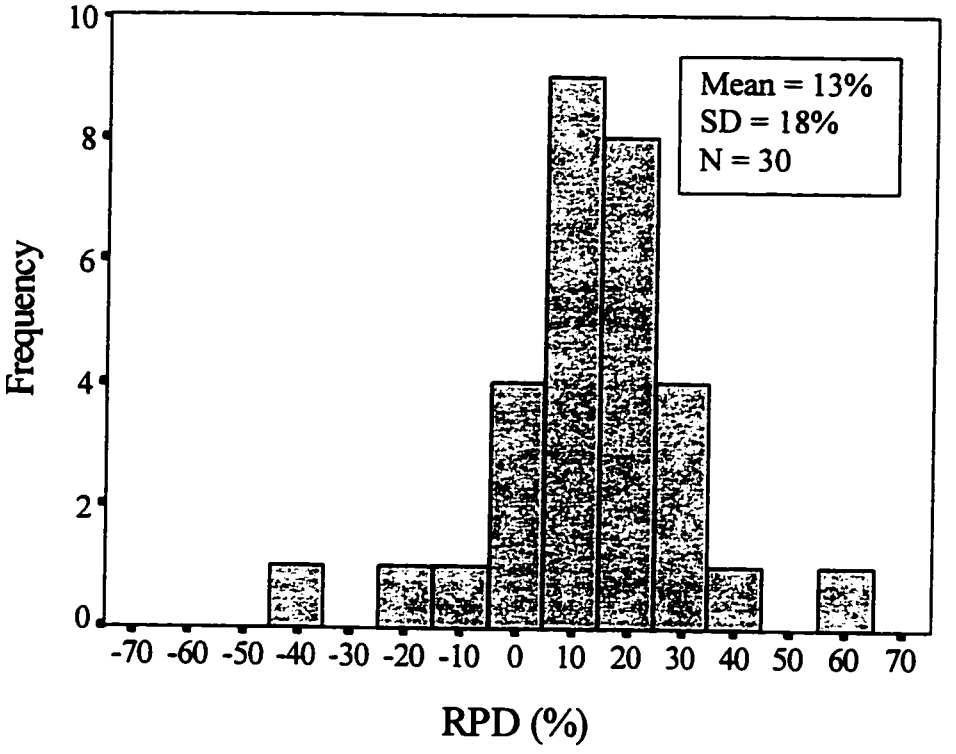


Figure 39: RPD of the total inhaled volume calculated via the Direct HR Method versus the Primary Standard.

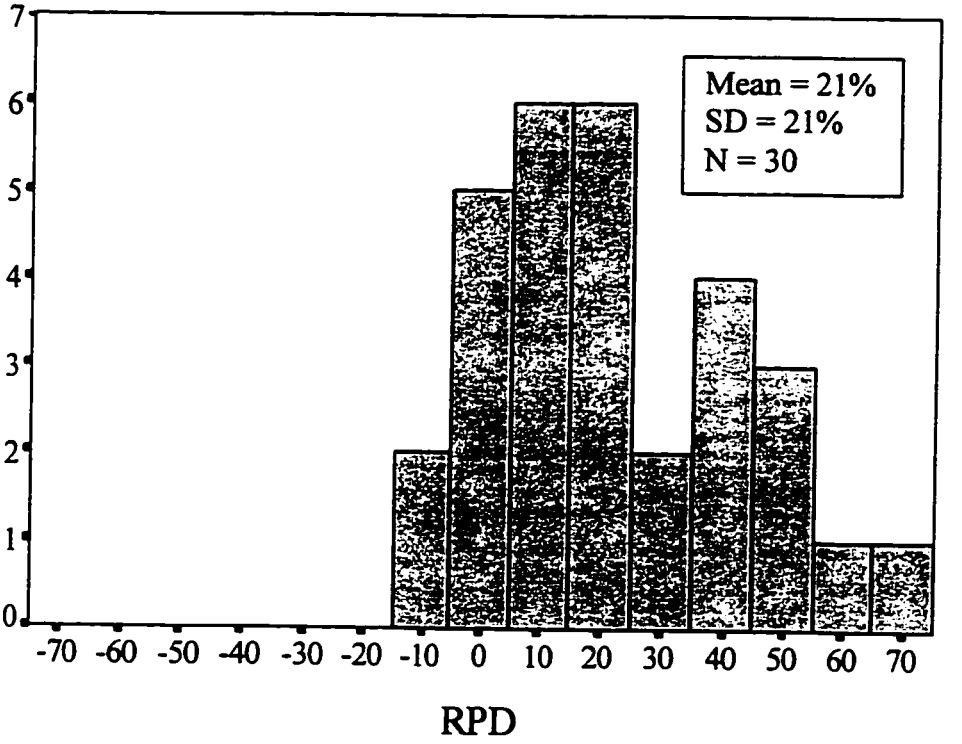


Figure 40: RPD of the total inhaled volume calculated via the Indirect HR Method versus the Primary Standard.

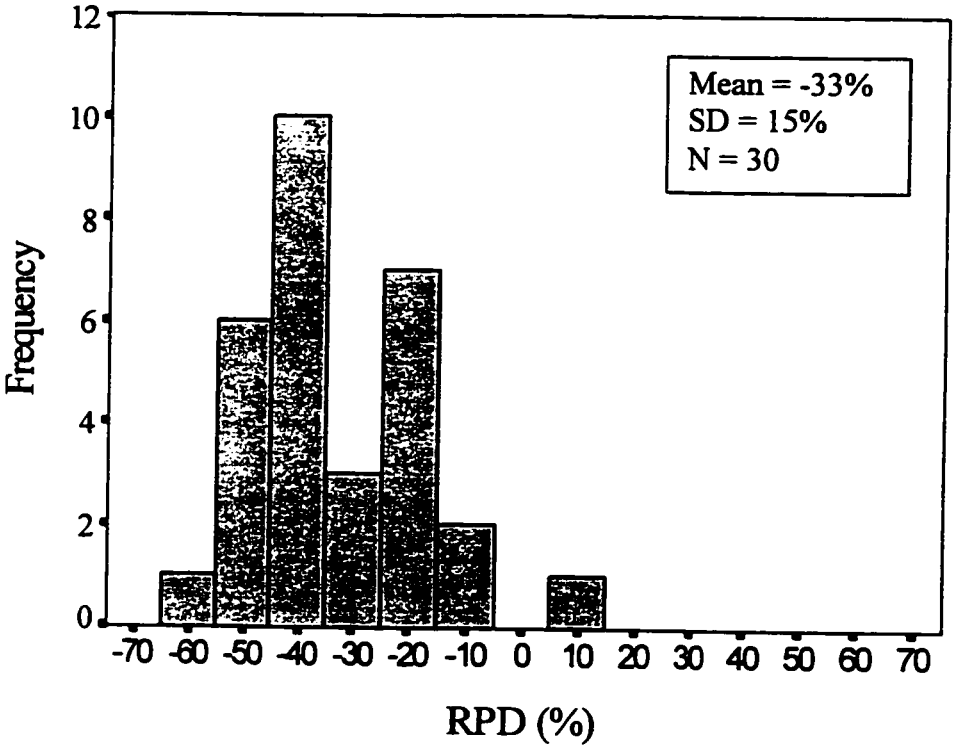


Figure 41: RPD of the total inhaled volume calculated via the Satoh's HR Method versus the Primary Standard.

CONCORDANCE ANALYSIS RESULTS

The concordance components relating the five methods for estimating minute ventilation to the Primary Standard are displayed in Table 15. In addition, this table compares the Indirect HR Method to the Direct HR Method, for purposes of assessing the ability of the Indirect HR Method to correspond to what was individually measured in the laboratory.

Table 15: Concordance component values for first session PSCU results. Statistically significant location and scale shifts are marked with asterisks (** for a P-value < 0.01, * for a P-value < 0.05).

Standard	Comparison	Location Shift	Scale Shift	Prec.	Accuracy	Concord.
Primary	Pneumo	-0.168**	0.984**	1.000	0.986	0.986
	Direct HR	0.768**	1.954**	0.544	0.654	0.356
	Indirect HR	1.645**	1.928**	0.401	0.388	0.156
	RIP	-0.381	1.303	0.525	0.903	0.473
	Satoh's	-2.845**	1.047**	-0.023	0.198	-0.005
Direct HR	Indirect HR	0.623	0.986*	0.405	0.837	0.339

The Pneumo Method had the highest concordance (0.986), as expected. It had a slight downward bias (as evidenced by a location shift of -0.168), but it had perfect precision. The Direct and Indirect HR Methods had a much lower concordance than the Pneumo Method (0.356 and 0.156, respectively). They both overestimated total inhaled volume by a significant magnitude and had essentially twice the standard deviation (reflected by a scale shift of approximately 2). Their concordance precision was about half that of the Pneumo Method. The RIP Method was not significantly different from the Primary Standard in terms of its location shift and scale shift, but its concordance precision was still unacceptably low (0.525). This method had a much higher concordance accuracy than the Direct and Indirect HR Methods, however. Satoh's HR Method significantly underestimated total inhaled volume and had a concordance precision of essentially zero.

The Indirect HR Method, compared to the Direct HR Method, was not significantly different in terms of its location shift. Its scale shift, on the other hand, was significant. The concordance accuracy was extremely high (0.939), but the overall concordance was only 0.377 due to its low concordance precision.

An additional comparison of the Direct HR Method to the Primary Standard was made when restricting the subjects in the analysis to Caucasian subjects (excluding Subjects #8 and #32). The results of this concordance analysis are displayed in Table 16.

Table 16: Comparison of concordance component values for first session PSCU results, of Direct HR Method to Primary Standard, when including all subjects versus only Caucasian subjects. Statistically significant location and scale shifts are marked with asterisks (** for a P-value < 0.01, * for a P-value < 0.05).

Subjects Included	Comparison to Primary Std.	Loc. Shift	Scale Shift	Prec.	Acc.	Concord.
All	Direct HR	0.768**	1.954**	0.544	0.654	0.356
Caucasian	Direct HR	0.847**	1.376**	0.677	0.709	0.480

Note the much lower scale shift when the analysis is restricted to Caucasian subjects. An F-test was performed to compare the variances of the total inhaled volume results for both sets of Direct HR data (all subjects versus Caucasians subjects only); the variance was significantly smaller when only Caucasians were included. The potential implication of these results will be addressed in the Discussion.

Table 17 uses concordance components to compare mass flowmeter estimates of total inhaled volume to PSCU estimates.

Table 17. Concordance component values for first session PSCU versus mass flowmeter results. The location and scale shifts were not statistically significant at the 0.01 level.

Standard (PSCU)	Comparison (mass flow)	Loc. Shift	Scale Shift	Prec.	Accuracy	Concordance
All	All Methods	0.025	1.056	0.967	0.998	0.965

Note that the location shift and scale shift were insignificant at the 0.01 level. This explains the almost perfect concordance accuracy (0.998). The concordance precision, while not perfect, was extremely high (0.967). The overall concordance was therefore 0.965.

Figure 42 displays the relationship of the mass flowmeter estimates of total inhaled volume for all Methods to the PSCU estimates. Note that the line in the graph represents a line of unity and is not a regression line.

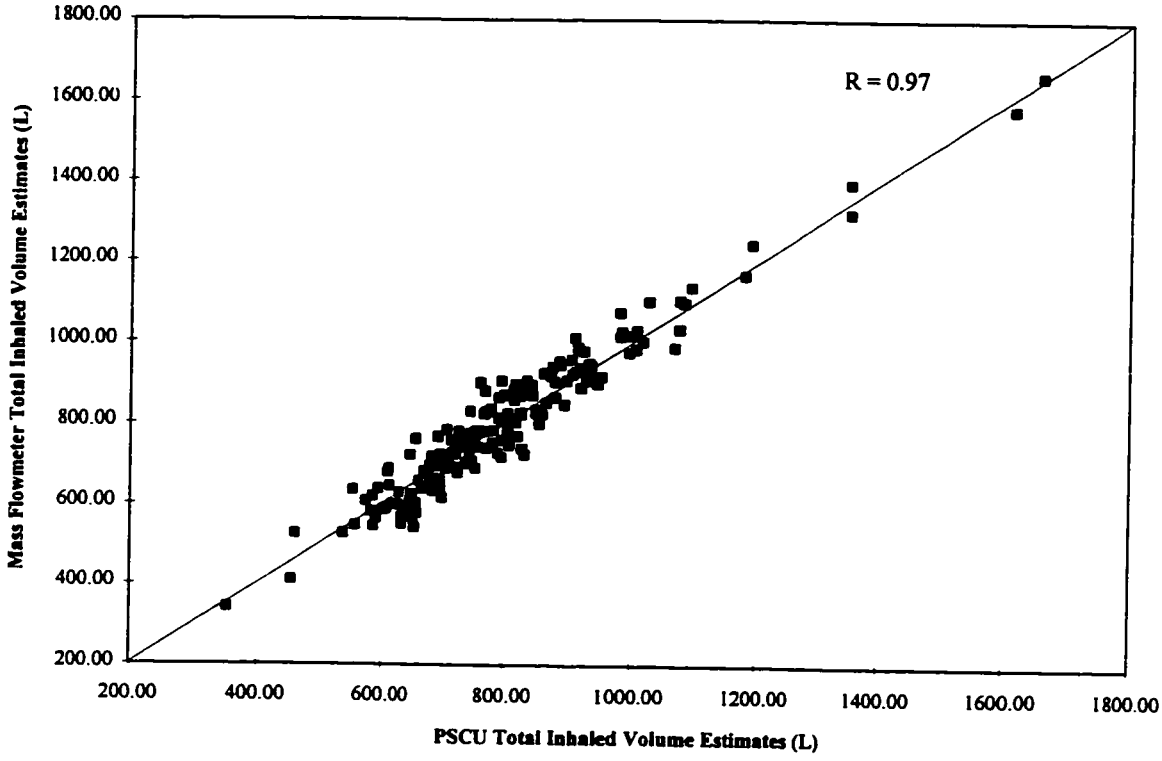


Figure 42. Relationship of mass flowmeter estimates for all Methods to PSCU total inhaled volume estimates. The line is a line of unity for the PSCU estimates.

DYNAMIC PUMP RESPONSE RESULTS

The original voltage sent to the PSCU is displayed in Figure 43.

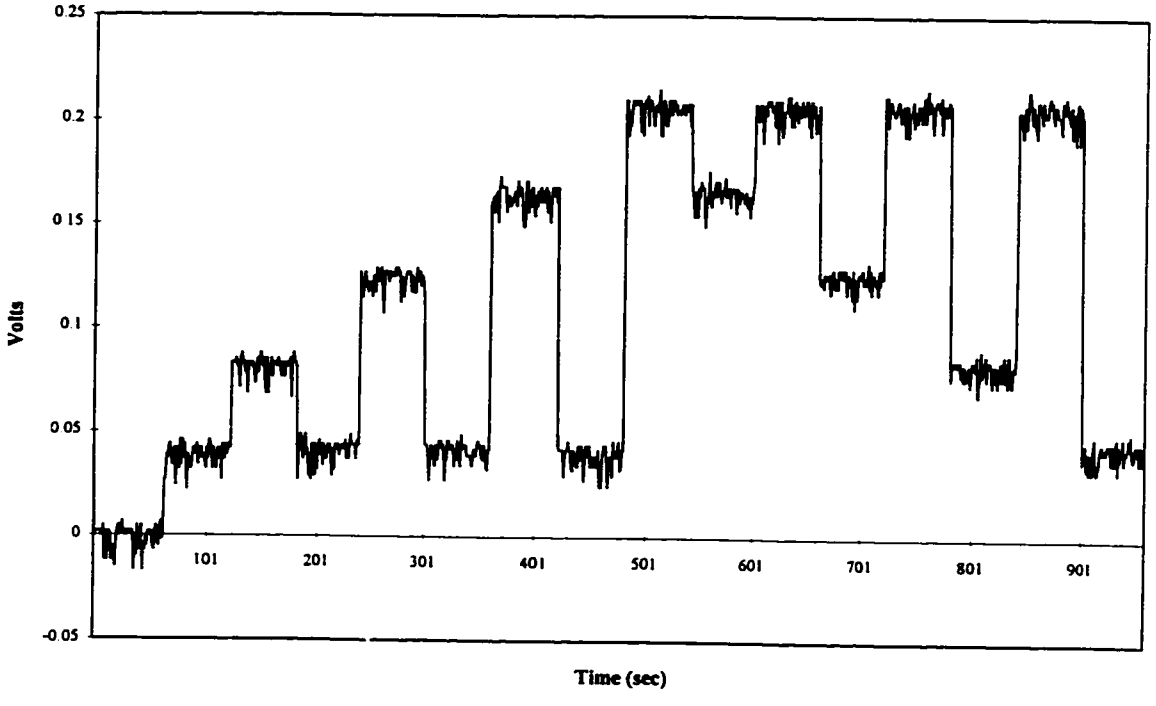


Figure 43: Original input voltage to PSCU (60 second steps).

Figure 44 shows a comparison of responses of all four pumps to the same input voltage, when the step length was 15 seconds. Note that except for an offset difference, the responses appear to be indistinguishable.

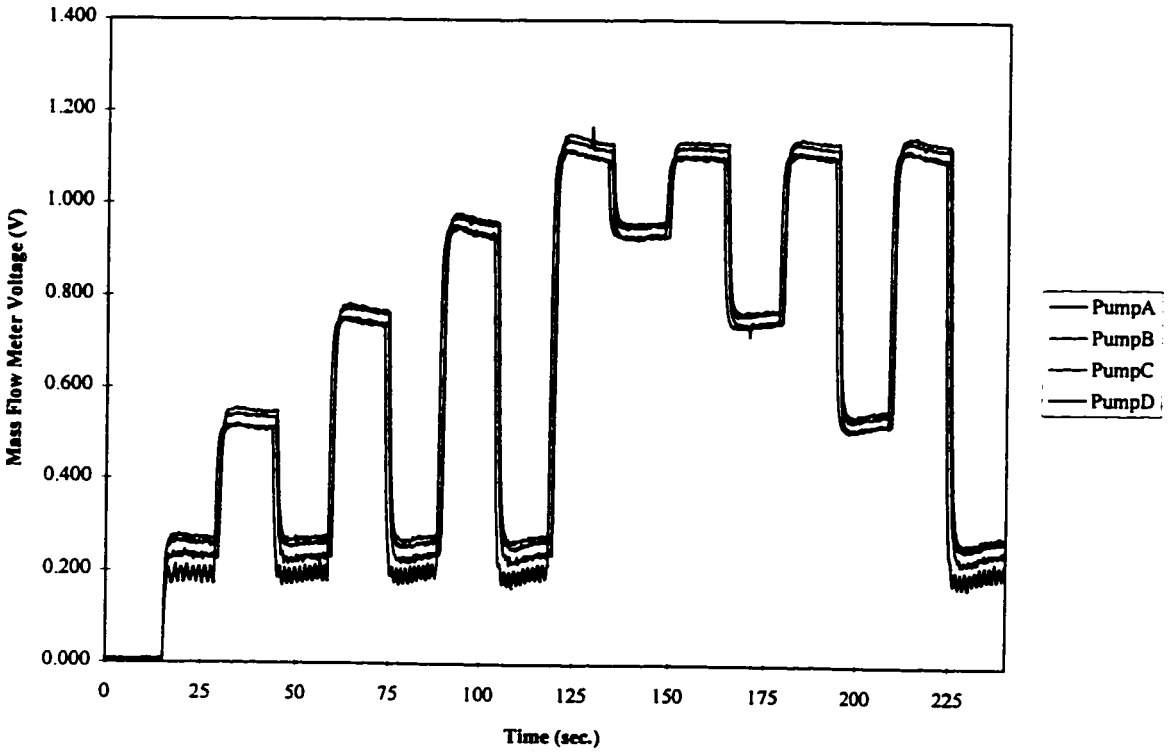


Figure 44: Mass flowmeter voltages, measuring the flows generated by the pump, for pumps A, B, C, and D, using 15-second steps.

Figure 45 shows the relationship between the mass flowmeter voltage, when using the pump labeled "A", during a 0.15 voltage positive step and a 0.15 voltage negative step. Note that the negative step was inverted in order to facilitate a visual comparison.

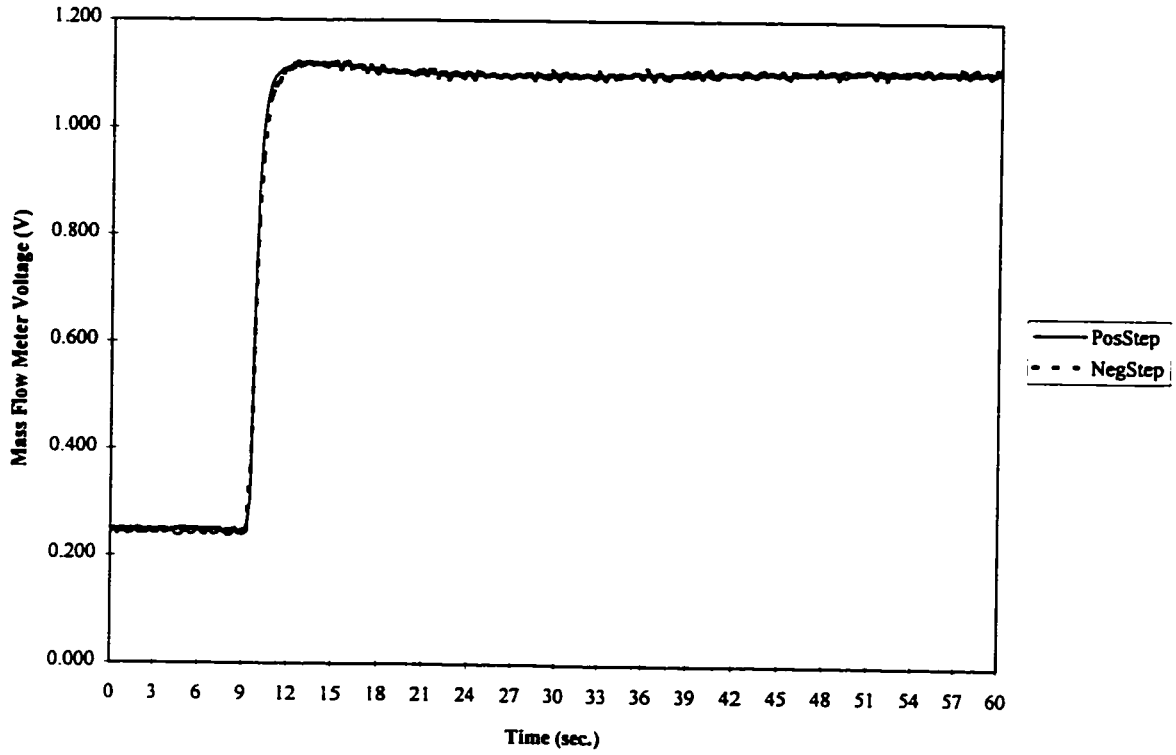


Figure 45: Comparison of mass flowmeter voltage in response to the same sized step using pump A. Solid line represents a positive step. Dashed line represents a negative step (that was inverted).

Figure 45 suggests that the pump and mass flowmeter system behaved in a symmetrical fashion with respect to transients for increasing and decreasing flow rates; i.e. there was no evidence of hysteresis.

Figure 46 displays the mass flowmeter voltage, using the pump labeled "A", under two conditions: (1) the mass flowmeter measured the pump flow throughout the procedure, and (2) the mass flowmeter was introduced at 30 seconds into each step.

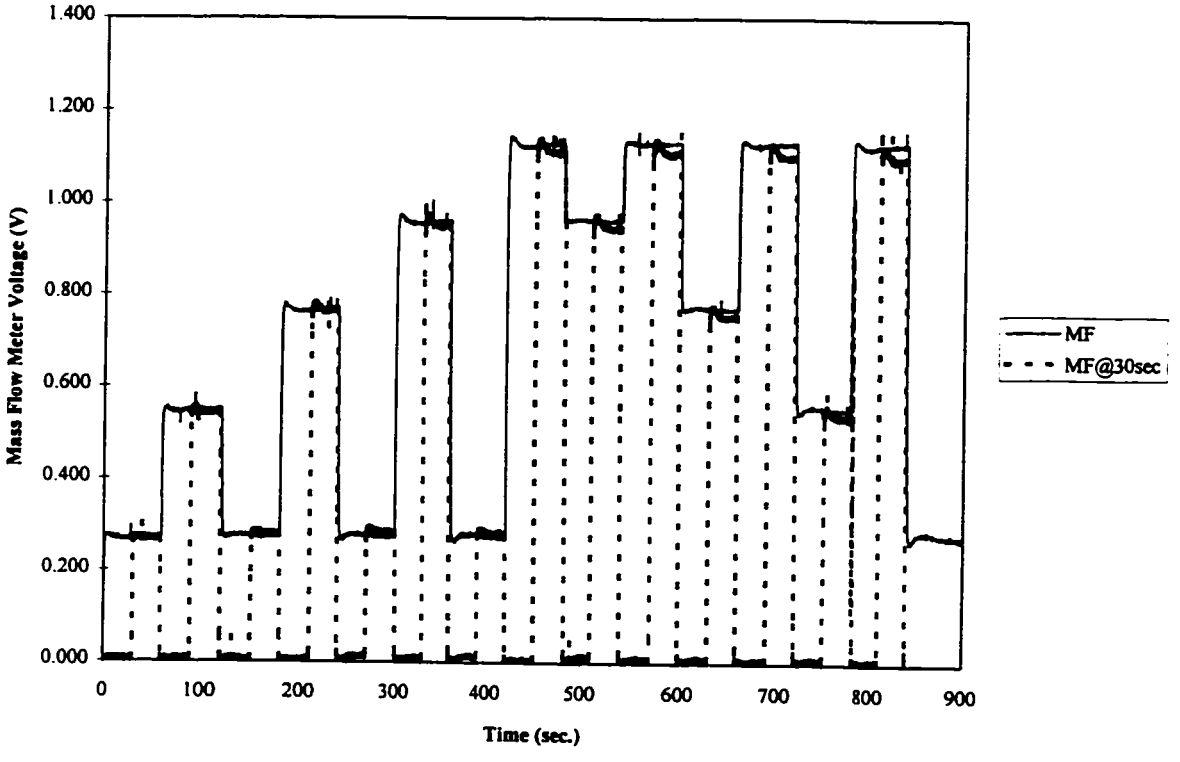


Figure 46: Mass flowmeter voltage during step input to the sampling pump, under two conditions: (1) flow through mass flowmeter was continuous (solid line), (2) flow through mass flowmeter was commenced at 30 seconds into each voltage step (dotted line).

In Figure 47, the fourth step in Figure 46 is examined more closely. In addition, the mass flowmeter response at 30 seconds (under Condition #2), was adjusted to overlay the mass flowmeter response at 0 seconds into the step (under Condition #1).

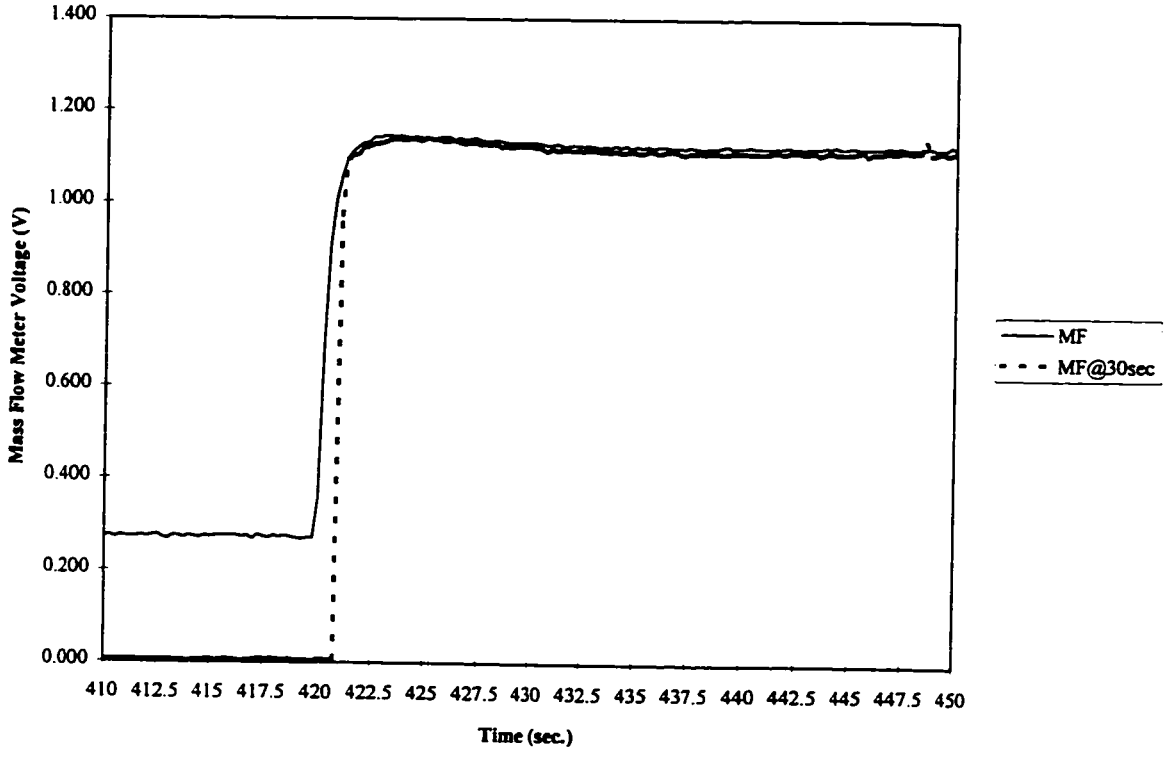


Figure 47: Same as Figure 46, except that the range is restricted to largest positive step and mass flowmeter response at 30 seconds into step shifted to overlap with original recorded values.

Based on the response of the pump-mass flowmeter system, it is predominantly a first order linear system, with a small second order component. Figure 47 suggests that the second order component can be attributed to the mass flowmeter response. The mass flowmeter documentation states that the mass flowmeter time constant is 0.08 seconds. The time constant for the mass flowmeter in this experiment was calculated using the data from Condition #2 for the step displayed in Figure 47; it was approximately 0.08

seconds, as expected. The time constant for the pump-mass flowmeter system was then calculated using the data from Condition #1 for the same step; this time constant was approximately 0.33 seconds. Using the law of superposition, the time constant for the pump was estimated to be 0.25 seconds.

Since it takes three to four time constants for a system to stabilize, we can assume that the pump stabilizes at each new flow rate within a second.

DISCUSSION

This set of experiments was successful in demonstrating that the PSP system works as expected. The average RPD between the Pneumo Method and the Primary Standard was –2% and the RPD SD was close to zero (0.18%). Note that the negative bias of –2% appears to be caused by an electrical bias in the PSCU, which could easily be corrected.

In addition, the pump responded rapidly to a voltage change, stabilizing at its new flow rate within a second. The magnitude of the error resulting from "wind-up time" was not significant for the protocol that the subjects performed; this conclusion was based on the fact that the mass flowmeter results were not significantly different from the PSCU results.

When the PSP used any of the Methods other than the Pneumo Method to estimate minute ventilation, the results did not fulfill the accuracy and precision goals. Both the RIP Method and the Direct HR Method met the accuracy goal (with average RPDs of -4.7% and 13%, respectively); however, none of the Methods met the precision goal (with RPD SDs ranging from 13 to 21%).

Concordance analysis demonstrated that the Direct HR Method had a significant location shift and scale shift. There are several possible reasons for this:

1. The calibration curves were generated at the start of the subject session, immediately after subjects were fitted with the sampling equipment. It is possible that the subjects tended to hyperventilate at a given work rate due to the effect of wearing a mouthpiece and other unfamiliar devices. This hypothesis is supported by Figure 24, which shows a larger range of minute ventilation values for the first two work rates during Session #1 when compared to Session #2.

2. The calibration curve was based on the data ranging from 0 to 80 Watts of work and the subjects performed work at 120 Watts during the sampling protocol.

3. The calibration curve was based on only three data points.

Figures 48 and 49 show cases that could be explained by reasons listed above. Each figure shows a timeline of minute ventilation estimates from both the Pneumo Method and the Direct HR Method. Subject #11's data is displayed in Figure 48 and Subject #18's data is displayed in Figure 49.

For Subject #11, the Direct HR Method overestimated minute ventilation for the entire length of the session (see Figure 48).

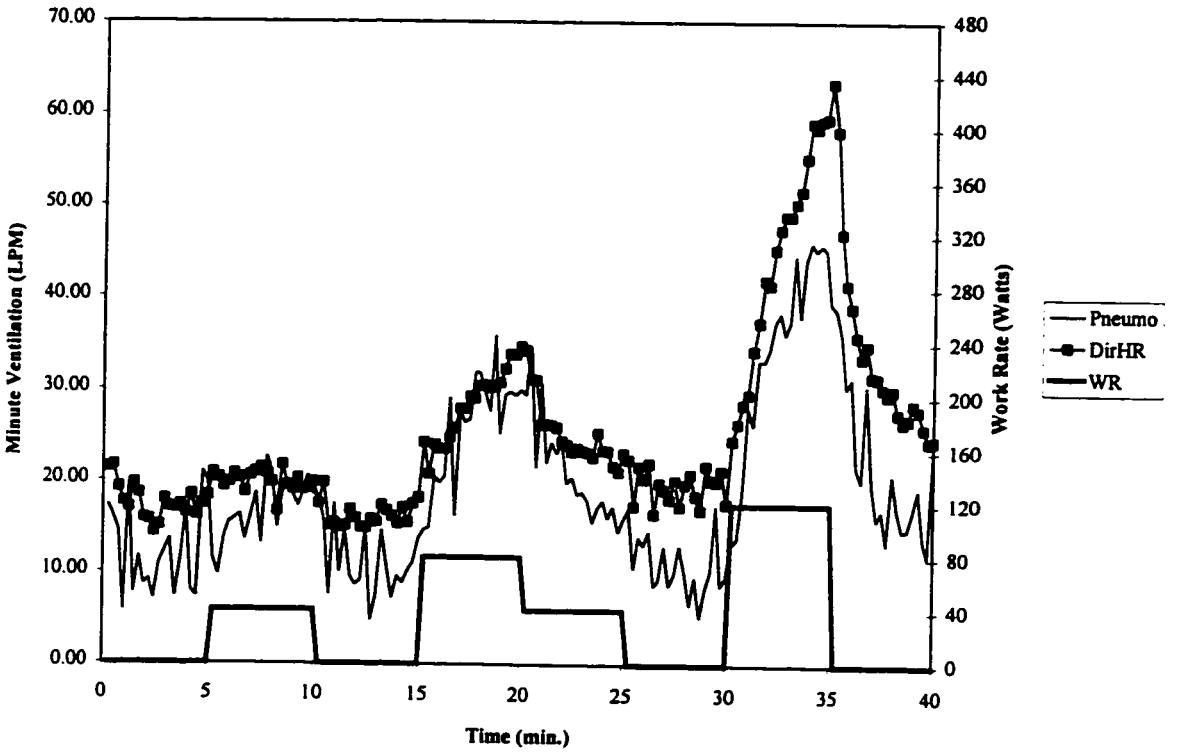


Figure 48. Timeline of minute ventilation estimates for Subject #11 using the Pneumo Method and Direct HR Method.

For Subject #18, the Direct HR Method does a much better job of tracking the Pneumo Method (see Figure 49). There are two interesting things to note in this graph. First, the Direct HR Method clearly overestimates minute ventilation when the subject is biking at 120 Watts (see the period from 30 to 35 minutes). Second, the lag of the respiratory system to a change in work rate is evident; the Direct HR Method's estimates always respond first. This leads to overestimation of minute ventilation at the start of a positive step (see 15 minutes on the graph) and an underestimation of minute ventilation at the start of a negative step (see 20 minutes on the graph).

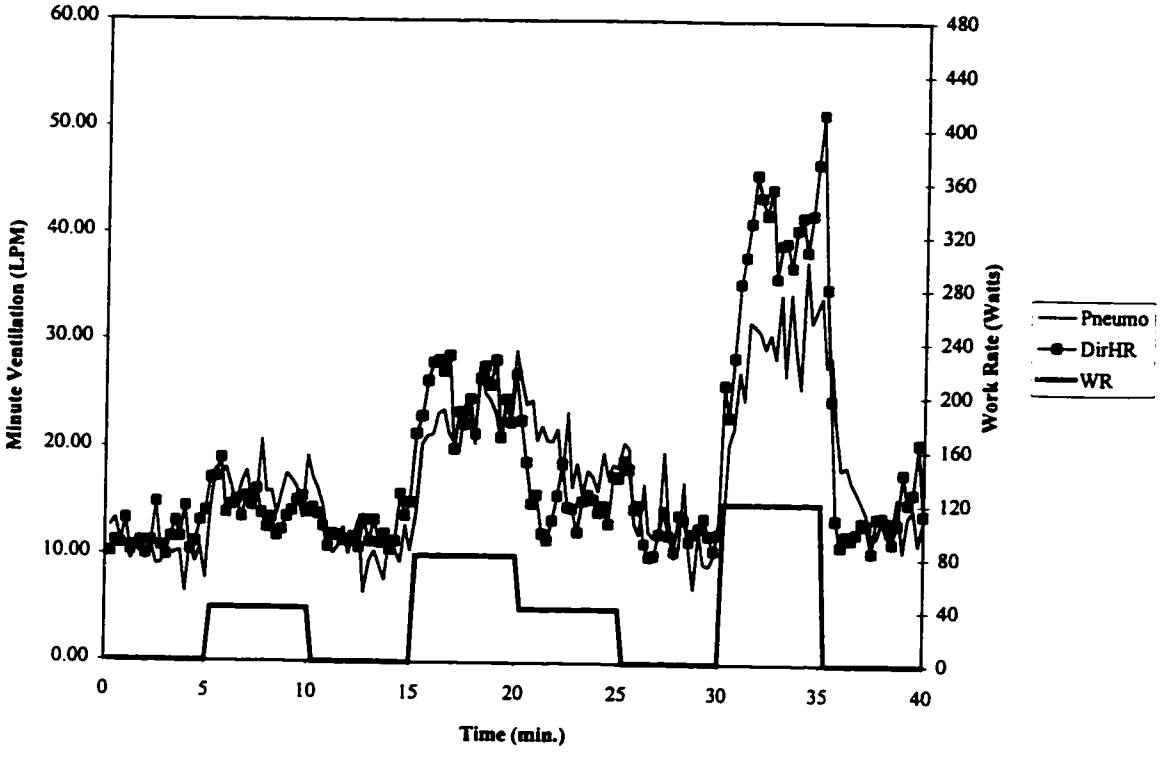


Figure 49. Timeline of minute ventilation estimates for Subject #18 using the Pneumo Method and Direct HR Method.

When examining the relationship of the Direct HR Method to the Primary Standard (as in Figure 32), there are two obvious outliers; these outliers represent the only non-Caucasian subjects. When the concordance analysis was restricted to Caucasian subjects (as in Table 16), both the concordance accuracy and precision of the Direct HR Method were increased such that the overall concordance increased from 0.356 to 0.480; while this could be attributed to coincidence, it is interesting to note that Satoh's HR Method, which was derived using Asian subjects, was essentially useless in predicting minute ventilation in this experiment. The implication is that perhaps people of different ethnicities have different heart rate to minute ventilation relationships, and that they cannot be characterized in the same manner. If this hypothesis were true, one would

expect that when Satoh's HR Method was used to estimate minute ventilation, the Asian subject in this experiment would have a total inhaled volume estimate closer to the Primary Standard value; when Satoh's HR Method was used, however, Subject #8's total minute ventilation was underestimated by 60%, as opposed to 43% when the Direct HR Method was used. One cannot reject the hypothesis, however, based on the results from one individual.

As expected, the RPD analysis showed that the Indirect HR Method had poorer accuracy and precision than the Direct HR Method. It would be unreasonable, after all, to expect a predictive equation to perform better than an individually determined calibration curve.

Concordance analysis demonstrated that the Indirect HR Method tended to overestimate minute ventilation compared to the Direct HR Method (as evidenced by the positive location shift). The Indirect HR Method actually had a concordance accuracy of 83.7%, when the Direct HR Method was used as its standard. The Indirect HR Method's low concordance precision, however, led to an overall concordance value to 0.339.

It should be noted that the Indirect HR Method did perform as well as, or in some cases, better than, the Direct HR Method for some subjects. Figure 50 shows a case where the Indirect HR Method performed almost identically to the Direct HR Method, although they both overestimated the true minute ventilation.

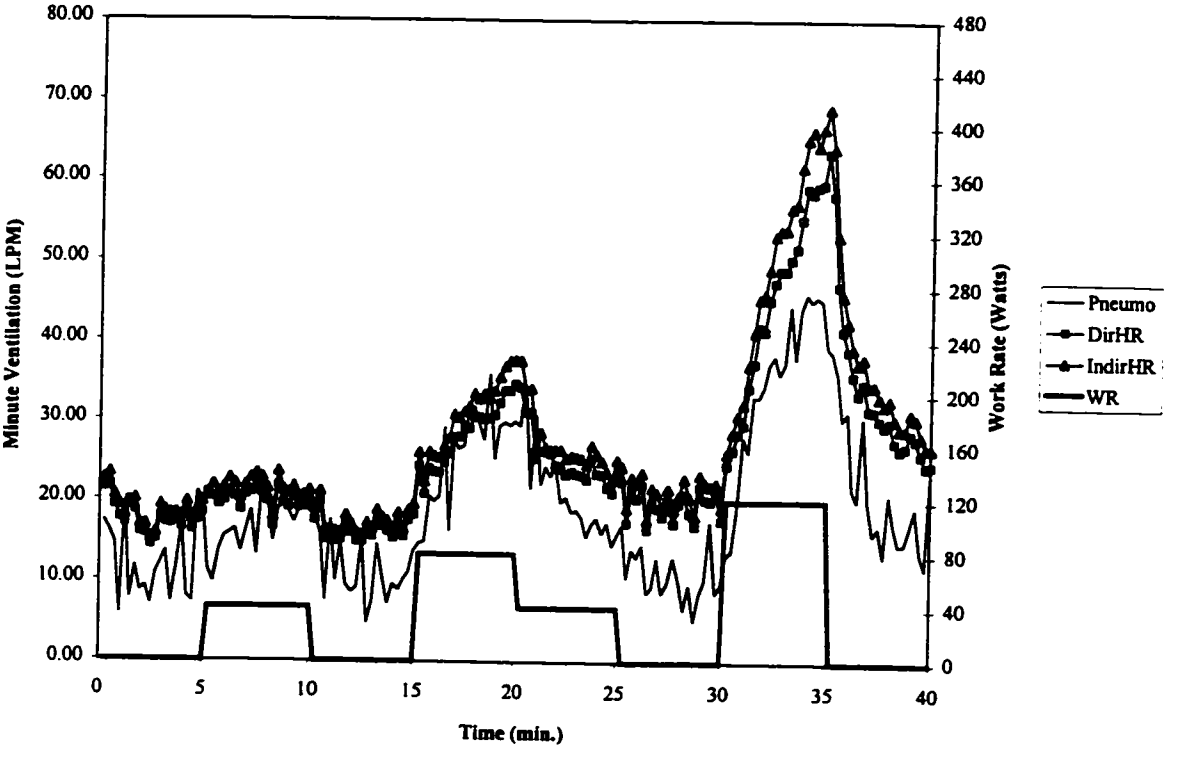


Figure 50. Time line of Pneumo Method, Direct HR Method, and Indirect HR Method, for Subject #11.

Figure 51 shows a case where the Indirect HR Method performed better than the Direct HR Method.

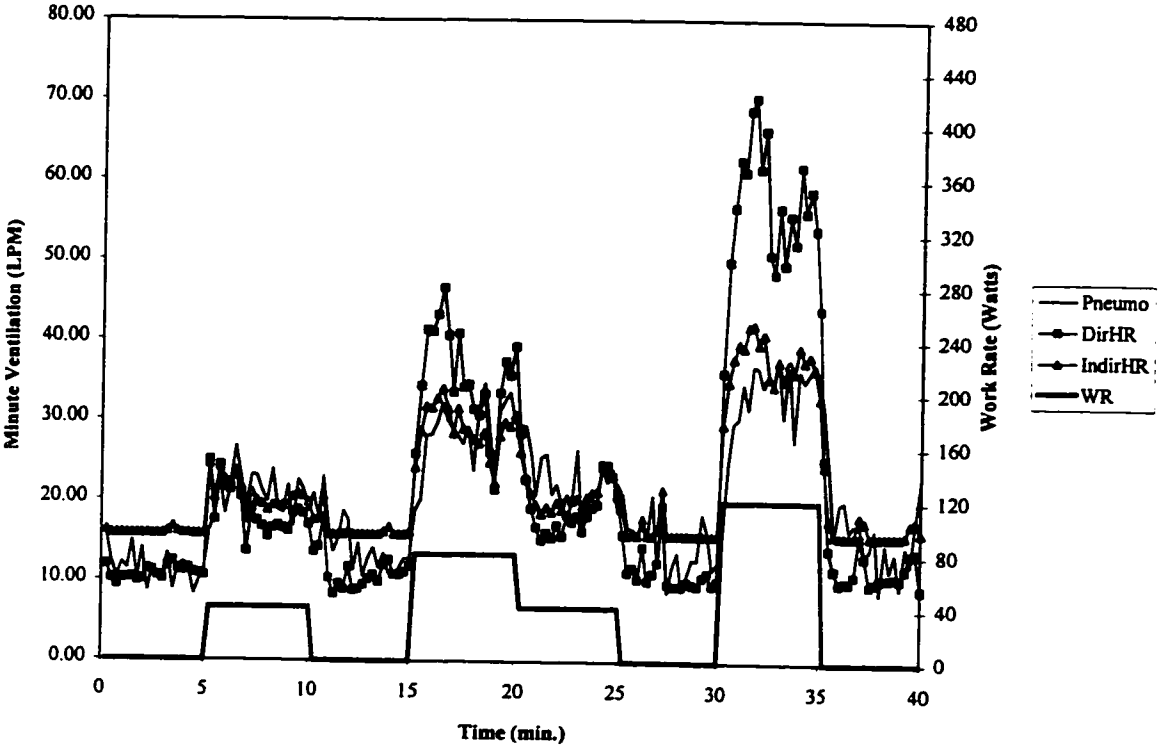


Figure 51. Time line of Pneumo Method, Direct HR Method, and Indirect HR Method, for Subject #30.

Figure 51 also provides an example of what happens when the estimated resting heart rate for a subject is actually higher than the true value. When the heart rate falls below the pre-determined resting heart rate, the minute ventilation estimates are clipped, as seen at each of the rest periods in Figure 51 (0 to 5 minutes, 10 to 15 minutes, 25 to 30 minutes, and 35 to 40 minutes). This effect could have contributed to (1) inflating the location shift and (2) decreasing the concordance precision of this method.

When compared to Satoh's HR Method, the Indirect HR Method's concordance accuracy and precision were approximately 20% and 40% higher, respectively; while this is a

significant improvement, it is very likely that a better predictive equation could be developed using data from a population that is more specifically representative of the population to which the predictive equation will be applied.

A subset of the database population was tested, for example, in which all of the severely diseased subjects as well as the subjects taking heart rate-influencing medication were excluded (in addition to some others); a total of 72 subjects were excluded. When the coefficients derived from this database subset were used instead of the coefficients presented in the Results, the concordance accuracy of the Indirect HR Method improved by 10 percent (the concordance precision remained unchanged). What this suggests is that the concordance accuracy is strongly influenced by the population from which the equation coefficients are derived. As mentioned in the Methods, the database population used for this study was not representative of a normal work population; many of the subjects performed the exercise test for medical purposes, and others had a history of Asbestos exposure. It is quite possible that an acceptable predictive equation could be developed if the following steps are taken:

1. Carefully select representative subjects from work environments.

2. Calibrate subjects at steady-state, over a work rate range up to at least 120 Watts.

3. Collect data at as many work rates as possible in order to accurately characterize the calibration curve.

4. Perform the calibration more than once per worker and do not use the first set of data to generate the predictive equation (due to the novelty of the situation, the first data set may not reflect the subject's normal heart rate to minute ventilation relationship).

5. Consider stratifying the database by a factor such as age, for example, and derive coefficients for the predictive equation for each stratum.

An additional advantage of collecting the data specifically for a predictive equation database is that possible predictive variables can be more accurately characterized. For example, the calibration curve comparison results suggest strongly that fitness should be a significant factor in predicting the heart rate to minute ventilation relationship (see Figures 28 through 30). The reason that it was not found to be significant in the predictive equation likely relates to the imprecise method of estimating fitness level for the database population; since the data was extracted from records, the subjects were separated into four fitness categories based on the written comments of the physician and/or the maximum oxygen consumption values. When the Direct HR Method was examined for a fitness effect, it was not determined to be significant either; this could have resulted from the limited characteristics of and/or size of the laboratory population. The method of categorizing fitness (by number of times the subject exercised per week) could also have led to enough misclassification of fitness to prevent an effect from being observed.

Of all the Methods (excluding the Pneumo Method), the RIP Method performed the best. After a potential outlier (Subject #21) was removed, it was able to meet the precision goal of a RPD SD less than 10%. Figures 52 and 53 display timelines of the RIP Method estimates that show the range over which it performed. For Subject #11, the RIP Method performed extremely well. Notice how closely it tracks the Pneumo Method's estimates.

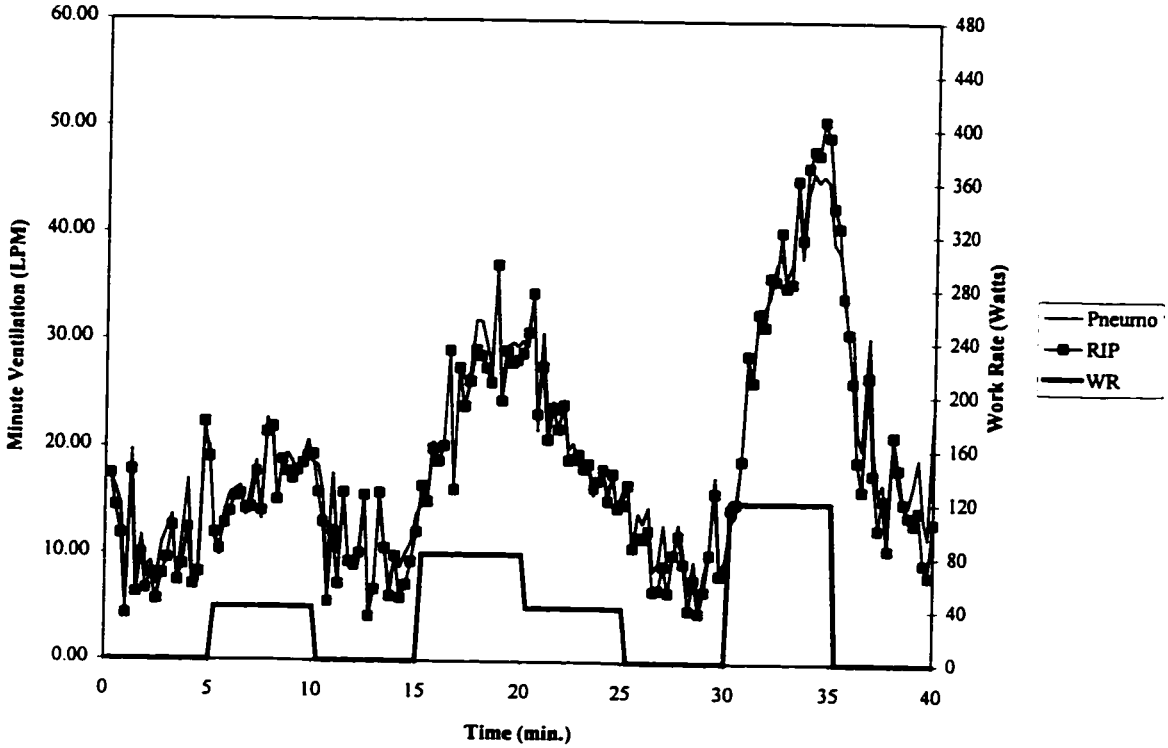


Figure 52. Timeline of minute ventilation estimates for Subject #11 using the Pneumo Method and the RIP Method.

For Subject #32, however, the RIP Method did not track the minute ventilation very accurately, as can be seen in Figure 53.

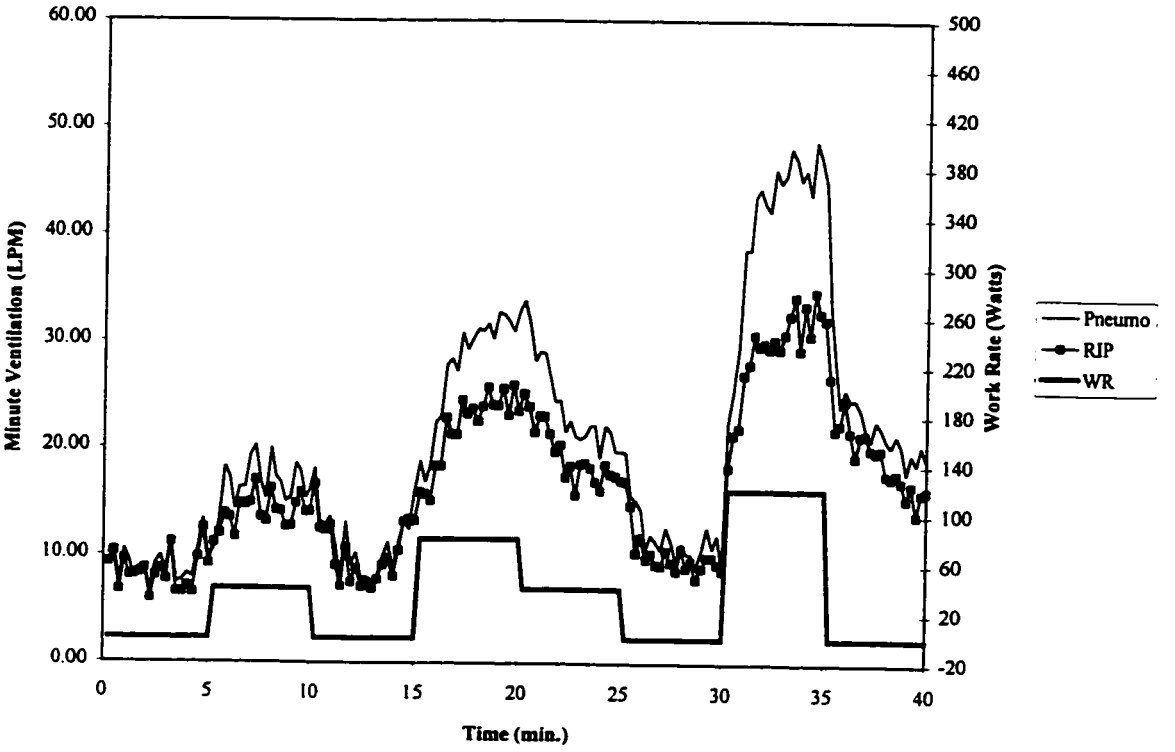


Figure 53. Timeline of minute ventilation estimates for Subject #32 using the Pneumo Method and the RIP Method.

The respiratory inductive plethysmograph had several characteristics that presumably contributed to the low concordance precision of the RIP Method. The RIP bands, even attached to the stress test shirt, tended to move. The abdominal band, in particular, would often ride up over a roll of abdominal tissue, thereby affecting the calibration of the system. The velcro was often stretched to the point of failure (and in some cases the velcro came off the bands altogether, thereby accounting for the missing RIP Method data for Subjects #8, 15, and 21).

The RIP Method is not a plausible Method for use in the field in its current form. One of the main reasons it was included in this study was to provide some reasonable alternative measure with which to compare the heart rate-based Methods. What the RIP Method results suggest is that achieving the specified accuracy and precision goals, using heart rate as a surrogate, could be difficult.

The question arises as to how to improve upon the heart rate-based Methods presented in this Chapter. The goal is, of course, a predictive equation; however, if we cannot fulfill the accuracy and precision goals using individually determined calibration curves, then developing a predictive equation would be pointless. Therefore, focusing on a Direct HR Method first would be prudent.

There are two possible approaches for improving upon the Direct HR Method described in this experiment. The first would be to assume that steady-state values remain the basis of the calibration curve, and refine the calibration protocol by: (1) increasing the number of work rate levels to increase the number of data points, and (2) increasing the range by adding work rates greater than 80 Watts.

The second approach would be to examine the transient relationship of minute ventilation to heart rate in order to derive the transfer function that relates heart rate (the input) to minute ventilation (the output). Transfer functions are used to relate inputs to outputs, without regard to the mechanism(s) by which the relationship is generated. The mechanism(s) are considered a "black box", and the transfer function mathematically generates the output for any given input into this black box. While transient relationships between heart rate and minute ventilation are not readily available, one can use work rate as a forcing factor for both variables; the transient relationships between work rate and heart rate and between work rate and minute ventilation have been investigated and should provide enough information to derive the transfer function in question.

Note, however, that in Chapter 1 it was emphasized that there is probably a difference in the relationship between heart rate and minute ventilation when performing upper body exercise versus lower body exercise. The subjects performed only lower body exercise in this study, thereby eliminating this difference as a factor. If a method is developed that relates heart rate to minute ventilation acceptably during lower body exercise, then it must be validated (and possibly adjusted) for upper body exercise.

One must keep in mind that the laboratory subject population was very restricted; the subjects were mostly Caucasian, in their 20s and 30s, and in good health. In addition, the exercise protocol that each subject performed was the same. The combination of a limited study population and identical protocol means that the ranges of the results are very restricted; these conditions could have artificially lowered the accuracy and precision, although this seems counter-intuitive. Figure 54 demonstrates how the cloud of data for a restricted population could differ from the true population; while this example is simplistic, it makes the point that under restricted circumstances, the overall trend may be more difficult to identify.

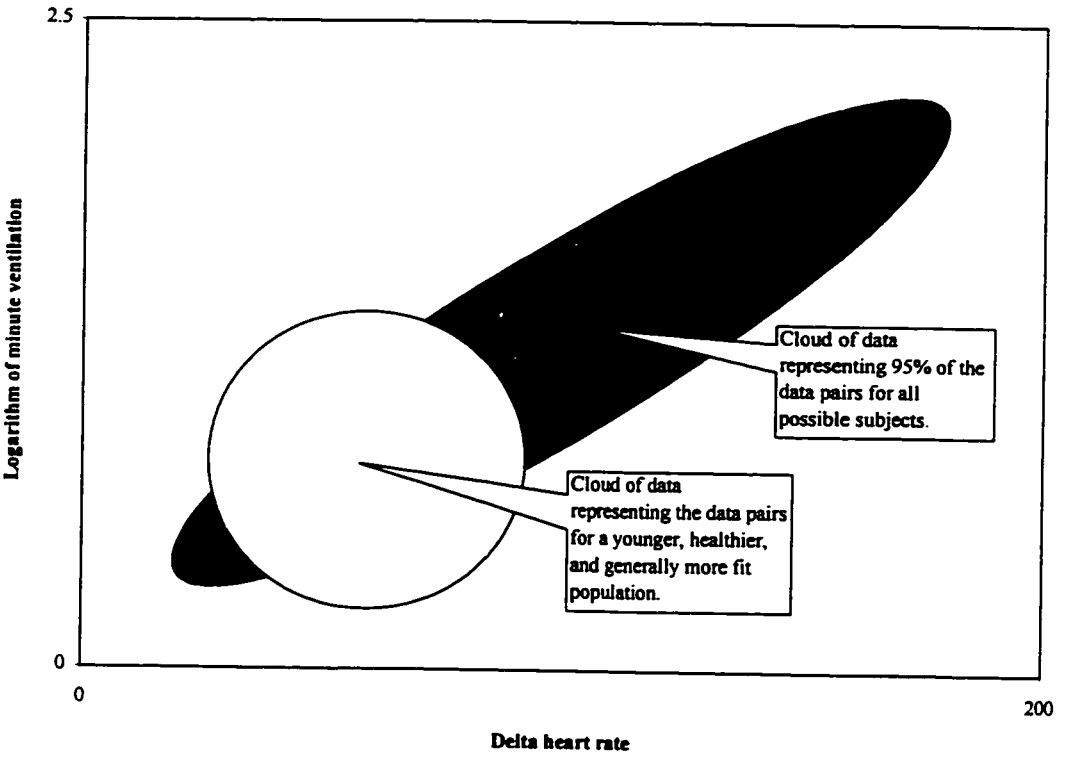


Figure 54: Example of how laboratory subject population may have differed from the true population.

CONCLUSIONS

The PSP itself performed extremely well in this laboratory study. The Methods that the PSP used to estimate minute ventilation, however, were not entirely acceptable for use of this device in the field.

The RIP Method performed the best, but the RIP itself could not withstand use in the field given its current form, which also requires complex calibration in the field.

While the Direct HR and Indirect HR Methods did not meet the a priori accuracy and precision goals, they each could be significantly improved. Protocol changes for calibration curve generation would likely improve the accuracy and precision of the

Direct HR Method, and the Indirect HR Method could be improved by generating a predictive equation using data collected from a subject population representative of workers on which PSP will be used.

The idea behind the Indirect HR Method warrants further study. The advantage of this calibration method is its inherent simplicity. One needs to know only the weight of the worker and measure his/her resting heart rate. A worker can wear a heart rate monitor during a work shift without interference, and this heart rate can either be logged via a wristwatch device (for estimating the total pulmonary ventilation over the entire work shift), or the detected signal can be sent to a device such as a physiologic sampling pump.

Finally, the accuracy and precision goals chosen for this study were very stringent. Methods that fall short of these goals are not necessarily worthless. Even in its current state, the Indirect HR Method, while not nearly optimal, does have some predictive value and may enable better stratification of workers for epidemiological or risk assessment purposes.

CHAPTER 4: PILOT STUDY OF CHARCOAL TUBE BREAKTHROUGH CHARACTERISTICS WHEN USING A PHYSIOLOGIC SAMPLING PUMP

INTRODUCTION

A physiologic sampling pump varies its pump rate in proportion to a worker's estimated inhalation rate. This raises the question, in regard to sampling strategy, of whether a variable flow rate affects the breakthrough point of a charcoal tube.

Charcoal tubes are divided into two sections: the front section and the back section. The front section contacts air first and is designed to capture organic compounds in the sampled air. The back section is necessary in order to validate that none of the contaminant of interest was able to pass through the front section without being captured. Therefore, in order to determine whether the mass collected on the front section of a charcoal tube is reflective of the true mass of contaminant in the air sampled, the back section must be analyzed. Breakthrough has occurred when the contaminant is found on the back section of the charcoal tube.

The primary purpose of this pilot study was to examine whether the breakthrough point of a charcoal tube is likely to be affected by the variability of the pump flow rate.

A secondary purpose of this pilot study was to examine the likelihood of breakthrough problems compromising the samples if the laboratory subjects had actually been exposed to toluene contaminated air (at realistic levels).

METHODS

For this experiment, the data from three laboratory subjects described in Chapter 3 were used. The subjects were selected based on the variance and mean of the flow rate of the

pump controlled by the predicted minute ventilation values (based on the Indirect HR Method). In addition, a constant flow condition was added (e.g. no variance) where the average flow rate was set to correspond to that of the subject that had the highest variance of flow rate (Subject #29). The statistics for these four conditions are displayed below in Table 18. While the average pump flow rates were much higher than the flow rates recommended in NIOSH Method 1500 (see Appendix Y), which ranged from 0.01 to 0.2 LPM, the charcoal tubes used in this experiment were four times the size of a standard charcoal tube (400mg/200mg versus 100mg/50mg).

Table 18: Four conditions tested in this experiment.

Condition #	Subject #	Average pump flow rate (LPM)	Variance of pump flow rate (LPM^2)
1	1	0.51	0.01
2	17	1.01	0.08
3	29	0.74	0.15
4	N/A	0.74	0

A solvent generator was used to create air of known toluene concentrations. This generator introduced toluene via a syringe pump into a mixing chamber with a known clean airflow rate. The syringe pump flow rate and clean airflow rate could be adjusted to provide a wide variety of known air concentrations.

A charcoal tube was attached to a sampling port downstream of the mixing chamber. The charcoal tube, in turn, was connected to a physiologic sampling pump (or traditional sampling pump, in the case of the Condition #4).

The sampling time for each charcoal tube was two hours. For Conditions #1-3, which entailed sending subject data to the PSP, the two hours of data were generated by sending the 40 minutes of data (collected for each laboratory subject) to the PSP three times in

series; the methods for the process of loading the PSCU with the program LABPSP and sending the subject data to the PSP were previously described in Chapter 3.

For Condition #4, the TSP was set to a flow rate of 0.74 liters per minute using a primary standard. The TSP then pulled air through each charcoal tube for a total of two hours.

Samples were collected for each Condition at varying air concentrations until two samples closely bracketing the breakthrough point were found. To ascertain when breakthrough occurred, the back section of each tube was desorbed and the desorption solution was sent to the University of Washington's Environmental Health Analytical Laboratory for qualitative analysis. If the solution contained a higher concentration of toluene than the desorption solution from a blank tube, breakthrough was determined to have occurred.

The expected masses of toluene on the two charcoal tubes bracketing the breakthrough point for each Condition was calculated using Equation 29.

$$\text{Expected Mass (mg)} = \frac{C \cdot V \cdot MW \cdot (1/24.04)}{1000}$$

Equation 29. Expected mass (in mg) collected on a charcoal tube. C= air concentration (in ppm), V= sampled air volume (in L), and MW= molecular weight.

To confirm that the expected masses on the charcoal tubes were accurate, the front sections of the charcoal tubes at each breakthrough point were desorbed and analyzed for toluene per NIOSH Method 1500 (see Appendix Y). The only variation on the method was that the charcoal from the front section of each tube was added to 2-ml of desorption solution, as opposed to 1-ml. The concentrations of the five standards ranged from 0.1-ml to 0.5-ml toluene per ml of desorption solution.

RESULTS

The expected masses of the two samples bracketing the breakthrough point will be referred to as the breakthrough mass.

The air concentrations and breakthrough masses for all four sampling conditions are listed in Table 19.

Table 19: Expected masses of toluene loaded onto charcoal tubes that were under and over the breakthrough point for each sampling condition.

Condition #	Under or Over Breakthrough Point	Average pump flow rate (LPM)	Variance of pump flow rate (LPM^2)	Air conc. (ppm toluene)	Expected mass (mg)
1	Under	0.51	0.01	366	86
1	Over	0.51	0.01	376	88
2	Under	1.01	0.08	112	52
2	Over	1.01	0.08	118	54
3	Under	0.74	0.15	250	86
3	Over	0.74	0.15	255	88
4	Under	0.74	0	255	86
4	Over	0.74	0	260	88

Note that Conditions #3 and #4 differ only in the variance of the flow rate (the most variance exhibited by all subjects versus no variance) and that the breakthrough mass was the same. The breakthrough mass was also the same for Conditions #1 and #3, where the average flow rates were 0.51 LPM and 0.74 LPM, respectively. The breakthrough mass for Condition #2, however, was much lower than that of the other Conditions (~ 60%).

The validity of the expected masses was confirmed by GC analysis of the front sections of the charcoal tubes. The recovered masses reported in Table 20 are consistent with

charcoal tube desorption characteristics. Note that not all the toluene collected by the charcoal will desorb off the charcoal into the desorption solution; this established characteristic adds to the error inherent in charcoal tube analysis.

Table 20: Recovered masses and apparent recovery efficiencies for the front sections of the charcoal tubes surrounding the breakthrough point for all four conditions.

Condition #	Under or Over Breakthrough Point	Expected mass (mg)	Recovered mass (mg)	Recovery Efficiency (%)
1	Under	86	81	94
1	Over	88	80	91
2	Under	52	40	77
2	Over	54	45	83
3	Under	86	N/A	N/A
3	Over	88	N/A	N/A
4	Under	86	82	95
4	Over	88	82	93

The recovered masses for Condition #3 are not available due to error during the desorption process; unfortunately it was not possible to repeat the samples. There is no reason to expect, however, that the recovered masses would not be similar to Conditions #1 and 3, which had the same expected masses.

DISCUSSION

The breakthrough mass was not affected by the variability of the pump flow rate.

There was an apparent non-linear relationship between the breakthrough mass and average flow rate. As only three flow rates were tested in this experiment, the nature of this curve was not well defined by the results. At flow rates of 0.51 and 0.74 LPM the

breakthrough mass remains stable. However, at some point between 0.74 and 1.01 LPM the breakthrough mass begins to drop. One could hypothesize that there is a critical flow rate above which the air does not have enough contact time with the charcoal in the front section to bind reliably; therefore, some organic contaminant is able to pass through to the back section where it may be captured.

Note that over a two hour period at a flow rate of 1.01 LPM, air concentrations less than 112 ppm did not lead to breakthrough. The high air concentrations used in this experiment were necessary in order to create a situation that would lead to breakthrough so that the breakthrough characteristics might be better understood. However, these air concentrations for toluene are not realistic in the real world. The PEL is toluene is only 50 ppm and the action level is 25 ppm; thus, even if one charcoal tube were used to sample for 8 hours, there should be no breakthrough if the toluene concentration equaled its action level.

CONCLUSIONS

This pilot study indicated that ability of a charcoal tube to capture toluene may be unaffected by the variability of the flow rate of the pump. In addition, it suggested that at toluene levels that are realistic in today's work environments, the odds of breakthrough occurring are low.

CHAPTER 5: FIELD TESTING OF THE PHYSIOLOGIC SAMPLING PUMP

INTRODUCTION

The physiologic sampling pump (PSP) was designed to be used in the field under real-world conditions without interfering with a worker's daily routine. In order to validate that it functions as intended, a small field study was performed. The Indirect HR Method described in Chapter 3 was used to estimate minute ventilation in this study.

Note that while the Indirect HR Method did not fulfill accuracy and precision goals (as discussed in Chapter 3), the primary goal of this field study was to assess the performance of the PSP in the field given some predictive equation based on heart rate, not to validate the predictive equation itself. This was also an opportunity to explore (non-statistically) some other questions of interest:

(1) Do the subjects' 15-second heart rate averages appear to be normally distributed?

(2) Do the subjects' 15-second minute ventilation averages appear to be normally distributed?

(3) Do the 15-second air concentration averages appear to be lognormally distributed?

(4) How do the TWAs compare to the P-VWAs for each subject?

(5) Does there seems to be any correlation between the minute ventilation estimates and the air concentration for any of the subjects?

(6) Do the charcoal tube results for the samples collected on the TSP and PSP suggest that samples collected on a PSP in the field are valid?

In order to explore these questions, two additional instruments were used in the field. The first instrument was a traditional sampling pump (TSP); by pairing the PSP with a TSP, the first three questions listed above were explored. The second instrument was a MIRAN 2B (Foxboro, East Bridgewater, MA), a continuous-reading adjustable wavelength infrared analyzer; by recording real-time air concentrations of toluene or xylene, the last two questions were examined. The answers to these questions may serve as guidance for further studies.

The methods used for this study will be presented first, followed by the results. The results will include: (1) descriptions of the workers and their tasks, (2) the PSP performance in the field, (3) histograms of heart rate, minute ventilation, and air concentration averages (3) the charcoal tube analysis results for six analytes (m-xylene, o-xylene, p-xylene, toluene, ethyl benzene, and benzene), (4) the sampling time and volume for each charcoal tube, (5) the calculated air concentrations, (5) the TWAs and P-VWAs, and (6) a graph of the air concentration versus minute ventilation. The results will be evaluated in the discussion section and a summary will be provided in the conclusions.

METHODS

The field study consisted of three subjects: a spray painter and two histologists. These subjects were selected by asking people who worked with solvents at the University of Washington if they would volunteer to be a subject. After receiving interest from a worker, as well as a signed Human Consent form, a sampling day was chosen in which their use of solvents was highest.

EQUIPMENT CALIBRATION

The PSP, the TSP, and the MIRAN were calibrated prior to the start of each worker's work shift. The PSP was calibrated using a program loaded into the PSCU via a Pentium laptop computer. A mass flowmeter was placed upstream of a blank charcoal tube; its voltage was fed into the PSCU. The pump flow rate was controlled by the PSCU program entitled PUMPCAL (see Appendix Z). This program sent five voltages to the pump, holding each voltage level for 20 seconds; the final mass flowmeter voltages were stored by the PSCU. After the last voltage was recorded, PUMPCAL calculated the calibration curve (a polynomial curve) and displayed the coefficients on the screen for the user to record. The calibration coefficients were then entered into the program entitled FIELDPSP (see Appendix AA) and saved, so that this step need not be repeated in the field.

The TSP was set to a flow rate of 0.200 liters per minute by comparing it to a primary standard.

The MIRAN was calibrated by laboratory personnel working in the University of Washington's Environmental Health Analytical Laboratory. They followed the closed-loop calibration method outlined in the MIRAN 2B manual. A chemically inert, bellows pump was placed in a closed circuit with a septum port and the MIRAN. A quantity of the toluene or xylene was injected through the septum in the fixed volume of the closed-loop so that atmospheres of 1, 5, 10, 15, and 20 ppm would be produced. The corresponding responses were entered into the MIRAN's memory and a calibration curve was created by the MIRAN. The MIRAN was used in the background subtraction mode.

EXPERIMENT PREPARATIONS

The following equipment was brought to each work site:

- a fanny pack with the PSCU in the pouch and two sampling pumps (a modified GilAir 5 for the PSP and a Gilian Low Flow Sampler for the

TSP) attached to the waist band on either side of the pouch. Tubing of approximately two feet in length was attached to their inlets.

- a Pentium laptop computer loaded with Txtools and the FIELDPSP program (which used the Indirect HR Method to estimate minute ventilation, using preliminary predictive equation coefficients based on the subject database).
- a Polar® Vantage XL heart rate monitor
- a MIRAN 2B
- a Rustrak Ranger datalogger (Gultan, East Greenwich, RI)
- pre-labeled charcoal tubes

At the work site, subjects were asked their age, height, weight, health status, and the number of days a week they usually exercised. In addition, they were asked if they were taking any medication.

Resting heart rate was measured by having each subject sit at rest for five minutes while wearing the Polar Monitor chest band and wristwatch. At the end of five minutes, the heart rate value displayed on the wristwatch was recorded.

The MIRAN instrument and the attached datalogger were started upon arrival to the site. For the first field session, a long hose was attached to the intake of the MIRAN; the MIRAN remained stationary while the intake location was moved in conjunction with the worker; this ensured that the toluene measurements would approximate the worker's personal toluene concentration values. For the second and third field sessions, however, this arrangement was not possible; the MIRAN's intake location remained stationary, thereby providing area measurements.

SAMPLE COLLECTION

Charcoal tubes were inserted into the ends of the tubing opposite their connection to the TSP and PSP inlets. The PSCU was connected to the laptop computer and the Txtools program named FIELDPSP was downloaded and initiated. This program was designed to pause two minutes before the PSP starts sampling. During this pause, the PSCU was disconnected from the computer, the fanny pack (with PSP and TSP) was placed on the worker, and the charcoal tubes were attached to the lapel of the worker (one charcoal tube on each side). When the PSP commenced sampling, the TSP was manually turned on.

Each charcoal tube was sampled for approximately two hours. During Subject #1's shift, three sets of charcoal tubes were sampled. For Subjects #2 and #3, only one set of charcoal tubes was sampled.

CHARCOAL TUBE ANALYSIS

The charcoal tubes were analyzed by the University of Washington's Environmental Health Analytical Laboratory for the following analytes: p-xylene, m-xylene, o-xylene, toluene, ethyl benzene, and benzene. Results were reported as mass of each analyte on each charcoal tube.

RESULTS

The three field subjects' characteristics are displayed in Table 21. The table includes the sample date, the subjects' job descriptions, the main task they performed the day of sampling, their age (in years), their sex, their weight (in kilograms), their height (in centimeters), their resting heart rate (in beats per minute), the average number of times they exercise per week, their current health status, and if they were taking any medication which could affect their heart rate.

The histologist that is identified as Subject #3 actually refused to wear the PSP soon after the experiment had started. In order to try to collect some potentially useful data, one of the researchers volunteered to wear the PSP equipment as a surrogate subject. The characteristics for Subject #3 that are listed in Table 21, other than job description and main task, apply to the researcher who wore the equipment, not to the intended subject (the histologist).

Table 21: Field subject characteristics. Note that Subject #3's characteristics that are in italics apply to the surrogate subject, not to the histologist.

Subject#	**1**	**2**	**3**
Date	3/18/97	4/10/97	*4/24/97*
Job	Painter	Histologist	Histologist
Main task	Spray booth	Coverslipping	Slide staining
Age (yr)	54	41	*41*
Sex	M	F	*M*
Wt (kg)	73	91	*79*
Ht (cm)	173	180	*188*
RestHR	93	78	*69*
Fitness	0	0	*3-4x/wk*
Health	Normal	Normal	*Normal*
Hrmed	None	None	*None*

SUBJECT #1

Subject #1 worked in the Carpentry Shop at the University of Washington as a painter. He worked primarily in a spray paint booth that was well ventilated. His main task for the day included spray painting chairs and doors with lacquer while in the spray paint booth (with the fan on and while wearing a respirator). Other tasks worth noting include sanding, applying lacquer by hand, applying stain by hand, and cleaning paint brushes.

The PSP system worked well in the field with Subject #1. The following tasks were easy to perform:

- pump calibration

- loading the PSCU with the FIELDPSP program tailored to Subject #1
- starting the pump
- downloading the data

The main complaint from the subject was that the fanny pack unit (which carried the PSCU and two sampling pumps) kept sliding down over his hips. Since he essentially stood all day, it did not get in his way due to its size or location. He would not, however, wear the unit during his lunch break due to embarrassment over its appearance and size.

Histograms of Subject #1's 15-second heart rate and minute ventilation averages are displayed in Figures 55 and 56, respectively. Note that the heart rate values extended below Subject #1's measured resting heart rate of 93 beats per minute. In addition, there seemed to be times in which the PSCU failed to pick up the heart rate signal, as evidenced by unreasonably low 15-second averages. It was decided that for the purpose of looking at the distributions of these values, heart rate averages less than 80 beats per minute would be eliminated. In addition, since the 15-second minute ventilation averages were based on the heart rate values, the minute ventilation estimates corresponding to these heart rate values were removed as well. A total of 88 sets of data points were removed, out of a total of 1816; therefore, approximately 5% of the heart rate averages were considered invalid.

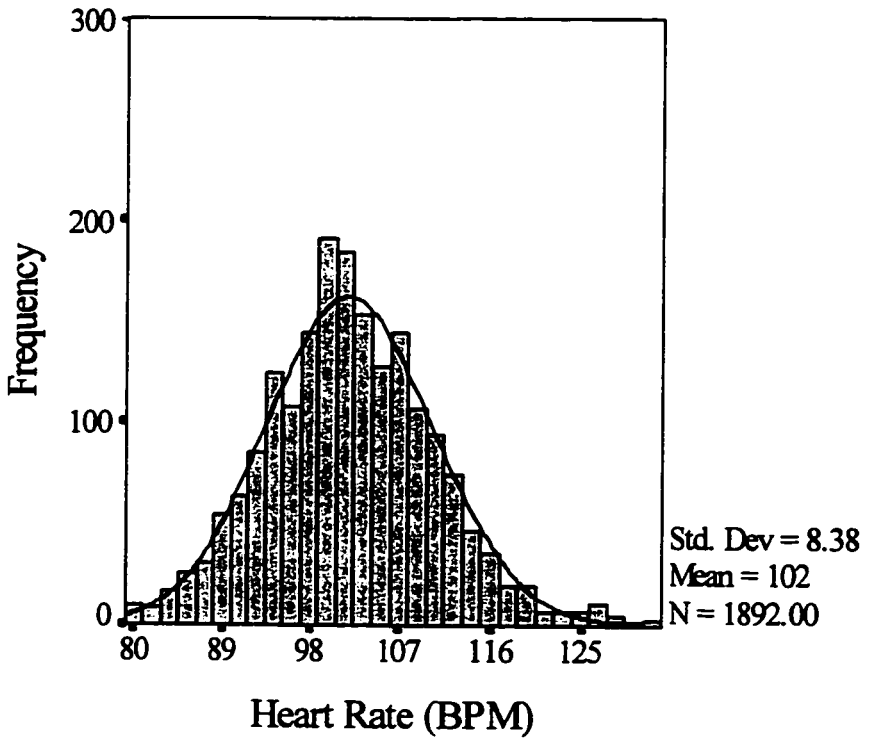


Figure 55: Histogram of 15-second averages of heart rate for Subject #1.

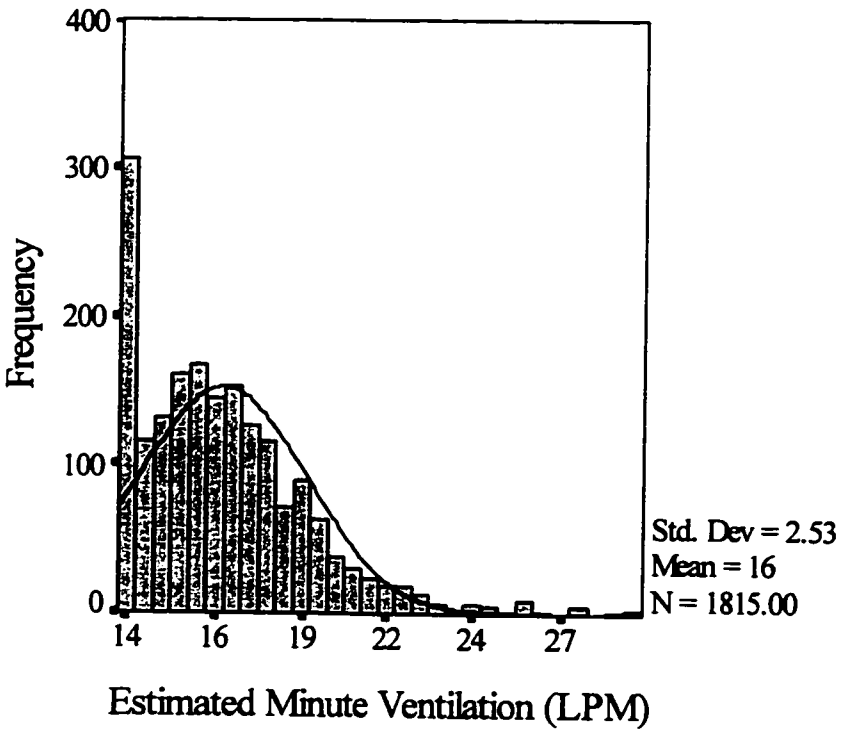


Figure 56: Histogram of 15-second averages of estimated minute ventilation for Subject #1.

A histogram of the toluene concentrations, recorded via the MIRAN's datalogger, is shown in Figure 57.

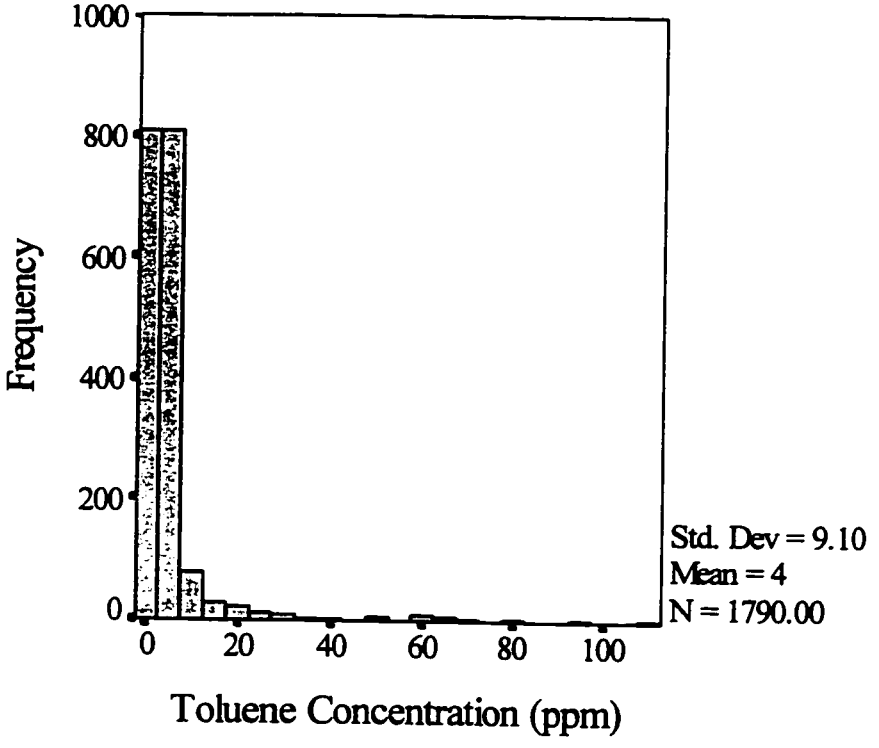


Figure 57: Histogram of 15-second averages of toluene concentration for Subject #1.

A timeline of Subject #1's minute ventilation averages and toluene concentrations are displayed in Figure 58. Subject #1 did take a lunch break from 12:20 P.M. until 1:08 P.M.; this portion of the timeline is not included in the graph.

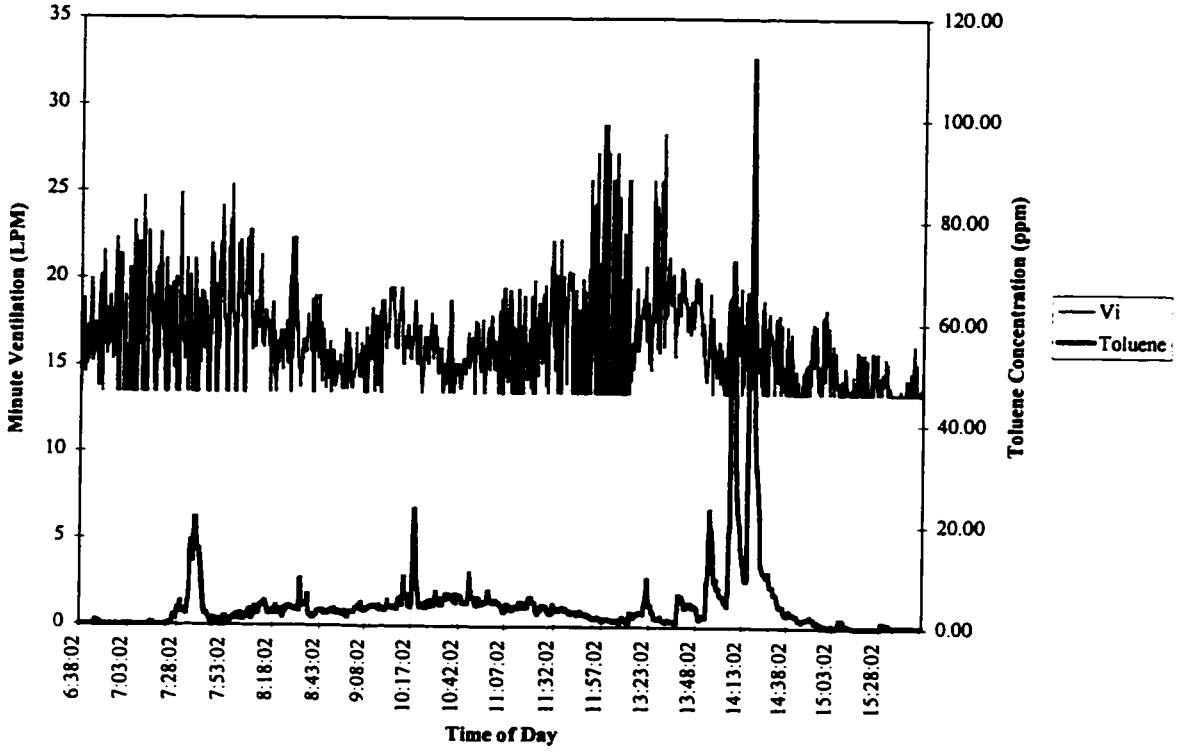


Figure 58: Minute ventilation and toluene concentration over time.

The charcoal tube laboratory results were reported as mass of each analyte per charcoal tube (see Table 22). The sample identification reflects both the pump type on which the sample was collected (TSP or PSP) and the sample collection progression (e.g. sample #1 collected first, followed by sample #2, etc.).

Table 22: Mass of analyte (in micrograms) on the charcoal tube samples collected on the TSP and PSP for Subject #1. The toluene results for the third PSP sample could not be assessed because the sample exceeded the upper quantitation limit (UQL).

Sample	p-xylene (μg)	m-xylene (μg)	o-xylene (μg)	toluene (μg)	ethyl benzene (μg)	benzene (μg)
TSP1	45.69	132.60	64.54	381.74	36.31	0.43
TSP2	24.52	66.16	36.75	188.99	19.38	0.36
TSP3	71.49	204.36	104.93	1291.88	59.43	0.51
PSP1	97.18	258.70	141.79	675.06	70.58	0.51
PSP2	58.49	173.13	95.77	469.90	46.45	0.63
PSP3	173.93	454.39	241.51	>UQL	137.16	0.73

The sample collection times and the volume of air sampled per charcoal tube are listed in Table 23.

Table 23: Sample times and volumes for each charcoal tube. Total time for the TSP samples and total volume for the PSP samples are also displayed.

Sample	Sample time (min.)	Sample (L)
TSP1	138	27.6
TSP2	156	31.2
TSP3	167	33.4
Total time	**461**	
PSP1	142	60.9
PSP2	158	63.0
PSP3	158	62.2
Total volume		**186.05**

The masses of each analyte listed in Table 22 were divided by these sample volumes in order to calculate the air concentrations listed in Table 24.

Table 24: Air concentration results (in milligrams per cubic meter) based on charcoal tube samples collected on the TSP and PSP for Subject #1. The toluene results for the third PSP sample could not be assessed because the sample exceeded the upper quantitation limit (UQL).

Sample	p-xylene (mg/m^3)	m-xylene (mg/m^3)	o-xylene (mg/m^3)	toluene (mg/m^3)	ethyl benzene (mg/m^3)	benzene (mg/m^3)
TSP1	1.66	4.80	2.34	13.83	1.32	0.02
TSP2	0.79	2.12	1.18	6.06	0.62	0.01
TSP3	2.14	6.12	3.14	38.68	1.78	0.02
PSP1	1.60	4.25	2.33	11.09	1.16	0.01
PSP2	0.93	2.75	1.52	7.46	0.74	0.01
PSP3	2.80	7.30	3.88	>UQL	2.20	0.01

The air concentrations based on the TSP samples are compared to those based on the PSP samples in the following graph (Figure 59). Each data point plots the air concentration of one of the analytes from a TSP sampled charcoal tube versus the air concentration of the same analyte for the corresponding PSP sampled charcoal tube. Therefore, for each analyte (except toluene) there are three data points plotted (one for each charcoal tube sampling period). For toluene, only two points are plotted since the PSP3 sample could not be quantified. Note that the benzene concentrations are all close to zero, and overlap one another on the graph. The line represents a line of unity (i.e. perfect agreement). Based on a paired T-test, these concentration values are not significantly different.

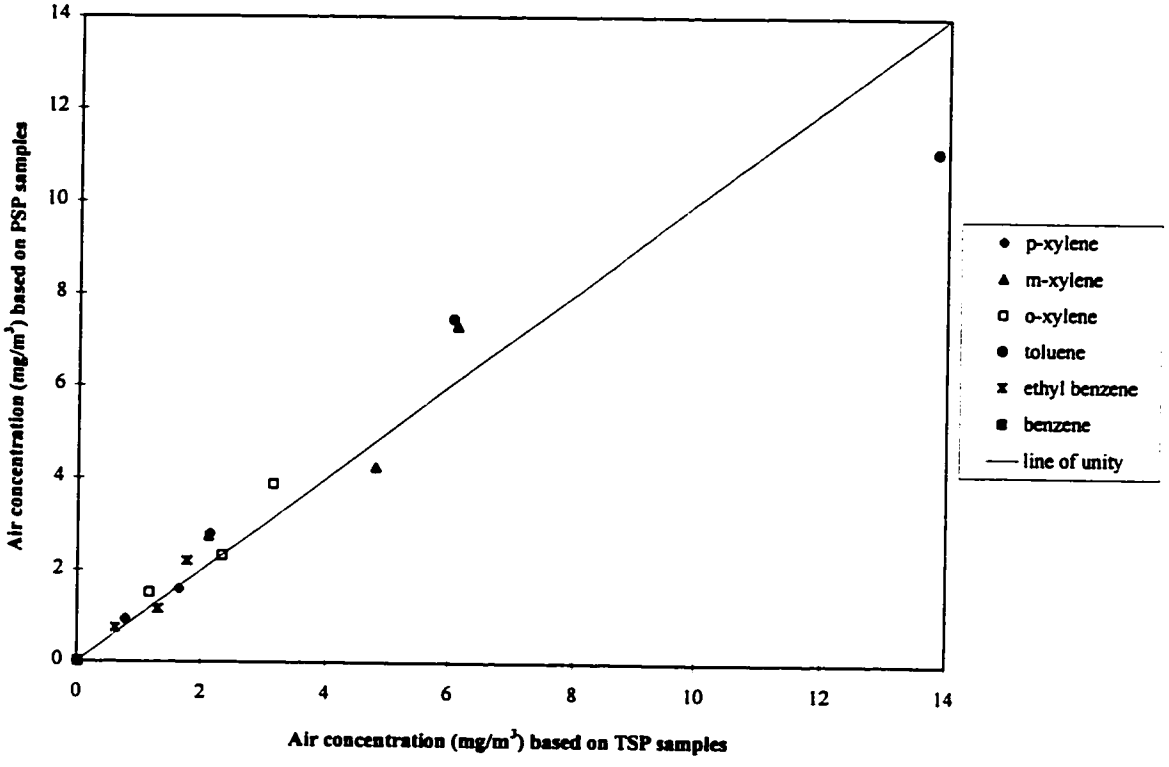


Figure 59: Comparison of analysis results between the three charcoal tubes sampled using the TSP versus the three sampled via the PSP. Estimated air concentrations of six analytes (p-xylene, m-xylene, o-xylene, toluene, ethyl benzene, and benzene) are expressed in milligrams per cubic meter or air.

The TWAs and PWAs for each analyte were calculated by dividing the sum of the air concentrations displayed in Table 24 by the total time from Table 23 (to arrive at the TWA) or the total volume from Table 23 (for the PWA). The one exception was the values for toluene, where the sum of the concentrations and sample times were restricted to samples #1 and #2. The TWAs and P-VWAs are shown in Table 25. A paired T-test was performed on this data and revealed that the P-VWAs were not significantly different from the TWAs.

Table 25: TWA and P-VWA results for Subject #1's work shift on 3/18/97. The toluene values are based on the first two charcoal tube results only.

Sample	p-xylene (mg/m^3)	m-xylene (mg/m^3)	o-xylene (mg/m^3)	toluene (mg/m^3)	ethyl benzene (mg/m^3)	benzene (mg/m^3)
TWA	1.54	4.37	2.24	9.71	1.25	0.01
P-VWA	1.77	4.76	2.57	9.25	1.37	0.01

It is important to note that the TWA for toluene based on the MIRAN data should agree with the TWA displayed in Table 25; unfortunately it does not agree. The TWA for toluene based on the MIRAN data (over the sampling period for the first two charcoal tubes) was almost twice as high, at 16.8 mg/m^3. One possible explanation for this discrepancy lies in the fact that the toluene concentrations sampled by the MIRAN exceeded the high end of the calibration range (20 ppm); this could have contributed to over-estimation of the air concentration at toluene concentrations above 20 ppm, thereby increasing the average concentration.

Figure 60 shows a comparison of these TWAs and P-VWAs.

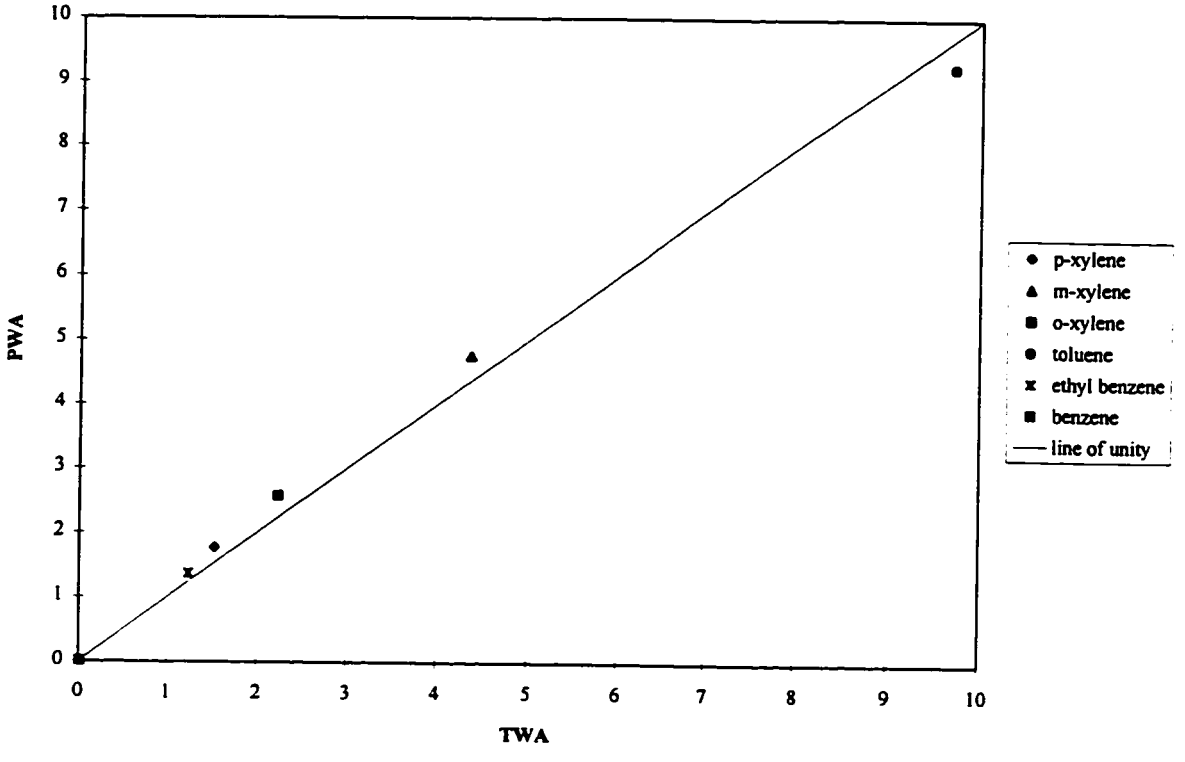


Figure 60: TWA versus P-VWA for Subject #1. Each data point represents a different chemical species.

The relationship between minute ventilation and toluene concentration is displayed in Figure 61. The correlation coefficient between these two variables is -0.02. As mentioned previously, however, it is possible that the MIRAN's calibration did not remain stable over the sampling period.

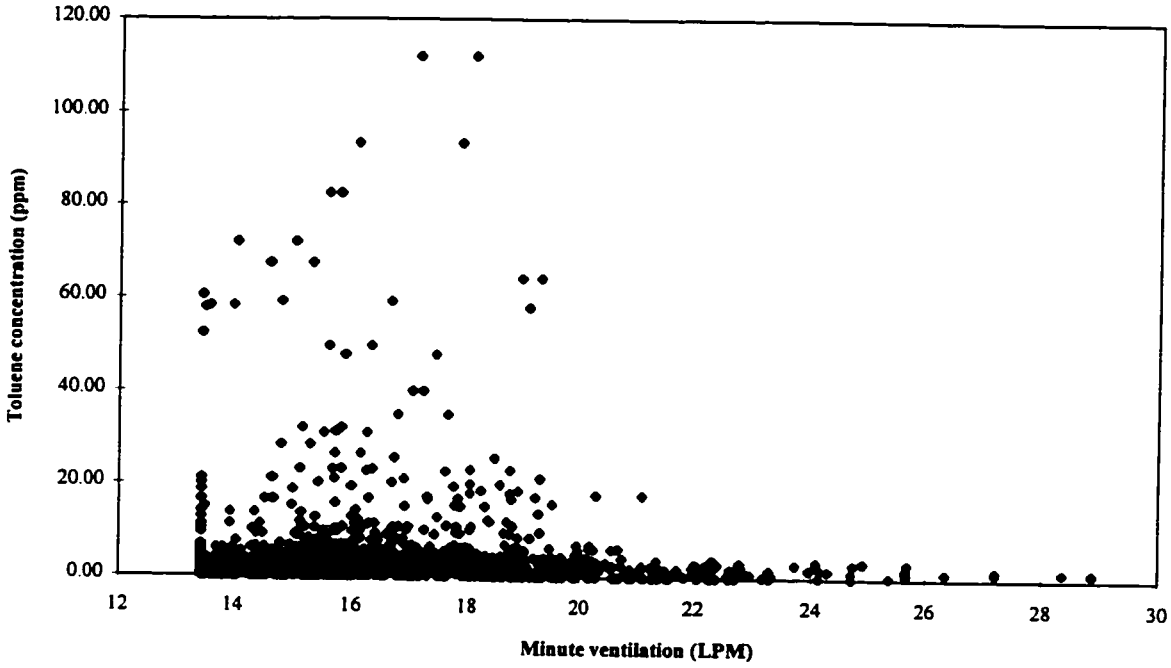


Figure 61: Relationship between air concentration of toluene (ppm) and minute ventilation (LPM) for Subject #1.

SUBJECT #2

Subject #2 worked in the Zoology Department at the University of Washington as a histologist. On the day of sampling her main task was to coverslip slides. In addition to coverslipping, she performed microscope work and cleaned dishes at a sink. She only worked two hours the day of sampling, so only one sample per pump was collected.

As in the experience with Subject #1, the use of the PSP from an industrial hygienist's perspective was very straightforward and easy. The subject did complain, however, that it was not easy to sit comfortably; she was not able to lean back properly and the bulk of the equipment pushed her too far forward on her chair. Suspenders were used to hold up the fanny pack, to avoid the slippage problem encountered with Subject #1; this solution was effective.

Figures 62 and 63 display the histograms of heart rate and minute ventilation for Subject #2. In this experiment, the critical heart rate value (below which data points were removed from these graphs) was 60 beats per minute. There were 11 values less than 60 beats per minute, out of a total of 487; therefore, about two percent of the 15-second heart rate averages were invalid.

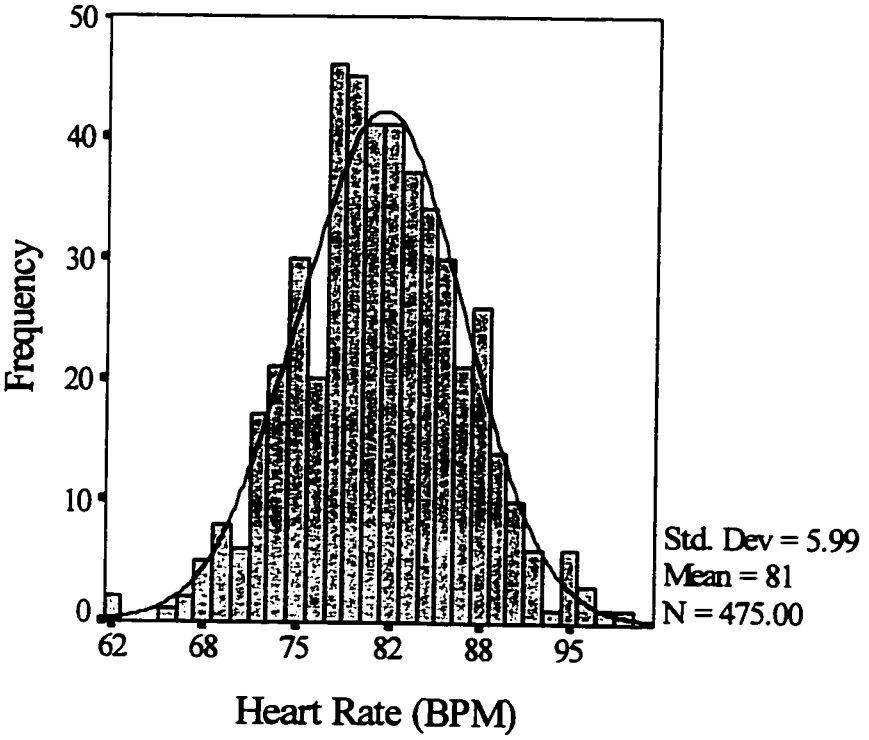


Figure 62: Histogram of Subject #2's 15-second heart rate averages.

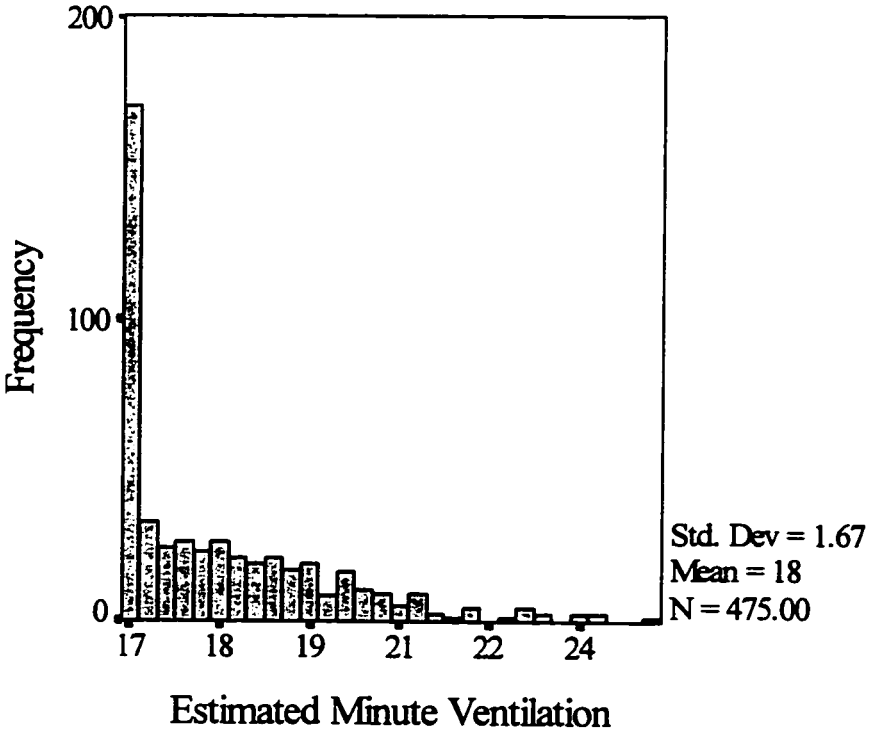


Figure 63: Histogram of Subject #2's 15-second estimated minute ventilation averages.

The MIRAN data for Subject #2 was not reliable due to an instrument drift problem. Therefore, the relationship of xylene concentration to the subject's minute ventilation can not be determined.

A timeline of the subject's minute ventilation is displayed in Figure 64.

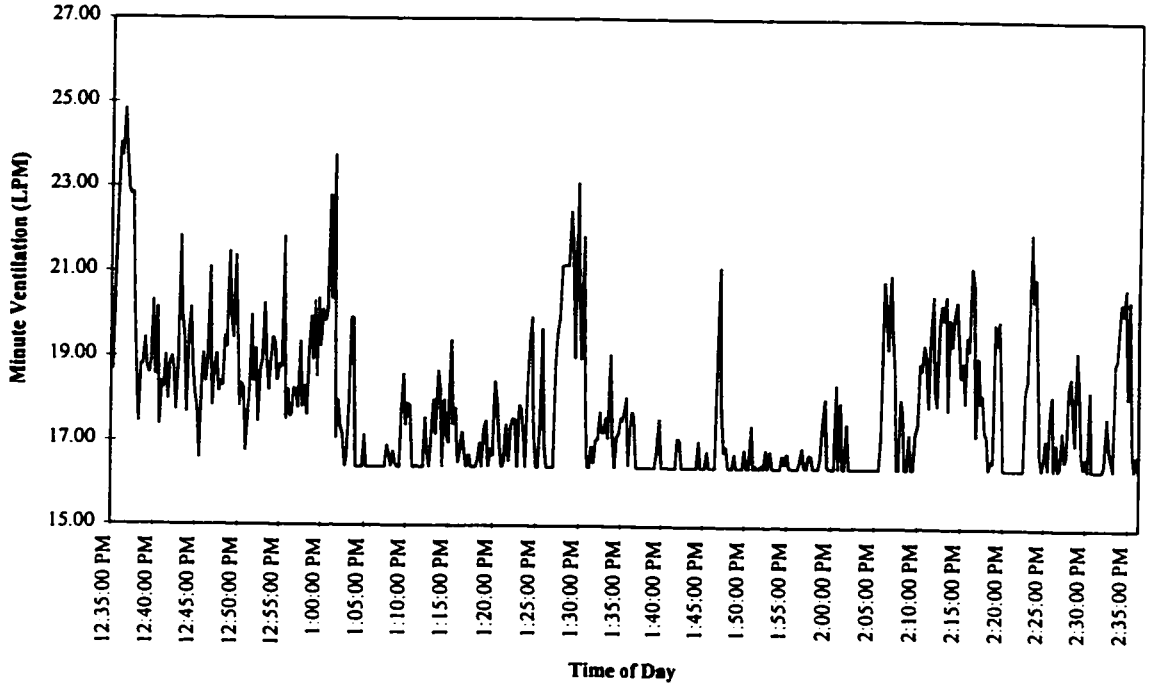


Figure 64: Minute ventilation over time.

The charcoal tube analysis results are summarized in Table 26.

Table 26: Mass of analyte (in micrograms) on the charcoal tube samples collected on the TSP and PSP for Subject #2.

Sample	p-xylene (μg)	m-xylene (μg)	o-xylene (μg)	toluene (μg)	ethyl benzene (μg)	benzene (μg)
TSP1	3.14	7.94	3.40	2.01	2.25	0.52
PSP1	7.49	18.47	7.96	3.56	5.20	0.89

The sample collection times and the volume of air sampled per charcoal tube are listed in Table 27.

Table 27: Sample times and volumes for both charcoal tube.

Sample	Sample time (min.)	Sample volume (L)
TSP1	120	24.0
PSP1	120	54.2

The air concentrations based on the two charcoal tube analyses are displayed in Table 28.

Table 28: Air concentration results (in milligrams per cubic meter) based on charcoal tube samples collected on the TSP and PSP for Subject #2.

Sample	p-xylene (mg/m^3)	m-xylene (mg/m^3)	o-xylene (mg/m^3)	toluene (mg/m^3)	ethyl benzene (mg/m^3)	benzene (mg/m^3)
TSP1	0.13	0.33	0.14	0.08	0.09	0.02
PSP1	0.14	0.34	0.15	0.07	0.10	0.02

The relationship of the PSP results to the TSP results is demonstrated in Figure 65. The correlation coefficient was 0.997 and a paired T-test showed no significant difference between the calculated air concentrations.

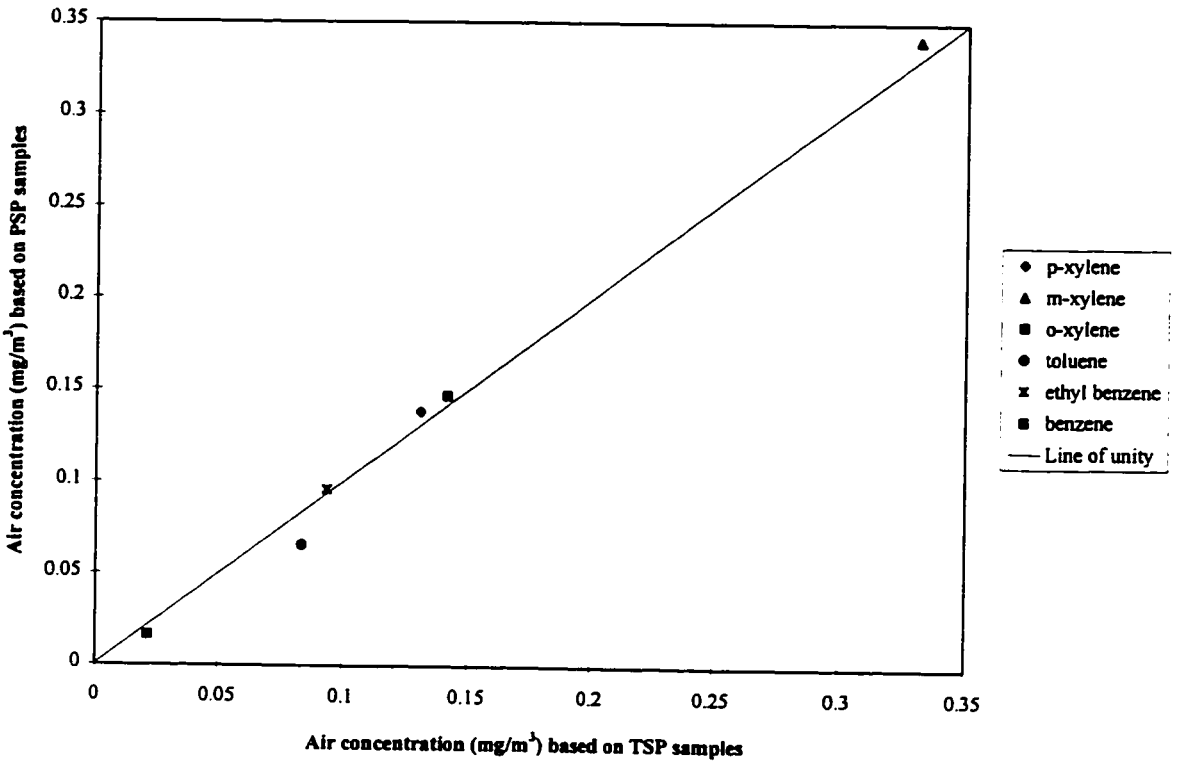


Figure 65: Comparison of analysis results between the charcoal tube sampled using the TSP versus the one sampled via the PSP. Estimated air concentrations of six analytes (p-xylene, m-xylene, o-xylene, toluene, ethyl benzene, and benzene) are expressed in milligrams per cubic meter or air.

SUBJECT #3

Subject #3 was a histologist at the University of Washington Medical Center. As mentioned previously, this worker refused to wear the PSP equipment after she was fitted with it (note: she had been shown the equipment prior to setting up the sampling appointment). She claimed that the equipment was too heavy and would look too embarrassing. None of the other workers in the laboratory would volunteer to wear it, so one of the researchers decided to wear it as a last resort. While the researcher was in the same room as the histologist, the room was too small to enable the researcher to mimic the histologist's movements. The main task performed the day of sampling was slide staining. This task was only performed during the morning shift, however, so only one charcoal tube sample per pump was collected.

The researcher who wore the sampling equipment did not voice any complaints about the system, and the system continued to operate flawlessly.

Figures 66 and 67 display the histograms of heart rate and minute ventilation for Subject #3. In this experiment, the critical heart rate value, below which data points were removed from these graphs, was 60 beats per minute. There were 2 values less than 60 beats per minute, out of a total of 614; therefore, about 0.3 percent of the 15-second heart rate averages were deemed invalid.

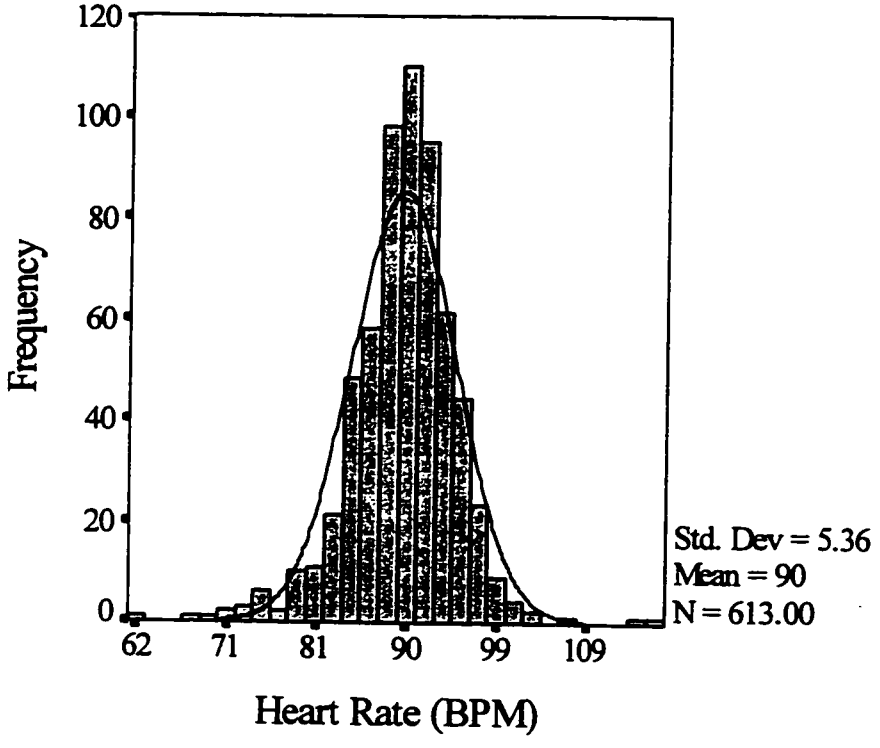


Figure 66: Histogram of Subject #3's 15-second heart rate averages.

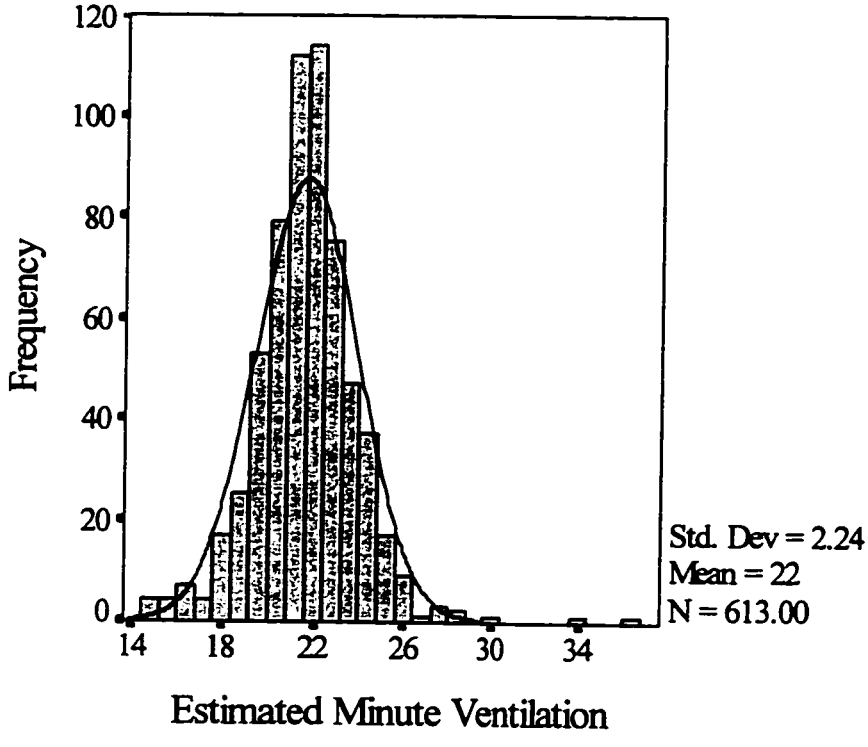


Figure 67: Histogram of Subject #3's 15-second estimated minute ventilation averages.

A histogram of the xylene concentrations, recorded via the MIRAN's datalogger, is displayed below in Figure 68.

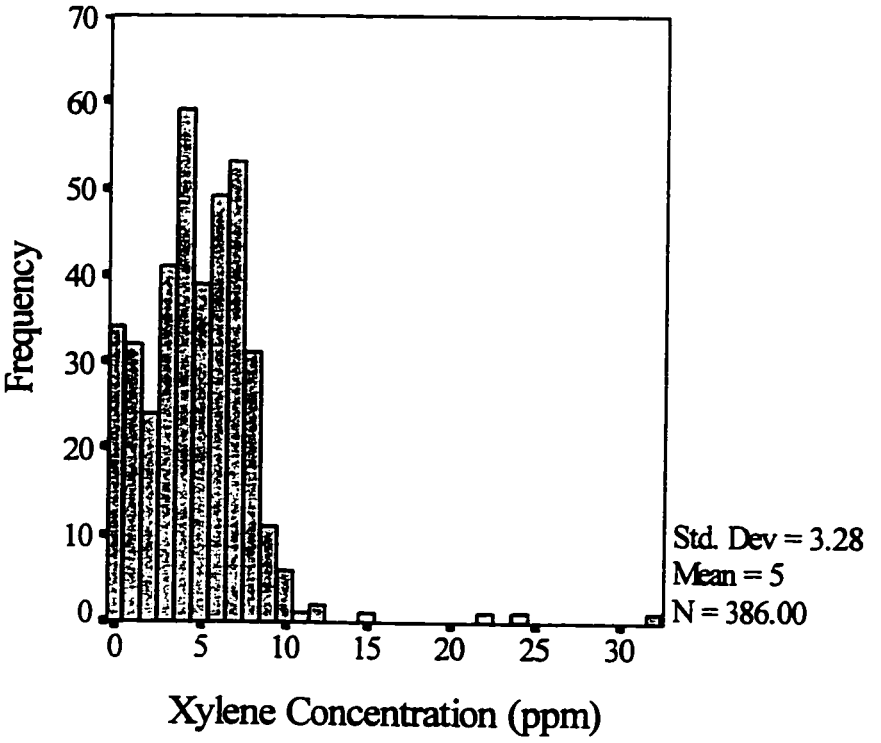


Figure 68: Histogram of 30-second xylene concentration averages.

A timeline of Subject #3's minute ventilation averages and the xylene concentration in the laboratory are displayed in Figure 69.

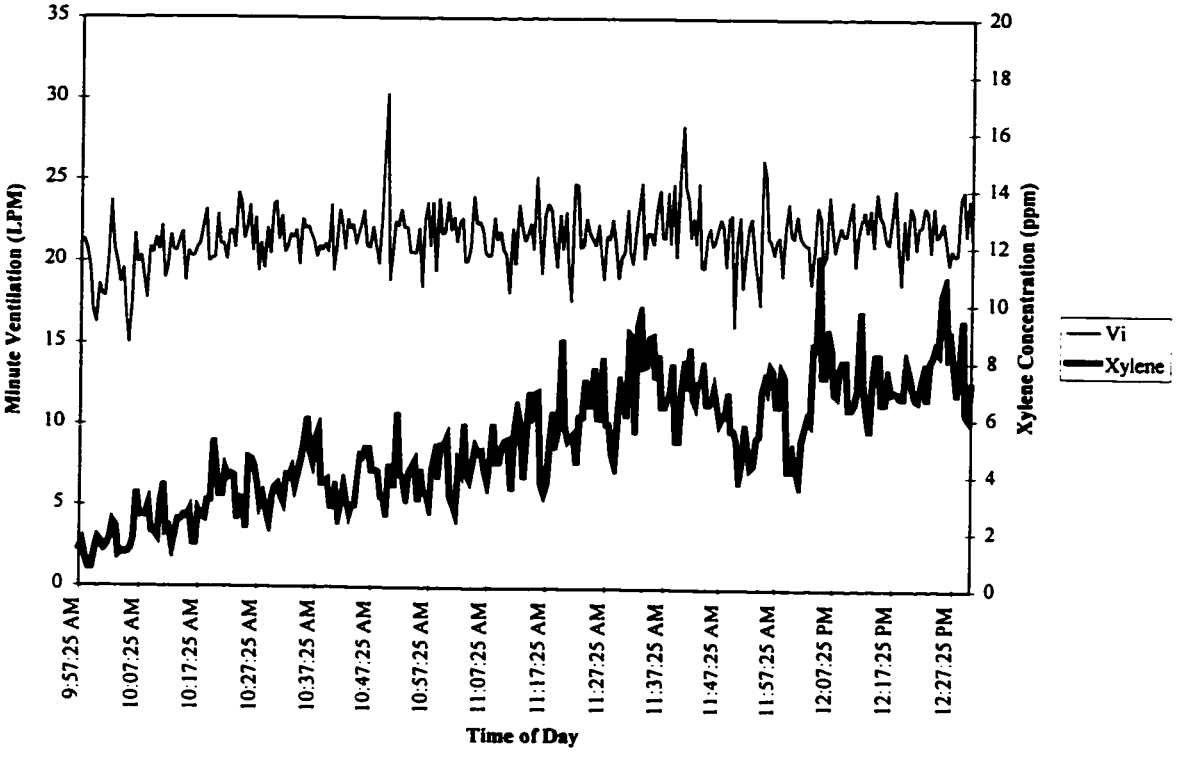


Figure 69: Minute ventilation and xylene concentration over time.

The masses of the six analytes on each charcoal tube, as reported by the University of Washington's Environmental Health Analytical Laboratory, are shown in Table 29.

Table 29: Mass of analyte (in micrograms) on the charcoal tube samples collected on the TSP and PSP for Subject #3. The benzene results for the PSP sample could not be assessed because it was less than the lower quantitation limit (LQL).

Sample	p-xylene (μg)	m-xylene (μg)	o-xylene (μg)	toluene (μg)	ethyl benzene (μg)	benzene (μg)
TSP1	1.31	3.01	1.51	3.70	0.89	0.11
PSP1	3.38	8.76	4.59	8.69	2.52	<LQL

The sample collection times and the volume of air sampled per charcoal tube are listed in Table 30.

Table 30: Sample times and volumes for both charcoal tubes.

Sample	Sample time (min.)	Sample volume (L)
TSP1	164	32.8
PSP1	164	82.7

The air concentrations based on the two charcoal tube analyses are displayed in Table 31.

Table 31: Air concentration results (in milligrams per cubic meter) based on charcoal tube samples collected on the TSP and PSP for Subject #3.

Sample	p-xylene (mg/m^3)	m-xylene (mg/m^3)	o-xylene (mg/m^3)	toluene (mg/m^3)	ethyl benzene (mg/m^3)	benzene (mg/m^3)
TSP1	0.034	0.092	0.046	0.113	0.027	0.003
PSP1	0.041	0.106	0.055	0.105	0.031	<LQL

The TWA for xylene based on the MIRAN data was not compared to the TWA based on the charcoal tubes because the MIRAN data represented an area sample while the charcoal tubes represented a personal sample.

The relationship of the PSP results to the TSP results is demonstrated in Figure 70. The correlation coefficient was 0.976 and a paired T-test showed no significant difference between the calculated air concentrations.

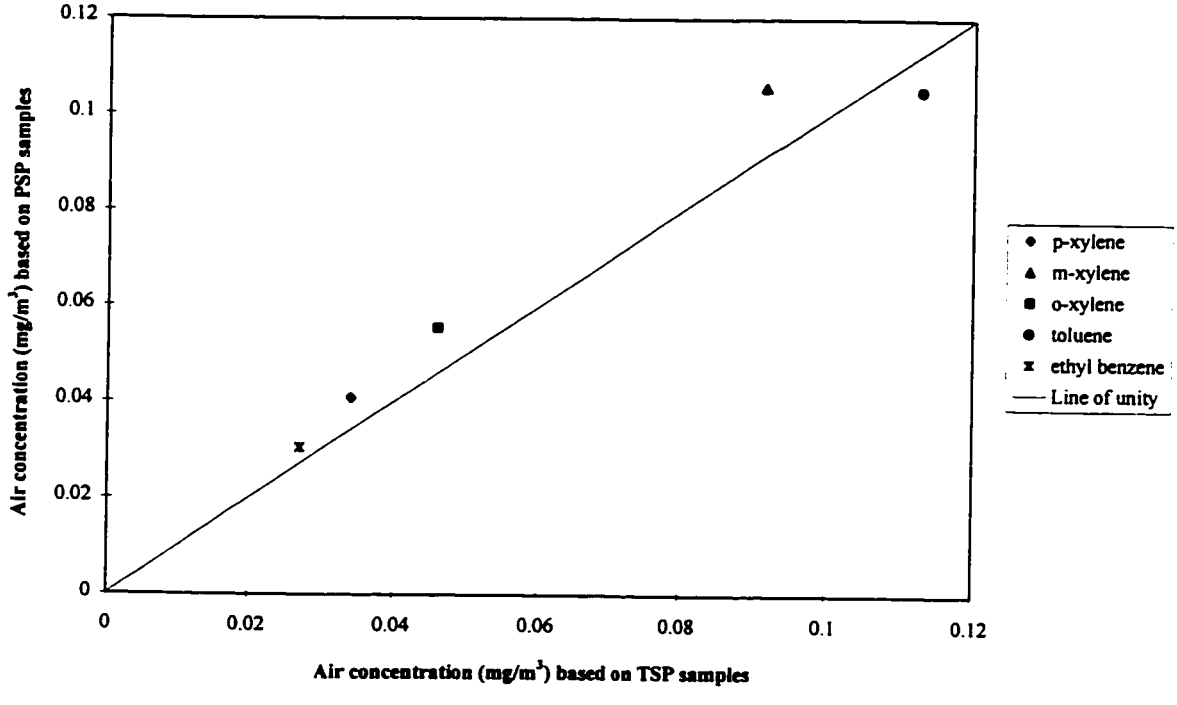


Figure 70: Comparison of analysis results between the charcoal tube sampled using the TSP versus the one sampled via the PSP. Estimated air concentrations of six analytes (p-xylene, m-xylene, o-xylene, toluene, ethyl benzene, and benzene) are expressed in milligrams per cubic meter or air.

A graph of xylene concentration versus minute ventilation for Subject #3 is presented in Figure 71. While there was a correlation coefficient of 0.27 between these two variables, conclusions regarding this relationship cannot be made since the xylene concentrations were measured from a stationary instrument and the minute ventilation values were from the surrogate subject, whose minute ventilation values did not represent those of the histologist.

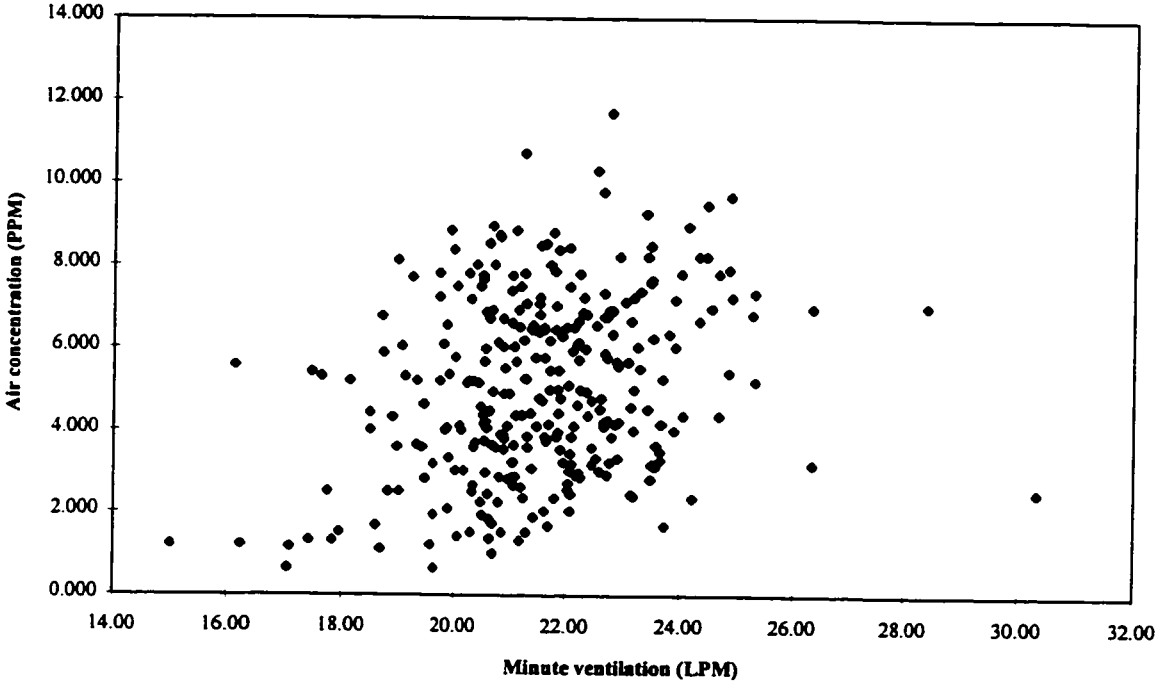


Figure 71: Relationship between air concentration of xylene (ppm) and minute ventilation (LPM) for Subject #3.

DISCUSSION

The overall usability of the PSP in the field was well within acceptable limits for industrial hygiene use. It was easy to calibrate, setup, collect samples with, and download data from.

The complaints from the subjects concerned the size, weight, and appearance of the equipment. It is important to remember that all of these factors would be much more agreeable on a manufactured device, rather than a prototype; it would be small, lighter, and more refined. In addition, the weight and size burden during this experiment was magnified by attaching the TSP to the same fanny pack that carried the PSP equipment

According to the Kolmogorov-Smirnov test for normality (using the Lilliefors significance correction), Subject #2 had the only heart rate average distribution that was normal (see Figure 62). The distribution for Subject #1 (see Figure 55) was positively skewed, probably due to the fact that the cut-off point (80 beats per minute) was set a little bit too high. Subject #3's heart rate average distribution was negatively skewed and had a kurtosis value of 3.7 (see Figure 66).

Because of the exponential relationship between heart rate and minute ventilation (see Equation 28), if the heart rate averages were normally distributed, then the minute ventilation average should be lognormally distributed. None of the subject's minute ventilation averages were lognormally distributed, however, as demonstrated by Figures 56, 63, and 67.

The peaks seen at the lowest x-axis values in Figures 56 and 63 are artifact; the minute ventilation equation (Equation 28) uses delta heart rate (HR-HR0) rather than the actual heart rate value. In order to incorporate this into the equation, the FIELDPSP program must be instructed to subtract the subject's resting heart rate from the calculated heart rate averages. When the heart rate averages fall below the resting heart rate, the delta heart

rate value is set to zero. Therefore, if the resting heart rate value entered into the FIELDPSP program is higher than the subject's true resting heart rate, any heart rate average less than the entered resting heart rate will result in a constant (and elevated) minute ventilation estimate. This phenomenon resulting in "clipped" minute ventilation estimates is reflected by the cited peaks in Figures 56 and 63. Since Figure 67 does not contain a similar peak, it is likely that the resting heart rate for Subject #3 was more accurate.

The existence of these peaks suggest that the method used to estimate resting heart rate in the field was not adequate; while five minutes is more than enough time for a person's heart rate to stabilize, it is likely that the anticipation of the experiment raised the subject's heart rate. A possible solution to this problem would be to have the intended subject wear the Polar Monitor (the chest band and watch) for a period of time prior to the sampling, during which they would likely become relaxed and forget about the monitoring equipment.

The obvious problem with this solution is that the resting heart rates in the predictive equation database were not acquired in such a manner. This logic would suggest that this would lower the estimated minute ventilation estimates for any given heart rate. Since the predictive equation seems to be over-estimating the true minute ventilation in general (see Chapter 3), this could be an advantageous by-product of this solution.

The real-time air concentration values from the MIRAN for Subjects #1 and #3 (for toluene and xylene, respectively) were not lognormally distributed (see Figures 57 and 68). They are both more positively skewed than would be expected in even a lognormal distribution.

Figures 58, 64, and 69, show the minute ventilation averages and MIRAN data over time for each of the subjects. The clipping of the minute ventilation data is apparent in

Figures 58 and 64. Figure 69 does not exhibit a similar clipping pattern, again reflecting the fact that the resting heart rate was more accurate for Subject #3.

The fact that Subject #1's heart rate data contained a higher percentage of invalid values than for Subject #2 or #3 indicates that either the Polar transmitter unit or the signal receiver in the PSCU was malfunctioning in an indeterminate manner. After this problem was noted, the system was tested vigorously in the laboratory in order to try to re-create it; this attempt was not successful. While the problem remains a unexplained, the fact that invalid values for Subject #2 and #3 were so much lower (2% and 0.3% versus 5% for Subject #1) suggests that it not a consistent problem.

Based on the fact that a high percentage of heart rate values were invalid for Subject #1, the data that was not thrown out is also put into question. If the PSCU was missing heartbeats, then some of the minute ventilation averages could have been erroneously low, but not low enough to be considered "invalid" (which was defined as below 80 beats per minute). The extensive change in heart rate values for Subject #1, seen in particular before 8:10 A.M. and between 11:15 A.M. and 12:15 P.M. (see Figure 58), might be due to missed heartbeats.

One point of interest to note about Subject #2's minute ventilation timeline is that the first half an hour (12:05 P.M. to 12:35 P.M.) is elevated compared to the rest of the shift, the same period of time in which the subject was performing the coverslipping task. It is unfortunate that the MIRAN data was not valid, because it is very plausible that the subject was exposed to higher xylene concentrations during the coverslipping task.

Subject #3's minute ventilation timeline reflects the fact that he was not performing any tasks other than sitting or standing quietly; it is very flat.

The air concentration timelines are interesting for the following reason: the toluene timeline (see Figure 58) shows peaks that were task-related while the xylene timeline (see Figure 69) shows a general background build-up of xylene over the work shift. The spray

painter, for example, was applying lacquer by hand when the large toluene peaks were measured. This is worth mentioning because task-related contaminant generation could lead to a positive correlation between minute ventilation and air concentration of a contaminant. The fact that the xylene timeline is much smoother does not necessarily signify that the inhaled xylene levels were not task-related; recall that the MIRAN was stationary during data collection for Subject #3, so it would not capture localized xylene peak concentrations around the histologist.

The fact that the air concentrations based on the TSP samples were not significantly different from the air concentrations based on the PSP samples (see Tables 24, 28, and 31) indicates the following:

1. There was no significant correlation between minute ventilation and air concentration for any of the subjects.

2. The PSP samples are valid. There appear to be no breakthrough problems related to the higher (and variable) flow rate of the PSP.

The lack of correlation suggested by the similar air concentration results is confirmed for Subject #1 by examining Figure 61.

Based on the fact that (1) the pump flow rate averages for all three subjects ranged from 0.4 to 0.5 liters per minute (less than the lowest flow rate tested in Chapter 4), and (2) the highest air concentration of toluene that was sampled was 10 parts per million (Subject #1, sample #3), the following conclusion can be drawn: a single charcoal tube sample would have been adequate to sample Subject #1's entire 8-hour shift (with very low probability of breakthrough occurring).

CONCLUSIONS

With the exception of the anomaly regarding missed heartbeats for Subject #1, the PSP performed exceptionally in the field. It was easy to use and easily calibrated prior to sampling.

One unexpected problem was the inability to accurately quantify a subject's resting heart rate in a five-minute period prior to sampling. One way to solve this problem would be to have each subject wear the Polar® monitor prior to the day of sampling; the 15-second heart rate averages could then be extracted from the watch receiving unit and examined. If the system is worn over a period of time where the subject is not active (a lunch break, for example), it is likely that their heart rate will settle at their true resting level.

Based on the size and protocol limitations of this field study, conclusions regarding a correlation between air concentration and minute ventilation of a worker cannot be drawn.

CHAPTER 6: SUMMARY AND CONCLUSIONS

Chapter 1 established the basis for this dissertation: that TSPs may not be accurately quantifying inhaled dose, due to differences in minute ventilation and potential correlation between minute ventilation and air concentration. The following topics were presented:

- The theory behind the use of physiologic sampling pumps (PSPs);
- Equations for calculating inhaled dose using samples collected on either a TSP or PSP;
- The criteria for an ideal PSP;
- Possible mechanisms that would lead to correlation between air concentration and minute ventilation;
- An equation for the physiologic volume-weighted average (P-VWA);
- Justification for using heart rate as a surrogate measure of minute ventilation.

These topics provide a theoretical foundation for PSPs, and a framework for transitioning them into mainstream use. The P-VWA, for example, is critical in order to relate PSP sample results to TWA concentrations currently used for guidelines and/or regulatory limits (i.e. TLVs and PELs). In addition, the criteria for an ideal PSP provides guidance for researchers attempting to develop a PSP for use in the real world.

Finally, in Chapter 1 it was emphasized that if correlation between air concentration and minute ventilation does not exist in work environments, then PSPs are not particularly

useful; rather, it would be just as accurate (and easier and cheaper) to measure minute ventilation independently of a sampling device.

Chapter 2 developed a computer simulation to examine how the Exposure Ratio was affected by (1) correlation between air concentration and minute ventilation, and (2) the distribution characteristics of these two variables. It was found that the GMs of the distributions had no effect on the Exposure Ratio; therefore, the computer simulation was designed to vary the GSDs and the correlation factor. It was hypothesized that the GSD[P], which combines all three of the factors known to have an effect on the Exposure Ratio, might therefore have a relationship with it. The simulation results demonstrated a definite, but unfortunately not simple, relationship. Tables and graphs were included in this chapter to help clarify the relationship, as well as to provide a tool for decision making. It was determined that the conditions necessary for an Exposure Ratio of 1.2 or greater may very well exist in some work environments.

In Chapter 3, the new PSP was tested in the laboratory using data collected from 30 human subjects. Methods were developed to calibrate heart rate (using both individual calibrations and a predictive equation) and a RIP. A total of five Methods for controlling the flow rate of the PSP were evaluated: the Pneumo Method, the Direct HR Method, the Indirect HR Method, the RIP Method, and Satoh's HR Method.

Compared to the Primary Standard, the Pneumo Method had an average RPD of only –2% and a RPD SD close to 0%. The Direct HR Method performed slightly better (in terms of the RPD values) than the Indirect HR Method; however, both Methods fell short of the accuracy and precision goals. The Direct HR Method had an average RPD of 13% and a RPD SD of 18%, while the Indirect HR Method had both an average RPD and a RPD SD of 21%. The RIP Method had a low average RPD (-4.8%) and a RPD SD of 13%. Note that one subject's RIP data could be an outlier; when that subject's data is removed the RPD SD falls below the precision goal of 10%. Satoh's HR Method had the poorest average RPD (-33%) and a RPD SD of 15%.

Mass flowmeters were used to validate that the flow through the pump corresponded to the values that the PSCU predicted. When compared to the PSCU results, the mass flowmeter values had almost perfect concordance accuracy and precision (0.998 and 0.967, respectively).

The dynamic response of the pump was also examined in Chapter 3, in order to assess the potential error due to wind-up time. It was found that the pump was a first-order linear system with a time constant of approximately 0.25 seconds. The pump, therefore, should stabilize at each new flow rate within one second, which is an adequate response time for a 15-second sampling period for each flow rate.

Chapter 4 examined an important issue for PSP use in the real world: whether a variable pump rate affects the breakthrough point of a charcoal tubes. It was found that in this pilot study, pump rate variability did not alter the breakthrough characteristics.

Chapter 5 described a pilot field study that used a PSP in conjunction with a TSP in order to test whether there was a significant difference between samples collected on these two devices. For the three subjects tested, there was no significant difference between the PSP and TSP samples; the probable reason for this is that there was no correlation between air concentration and minute ventilation in the workers tested. The only worker who appeared to change work rates was Subject #1, and a plot of his minute ventilation versus the corresponding real-time toluene concentrations confirmed the lack of correlation. This Chapter validated that the pump is easy to use in the field and can be easily calibrated, and provided evidence to support the physiologic sampling pump theory that was presented in Chapter 1.

In conclusion, this dissertation presented a solid theoretical foundation for a PSP and its applicability to real-world scenarios. A computer simulation was developed which demonstrated that TSPs may significantly underestimate worker exposure if a correlation exists between air concentration and minute ventilation.

The PSP prototype met all requirements for usability in the field. While the predictive equation upon which the pump's variable flow rate was based did not meet goals for precision and accuracy, the design and implementation of the pump itself was clearly validated by its perfect precision and near-perfect accuracy when pneumotachometer data was sent to the PSCU. Two factors must be considered when evaluating the PSP for use in the field: (1) the accuracy and precision goals used in this study were very high (the Indirect HR Method did have some predictive value), and (2) a more accurate and precise predictive equation could be developed (specific suggestions for such improvement were included in Chapter 3).

Future directions for research include:

1. Investigating the issue of potential correlation between air concentration and minute ventilation using a direct reading setup (as used by Satoh *et al.*[4]). Such a device would log data from both a real-time air monitoring instrument, as well as a heart rate monitor. Ideally, the subject being monitored would also be calibrated in the laboratory, such that his/her minute ventilation could be predicted as accurately as possible. This study could provide the information necessary for: (1) more precisely characterizing the range of GSD values for 15-second averages of both air concentration and minute ventilation, and (2) characterizing the level of correlation between air concentration and minute ventilation that exists in real-world environments. This, in turn, would assist in determining the need for PSPs for certain worker populations;

2. Improving the Indirect HR Method's predictive equation. This could be accomplished by: (1) calibrating workers from potential target populations, and (2) collecting steady-state data from many work rates.

3. Developing a RIP system that could be used in the field. Areas for improvement should include: (1) developing a method to fasten the two ends of each band firmly

together to form secure loops, (2) adding a signal to account for flexing movements (i.e. bending forward), (3) developing a system to calibrate the subject quickly and reliably in the field, (4) developing a method to hold the bands in place reliably, and (5) verifying that a calibration performed at the start of the day is valid over the entire work shift. This is a formidable list, but if such a system could be developed then it is very likely that a RIP Method could be used successfully to control a PSP in the real world.

BIBLIOGRAPHY

[1]Wasserman, K., Hansen J.E., Sue, D.Y., Whipp, B.J., and R. Casaburi. Principles of exercise testing and interpretation, 2nd Ed. Lea & Febiger, 1994.

[2]Kucharski, R. *A personal dust sampler simulating variable human lung function.* British Journal of Industrial Medicine, 37:194-196, 1980.

[3]Hinds, W.C. Aerosol technology: Properties, behavior, and measurement of airborne particles, 1st Ed. John Wiley and Sons, Inc., 1982.

[4]Satoh, T., Higashi, T., Sakurai, H., and Omae K. *Development of a new exposure monitoring system considering pulmonary ventilation (DEM 1).* Keio Journal of Medicine, 38(4):432-443, 1989.

[5]Levine, M.S. *A respiration-modulated personal air sampling pump.* Applied Occupational Environmental Hygiene, 9(12):994-1005, 1994.

[6]Åstrand, P. and K. Rodahl. Textbook of work physiology: Physiological bases of exercise, 3rd Ed. McGraw-Hill Book Company, 1986.

[7]Miyamoto, Y., Tatsuhiko, T., Takahashi, T., and T. Mikami. *Transient changes in ventilation and cardiac output at the start and end of exercise.* Japanese Journal of Physiology, 31:153-168, 1981.

[8]Bakker, H.K., Struikenkamp, R.S., and G.A. de Vries. *Dynamics of ventilation, heart rate, and gas exchange: sinusoidal and impulse work loads in man.* Journal of Applied Physiology: Respiratory and Environmental Exercise Physiology, 48(2):289-301, 1980.

[9]Treese, N., MacCarter, D., Adbulut, O., Coutinho, M., Baez, M., Liebrich, A. and J. Meyer. *Ventilation and heart rate response during exercise in normals: relevance for rate variable pacing.* PACE 16:1693-1700, 1993.

[10]Mermier, C.M., Samet, J.M., and Lambert, W.E. *Evaluation of the relationship between heart rate and ventilation for epidemiologic studies.* Archives of Environmental Health, 48(4):263-269, 1993.

[11]McCool, F.D. and Paek, D. *Measurements of ventilation in freely ranging subjects.* Health Effects Institute Research Report, 59:1-17, 1993.

[12]Samet, J.M. *Assessment of heart rate as a predictor of ventilation.* Health Effects Institute Research Report, 59:19-55, 1993.

[13]Taguchi, S. and S.M Horvath. *Metabolic responses to light arm and leg exercise when sitting.* European Journal of Applied Physiology, 56:53-57, 1987.

[14]Droz, P.O., Wu, M.M., and W.G. Cumberland. *Variability in biological monitoring of organic solvent exposure. II. Application of a population physiological model.* British Journal of Industrial Medicine, 46: 547-558, 1989.

[15]Buringh, E., and R. Lanting. *Exposure variability in the workplace: its implications for the assessment of compliance.* American Industrial Hygiene Journal, 52(1): 6-13, 1991.

[16]Kromhout, H., Symanski, E., and S.M. Rappaport. *A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents.* Annals of Occupational Hygiene, 37(3): 253-270, 1993.

[17]Spielholz, Peregrin. *Heart measurement and metabolic workload estimate of drywallers.* Laboratory rotation report prepared for Associate Professor Michael Yost, University of Washington, 1995.

[18]Droz, P.O., Berode, M., and M.M. Wu. *Evaluation of Concomitant Biological and Air Monitoring Results.* Applied Occupational Environmental Hygiene, 6(6): 465-474, 1991.

[19]Droz, P.O., Wu, M.M., and W.G. Cumberland, and M. Berode. *Variability in biological monitoring of solvent exposure. I. Development of a population physiological model.* British Journal of Industrial Medicine, 46: 447-460, 1989.

[20]Thomas, R.S., Bigelow, P.L., Keefe, T.J, and R.S. Yang. *Variability in biological exposure indices using physiologically based pharmacokinetic modeling and Monte Carlo simulation.* American Industrial Hygiene Association Journal, 57: 23-32, 1996.

[21]Efron, B. *Bootstrap methods: another look at the jackknife.* The Annals of Statistics, 7(1):1-26, 1979.

[22]LePage, R. and K. Podgorski. *Giving the boot, block and shuffle to statistics.* Scientific Computing and Automation, March:28-34, 1994.

[23]van Belle, G., Gibson, K., Nochlin, D., Sumi, M., and E.B. Larsen. *Counting plaques and tangles in Alzheimer's disease: Concordance of technicians and pathologists.* Journal of the Neurological Sciences, 145: 141-146, 1997.

[24]Basic Electronic Circuit Analysis. Johnson and Hilburn. Inglewood Cliffs, NJ: Prentice-Hall, 1978.

APPENDIX A: YOST/HART/LOPEZ MANUSCRIPT

HARDWARD DESIGN AND PERFORMANCE OF A NEW PHYSIOLOGIC SAMPLING PUMP

BY

M. G. YOST, C. HART, P. LOPEZ

DEPARTMENT OF ENVIRONMENTAL HEALTH, UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON 98195

INTRODUCTION

Accurately measuring a worker's exposure to airborne contaminants is a fundamental aspect of Industrial Hygiene (IH) practice, both for hazard assessment purposes and for epidemiological studies. Exposure assessment aims to determine the mass of the compound of interest at the exposed body surface, defined here as the barrier dose. Of the three main routes of entry for hazardous chemicals (inhalation, dermal, and ingestion), inhalation is considered to be of greatest significance.

Personal sampling pumps have emerged as standard tools for exposure assessment. These devices yield an improvement over area sampling and account for individual mobility and proximity to sources. However, they fail to account for a critical factor contributing to individual differences in barrier dose: variations in breathing rate. When a person works harder, the pulmonary ventilation, cardiac output, and metabolism all increase to meet the additional energy requirements[1]. As pulmonary ventilation increases, the gas-exchanging surfaces in the lung are exposed to more of the hazardous contaminants in the air. A "traditional" sampling pump (TSP) has a constant sampling rate so the ability to accurately estimate inhaled dose for a physically active individual is limited. In contrast, a physiologic sampling pump, or PSP, eliminates this constraint. A PSP is a personal sampling pump that has been modified such that the sampling rate varies in proportion to pulmonary ventilation for the individual worker.

Three physiologic sampling pumps have been reported in the literature. It is important to examine these past efforts in order to demonstrate how the physiologic sampling pump evaluated in this paper is unique and builds upon this past work. In 1980, Kucharski described a personal dust sampler that used heart rate to vary the pump rate in proportion to minute ventilation.[2] While the signal used to control this pump was similar to the an input in our PSP, this early paper did not describe the control system in any detail nor did it validate the pump operation in controlled testing. Although Kucharski's device was designed to sample dust it did not account for size selective changes in sampling efficiency due to changes in flow rate.

For nearly a decade the PSP concept dropped out of the literature. A paper by Satoh *et al.* in 1989 revived this idea in a real-time instrument designed to simultaneously record both heart rate and air concentrations.[3] Theoretically, this approach could account for changes in minute ventilation post-sampling by weighting the instantaneous air concentrations by the corresponding minute ventilation (as estimated by pulse rate). Satoh's 1989 paper only addressed the use of heart rate as a surrogate measure of minute ventilation. The authors never published an article evaluating the actual sampling device. Their approach has several advantages over our proposed PSP, all deriving from the detail of data one can acquire with real-time instruments. For example, if one can gather task information in addition to performing data logging, it is possible to determine the relative contribution of both minute ventilation and air concentration to exposures.

In principle, Satoh's approach to physiologic sampling could be implemented with any existing real-time monitor, simply by collecting personal heart rate data at the same time. A number of inexpensive (<$400) data logging heart-rate monitors are commercially available that could be used for this purpose. Real-time instruments are available that can provide detailed time-series data, but they restrict the number and types of compounds that one can measure. In addition, offline data processing would be needed to convert the heart rate data to estimates of minute ventilation for each subject, and to synchronize with the air concentration data with the minute ventilation estimates.

More recently, Levine developed a physiologic sampling pump that used thoracic impedance to estimate minute ventilation[4]. Levine conducted a strictly laboratory

investigation; the hardware used a personal computer as the pump controller so device was not designed to be a portable field unit. A major limitation of this design lies in its use of thoracic impedance, which is measured by passing a low voltage, low-frequency AC electric current through the chest cavity via electrodes attached to the skin. Levine identified two problems with his device: (1) false signals introduced due to sharp body motions of the subjects, and (2) the difficulty of field calibration which would pose a significant hindrance to use by Industrial Hygienists. Thoracic impedance also is known to change with body position, which introduces error into minute ventilation estimates[5]. Levine did not evaluate how long a calibration would remain stable during a real work shift and only calibrated his subjects up to 100 Watts. It may be critical to understand the calibration relationship above 100 Watts, as workers undoubtedly work harder than this level at times, sometimes for extended periods.

A critical problem in developing a PSP is selecting and adapting a physiologic signal to control the pump rate. The PSP described in this document utilizes state-of-the-art technology based on a user-programmable digital micro-controller to accommodate several input signals. Until now, these techniques only have been attempted in a limited fashion. The micro-controller approach allows the hardware to remain fixed, while selection of the flow control algorithm and pump functions are implemented in software. A micro controller provides great flexibility in signal inputs, and can be adapted to produce a variety of linear or non-linear control signals. With appropriate software the pump can be readily calibrated for a subject in the field, without the need for extensive

training on the part of the operator. Most importantly, the physiologic sampling pump described here can change its pump rate in proportion to pulmonary ventilation, using a direct measure of minute ventilation rate, by using heart rate, or using other signals as a surrogate measure.

PSP HARDWARE DESCRIPTION

Our PSP design builds upon the earlier work by using elements from both to develop a novel instrument. The PSP was designed as a general-purpose device for sampling gases and vapors using standard adsorbent (charcoal) tube media. The PSP functions much like a traditional sampling pump; it is calibrated, fitted on a person during sample collection, and the tube is analyzed with standard methods. The main difference is that the sampling rate constantly changes, so the PSP logs a record of both the flow rate over time and the total air volume sampled for that subject. This additional data allows the hygienist to evaluate the work activity induced changes in ventilation for the subject, and to establish a volume-weighted average air concentration for the session.

The instrument package consists of two parts, a modified commercial personal sampling pump (GilAir-5, Gilian Corp.) and a custom-built controller unit. The pump was selected because it has the capability to give a wide dynamic range of flow rates, from approximately 75 ml per minute to a maximum of about 7.5 liters per minute. In addition, the analog control loop provided very stable and tight control of the pump flow over this range. The manufacturer generously provided us with complete schematics for the pump;

by agreement we cannot reveal details of the internal control circuit because the manufacture considers this information proprietary.

A simple modification was devised that adapted the pump into a variable flow device. A resistor network was attached in the control loop at a point in the circuit nominally connected to the wiper of the flow control potentiometer. A shielded signal lead attached to the network was brought outside the case. An analog DC control signal between 0 and 100 mV injected via the signal lead into the pump feedback control loop varies the pump rate. This external control signal overrides the normal flow control setting when supplied by a low impedance source, such as the output of a linear amplifier. A benefit of this modification scheme was that the pump would continue to operate normally if no connection was made to the external signal input. Further, when the pump was under external voltage control, the feedback circuit remained closed loop, so that the flow regulation, pressure compensation, and fault sensing circuits continued to function normally. The modified pump functions as a DC proportional controlled sampling device.

The Physiologic Sampler Control Unit (PSCU) forms the second component of the system. The PSCU accepts a physiologic input signal and produces an analog output signal to control the pump. The PSCU provides a number of important system functions including: (1) conditioning the analog or digital physiologic input signals; (2) converting the input signals to a calibrated measure of minute ventilation for the individual subject; (3) converting the minute ventilation to a proportional calibrated flow rate for the pump; (4) data logging the pump flow rate over time to record the total sample volume; (5)

sending an appropriate output signal to control the pump at specified time intervals. These functions were implemented with a digital micro-controller and custom designed input-output hardware.

The micro-controller used for the PSCU was a Tattletale model 4A (Onset Computer, Falmouth MA). The Tattletale 4A provides the following standard features: 8-channel, 12-bit analog-to-digital conversion, onboard clock/timer, 16 digital input/output lines, 2 channel serial interface (programmable UART), 32 K of non volatile RAM data storage, and a regulated 5 volt supply. Operating current is strongly dependent on the operating program activities; typical power requirements are 7 to 15 volts DC at a current of about 30 mA (lower power operation is possible).

Functions for this compact (2.25x 3.75 inch) micro-controller device are programmable in TxBASIC, a compiled BASIC-style language customized for the controller features. Control programs are written on a laptop or desktop PC, compiled, and downloaded into the Tattletale via a standard RS-232 interface connection. Functions in TxBASIC allow the user to provide input or control the operating program when the terminal is connected. When the terminal is disconnected, the program will operate unattended until the terminal connection is re-established. Data logging is done in a compact binary format, with time stamping of each record. The binary format is expanded to ASCII format for downloading. Retrieval of logged data is accomplished over the same RS-232 serial interface with a PC attached as a terminal.

In addition to the standard features of the Tattletale 4A, a number of auxiliary circuit functions were needed. The four main functions were: 1) input signal conditioning; 2) digital to analog conversion for the pump control signals; 3) output signal drivers; 4) and auxiliary power supply. These functions were developed on auxiliary "piggy-back" circuit boards that plugged directly onto a buss connector for the Tattletale 4A.

The digital-to-analog converter for the PSCU used an Analog Devices AD7225. This chip provides four independent analog outputs with 8-bit resolution, which allows the PSCU to independently control up to 4 sampling pumps. Input lines are addressed sequentially via a 74HC595 serial shift register and the four outputs are updated all at once when latched. These outputs are buffered by a CA660 CMOS quad operational amplifier to provide low impedance signals required by the pumps.

The input signal conditioning board accepts signals from a pneumo-tachometer, a Respiratory Inductive Plethysmograph (RIP), or a cardio-tachometer to measure heart rate. The pneumo-tachometer and RIP signals are analog voltages, which are processed by a LM324 quad operational amplifier. The pneumo-tachometer signal resembles a positive rectified sinusoid. The DC offset of the pneumo-tachometer signal is level-shifted to +0.5 volts and the signal range is adjusted to span from zero to +4.5 volts at the A/D converter. The RIP provides two bipolar sinusoidal analog signals, one signal for each band, corresponding to changes in the rib cage or abdominal cross section area. The DC zero crossing of the RIP signals are level-shifted to an offset of +2.5 volts and the signal range is adjusted to span from 0 to +5 volts at the A/D converter.

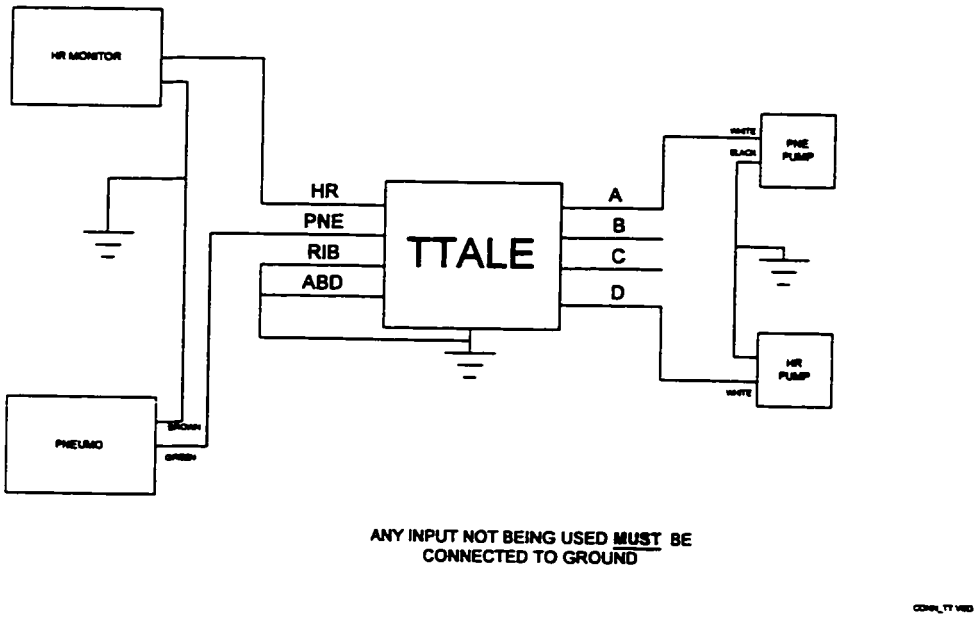


Figure 72: Block diagram of Physiologic Sampler showing PSCU and Pump connections

The beat by beat pulse data is provided by a commercially available heart rate monitor sold for exercise training (Polar CIC Inc, Port Washington, NY). The cardio-tachometer signal is broadcast from a lightweight, wireless transmitter unit (Polar Inc) held on the subject's chest with an elastic band. The PSCU heart rate input accepts 15 ms digital pulses corresponding to the R-wave detected from the electrocardiogram. The receiver unit inside the PSCU can detect the 5 kHz transmitter signal up to 3 feet away. The heart rate input is processed by a 7473 flip flop operating in toggle mode to provide an alternating-stepped 5V transition for each heart beat. This arrangement eliminates the need for interrupts or high-speed digitization to reliably detect the heart beat. Software timing of the inter-beat interval using the internal clock provides a beat-by-beat estimate of heart rate.

Figure 1 shows a block diagram of the PSP system. Output lines "A" through "D" represent analog control signals connected to the modified sampling pumps (two are illustrated in the diagram). The heart rate input is always connected and automatically enabled when a pulse from the wireless link is detected. The pneumo-tachometer and RIP inputs normally are grounded and enabled only when connected to an input signal.

STEADY STATE AND DYNAMIC PUMP RESPONSE

An important step in the development if this PSP system was to assess the ability of the pump hardware to accurately respond with the appropriate flow rate to the control program signal. Two types of testing were conducted, steady state tests to evaluate the ability of the PSP produce a repeatable calibrated flow response, and dynamic testing to evaluate the ability of the PSP to follow transient changes in flow. Both tests are important since the pump must respond in a controlled manner in order to obtain an accurate estimate of the total air volume collected during a sampling session.

The steady state response of the PSP system was evaluated by calibrating the pumps compared to a bubble meter (Gilibrator) primary standard. A TxBASIC program generated several different control voltages at the D/A converter and sent these to the pump. The actual control voltage was monitored with a digital voltmeter (Fluke Inc.) during calibration. Only a limited range of flow rates was needed so the primary calibration was limited to the range of zero to 3.5 l/min. At low flow rates, the non-linear pump response was best described by a quadratic function. Figure 2 shows the calibration

data of all four pumps, along with a calibration curve fitted to the response for Pump B. Individual pump coefficients were stored in the tattletale non-volatile memory.

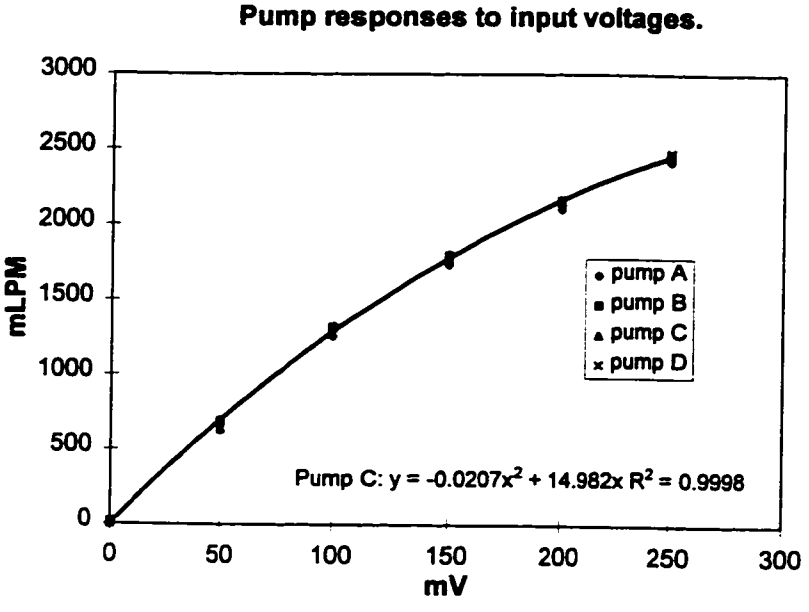


Figure 2: Steady state calibration response of pumps, calibration line shown for pump B.

In addition, a self-calibrating software procedure was implemented for routine field applications. A mass flow meter connected downstream of the pump provided the flow rate standard, and this electrical signal flow rate signal was connected into an analog input (Pneumo-tachometer input). After the user entered the linear calibration equation for the mass flow meter signal, the software would automatically step the pump through a range of control voltages while monitoring the flow rate signal. The software would then automatically generate the pump calibration coefficients from a least squares fit of the data and save the coefficients in the micro-controller memory.

DYNAMIC RESPONSE

It is important to understand the dynamic behavior of the pump to assess the error associated with the pump's response to an input voltage. The inability of the pump to respond instantaneously to a new input voltage is a characteristic of many Electro-mechanical systems. Usually these systems can be described as either a first order linear system or a second order linear system. A first order linear system can be characterized by the differential equation:

$$\tau \frac{dx}{dt} + x(t) = y(t)$$

Where y(t) is the input at time t, x(t) is the output at time t, and τ is the time constant of the system.[5] The third component in this equation, τ dx/dt, is the transient response, where τ is the time constant of the system.

The equation relating the response of a first order linear system to a rising step input is:

$$x(t) = X_f + (X_i - X_f) \cdot e^{-t/\tau}$$

Where x(t) is the output response, X_f is the final value of x after the transients have died out, X_i is the initial value of x, just after the step was imposed, and τ is the time constant of the system.[5] Graphically, the output of a first order system in response to a positive step input is an exponentially rising curve.

A series of experiments was conducted to determine whether the pumps responded to the changes in voltage from the PSCU as a first order system. In the process, we also evaluated the approximate time constant of the system response. These tests were complicated by the fact that the pump flows were measured using mass flow meters, instruments that also had a first order response.

Two Txtools programs were written that would instruct the PSCU to generate of a series of voltage steps. The first program had voltage steps that were 15 seconds in length, and the second program was identical except that the voltage steps were 60 seconds long. The PSCU output voltages to each pump (A through D) were first calibrated with a digital voltmeter. Next, each pump was tested three times using Vtest15s. A mass flow meter was attached upstream and its output voltage was digitally recorded at 100 Hz sampling rate on a personal computer using Dataq software and hardware (Dataq Inc.). Finally, the Vtest60 program was sent to pumps "A" and "C" two times, hereafter referred to as "Run 1" and "Run 2". During Run 1, the mass flow meter was located upstream of the pump. For Run 2, however, a valve was used to select whether the pump pulled air through the mass flow meter or not. For each 60-second voltage step in the Vtest60 program, the air was pulled through the mass flow meter for only the last 30 seconds to decouple the mass flow meter transient response from the pump response.

The mass flow meter output voltages for Runs 1 and 2 were overlaid to allow for a comparison of the combined pump - mass flow meter responses and the mass flow meter alone. Note that after 30 seconds into each step, we assumed the pump response had

stabilized so that the mass flow meter response would not be linked to the pump response.

The time constant for the mass flow meter (whose published value was 0.08 seconds) and the time constant for the pump and mass flow meter in series were calculated. The time constant for the pump could then be estimated by assuming the time constants would be additive.

Figure 3 shows the responses recorded from all four pumps to the same input voltage applied for 15 seconds per step. Note that except for an offset difference (due to the mass flow meters), the responses appear to be indistinguishable. The signals were further examined to compare pump responses during a positive step and negative step. These tests indicated that the pump and mass flow meter system behaved in a symmetrical fashion and there was no evidence of hysteresis.

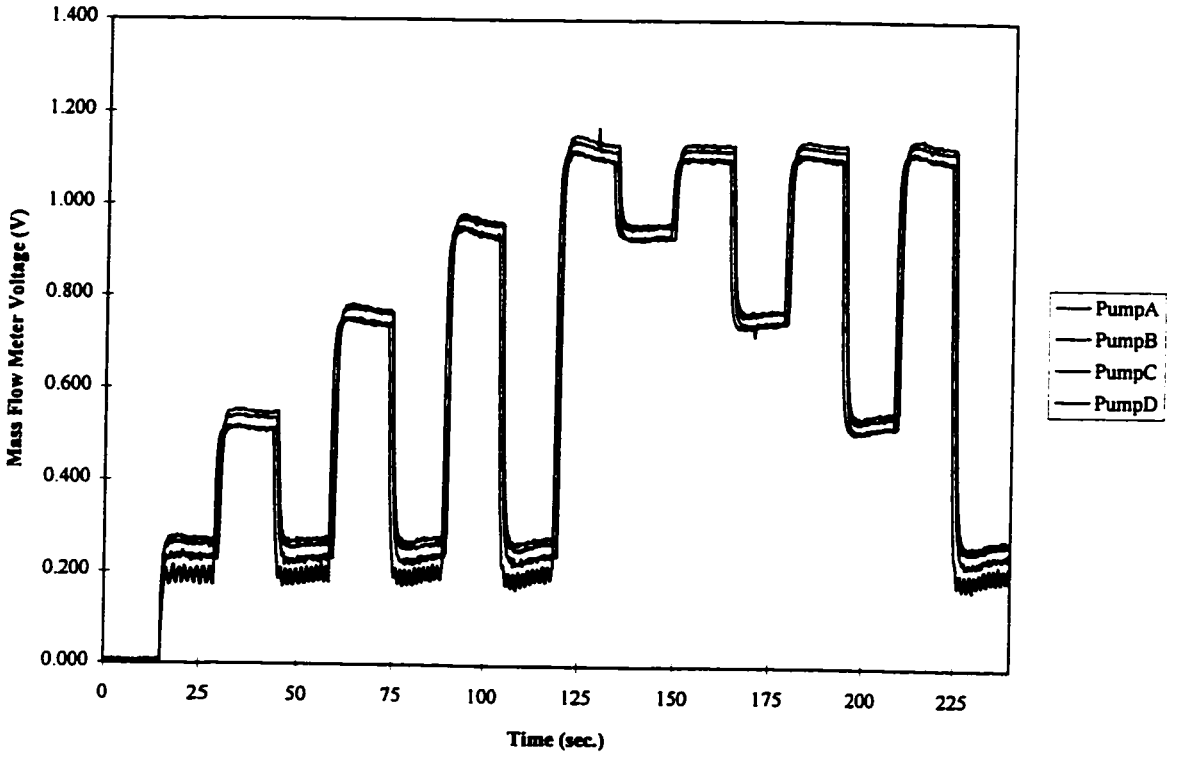


Figure 3: Mass flow meter voltages for pumps A, B, C, and D, using 15-second steps.

Figure 4 displays the mass flow meter data obtained using the pump "A". The test results illustrated two test conditions: (1)when the mass flow meter measured the pump flow throughout the procedure, and (2) when the mass flow meter was introduced at 30 seconds into each step. Figure 5, shows one of the steps shown in the previous Figure in more detail. In Figure 5 the mass flow meter response alone at 30 seconds is shifted in time to overlay the combined pump and mass flow meter response at 0 seconds into the step (under Condition #1).

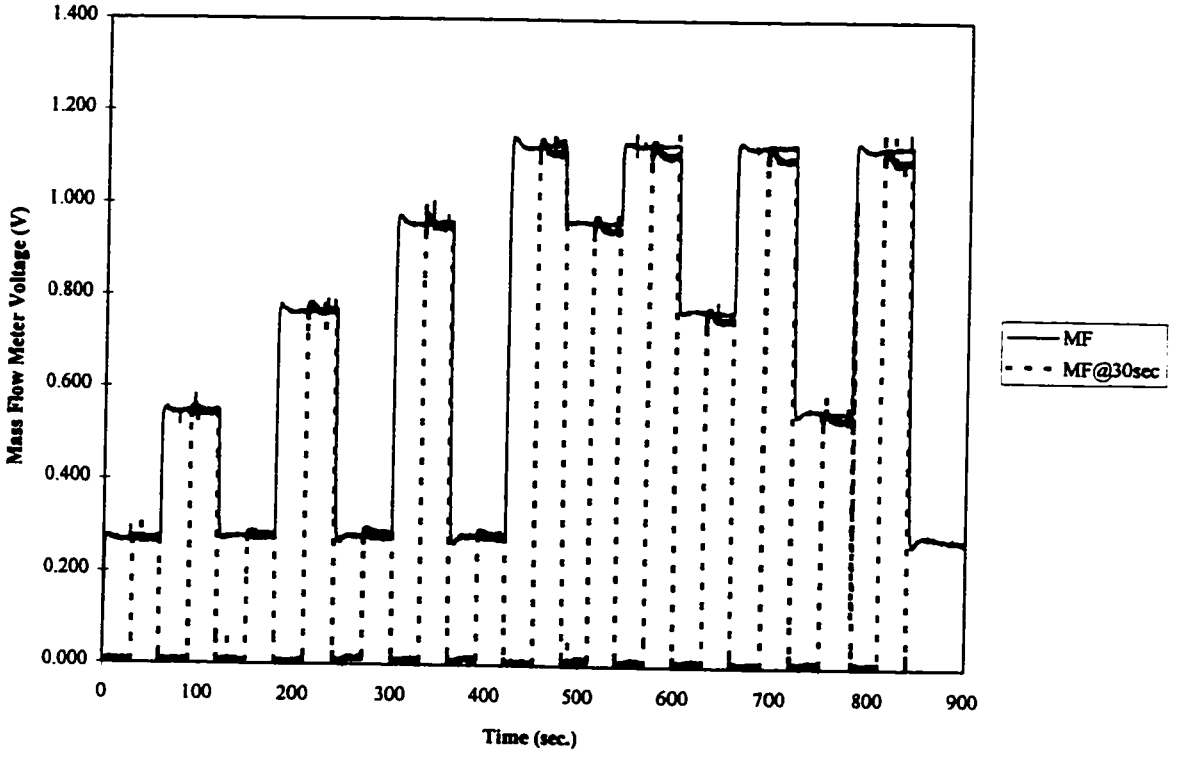


Figure 4: Mass flow meter voltage during step input to GilAir5, under two conditions: (1) flow through mass flow meter was continuous (solid line), (2) flow through mass flow meter was commenced at 30 seconds into each voltage step (dotted line).

Based on these tests the response of the pump-mass flow meter system is predominantly a first order linear system, with at most a very small second-order component. Figures 4 and 5 suggests that the second order component can be attributed to the mass flow meter response. The mass flow meter documentation states that the mass flow meter time constant is 0.08 seconds. The time constant for the mass flow meter in this experiment was calculated to be approximately 0.08 seconds, as expected. The time constant for the

pump-mass flow meter system was then calculated using the data from Condition #1 for the same input step. This time constant for the combined flow meter and pump system was approximately 0.33 seconds and by superposition, the time constant for the pump was estimated to be 0.25 seconds.

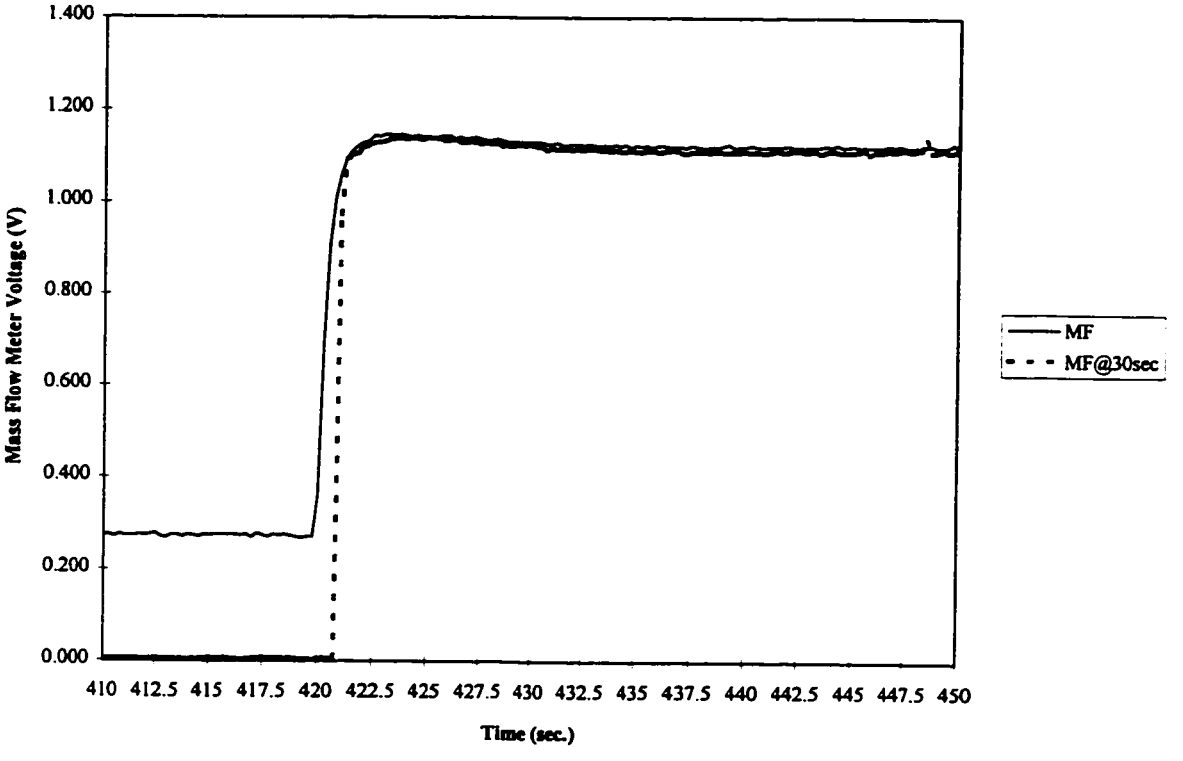


Figure 5: The data from the largest step in Figure 4, with mass flow meter response at 30 seconds into step shifted to overlap with original recorded values.

LABORATORY VALIDATION

A laboratory simulation was performed in order to collect human subject data to send to the PSP. The subjects bicycled on an ergometer in order to simulate workers changing work rates on the job. In order to evaluate the performance of the PSP and PSCU we needed a measure of the "true" minute ventilation from active human subjects. For this purpose, a "Primary Standard" reference signal was generated in a controlled laboratory experiment. The Primary Standard was the integrated, positive half-wave rectified signal from a calibrated pneumotachometer fitted to an exercising subject (corresponding to inhaled minute ventilation). This reference signal was digitally recorded at 125 Hz sampling rate using a personal computer. Since this reference signal was digitally recorded with high fidelity, it could be replayed later to simulate an input to the PSCU from a subject. For comparison with the Primary Standard, the same signal was processed by the PSCU and sent to the pump as the control signal, to create the "Pneumo Method" for pump control. The only difference between the Primary Standard and the Pneumo Method was the signal processing location (Dataq Software for the Primary Standard and the PSCU for the Pneumo Method). Differences between the Primary Standard and the Pneumo Method, therefore, could be specifically attributed to the PSCU processing of the pneumo-tachometer signal.

Fifteen males and fifteen females participated in the study. Summary statistics for the age, height, weight, and BMI for males and females are provided in Tables 1. All subjects provided written informed consent prior to participating in the session.

Minute ventilation, heart rate, and RIP data were recorded for all 30 subjects over a 40-minute sampling period that involved four work rates. The subjects breathed through a pneumo-tachometer mouth unit and wore a nose clip to assure that all breaths were recorded. The Pneumo-tachometer was calibrated immediately before the session using a spirometer as the primary standard. Calibrations were within 2% of the expected values.

Table 32: Summary of subjects' age, height, weight, and body mass index.

MALES	**Mean**	**Minimum**	**Maximum**
Age (yr.)	31.1	26	43
Height (cm)	182.5	165	193
Weight (lb.)	186.9	152	285
BMI (kg/m^2)	25.5	21	40
FEMALES			
Age (yr.)	29.5	20	60
Height (cm)	166.0	155	173
Weight (lb.)	137.7	115	190
BMI (kg/m^2)	22.6	19	32

The 40 minute sampling period was broken into 8 five-minute sections, with the following work rates assigned to each section: rest, 40 Watts, rest, 80 Watts, 40 Watts, rest, 120 Watts, and rest. When biking, subjects were required to pedal at 50 rotations per minute; a metronome set to 100 beats per minute provided an auditory cadence, and the pedaling rate could be visually confirmed via a display on the ergometer itself. Subjects were allowed to stand during the periods at rest. In addition, they were allowed to read if

they wished to and could do so while biking steadily at 50 rotations per minute. (Music was not allowed.)

A batch file was used to process the digitally recorded data and send the subject data to the PSCU via a Pentium computer and digital to analog converter board controlled by a QuickBasic program. This program sends the data from a comma-delimited file to a digital-to-analog converter at a rate of 125 samples per second per channel. Prior to sending a subject's data, this program was used to send zero volts to both the RIP and pneumo-tachometer channels of the PSCU. Three Txtools programs were run on a laptop Pentium computer to measure the voltage offsets of the PSCU.

A TxBASIC program titled LABPSP was downloaded into the PSCU and started. This program prompted the user to input several values, including the PSCU offsets and the four pump calibration coefficients. The program was initiated manually, to correspond with the start of the Pentium computer sending a subject's data file to the PSCU. The PSCU program then calculated and recorded 15-second averages of minute ventilation estimates for the Pneumo Method. After each subject's 40-minute sampling period had been sent to the PSCU, the 15-second averages were downloaded from the PSCU to the laptop Pentium computer and saved as an Excel file. The data was integrated over the sampling period to attain the total inhaled volume estimated for that subject.

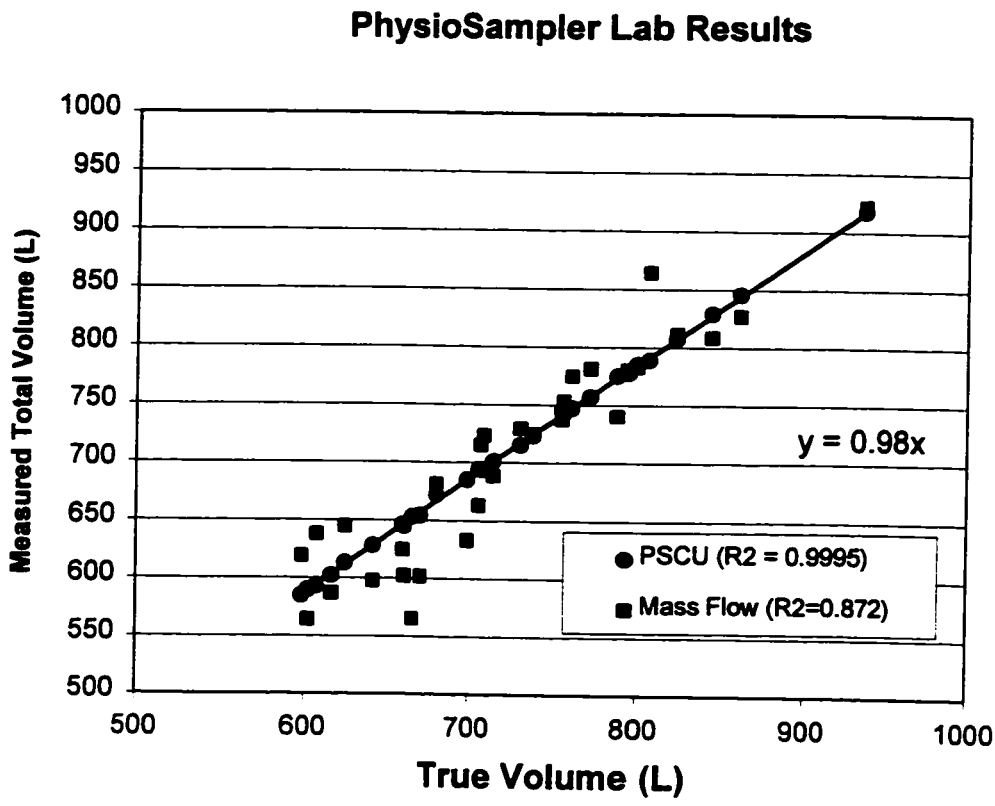


Figure 6: Scatter plot of total inhaled volume recorded from for 30 subjects during exercise.

The raw pneumo-tachometer data were processed in order to calculate the Primary Standard. The processing of this signal involved summing the positively rectified data over the sampling period and applying the pneumo-tachometer calibration curve.

In addition to the recorded pneumo-tachometer signal and the PSCU data record, mass flow meters were used to measure the airflow rate pulled through charcoal tubes attached to the PSPs. The flow rate signals from the mass flow meters were digitally recorded

using the Dataq Instruments software and hardware, and integrated to estimate the total sample volume drawn by the PSP. The mass flow meters were calibrated against a bubble meter (Gilibrator). Mass flow meters were attached upstream of the charcoal tube and allowed to warm up for at least 15 minutes prior to use. The pumps and flow meters were randomly assigned each data set, in order to eliminate any bias.

Figure 6 shows the results from the laboratory testing with active subjects. The PSCU data points (circles) represent the volume data predicted by the PSCU (i.e. the Pneumo Method) compared to the primary standard on the X axis. The mass flow results (squares) represent the integrated volume recorded from the mass flow meters and charcoal tubes compared to the primary standard. Both data sets had a regression line that passed through the origin with a slope very close to unity (0.98), however the mass flow results displayed more scatter over the 30 subjects tested.

DISCUSSION

This set of experiments was successful in demonstrating that the PSP system works as expected. A concordance analysis revealed that the Pneumo Method estimates had 100 percent precision (as measured by the correlation coefficient) and 98.6 percent accuracy when compared to the Primary Standard. The slight loss of accuracy appears to be caused by an electrical bias in the PSCU analog circuits; this bias could easily be corrected.

The mass flow results for to PSP testing had the lower precision (0.934 correlation) as expected but had a similar slope to the line. Based on the results from the dynamic testing, these errors in total volume probably reflect errors introduced by the mass flow meters themselves. Since it takes approximately three to four time constants for the system to reach a new steady-state operating condition, we can assume that the pump stabilizes at each new flow rate within a second. The pump responded quickly and smoothly to a step voltage change, stabilizing at the new flow rate within less than a second. There was no evidence of overshoot in the response or of oscillation in the feedback control. The magnitude of the error resulting from "wind-up time" was not significant compared to the 15 second update rate. This conclusion follows from the data indicating that the mass flow meter results were not significantly different from the PSCU data logged results.

One notable PSP design issue is the balance between the averaging time for the input signals and the update period for the output voltage to the pump. The shorter the averaging period, the better the PSP will be able to capture peak exposures; unfortunately, this must be balanced against the error introduced by "wind-up time". This particular error is inherent in a pump that changes pump rate; there must be a transient period during which the pump rate approaches the steady-state pump rate for a given input voltage. The more frequently the pump rate is altered, the more error will accumulate. A 15-second averaging period for the input signal was selected for two reasons: (1) it would be long enough to limit the error associated with pump response

transients just mentioned, and (2) research performed by Mermier *et al.* concluded that 15-second averages of heart rate and minute ventilation were closely linked.[6]

We have presented a hardware design for a new PSP that is suitable for field use by industrial hygienists. The practical application of this device for exposure assessment remains to be developed along with a theoretical framework theory for interpreting the data from a PSP.

BIBLIOGRAPHY

[1]Wasserman, K., Hansen J.E., Sue, D.Y., Whipp, B.J., and R. Casaburi. Principles of exercise testing and interpretation, 2nd Ed. Lea & Febiger, 1994.

[2]Kucharski, R. *A personal dust sampler simulating variable human lung function.* British Journal of Industrial Medicine, 37:194-196, 1980.

[3]Satoh, T., Higashi, T., Sakurai, H., and Omae K. *Development of a new exposure monitoring system considering pulmonary ventilation (DEM 1).* Keio Journal of Medicine, 38(4):432-443, 1989.

[4]Levine, M.S. *A respiration-modulated personal air sampling pump.* Applied Occupational Environmental Hygiene, 9(12):994-1005, 1994.

[5]Basic Electronic Circuit Analysis. Johnson and Hilburn. Inglewood Cliffs, NJ: Prentice-Hall, 1978.

[6]Mermier, C.M., Samet, J.M., and Lambert, W.E. *Evaluation of the relationship between heart rate and ventilation for epidemiologic studies.* Archives of Environmental Health, 48(4):263-269, 1993.

Acknowledgment

Partial financial support for this research was provided by the National Institute of Occupational Safety and Health, grant RO1-OH-03134, and the ALCOA Science Foundation

APPENDIX B: COMPUTER SIMULATION VISUAL C++ CODE

File #1

```
#include <stdio.h>
#include <math.h>
#include <nr.h>

FILE *tmp;

void main()
{
        int i,m,j1,j2,j3,pflag;
        int knum,g1num,g2num;
        long sim, sample;
        long ngsdvi,ngsdc,nk;
        long ii,*int_ptr;
        float kcorr,corrsum,vtsum;
        float gmvi, gsdvi, lngmvi, lngsdvi;
        float gmc, gsdc, lngmc, lngsdc;
        float rand1,rand2,data;
        float a, bdenom, bnum;
        float b, vtratio[5000],corr[5000];
        float c[5000],vdot[5000];
        float kmin,kmax,g1min,g1max,g2min,g2max;
        char idstring[81],fname[81];

        int_ptr=&ii;
        ii= -1;

        printf("C.K. Hart thesis simulator, 1998\n\n");

        printf("Enter output filename (including path if you want):\n");
        scanf("%s",fname);
        printf("Would you like to output the results of EACH simulation?\n");
        printf("    1=Yes\n");
        printf("    0=No, report only the average from each simulation\n");
        scanf("%d",&pflag);
        tmp=fopen(fname,"w");
        printf("Enter file/run identifier < 81 characters\n");
        printf("use characters other than space please:\n");
```

```
 scanf("%s",idstring);
 fprintf(tmp,"%s\n",idstring);
 fprintf(tmp,"kvalue--, pearsonr, gsdvi---, gsdc----, vtratio-\n");
 printf("Enter number of samples (<5000): ");
 scanf("%d",&sample);
 printf("Enter number of simulations (<5000): ");
 scanf("%d",&sim);
 printf("Enter negative integer: ");
 scanf("%d",&ii);
 printf("Enter GMVi: ");
 scanf("%f",&gmvi);
 printf("Enter GMC: ");
 scanf("%f",&gmc);

 printf("Enter min k: ");
 scanf("%f",&kmin);
 printf("Enter max k: ");
 scanf("%f",&kmax);
 printf("Enter number of k values you want: ");
 scanf("%d",&knum);

 printf("Enter min GSDVi: ");
 scanf("%f",&g1min);
 printf("Enter max GSDVi: ");
 scanf("%f",&g1max);
 printf("Enter number of GSDVi values you want: ");
 scanf("%d",&g1num);

 printf("Enter min GSDC: ");
 scanf("%f",&g2min);
 printf("Enter max GSDC: ");
 scanf("%f",&g2max);
 printf("Enter number of GSDC values you want: ");
 scanf("%d",&g2num);

 lngmvi=log(gmvi);
 lngmc=log(gmc);

 for (j1=0;j1<=(knum-1);j1++)
 {kcorr=kmin+(kmax-kmin)*j1/(knum-1);
         for (j2=0;j2<=(g1num-1);j2++)
         {gsdvi=g1min+(g1max-g1min)*j2/(g1num-1);
             for (j3=0;j3<=(g2num-1);j3++)
```

```
{gsdc=g2min+(g2max-g2min)*j3/(g2num-1);
    lngsdvi=log(gsdvi);
    lngsdc=log(gsdc);
    corrsum=0;
    vtsum=0;
    for(m=1;m<=sim;m++)
    {
        a=0;
        bnum=0;
        bdenom=0;
                for(i=1;i<=sample;i++)
                {
                        rand1=gasdev(int_ptr);
                        rand2=gasdev(int_ptr);
                        data=kcorr*rand1+(1.00-kcorr)*rand2;
                        vdot[i]=exp(lngmvi+lngsdvi*rand1);
                        c[i]=exp(lngmc+lngsdc*data);
                        a += c[i];
                        bdenom += vdot[i];
                        bnum += vdot[i]*c[i];
                }
            b=bnum/bdenom;
            vtratio[m]=sample*b/a;
            vtsum += vtratio[m];
            corr[m] = prsn(c, vdot, sample);
            corrsum += corr[m];
            if(pflag==1)
            {fprintf(tmp,"%f,        %f,        %f,        %f,
%f\n",kcorr,corr[m],gsdvi,gsdc,vtratio[m]);
                printf("%f,        %f,        %f,        %f,
%f\n",kcorr,corr[m],gsdvi,gsdc,vtratio[m]);
            }
            }
        if(pflag==0)
        {corrsum /= sim;
                vtsum /= sim;
                fprintf(tmp,"%f,        %f,        %f,        %f,
%f\n",kcorr,corrsum,gsdvi,gsdc,vtsum);
                printf("%f,        %f,        %f,        %f,
%f\n",kcorr,corrsum,gsdvi,gsdc,vtsum);
        }
        }
    }
```

```
        }
        fclose(tmp);
}
```

File #2

```
#include <math.h>
#define TINY 1.0e-20

float prsn(float x[], float y[], long n)
{
        unsigned long j;
        float yt,xt;
        float syy=0.0,sxy=0.0,sxx=0.0,ay=0.0,ax=0.0;

        for (j=1;j<=n;j++) {
                ax += x[j];
                ay += y[j];
        }
        ax /= n;
        ay /= n;
        for (j=1;j<=n;j++) {
                xt=x[j]-ax;
                yt=y[j]-ay;
                sxx += xt*xt;
                syy += yt*yt;
                sxy += xt*yt;
        }
        return(sxy/(sqrt(sxx*syy)+TINY));
}
#undef TINY
```

File #3:

```
#include <math.h>

float gasdev(long *idum)
{
        float ran1(long *idum);
        static int iset=0;
        static float gset;
        float fac,rsq,v1,v2;

        if (*idum < 0) iset=0;
        if  (iset == 0) {
                do {
                        v1=2.0*ran1(idum)-1.0;
                        v2=2.0*ran1(idum)-1.0;
                        rsq=v1*v1+v2*v2;
                } while (rsq >= 1.0 || rsq == 0.0);
```

```
            fac=sqrt(-2.0*log(rsq)/rsq);
            gset=v1*fac;
            iset=1;
            return v2*fac;
      } else {
            iset=0;
            return gset;
      }
}
```

File #4:

```
/* note #undef's at end of file */
#define IA 16807
#define IM 2147483647
#define AM (1.0/IM)
#define IQ 127773
#define IR 2836
#define NTAB 32
#define NDIV (1+(IM-1)/NTAB)
#define EPS 1.2e-7
#define RNMX (1.0-EPS)

float ran1(long *idum)
{
      int j;
      long k;
      static long iy=0;
      static long iv[NTAB];
      float temp;

      if (*idum <= 0 || !iy) {
            if (-(*idum) < 1) *idum=1;
            else *idum = -(*idum);
            for (j=NTAB+7;j>=0;j--) {
                  k=(*idum)/IQ;
                  *idum=IA*(*idum-k*IQ)-IR*k;
                  if (*idum < 0) *idum += IM;
                  if (j < NTAB) iv[j] = *idum;
            }
            iy=iv[0];
      }
      k=(*idum)/IQ;
      *idum=IA*(*idum-k*IQ)-IR*k;

if (*idum < 0) *idum += IM;
      j=iy/NDIV;
      iy=iv[j];
      iv[j] = *idum;
      if ((temp=AM*iy) > RNMX) return RNMX;
      else return temp;
}
#undef IA
```

```
#undef IM
#undef AM
#undef IQ
#undef IR
#undef NTAB
#undef NDIV
#undef EPS
#undef RNMX
```

APPENDIX C: PHONE INTERVIEW

"My name is Cheryl and I am a Ph.D. student that has developed a new sampling pump for my dissertation research. This pump will improve our ability to assess workers' exposure to toxic chemicals that they work with, solvents like paint, for example. My new pump changes its pump rate along with the workers breathing rate. So in order to test my new pump, I need to collect breathing and heart rate data from real people while they bicycle on a stationary bike for me.

"Does this sound like something you're interested in?"

If YES: "Let me describe exactly what you would do. When you arrived, I would go over this procedure again with you, then you would read a form which explains it again. If you still want to participate, then you sign the consent form.

"Then, I'll hook up some equipment to you that will monitor your breathing and heart rate for me. I'll have to calibrate this equipment first, to make sure that it's telling me the right values. This will entail 2 procedures. In the first one you'll just sit at rest for 2 minutes and then do three "breathing maneuvers" that are very quick and easy. You just hold your breath for a few seconds and move your abdomen in and out. I'll demonstrate for you first, don't worry. Next, I'll have you sit for 2 MORE minutes at rest, then have you bike for 2 five-minute periods. If all goes well, then we'll start the actual data collection. This part shouldn't be difficult, it just gets a little boring. You'll be sitting on the bike for 40 minutes, but you'll only have to be biking for half the time. Also, there's one point where I'll offer you a water break if you want it. You'll have this mouthpiece in your mouth, so I can measure your breathing, and sometimes your mouth gets a little dry.

"So, after all that I'll just need you to do 3 more of those breathing maneuvers for me, and then you're finished.

"Does this STILL sound like something you're interested in participating in?"

If YES: "OK. Then I need to ask you a few questions to see if you're eligible to participate.

1. How old are you?
2. Do you have any lung or cardiovascular disease that you're aware of (for example, asthma, bronchitis, emphysema, high blood pressure...)?

3. Do you take any medication?
4. Do you exercise? If so, how often?
5. Have you ever had any problems during exercise, such as light-headedness or sore muscles?
6. Do you know of any reason why you should not participate in this study?
7. How tall are you?
8. How much do you weigh?

"Based on the information you've just given me, you are/are not eligible to participate...."

APPENDIX D: SUBJECT DATA COLLECTION AND PROCESSING CHECKLIST

Subject# ___ Session# ___ Date _____

SUBJECT DATA COLLECTION

EQUIPMENT PREP
__ turn on power source and check voltages (A and B)
__ turn on pneumotach heater
__ check pneumotach calibration
slope= _____
intcpt= _____
__ get water for subject

SUBJECT PREP
__ describe protocol to subject
__ have subject read and sign human subject's form
__ have subject complete SubjHx macro
__ announce last bathroom break
__ attach HR band
__ secure RIP bands to stress test shirt
__ have subject sit on bike at rest during calibration period for Respitrace

RIP CALIBRATION (pre-data collection)
__ D:\CODAS\ acodas S_N_RC1 -four
__ when "cal" reached perform 3 isovolume maneuvers (low, med, high)
__ attach headgear with pneumotach, add noseclips
__ collect 2 minutes of breathing data at rest

HR-VI CALIBRATION
__ explain HR calibration protocol
__ D:\CODAS acodas S_N_HC 1000K -four
__ Bike: 2 min. @ rest, 5 min. @ 40 W, 5 min. @ 80 W

DATA COLLECTION
__ explain WR protocol for data collection (on mouthpiece entire time)
__ D:\CODAS acodas CS_N_DT 4000K -four
__ Bike: 5 min. @ rest __, 40W __, rest __, 80W __, 40W __, rest __, 120W __, rest __

RIP CALIBRATION (post-data collection)

__ D:\CODAS\ acodas S_N_RC2 -four
__ collect 2 minutes of data at rest
__ remove pneumotach/headgear/noseclips
__ perform 3 isovolume maneuvers (low, med, high)

POST PNEUMOTACH CALIBRAITON
__ check pneumotach calibration

DATA PROCESSING

RIP CALIBRATION (pre-data collection)
__ RIPCAL S_N_RC1 (batch file)
__ D:\CODAS post S_N_RC1
__ cut regular breathing @ rest = S_N_C1B.xlw (Lotus format=5)
__ cut each isovolume maneuver = S_N_C1_.xlw (ctrl+arrow to skip to event markers)
__ --> windows --> excel: run SubjIso macro
__ record slope and intercept
slope= _____
intcpt= _____
__ run RIPcal macro
slope= _____
intcpt= _____
R^2= _____ acceptable? Y N

HR-VI CALIBRATION
__ HRATE S_N_HC (batch file)
__ D:\CODAS AdvPost S_N_HC
__ F3, Shift+9 (check for "C" line)
__ B, Select TFB 45-105, Ctrl + T, Y, save as XS_N_HC.xlw, comment "rest,pne"
__ C, Select TFB 45-105, Ctrl + T, Y, comment "rest,HR"
__ B, Select TFB 345-405, Ctrl + T, Y, comment "40W,pne"
__ C, Select TFB 345-405, Ctrl + T, Y, comment "40W,HR"
__ B, Select TFB 645-705, Ctrl + T, Y, comment "80W,pne"
__ C, Select TFB 645-705, Ctrl + T, Y, comment "80W,HR"
__ --> windows --> excel, run HRcal macro
slope= _____
intcpt= _____
R^2= _____ acceptable? Y N
__ run PredHRVI macro

RIP CALIBRATION (post-data collection)
__ RIPCAL S_N_RC2 (batch file)

__ D:\CODAS post S_N_RC2
__ cut regular breathing @ rest = S_N_C2B.xlw (Lotus format=5)
__ cut each isovolume maneuver = S_N_C2_.xlw (ctrl+arrow to skip to event markers)
__ --> windows --> excel: run SubjIso macro
__ run RIPcal macro
slope= _____
intcpt= _____
R^2= _____ acceptable? Y N

TATTLETALE SAMPLE ONLY

PROCESS EXERCISE DATA
__ D:\Codas TXDAT S_N_DT_
__ D:\Codas POST S_N_DAT_, Home, F4, End, F6 (to copy entire session)
__ Name new file XS_N_DT.XLW, file type 5, open new file in Word, save as text file

MEASURE SIGNAL OFFSETS
__ Open Analog in TXTOOLS on laptop
__ On Lab pentium C:\QB4, QB /l pro, B:
__ Open D:\Cheryl\Senddat4, run: D:\Cheryl\zeroanlg.txt

INPUT TO TATTLETALE PROGRAM
Offsets
____Channel 0 (~2.5)
____Channel 2 (~0.2)
____Channel 6 (~2.5)
Pneumotach calcurve
____M
____B
Ventilation vs. HR
____M
____B
Isovolume curve
____M
____B
Ventilation vs. Sum Derivatives
____M
____B
Subject age
____yr
Subject height
____cm

Subject weight
____lbs
Subject resting heart rate
____BPM

RUN TATTLETALE PROGRAM
____Run C:\txtools\txtools, SMPLONLY
____Input blanks for pump data, input values for remaining prompts
____On pentium (QB4,Senddat4) run filename D:\CODAS\S_N_DT

OFFLOAD DATA
____Offload to file Offload.dat, open Excel, run macro "Offload"

APPENDIX E: SUMMARY OF NOTATIONS MADE ON CHECKLIST

Subject	Session	Notations
1	1	No water break during data collection.
1	2	HR unit came off waistband and added false signal during first 40 Watt period. 40 Watt period repeated. Bad data cut from data file post-sampling.
2	1	RIP bands worn over t-shirt.
5	1	First isovolume maneuver repeated. HR band slipped during HR-Vi calibration (at beginning of 80 Watt period); will not affect calibration curve.
6	1	None.
7	1	No water break during data collection.
8	1	RC wire noisy due to tear in wire. Do not use RIP data.
9	1	Accidentally biked 120 Watts for first 40 seconds of the first 40 Watt period. During 120 Watt period, first 2 minutes without nose clip. 120 Watt period repeated. Bad data cut from data file post-sampling.
10	1	None.
10	2	First 40 Watt period was greater than 5 minutes. Extra data cut from data file post-sampling.
11	1	First RIP calibration data not accurate; nose clips not worn. Use Second RIP calibration data.
11	2	Noticed periodic HR signal additions (looks like possible noise).
12	1	At 3 minutes into the 120 Watt period, wattage lowered to 100 due to saturation of the pneumotachometer signal.
13	1	None.
13	2	Subject removed nose clips before data collection paused for water break. Data file corrected post-sampling.
14	1	During first rest period, adjusted the head strap holding the mouthpiece up.
14	2	None.
15	1	Both RIP bands became unhooked. Not discovered until after data collection finished (under a t-shirt). Do not use RIP data.
16	1	Nose clips forgotten after water break. 120 Watt period re-started. Data file adjusted post-sampling.
16	2	Reported change in caffeine intake since Session 1. Now 1 espresso and 24 ounces coffee per day. Was 64 oz. Pepsi per day.

Subject	Session	Notations
17	1	None.
18	1	None.
19	1	Spit plugged pneumotachometer during data collection. Data cannot be used.
19	2	Pneumotachometer pressure line hose barbs not tightened properly. However, the calibration was the same as for previous subjects and remained stable across the session.
20	1	None.
20	2	Room was very hot; may have affected HR-Vi relationship. Stopwatch stopped during 80 Watt period; subject instructed to bike a little longer to ensure 5 minutes of data was collected. Extra data removed from data file post-sampling.
21	1	None.
21	2	Ribcage band velcro popped off during second RIP calibration. Sample data unaffected.
22	1	Subject slightly ill (a cold). Pedal strap broken, lack of resistance a problem with biking continuity. Pedal repaired with duct tape.
22	2	None.
23	1	Abdominal band rode up during session, over a roll of adipose tissue. Wattage during 120 Watts readjusted to 100 Watts.
24	1	None.
25	1	None.
25	2	First 40 Watt period longer than 5 minutes. Extra data removed from data file post-sampling.
26	1	First 30 seconds of 120 Watt period without nose clips. Data file adjusted post-sampling.
27	1	120 Watt period too long. Extra data removed from data file post-sampling.
28	1	None.
28	2	Second isovolume maneuver (during first RIP calibration) did not record for an unknown reason.
29	1	None.
29	2	None.
30	1	Pause during second rest period in order to switch nose clips.
30	2	Second 40 Watt period 15 seconds too long. Extra data removed from data file post-sampling.
31	1	None.
31	2	None.
32	1	None.

APPENDIX F: HUMAN SUBJECT CONSENT FORM

UNIVERSITY OF WASHINGTON

CONSENT FORM

PHYSIOLOGIC SAMPLER FOR AIRBORNE HEALTH HAZARDS

Investigators:

Michael S. Morgan, Sc.D., Associate Professor,
Department of Environmental Health
685-3221

Michael G Yost, Ph.D., M.S., Assistant Professor,
Department of Environmental Health
685-7243

William Daniell, M.D., Assistant Professor,
Department of Medicine and Environmental Health
685-3160

24-Hour Emergency Telephone Number: 548-6190
(University Hospital Paging Operator: Page Dr. William Daniell)

Investigators' Statement

PROCEDURES

Prior to the study, potential subjects will be asked to fill out a questionnaire. The questionnaire asks about health, smoking history and exercise patterns.

Subjects who are eligible and agree to participate will be scheduled for either one or two sessions (at least a week apart). All sessions will follow the same protocol. The subject will be fitted with equipment to monitor their ventilation and heart rate. They will perform two calibration procedures in order to calibrate (1) the equipment that monitors their ventilation and (2) their heart rate versus their ventilation. The first calibration procedure involves the subject sitting on the exercise bicycle at rest for 2 minutes, followed by three breathing maneuvers. The breathing maneuvers are performed by closing the mouth and pinching the nose while moving the abdomen in and out several times. The second calibration procedure involves the subject sitting on the exercise bicycle at rest for 2 minutes, followed by two 5-minute periods where the subject bikes at 40 Watts and 80 Watts.

After the calibrations, the actual exercise session will start. The subject will be on the exercise bicycle for 40 minutes, but will only be biking for a total of 20 minutes. The remaining 20 minutes the subject will be at rest. The 40 minute period is broken up into eight 5-minute periods with the following work rates assigned to each period: rest, 40 Watts, rest, 80 Watts, 40 Watts, rest, 120 Watts, rest. At the end of the sixth period, the subject will be given the opportunity to stop, stand, and have a drink of water.

After the exercise session, one last calibration will be performed in order to validate the equipment measuring the subject's ventilation. This calibration is identical to the first calibration described above (2 minutes at rest followed by 3 breathing maneuvers).

RISK, STRESS AND DISCOMFORT

The bicycle exercise will require riding at a pace of 50 rotations per minutes. The exercise bicycle will permit adjustment of the saddle height and handlebar position in order to maximize comfort. The exercise may cause fatigue, shortness of breath, muscle soreness, and/or joint or muscle injury. If the subject feels any pain or discomfort, he/she should inform the investigators immediately and the exercise will be stopped at once.

OTHER INFORMATION

The identity of the subjects will remain confidential. The subject's identity will not be retained in the data for this study, and no records will be kept as to a specific participant. The investigators named above will have access to the data, which will be retained indefinitely. The results of this study will be published in the scientific literature, but only in a summary form so that no individual data can be identified with any participant. In the event of a physical injury as a direct result of the study procedures, subjects will be referred for appropriate treatment, at no cost, within the limits of the University's compensation plan.

Subjects may refuse to participate and may withdraw from the study at any time without penalty or loss of benefits to which they are otherwise entitled. Subjects scheduled for one session will receive $40 for their participation. Subjects scheduled for two sessions will receive $40 for their first session and $60 for their second session.

Investigator's signature _____ Date _____

SUBJECT'S STATEMENT

The study described above has been explained to me, and I voluntarily agree to participate in this activity. I have had an opportunity to ask questions. I understand that further questions I may have about the research or about my rights as a subject will be answered by one of the investigators listed above.

Signature of the Subject _____ Date _____

Rev August 20, 1998
Copies to: Subject
Investigators' File

APPENDIX G: SUBJHX EXCEL MACRO

```
'
' SubjHx1 Macro
' Macro recorded 6/24/96 by Cheryl Hart
'
'
Sub SubjHx1()
'save file in subject's folder
    Workbooks.Add
    Cells.Select
    With Selection
        .HorizontalAlignment = xlCenter
        .VerticalAlignment = xlBottom
        .WrapText = False
        .Orientation = xlHorizontal
    End With
    Rows("1:1").Select
    With Selection.Font
        .Name = "Arial"
        .FontStyle = "Bold"
        .Size = 10
        .Strikethrough = False
        .Superscript = False
        .Subscript = False
        .OutlineFont = False
        .Shadow = False
        .Underline = xlNone
        .ColorIndex = xlAutomatic
    End With

    ActiveCell.FormulaR1C1 = "Subject#"
    Range("B1").Select
    ActiveCell.FormulaR1C1 = "Age"
    Range("C1").Select
    ActiveCell.FormulaR1C1 = "Sex"
    Range("D1").Select
    ActiveCell.FormulaR1C1 = "Ht_ft"
    Range("E1").Select
    ActiveCell.FormulaR1C1 = "Ht_in"
```

```
Range("F1").Select
ActiveCell.FormulaR1C1 = "Ht_cm"
Range("G1").Select
ActiveCell.FormulaR1C1 = "Wt_lb"
Range("H1").Select
ActiveCell.FormulaR1C1 = "Wt_kg"
Range("I1").Select
ActiveCell.FormulaR1C1 = "BMI"
Range("J1").Select
ActiveCell.FormulaR1C1 = "Self_Fit"
Range("K1").Select
ActiveCell.FormulaR1C1 = "Sweat_No."
Range("L1").Select
ActiveCell.FormulaR1C1 = "ExType"
Range("M1").Select
ActiveCell.FormulaR1C1 = "ExDur"
Range("N1").Select
ActiveCell.FormulaR1C1 = "ExFreq"
Range("O1").Select
ActiveCell.FormulaR1C1 = "RespD"
Range("P1").Select
ActiveCell.FormulaR1C1 = "CardioD"
Range("Q1").Select
ActiveCell.FormulaR1C1 = "HRmed"
Range("R1").Select
ActiveCell.FormulaR1C1 = "Meds"
Range("S1").Select
ActiveCell.FormulaR1C1 = "OtherD"
Range("T1").Select
ActiveCell.FormulaR1C1 = "Smoker"
Range("U1").Select
ActiveCell.FormulaR1C1 = "SmokeHx"
Range("V1").Select
ActiveCell.FormulaR1C1 = "SelfHealth"
Range("W1").Select
ActiveCell.FormulaR1C1 = "Cyclefactor"
Range("X1").Select
ActiveCell.FormulaR1C1 = "PWt"
Range("Y1").Select
ActiveCell.FormulaR1C1 = "PVO2Max"
Range("Z1").Select
ActiveCell.FormulaR1C1 = "Health1"
Range("AA1").Select
```

```
    ActiveCell.FormulaR1C1 = "Health2"
    Range("AB1").Select
    ActiveCell.FormulaR1C1 = "Health3"

    Subno = InputBox("Input subject#:", "Subject#", "1")
    age = InputBox("Age (in years):", "Age", "20")
    sx = InputBox("Sex (M=male, F=female):", "Sex", "F")
    If sx = "F" Then sex = 1
    If sx = "M" Then sex = 0
    If sx = "f" Then sex = 1
    If sx = "m" Then sex = 0
    Ht_ft = InputBox("Height (enter feet here, inches next screen):", "Height", "5")
    Ht_in = InputBox("Height (inches):", "Height", "8")
    Ht_cm = (Ht_ft * 12 + Ht_in) * 2.54
    Ht_m = Ht_cm / 100
    Wt_lb = InputBox("Weight (lbs):", "Weight", "150")
    Wt_kg = Wt_lb / 2.2
    BMI = Wt_kg / (Ht_m ^ 2)
    Self_Fit = InputBox("How would you rate yourself in terms of fitness: 0=sedentary
2=exercise 4-5x per wk          1=exercise 1-3x per wk     3=exercise 6-7x per wk",
"Fitness", "")
    Sweat_no = InputBox("How many times a week do you think you sweat as a result of
exercise:", "#Sweat/wk", "0")
    ExType = InputBox("What type of exercise do you participate in?", "Type of
Exercise", "")
    ExDur = InputBox("How long do you usually exercise (in minutes)?", "Exercise
Duration", "")
    ExFreq = InputBox("How often do you exercise per week?", "Exercise Frequency", "")
    RespD = InputBox("Have you been diagnosed with any chronic respiratory disease,
such as asthma, emphysema, or chronic bronchitis?", "Respiratory Disease", "No")
    CardD = InputBox("Have you ever been diagnosed with any cardiovascular disease?",
"Cardiovascular Disease", "No")
    HRmed = InputBox("Do you take any beta-blockers (heart medication)?", "HR
Medication", "No")
    OtherD = InputBox("Do you have any other disease?", "Other Disease", "")
    Meds = InputBox("Please list all medication you are currently taking.", "Medication",
"")
    Smk = InputBox("Are you currently a smoker?", "Current Smoking Status", "No")
    SmkHx = InputBox("Were you ever a smoker? If so, for how many years and how
many packs/day?", "Smoking History", "")
    Self_Health = InputBox("How would you rate your overall health:          0=above
average   2=slightly diseased     1=average          3=severely diseased", "Overall
Health Status", "")
```

```
    If Self_Health = 1 Then Health1 = 1
    If Self_Health = 1 Then Health2 = 0
    If Self_Health = 1 Then Health3 = 0
    If Self_Health = 2 Then Health1 = 0
    If Self_Health = 2 Then Health2 = 1
    If Self_Health = 2 Then Health3 = 0
    If Self_Health = 3 Then Health1 = 0
    If Self_Health = 3 Then Health2 = 0
    If Self_Health = 3 Then Health3 = 1

    If sex = 0 Then cyclefactor = 50.72 - 0.372 * age
    If sex = 1 Then cyclefactor = 22.78 - 0.17 * age
    If sex = 0 Then PWt = 0.79 * Ht_cm - 60.7
    If sex = 1 Then PWt = 0.65 * Ht_cm - 42.8
    If sex = 0 And PWt = Wt_kg Then PVO2Max = (Wt_kg * cyclefactor) / 1000
    If sex = 1 And PWt = Wt_kg Then PVO2Max = ((Wt_kg + 43) * cyclefactor) / 1000
    If sex = 0 And Wt_kg < PWt Then PVO2Max = (((PWt + Wt_kg) / 2) * cyclefactor) /
1000
    If sex = 1 And Wt_kg < PWt Then PVO2Max = (((PWt + Wt_kg + 86) / 2) *
cyclefactor) / 1000
    If sex = 0 And Wt_kg > PWt Then PVO2Max = ((PWt * cyclefactor) + 6 * (Wt_kg -
PWt)) / 1000
    If sex = 1 And Wt_kg > PWt Then PVO2Max = ((PWt + 43) * cyclefactor + 6 *
(Wt_kg - PWt)) / 1000

    If HRmed = "No" Then Hmed = 0
    If HRmed = "no" Then Hmed = 0
    If HRmed = "N" Then Hmed = 0
    If HRmed = "n" Then Hmed = 0
    If HRmed = "Yes" Then Hmed = 1
    If HRmed = "yes" Then Hmed = 1
    If HRmed = "Y" Then Hmed = 1
    If HRmed = "y" Then Hmed = 1

    Range("A2").Select
    ActiveCell.FormulaR1C1 = Subno
    Range("B2").Select
    ActiveCell.FormulaR1C1 = age
    Range("C2").Select
    ActiveCell.FormulaR1C1 = sex
    Range("D2").Select
    ActiveCell.FormulaR1C1 = Ht_ft
    Range("E2").Select
```

```
ActiveCell.FormulaR1C1 = Ht_in
Range("F2").Select
ActiveCell.FormulaR1C1 = Ht_cm
Range("G2").Select
ActiveCell.FormulaR1C1 = Wt_lb
Range("H2").Select
ActiveCell.FormulaR1C1 = Wt_kg
Range("I2").Select
ActiveCell.FormulaR1C1 = BMI
Range("J2").Select
ActiveCell.FormulaR1C1 = Self_Fit
Range("K2").Select
ActiveCell.FormulaR1C1 = Sweat_no
Range("L2").Select
ActiveCell.FormulaR1C1 = ExType
Range("M2").Select
ActiveCell.FormulaR1C1 = ExDur
Range("N2").Select
ActiveCell.FormulaR1C1 = ExFreq
Range("O2").Select
ActiveCell.FormulaR1C1 = RespD
Range("P2").Select
ActiveCell.FormulaR1C1 = CardD
Range("Q2").Select
ActiveCell.FormulaR1C1 = Hmed
Range("R2").Select
ActiveCell.FormulaR1C1 = Med
Range("S2").Select
ActiveCell.FormulaR1C1 = OtherD
Range("T2").Select
ActiveCell.FormulaR1C1 = Smk
Range("U2").Select
ActiveCell.FormulaR1C1 = SmkHx
Range("V2").Select
ActiveCell.FormulaR1C1 = Self_Health

Range("W2").Select
ActiveCell.FormulaR1C1 = cyclefactor
Range("X2").Select
ActiveCell.FormulaR1C1 = PWt
Range("Y2").Select
ActiveCell.FormulaR1C1 = PVO2Max
Range("Z2").Select
```

```
    ActiveCell.FormulaR1C1 = Health1
    Range("AA2").Select
    ActiveCell.FormulaR1C1 = Health2
    Range("AB2").Select
    ActiveCell.FormulaR1C1 = Health3

    Columns("W:Y").Select
    Selection.NumberFormat = "0.00"
    Columns("H:I").Select
    Selection.NumberFormat = "0.00"

    Rows("1:2").Select
    Selection.EntireColumn.AutoFit

'save file in subject's folder
   ActiveWorkbook.SaveAs  Filename:="D:\CHERYL\LABSUBJ\SUBJ" & Subno &
"\S" & Subno & "HX1.XLW", FileFormat:=xlNormal, _
      CreateBackup:=False

End Sub
```

APPENDIX H: INFORMATION COLLECTED VIA THE SUBJECT HISTORY MACRO

Subject#	Age	Sex	Ht_ft	Ht_in	Ht_cm	Wt_lb	Wt_kg
1	35	0	6	4	193.04	225	102.27
2	25	1	5	1	154.94	119	52.27
5	26	1	5	8	172.72	150	68.18
6	33	0	6	1	185.42	185	84.09
7	27	0	6	2	187.96	205	93.18
8	26	0	5	9	175.26	198	90.00
9	32	0	6	2	187.96	160	72.73
10	26	1	5	7.5	171.45	145	65.91
11	27	0	6	2	187.96	163	74.09
12	28	1	5	6	167.64	116	52.73
13	40	1	5	5	165.1	130	59.09
14	22	1	5	6	167.64	120	54.55
15	27	1	5	8	172.72	135	61.36
16	31	0	6	2	187.96	160	72.73
17	27	0	5	11	180.34	285	129.55
18	27	0	5	8	172.72	152	69.09
19	34	0	6	1	185.42	240	109.09
20	23	1	5	5	165.1	120	54.55
21	26	1	5	4	162.56	123	55.91
22	31	0	5	11	180.34	155	70.45
23	60	1	5	5	165.1	142	64.55
24	23	1	5	7	170.18	155	70.45
25	24	1	5	7	170.18	145	65.91
26	29	0	5	5	165.1	165	75.00
27	35	0	5	11	180.34	160	72.73
28	43	0	5	10	177.8	170	77.27
29	30	1	5	5	165.1	160	72.73
30	30	0	6	3	190.5	180	81.82
31	43	1	5	5	165.1	190	86.36
32	20	1	5	1	154.94	115	52.27

Subject#	BMI	Self_Fit	Sweat_No.	ExType	ExDur	ExFreq
1	27.45	3	7	stairmaster,weights, biking	60	1
2	21.77	1	5	running, aerobics, biking	60	5
5	22.86	3	6	aerobics,wts.,stairmaster	90	6
6	24.46	0	0	0	0	0
7	26.38	1	2	basketball,bicycle riding	45	3
8	29.30	2	5	wts.,tennis,basketball,aikido	120	5
9	20.59	1	3	cycling	30	4
10	22.42	2	5	soccer, running	45	5
11	20.97	1	2	basketball, running, softball	30	3
12	18.76	1	2	yoga; treadmill;stairmaster	90	2
13	21.68	2	2	bicycling, hiking, climbing	120	2
14	19.41	1	3	aerobics, walking	45	3
15	20.57	2	2	run,bike,walk,hike,tai chi	35	4
16	20.59	1	3	running,swimming,	60	3
17	39.83	0	0	0	0	0
18	23.16	3	6	running,basketball,weightlift	55	7
19	31.73	2	4	walking/stair	25	4
20	20.01	1	3	running, stairs, biking	30	3
21	21.16	2	4	speed walking primarily	30	5
22	21.66	2	4	weight lifting, running	45	4
23	23.68	1	0	walking/biking	30	3
24	24.33	1	3	raquetball	35	3
25	22.76	1	1	biking, walking	30	1.5
26	27.51	1	0	bike	15	4
27	22.36	2	4	swimming,hiking,yoga,	45	5
28	24.44	3	6	swim,bike,walk,sea kayak	75	7
29	26.68	0	0	walking,rollerblading,	30-45	1
30	22.55	2	5	stairmaster,bicycle,wts,row.	100	5
31	31.68	1	3	treadmill,weights	45	3
32	21.77	2	4	aerobics	75	4

Subject#	RespDisease	CardioDisease	HRmed	OtherDisease
1	No	No	0	No
2	No	No	0	No
5	No	No	0	No
6	mild asthma	No	0	No
7	No	No	0	No
8	No	No	0	No
9	No	No	0	No
10	No	No	0	No
11	No	No	0	No
12	No	No	0	No
13	No	No	0	No
14	No	No	0	No
15	No	No	0	No
16	No	No	0	No
17	No	No	0	No
18	No	No	0	No
19	No	No	0	No
20	No	No	0	No
21	No	No	0	No
22	No	No	0	No
23	No	No	0	No
24	No	No	0	No
25	No	No	0	hypothyroidism
26	No	No	0	No
27	No	No	0	No
28	No	No	0	No
29	No	No	0	No
30	No	No	0	No
31	No	No	0	hypothyroidism
32	No	No	0	No

Subject#	Smoker	SmokeHx	SelfHealth	Race
1	No	No	1	Caucasian
2	No	No	1	Caucasian
5	No	No	0	Caucasian
6	No	No	1	Caucasian
7	No	No	1	Caucasian
8	No	No	1	Asian
9	No	No	1	Caucasian
10	No	No	1	Caucasian
11	No	No	1	Caucasian
12	No	No	0	Caucasian
13	No	No	0	Caucasian
14	No	No	1	Caucasian
15	No	No	1	Caucasian
16	No	No	0	Caucasian
17	No	No	1	Caucasian
18	No	No	0	Caucasian
19	No	No	1	Caucasian
20	No	No	0	Caucasian
21	No	No	0	Caucasian
22	No	No	0	Caucasian
23	No	No	0	Caucasian
24	No	No	1	Caucasian
25	No	No	1	Caucasian
26	No	No	0	Caucasian
27	No	No	0	Caucasian
28	No	No	0	Caucasian
29	No	No	1	Caucasian
30	Yes	12 years, 1 pack per day	0	Caucasian
31	No	No	0	Caucasian
32	No	4 years, 2 packs/month	0	Filipino

APPENDIX I: RIPCAL.BAT CODE

```
deriv %1 1 5 10
deriv %1 2 6 10
```

APPENDIX J: ISO EXCEL MACRO

```
'
' SubjIso Macro
' Macro recorded 6/4/96 by Cheryl Hart
'
' Keyboard Shortcut: Ctrl+i
'
Sub SubjIso()

    Subno = InputBox("Input subject#:", "Subject#", "1")
    sessno = InputBox("Input session#:", "Session#", "1")
    calno = InputBox("Input HR-Vi calibration#:", "Calibration#", "1")
    ISO1 = "D:\CODAS\S" & Subno & "N" & sessno & "C" & calno & "I1.xlw"
    Workbooks.OpenText Filename:=ISO1, Origin:=xlWindows _
      , StartRow:=1, DataType:=xlDelimited, TextQualifier:= _
      xlDoubleQuote, ConsecutiveDelimiter:=False, Tab:=True, Semicolon _
       :=False, Comma:=True, Space:=False, Other:=False, FieldInfo _
      :=Array(Array(1, 1), Array(2, 1), Array(3, 1))
    ActiveWorkbook.SaveAs Filename:="TEMPISO1.XLW", FileFormat:=xlText, _
      CreateBackup:=False
    Cells.Select
    Selection.NumberFormat = ".000"

    ISO2 = "D:\CODAS\S" & Subno & "N" & sessno & "C" & calno & "I2.xlw"
    Workbooks.OpenText Filename:=ISO2, Origin:=xlWindows _
      , StartRow:=1, DataType:=xlDelimited, TextQualifier:= _
      xlDoubleQuote, ConsecutiveDelimiter:=False, Tab:=True, Semicolon _
       :=False, Comma:=True, Space:=False, Other:=False, FieldInfo _
      :=Array(Array(1, 1), Array(2, 1), Array(3, 1))
    ActiveWorkbook.SaveAs Filename:="TEMPISO2.XLW", FileFormat:=xlText, _
      CreateBackup:=False
    Cells.Select
    Selection.NumberFormat = ".000"

    ISO3 = "D:\CODAS\S" & Subno & "N" & sessno & "C" & calno & "I3.xlw"
    Workbooks.OpenText Filename:=ISO3, Origin:=xlWindows _
      , StartRow:=1, DataType:=xlDelimited, TextQualifier:= _
      xlDoubleQuote, ConsecutiveDelimiter:=False, Tab:=True, Semicolon _
       :=False, Comma:=True, Space:=False, Other:=False, FieldInfo _
```

```
    :=Array(Array(1, 1), Array(2, 1), Array(3, 1))
  ActiveWorkbook.SaveAs Filename:="TEMPISO3.XLW", FileFormat:=xlText, _
    CreateBackup:=False
  Cells.Select
  Selection.NumberFormat = ".000"

' move all sheets to one workbook
  Windows("TEMPISO2.XLW").Activate
  ActiveWindow.WindowState = xlNormal
  Sheets("Sheet1").Select
  Sheets("Sheet1").Move After:=Workbooks("TEMPISO1.xlw").Sheets(1)
  Windows("TEMPISO3.xlw").Activate
  Sheets("Sheet1").Select
  Sheets("Sheet1").Move After:=Workbooks("TEMPISO1.xlw").Sheets(2)
  Sheets("Sheet1").Select
  Sheets("Sheet1").Name = "iso1"
  Sheets("Sheet1 (2)").Select
  Sheets("Sheet1 (2)").Name = "iso2"
  Sheets("Sheet1 (3)").Select
  Sheets("Sheet1 (3)").Name = "iso3"

' add titles
  Sheets("iso1").Select
  Range("A3").Select
  ActiveCell.FormulaR1C1 = "HR"
  Range("B3").Select
  ActiveCell.FormulaR1C1 = "PNEUMO"
  Range("C3").Select
  ActiveCell.FormulaR1C1 = "ABD"
  Range("D3").Select
  ActiveCell.FormulaR1C1 = "RC"
  Range("E3").Select
  ActiveCell.FormulaR1C1 = "ABDder"
  Range("F3").Select
  ActiveCell.FormulaR1C1 = "RCder"
  Range("A3:F3").Select
  Selection.Copy
  Sheets("iso2").Select
  Range("A3").Select
  ActiveSheet.Paste
  Sheets("iso3").Select
  Range("A3").Select
  ActiveSheet.Paste
```

```
Sheets("iso1").Select
Range("A3").Select
Selection.End(xlDown).Select
Selection.ClearContents

' create graphs
Rows("1:2").Select
Selection.Delete Shift:=xlUp
Columns("C:D").Select
Charts.Add
ActiveChart.ChartWizard Source:=Sheets("iso1").Range("C:D"), _
  Gallery:=xlXYScatter, Format:=1, PlotBy:=xlColumns, _
  CategoryLabels:=1, SeriesLabels:=1, HasLegend:=1
ActiveChart.PlotArea.Select
With Selection.Border
  .ColorIndex = 16
  .Weight = xlThin
  .LineStyle = xlContinuous
End With
Selection.Interior.ColorIndex = xlNone
ActiveChart.SeriesCollection(1).Select
ActiveChart.SeriesCollection(1).Trendlines.Add(Type:=xlLinear, Forward _
  :=0, Backward:=0, DisplayEquation:=True, DisplayRSquared:= _
  True).Select
ActiveChart.SeriesCollection(1).Trendlines(1).DataLabel.Select
Selection.Left = 453
Selection.Top = 85
ActiveChart.PlotArea.Select
ActiveChart.SeriesCollection(1).Trendlines(1).DataLabel.Select
Selection.Left = 450
Selection.Top = 72
Sheets("Chart1").Select
Sheets("Chart1").Name = "iso1.cht"

Sheets("iso2").Select
Range("A3").Select
Selection.End(xlDown).Select
Selection.ClearContents
Rows("1:2").Select
Selection.Delete Shift:=xlUp
Columns("C:D").Select
Charts.Add
ActiveChart.ChartWizard Source:=Sheets("iso2").Range("C:D"), _
```

```
    Gallery:=xlXYScatter, Format:=1, PlotBy:=xlColumns, _
    CategoryLabels:=1, SeriesLabels:=1, HasLegend:=1
  ActiveChart.PlotArea.Select
  ActiveChart.SeriesCollection(1).Select
  ActiveChart.SeriesCollection(1).Trendlines.Add(Type:=xlLinear, Forward _
     :=0, Backward:=0, DisplayEquation:=True, DisplayRSquared:= _
    True).Select
  ActiveChart.SeriesCollection(1).Trendlines(1).DataLabel.Select
  Selection.Left = 457
  Selection.Top = 73
  ActiveChart.PlotArea.Select
  With Selection.Border
    .ColorIndex = 16
    .Weight = xlThin
    .LineStyle = xlContinuous
  End With
  Selection.Interior.ColorIndex = xlNone
  Sheets("Chart2").Select
  Sheets("Chart2").Name = "iso2.cht"

  Sheets("iso3").Select
  Range("A3").Select
  Selection.End(xlDown).Select
  Selection.ClearContents
  Rows("1:2").Select
  Selection.Delete Shift:=xlUp
  Columns("C:D").Select
  Charts.Add
  ActiveChart.ChartWizard Source:=Sheets("iso3").Range("C:D"), _
    Gallery:=xlXYScatter, Format:=1, PlotBy:=xlColumns, _
    CategoryLabels:=1, SeriesLabels:=1, HasLegend:=1
  ActiveChart.SeriesCollection(1).Select
  ActiveChart.SeriesCollection(1).Trendlines.Add(Type:=xlLinear, Forward _
     :=0, Backward:=0, DisplayEquation:=True, DisplayRSquared:= _
    True).Select
  ActiveChart.PlotArea.Select
  With Selection.Border
    .ColorIndex = 16
    .Weight = xlThin
    .LineStyle = xlContinuous
  End With
  Selection.Interior.ColorIndex = xlNone
  ActiveChart.SeriesCollection(1).Trendlines(1).DataLabel.Select
```

```
    Selection.Left = 448
    Selection.Top = 78
    Sheets("Chart3").Select
    Sheets("Chart3").Name = "iso3.cht"

'   create composite graph
    Sheets("iso1").Select
    Columns("C:D").Select
    Charts.Add
    ActiveChart.ChartWizard Source:=Sheets("iso1").Columns("C:D"), Gallery _
        :=xlXYScatter, Format:=1, PlotBy:=xlColumns, CategoryLabels _
        :=1, SeriesLabels:=1, HasLegend:=1, Title:= _
        "Isovolume curves: Subject #" & Subno & ", Session #" & sessno & ", Calibration
#" & calno & ".", CategoryTitle:="Abd. signal", ValueTitle:= _
        "RC signal", ExtraTitle:=""
    ActiveChart.PlotArea.Select
    With Selection.Border
        .ColorIndex = 16
        .Weight = xlThin
        .LineStyle = xlContinuous
    End With
    Selection.Interior.ColorIndex = xlNone
    ActiveChart.SeriesCollection(1).Select
    ActiveChart.SeriesCollection(1).Trendlines.Add(Type:=xlLinear, Forward _
        :=0, Backward:=0, DisplayEquation:=True, DisplayRSquared:= _
        True).Select
    ActiveChart.SeriesCollection(1).Trendlines(1).DataLabel.Select
    Selection.Left = 49
    Selection.Top = 382

    ActiveChart.SeriesCollection.Add Source:="iso2!$C:$D", Rowcol:= _
        xlColumns, SeriesLabels:=True, CategoryLabels:=True, Replace _
        :=False
    ActiveChart.SeriesCollection(2).Select
    ActiveChart.SeriesCollection(2).Trendlines.Add(Type:=xlLinear, Forward _
        :=0, Backward:=0, DisplayEquation:=True, DisplayRSquared:= _
        True).Select
    ActiveChart.SeriesCollection(2).Trendlines(1).DataLabel.Select
    ActiveChart.SeriesCollection(2).Trendlines(1).DataLabel.Select
    Selection.Left = 570
    Selection.Top = 163
    ActiveChart.Legend.Select
    Selection.Delete
```

```
    ActiveChart.SeriesCollection(1).Trendlines(1).DataLabel.Select
    Selection.Left = 567
    ActiveChart.SeriesCollection(2).Trendlines(1).DataLabel.Select
    Selection.Top = 224

    Sheets("iso3").Select
    Columns("C:D").Select
    Sheets("Chart4").Select
    ActiveChart.SeriesCollection.Add Source:="iso3!$C:$D", Rowcol:= _
      xlColumns, SeriesLabels:=True, CategoryLabels:=True, Replace _
      :=False
    ActiveChart.SeriesCollection(3).Select
    ActiveChart.SeriesCollection(3).Trendlines.Add(Type:=xlLinear, Forward _
       :=0, Backward:=0, DisplayEquation:=True, DisplayRSquared:= _
      True).Select
    ActiveChart.SeriesCollection(3).Trendlines(1).DataLabel.Select
    Selection.Left = 573
    Selection.Top = 59
    ActiveChart.SeriesCollection(2).Trendlines(1).DataLabel.Select
    Selection.Left = 573
    Selection.Top = 217
    ActiveChart.SeriesCollection(1).Trendlines(1).DataLabel.Select
    Selection.Left = 573
    Selection.Top = 387

    Sheets("Chart4").Select
    Sheets("Chart4").Name = "isocurves.cht"

'save file in subject's folder
   ActiveWorkbook.SaveAs Filename:="D:\CHERYL\LABSUBJ\SUBJ" & Subno &
"\S" & Subno & "N" & sessno & "IC" & calno & ".XLW", FileFormat:=xlNormal, _
      CreateBackup:=False

End Sub
```

APPENDIX K: RIPCAL EXCEL MACRO

```
'
' RIPcal Macro
' Macro recorded 6/26/96 by Michael Yost
'
'
Sub ripcal()

    Subno = InputBox("Input subject#:", "Subject#", "1")
    sessno = InputBox("Input session#:", "Session#", "1")
    calno = InputBox("Input HR-Vi calibration#:", "Calibration#", "1")

    Slope = InputBox("Enter the slope from the isocurve of choice:", "Isocurve Slope")
    Intercept = InputBox("Enter the intercept from this curve:", "Isocurve Intercept")

    ABcoef = -Slope / Intercept
    RCcoef = 1 / Intercept

    CalDat = "D:\CODAS\S" & Subno & "N" & sessno & "C" & calno & "B.xlw"
    Workbooks.OpenText Filename:=CalDat, Origin:=xlWindows _
      , StartRow:=1, DataType:=xlDelimited, TextQualifier:= _
      xlDoubleQuote, ConsecutiveDelimiter:=False, Tab:=True, Semicolon _
       :=False, Comma:=True, Space:=False, Other:=False, FieldInfo _
      :=Array(Array(1, 1), Array(2, 1), Array(3, 1))
    ActiveWorkbook.SaveAs Filename:="TEMPCAL.XLW", FileFormat:=xlText, _
      CreateBackup:=False
    Cells.Select
    Selection.NumberFormat = ".000"

'   add titles
    Range("A3").Select
    ActiveCell.FormulaR1C1 = "HR"
    Range("B3").Select
    ActiveCell.FormulaR1C1 = "PNEUMO"
    Range("C3").Select
    ActiveCell.FormulaR1C1 = "AB"
    Range("D3").Select
    ActiveCell.FormulaR1C1 = "RC"
    Range("E3").Select
```

```
    ActiveCell.FormulaR1C1 = "ABDder"
    Range("F3").Select
    ActiveCell.FormulaR1C1 = "RCder"
    Range("G3").Select
    ActiveCell.FormulaR1C1 = "adjABDder"
    Range("H3").Select
    ActiveCell.FormulaR1C1 = "adjRCder"
    Range("I3").Select
    ActiveCell.FormulaR1C1 = "SUMder"
    Range("J3").Select
    ActiveCell.FormulaR1C1 = "Vi"

    Rows("1:2").Select
    Selection.Delete Shift:=xlUp

    Range("L1").Select
    ActiveCell.FormulaR1C1 = "Slope"
    Range("M1").Select
    ActiveCell.FormulaR1C1 = "Intercept"
    Range("N1").Select
    ActiveCell.FormulaR1C1 = "ABcoef"
    Range("O1").Select
    ActiveCell.FormulaR1C1 = "RCcoef"

    Range("L2").Select
    ActiveCell.FormulaR1C1 = Slope
    Range("M2").Select
    ActiveCell.FormulaR1C1 = Intercept
    Range("N2").Select
    ActiveCell.FormulaR1C1 = ABcoef
    Range("O2").Select
    ActiveCell.FormulaR1C1 = RCcoef

'   copy and paste formulas for adjusted ab and rc values
    Range("G2") = "=E2*$N$2"
    Range("H2") = "=F2*$o$2"
    Range("I2") = "=G2+H2"
    Range("J2") = "=B2*30.091-0.6668"
    Range("G2:J2").Select
    Range("G2:J" & Range("E2").End(xlDown).Row).FillDown

'   create graph
    Columns("I:J").Select
```

```
    Charts.Add
    ActiveChart.ChartWizard Source:=Sheets("Sheet1").Columns("I:J"), _
      Gallery:=xlXYScatter, Format:=1, PlotBy:=xlColumns, _
      CategoryLabels:=1, SeriesLabels:=1, HasLegend:=1, Title:= _
      "RIP Calibration: Subject #" & Subno & ", Session #" & sessno & ", Calibration #"
& calno & ".", CategoryTitle:= _
      "Sum of AbDeriv. and RcDeriv.", ValueTitle:= _
      "Minute ventilation (LPM)", ExtraTitle:=""
    ActiveChart.Legend.Select
    Selection.Delete
    ActiveChart.SeriesCollection(1).Select
    ActiveChart.SeriesCollection(1).Trendlines.Add(Type:=xlLinear, Forward _
      :=0, Backward:=0, DisplayEquation:=True, DisplayRSquared:= _
      True).Select
    ActiveChart.SeriesCollection(1).Trendlines(1).DataLabel.Select
    Selection.Left = 560
    Selection.Top = 8
    ActiveChart.PlotArea.Select
    With Selection.Border
      .ColorIndex = 16
      .Weight = xlThin
      .LineStyle = xlContinuous
    End With
    Selection.Interior.ColorIndex = xlNone
    Sheets("Chart1").Select
    Sheets("Chart1").Name = "ripcal.gr"
    Sheets("Sheet1").Select
    Sheets("Sheet1").Name = "ripcal.dat"

'save file in subject's folder
    ActiveWorkbook.SaveAs Filename:="D:\CHERYL\LABSUBJ\SUBJ" & Subno &
"\S" & Subno & "N" & sessno & "RC" & calno & ".XLW", FileFormat:=xlNormal, _
      CreateBackup:=False

End Sub
```

APPENDIX L: HRATE.BAT CODE

```
peak %1 3 3 p 2
arith %1 5 #3*0+1
integral %1 5 6 + 1 3 + H
arith %1 7 #6+0.01
arith %1 8 60/#7
rectify %1 4 9 A
average %1 9 10 1000
average %1 8 11 1000
```

APPENDIX M: HRCAL EXCEL MACRO

```
'
' HRcal Macro
' Macro recorded 7/8/96 by Cheryl Hart
'
'
Sub HRcal()

    Subno = InputBox("Input subject#:", "Subject#", "1")
    sessno = InputBox("Input session#:", "Session#", "1")
    calno = InputBox("Input HR-logVi calibration#:", "Calibration#", "1")

    HC = "D:\CODAS\XS" & Subno & "N" & sessno & "HC" & calno & ".xlw"
    Workbooks.OpenText Filename:=HC, Origin:=xlWindows _
      , StartRow:=1, DataType:=xlDelimited, TextQualifier:= _
      xlDoubleQuote, ConsecutiveDelimiter:=False, Tab:=True, Semicolon _
       :=False, Comma:=True, Space:=False, Other:=False, FieldInfo _
      :=Array(Array(1, 1), Array(2, 1), Array(3, 1), Array(4, 1), Array(5, 1), _
      Array(6, 1), Array(7, 1), Array(8, 1), Array(9, 1), Array(10, 1), Array(11, 1), _
      Array(12, 1), Array(13, 1), Array(14, 1), Array(15, 1), Array(16, 1), Array(17 _
      , 1))
    Sheets("Sheet1").Select
    Cells.Select
    Selection.NumberFormat = "0.00"
    Sheets("Sheet1").Name = "Raw.dat"
    Sheets.Add
    Sheets("Sheet2").Select
    Sheets("Sheet2").Name = "HRcal.dat"
    Sheets("Raw.dat").Select
    Range("C2:C9").Select
    Selection.Copy
    Sheets("HRcal.dat").Select
    Range("A2").Select
    ActiveSheet.Paste
    Cells.Select
    Application.CutCopyMode = False
    Selection.NumberFormat = "0.00"
    Range("A3").Select
    Selection.Cut Destination:=Range("C2")
```

```
    Range("A5").Select
    Selection.Cut Destination:=Range("C3")
    Range("A7").Select
    Selection.Cut Destination:=Range("C4")
    Range("A9").Select
    Selection.Cut Destination:=Range("C5")
    Range("A4").Select
    Selection.Cut Destination:=Range("A3")
    Range("A6").Select
    Selection.Cut Destination:=Range("A4")
    Range("A8").Select
    Selection.Cut Destination:=Range("A5")
    Range("A1").Select
    ActiveCell.FormulaR1C1 = "Pne"
    Range("B1").Select
    ActiveCell.FormulaR1C1 = "LPM"
    Range("C1").Select
    ActiveCell.FormulaR1C1 = "HR"
    Range("D1").Select
    ActiveCell.FormulaR1C1 = "logVI"
    Range("B2").Select
    ActiveCell.FormulaR1C1 = "=RC[-1]*30.091-0.6668"
    Range("B2").Select
    Selection.AutoFill Destination:=Range("B2:B4"), Type:=xlFillDefault
    Range("D2").Select
    ActiveCell.FormulaR1C1 = "=log(RC[-2])"
    Range("D2").Select
    Selection.AutoFill Destination:=Range("D2:D4"), Type:=xlFillDefault
    Columns("C:D").Select
    Charts.Add
    ActiveChart.ChartWizard Source:=Sheets("HRcal.dat").Columns("C:D"), _
      Gallery:=xlXYScatter, Format:=1, PlotBy:=xlColumns, _
      CategoryLabels:=1, SeriesLabels:=1, HasLegend:=1, Title:= _
      "HR-Vi Calibration: Subject #" & Subno & ", Session #" & sessno & ", Calibration
#" & calno & ".", CategoryTitle:="", _
      ValueTitle:="", ExtraTitle:=""
    ActiveChart.PlotArea.Select
    With Selection.Border
      .ColorIndex = 16
      .Weight = xlThin
      .LineStyle = xlContinuous
    End With
    Selection.Interior.ColorIndex = xlNone
```

```
    ActiveChart.ChartArea.Select
    With ActiveChart
       .HasTitle = True
       .Axes(xlCategory, xlPrimary).HasTitle = False
       .Axes(xlValue, xlPrimary).HasTitle = True
       .Axes(xlValue, xlPrimary).AxisTitle.Select
    End With
    Selection.Characters.Text = "LPM"
    With ActiveChart
       .HasTitle = True
       .Axes(xlCategory, xlPrimary).HasTitle = True
       .Axes(xlValue, xlPrimary).HasTitle = True
       .Axes(xlCategory, xlPrimary).AxisTitle.Select
    End With
    Selection.Characters.Text = "Heart Rate (BPM)"
    Sheets("Chart1").Select
    Sheets("Chart1").Name = "HRcal.cht"
    ActiveChart.SeriesCollection(1).Select
    ActiveChart.SeriesCollection(1).Trendlines.Add(Type:=xlLinear, _
       Forward:=0, Backward:=0, DisplayEquation:=True, _
       DisplayRSquared:=True).Select
    ActiveChart.SeriesCollection(1).Trendlines(1).DataLabel.Select
    ActiveChart.Legend.Select
    Selection.Delete
    ActiveChart.SeriesCollection(1).Trendlines(1).DataLabel.Select
    Selection.Left = 580
    Selection.Top = 206
    With ActiveWorkbook
       .Title = ""
       .Subject = ""
       .Author = "Cheryl Hart"
       .Keywords = ""
       .Comments = ""
    End With

    'save file in subject's folder
    ActiveWorkbook.SaveAs Filename:="D:\CHERYL\LABSUBJ\SUBJ" & Subno &
"\S" & Subno & "N" & sessno & "HC" & calno & ".XLW", FileFormat:=xlNormal, _
       CreateBackup:=False
End Sub
```

APPENDIX N: PREDHRLOGVI EXCEL MACRO

```
'
' PredHRlogVi Macro
' Macro recorded 7/17/96 by Cheryl Hart
'
'
Sub PredHRlogVi()
    Subno = InputBox("Input subject#:", "Subject#", "1")
    sessno = InputBox("Input session#:", "Session#", "1")
    calno = InputBox("Input HR-Vi calibration#:", "Calibration#", "1")
    calM = InputBox("Input the calcurve slope", "Slope", "")
    calB = InputBox("Input the calcurve intercept", "Intercept", "")

    SHX = "S" & Subno & "Hx1.xlw"
    HC = "D:\CHERYL\LABSUBJ\SUBJ" & Subno & "\S" & Subno & "N" & sessno &
"HC" & calno & ".xlw"
    Workbooks.Open Filename:=HC

    Sheets.Add
    Sheets("Sheet1").Select
    Sheets("Sheet1").Name = "PredHRlogVi"
    Sheets("HRcal.dat").Select
    Columns("C:D").Select
    Selection.Copy
    Sheets("PredHRlogVi").Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _
      SkipBlanks:=False, Transpose:=False
    Cells.Select
    Application.CutCopyMode = False
    With Selection
      .HorizontalAlignment = xlCenter
      .VerticalAlignment = xlBottom
      .WrapText = False
      .Orientation = xlHorizontal
    End With

    Range("A1").Select
    ActiveCell.FormulaR1C1 = "HR"
    Range("B1").Select
```

```
    ActiveCell.FormulaR1C1 = "Act.LogVi"

    Range("D1").Select
    ActiveCell.FormulaR1C1 = "HR"
    Range("E1").Select
    ActiveCell.FormulaR1C1 = "Act.Vi"
    Range("F1").Select
    ActiveCell.FormulaR1C1 = "Cal.Vi"
    Range("G1").Select
    ActiveCell.FormulaR1C1 = "Pred.Vi"

    Range("I1").Select
    ActiveCell.FormulaR1C1 = "CalM"
    Range("J1").Select
    ActiveCell.FormulaR1C1 = "CalB"
    Range("I2").Select
    ActiveCell.FormulaR1C1 = calM
    Range("J2").Select
    ActiveCell.FormulaR1C1 = calB

    Range("D2").Select
    ActiveCell.FormulaR1C1 = "=(RC[-3])"
    Range("E2").Select
    ActiveCell.FormulaR1C1 = "=10^(RC[-3])"
    Range("F2").Select
    ActiveCell.FormulaR1C1 = "=10^(RC[3]*(RC[-2])+RC[4])"
    Range("F3").Select
    ActiveCell.FormulaR1C1 = "=10^(R[-1]C[3]*(RC[-2])+R[-1]C[4])"
    Range("F4").Select
    ActiveCell.FormulaR1C1 = "=10^(R[-2]C[3]*(RC[-2])+R[-2]C[4])"

    ChDrive "D"
    ChDir "D:\CHERYL\LABSUBJ\SUBJ" & Subno & ""
    Range("G2").Select
    ActiveCell.FormulaR1C1 = "=10^(0.0041*[" & SHX & "]Sheet1!R2C8+0.791)"
    Range("G3").Select
    ActiveCell.FormulaR1C1 = "=10^(0.00911*((RC[-6])-(R[-1]C[-6]))+0.0041*[" &
SHX & "]Sheet1!R2C8+0.791)"
    Range("G4").Select
    ActiveCell.FormulaR1C1 = "=10^(0.00911*((RC[-6])-(R[-2]C[-6]))+0.0041*[" &
SHX & "]Sheet1!R2C8+0.791)"

    Range("D2:F2").Select
```

```
    Selection.AutoFill Destination:=Range("D2:F4"), Type:=xlFillDefault
    Range("D1:G4").Select
    Application.CutCopyMode = False
    Charts.Add
    ActiveChart.ChartWizard Source:=Sheets("PredHRlogVi").Range("D1:G4"), _
      Gallery:=xlXYScatter, Format:=1, PlotBy:=xlColumns, _
      CategoryLabels:=1, SeriesLabels:=1, HasLegend:=1, Title:= _
      "Predictive equation assessment: Subject #" & Subno & ", Session #" & sessno & ",
Calibration #" & calno & "." _
      , CategoryTitle:="HR", ValueTitle:="LPM", ExtraTitle:=""
    ActiveChart.PlotArea.Select
    With Selection.Border
      .ColorIndex = 16
      .Weight = xlThin
      .LineStyle = xlContinuous
    End With
    Selection.Interior.ColorIndex = xlNone
    ActiveChart.SeriesCollection(2).Trendlines.Add(Type:=xlLinear, _
      Forward:=0, Backward:=0, DisplayEquation:=False, _
      DisplayRSquared:=False).Select
    ActiveChart.SeriesCollection(3).Trendlines.Add(Type:=xlLinear, _
      Forward:=0, Backward:=0, DisplayEquation:=False, _
      DisplayRSquared:=False).Select
    ActiveChart.SeriesCollection(3).Trendlines(1).Select
    With Selection.Border
      .ColorIndex = 1
      .Weight = xlHairline
      .LineStyle = xlGray50
    End With
    ActiveChart.PlotArea.Select
    ActiveChart.ChartTitle.Select
    Selection.Left = 56
    Selection.Top = 13
    ActiveChart.SeriesCollection(1).Select
    With Selection.Border
      .Weight = xlHairline
      .LineStyle = xlNone
    End With
    With Selection
      .MarkerBackgroundColorIndex = 25
      .MarkerForegroundColorIndex = 25
      .MarkerStyle = xlStar
      .Smooth = False
```

```
    End With
    ActiveChart.PlotArea.Select
    With ActiveChart.PageSetup
      .LeftHeader = ""
      .CenterHeader = ""
      .RightHeader = ""
      .LeftFooter = ""
      .CenterFooter = "S" & Subno & "N" & sessno & "HC" & calno & ".XLW"
      .RightFooter = ""
      .LeftMargin = Application.InchesToPoints(0.75)
      .RightMargin = Application.InchesToPoints(0.75)
      .TopMargin = Application.InchesToPoints(1)
      .BottomMargin = Application.InchesToPoints(1)
      .HeaderMargin = Application.InchesToPoints(0.5)
      .FooterMargin = Application.InchesToPoints(0.5)
      .ChartSize = xlFullPage
      .PrintQuality = -4
      .CenterHorizontally = False
      .CenterVertically = False
      .Orientation = xlLandscape
      .Draft = False
      .PaperSize = xlPaperLetter
      .FirstPageNumber = xlAutomatic
      .BlackAndWhite = False
      .Zoom = 100
    End With
    Sheets("Chart1").Select
    Sheets("Chart1").Name = "Pred.cht"

ActiveWorkbook.Save

End Sub
```

APPENDIX O: S-PLUS CODE FOR BOOTSTRAP ANALYSIS

```
function()
{
# Must increment set.seed input (to next prime) when re-launching
# this unless you want to duplicate a run.  Increment ppBase
# for each run to avoid overwriting files.  Restart SPlus before
# each run.
        set.seed(41)
        ppBase <- 1
        iter <- 200
        pp <- 100
        itot <- 0
        ii <- 1
        cfout.mat <- matrix(nrow = pp, ncol = 3, 0)
        vals.mat <- matrix(nrow = 175, ncol = 3, 0)
        while(itot < iter) {
                crand.dat <- runif(175, min = 0, max = 1)
                new.mat <- cbind(rownum, crand.dat)
                new.mat <- cbind(new.mat, crow = (new.mat[, "X"] * new.mat[, c(1)] +
                        new.mat[, c(2)]))
                rrow.mat <- round(new.mat[, "crow"], 0)
                tt <- 1
                while(tt < 176) {
                        rr <- rrow.mat[tt]
                        vals.mat[c(tt), c(1, 2, 3)] <- unlist(alldat[c(rr), c(2, 3, 4)],
                                use.names = FALSE)
                        tt <- tt + 1
                }
                dimnames(vals.mat) <- list(NULL, c("minvi", "deltahr", "wtkg"))
                newvals <- as.data.frame(vals.mat)
                out <- glm(minvi ~ deltahr + wtkg, family = poisson(link = log), data
                        = newvals)
                ckh <- list(coef(out))
                xint <- as.numeric(ckh[[1]][[1]])
                xdhr <- as.numeric(ckh[[1]][[2]])
                xwt <- as.numeric(ckh[[1]][[3]])
                cf <- c(xdhr, xwt, xint)
                cfout.mat[c(ii), c(1, 2, 3)] <- cf
                print(itot + ii)
                ii <- ii + 1
```

```
            if(ii > pp) {
                itot <- itot + pp
                ii <- 1
                dimnames(cfout.mat) <- list(NULL, c("cf.dhr", "cf.wt", "cf.int"
                    ))
                write(t(cfout.mat), paste("cf", ppBase, itot, ".mat", sep = ""),
                    ncol = 3)
            }
        }
        print("All done")
}
```

#NOTE: Alldat contained three columns of data: minute ventilation, delta heart rate, and #weight (in kg). Rownums contained two columns of data: the number of rows of data #for each subject, and the row in which each subject's data started in alldat.

APPENDIX P: TXDAT.BAT CODE

```
rectify %1 2 2 +
arith %1 2 #2*0.95-0.8
deriv %1 3 3 64
deriv %1 4 4 64
arith %1 3 #3/2
arith %1 4 #4/2
```

APPENDIX Q: SENDDAT.BAS CODE

```
'*************************************************************************
*****
'SENDDAT.BAS                                          DDA-06
'
'     This program is an example program that demonstrates how to write out
'to six channels of DDA-06 and simultaneous update. Make sure the jumpers
'are on X position. The program will prompt user for the Base Address, channel
'and the value he wants to write. The output voltage shown on DDA-06 will
'depend on the range setting on the board. The user has to refer to the DDA-06
'manual about switch setting.
'
'*************************************************************************
*****
' declare Pause2%
 DIM chlval!(4)
 DIM chsignal!(4)
 COMMON SHARED chlval!()
 COMMON SHARED chsignal!()
 BaseAddr = 768

 CLS
 LOCATE 1, 1:
 PRINT "SendData                                          DDA-06"
 PRINT ""
 PRINT "              Sends Subject Data to Tattletale Using DDA-06"
 PRINT ""
 PRINT ""

 INPUT "Name of codas file to be sampled"; codas$
 OPEN codas$ + ".txt" FOR INPUT AS #1

 FOR k = 0 TO 1000000
    i = 0
    DO UNTIL EOF(1) OR i = 4
        INPUT #1, chsignal!(i)              'Get value for channel I
        chlval!(i) = INT(((chsignal!(i) / 10) + .5) * 4095)
        HighByte% = INT(chlval!(i) / 256)      'Convert to high byte
        LowByte% = chlval!(i) AND &HFF         'Convert to low byte
```

```
        OUT BaseAddr + i * 2, LowByte%          'Write low byte
        OUT BaseAddr + i * 2 + 1, HighByte%     'Write high byte
        i = i + 1
    LOOP
    CALL Pause2(7620)
    Dummy% = INP(BaseAddr)    'Read dummy to update
    IF EOF(1) THEN GOSUB Ending
 NEXT k

Ending:
    CLOSE #1
    END
```

APPENDIX R: ANALOG0 PSCU CODE

```
//*****************************************************************************************
***
//
//                                    ANALOG0
// THIS PROGRAM IS SET TO READ THE ABD SIGNAL
// THE OFFSET SHOULD BE ABOUT 2.5.
// THE BEST WAY TO KNOW THE EXACT VALUE OF THE OFFSET NEEDED IS TO PUT A
 // ZERO OFFSET AND FEED ZERO VOLTS, AND READ THE VALUE . THAT VALUE WILL
// BE THE OFFSET. THIS PROGRAM SHOULD BE USEFUL TO TEST THE INPUT
// CHANNELS.
//
//
//*****************************************************************************************
*****



MODEL 400
DFSIZE 1024

old_sum!=0
ii=1
for ii=1 to 100
        X!=(4.94*CHAN(0)/65536)-0          //4.94 IS THE VALUE OF THE REFERENCE
                                           //(BOARD SUPPLY).
                                           //0.985 IS THE OFFSET INTRODUCED
        sum!=X+old_sum                                  //BY THE INPUT STAGE
        old_sum!=sum
        average!=sum/ii
next ii

PRINT "AVERAGE OFFSET FOR CHANNEL 0 =  ",#6.3F,average
```

APPENDIX S: ANALOG2 PSCU CODE

```
//*****************************************************************************************
***
//
//                                    ANALOG 2
// THIS PROGRAM CAN BE USED TO READ THE PNEUMO SIGNAL,
// THE OFFSET SHOULD BE ABOUT 0.985
// THE BEST WAY TO KNOW THE EXACT VALUE OF THE OFFSET NEEDED IS TO PUT A
// ZERO OFFSET AND FEED ZERO VOLTS, AND READ THE VALUE . THAT VALUE WILL
// BE THE OFFSET. THIS PROGRAM SHOULD BE USEFUL TO TEST THE INPUT
// CHANNELS.
//
//
//*****************************************************************************************
*****


MODEL 400
DFSIZE 1024

old_sum!=0
ii=1
for ii=1 to 100
        X!=(4.94*CHAN(2)/65536)-0          //4.94 IS THE VALUE OF THE REFERENCE
                                           //(BOARD SUPPLY).
                                           //0.985 IS THE OFFSET INTRODUCED
        sum!=X+old_sum                                  //BY THE INPUT STAGE
        old_sum!=sum
        average!=sum/ii
next ii

PRINT "AVERAGE OFFSET FOR CHANNEL 2 =  ",#6.3F,average
```

APPENDIX T: ANALOG6 PSCU CODE

```
//*****************************************************************************************
***
//
//                                    ANALOG 6
// THIS PROGRAM IS SET TO READ THE RIB SIGNAL
// THE OFFSET SHOULD BE ABOUT 2.5.
// THE BEST WAY TO KNOW THE EXACT VALUE OF THE OFFSET NEEDED IS TO PUT A
 // ZERO OFFSET AND FEED ZERO VOLTS, AND READ THE VALUE . THAT VALUE WILL
// BE THE OFFSET. THIS PROGRAM SHOULD BE USEFUL TO TEST THE INPUT
// CHANNELS.
//
//
//*****************************************************************************************
*****



MODEL 400
DFSIZE 1024

old_sum!=0
ii=1
for ii=1 to 100
        X!=(4.94*CHAN(6)/65536)-0       //4.94 IS THE VALUE OF THE REFERENCE
                                        //(BOARD SUPPLY).
                                        //0.985 IS THE OFFSET INTRODUCED
        sum!=X+old_sum                  //BY THE INPUT STAGE
        old_sum!=sum
        average!=sum/ii
next ii

PRINT "AVERAGE OFFSET FOR CHANNEL 6 =  ",#6.3F,average
```

APPENDIX U: LABPSP PSCU CODE

```
// LABPSP      written by Pablo Lopez, modified by Cheryl Hart 2/5/97
    MODEL 400
    DFSIZE 32000
    FILE_COUNT!=1
    OLD_HR_SUM!=0:OLD_HR_COUNT!=0
    OLD_PNE_SUM!=0:OLD_PNE_COUNT!=0
    OLD_RIP_SUM!=0:OLD_RIP_COUNT!=0
        // INITIALIZE FILTER
    HR_OUT!=0:PNE_OUT!=0:RIP_OUT!=0:PRED_OUT!=0:PNE!=0
        // INITIALIZE OUTPUTS
    POINTER=0
        // INITIALIZE POINTER TO FILE
    STORE POINTER,"PNE_AVE, ","RIP_AVE, ","HR_VI, ","PRED_VI, ","SATO_VI, ","HR_AVE

//  FOR PUMPS, EQUATIONS ARE mLPM vs mV
//  FOR PNEUMOTACH PUMP FLOW vs. VOLTAGE,
        // ENTER CONSTANTS: AX^2 +BX +C
    Apne!=0.00002006
    Bpne!=0.03189
    Cpne!=19.71

//   FOR RIP PUMP FLOW vs. VOLTAGE,
        //ENTER CONSTANTS: AX^2 +BX +C
    Arip!=0.00002231
    Brip!=0.03420
    Crip!=15.83

//  FOR HEART RATE PUMP FLOW vs. VOLTAGE,
        // ENTER CONSTANTS: AX^2 +BX +C
    Ahr!=0.00001896
    Bhr!=0.03100
    Chr!=17.7

//  FOR PREDICTED VENTILATION PUMP FLOW vs. VOLTAGE,
        // ENTER CONSTANTS: AX^2 +BX +C
    Apvent!=0.00002291
    Bpvent!=0.03181
    Cpvent!=22.82

PRINT ""
PRINT ""
PRINT ""
    INPUT "ENTER OFFSET FOR CHANNEL 0 "OFFSET_0!
    INPUT "ENTER OFFSET FOR CHANNEL 2 "OFFSET_2!
    INPUT "ENTER OFFSET FOR CHANNEL 6 "OFFSET_6!
```

```
PRINT "ENTER PNEUMOTACH CALCURVE CONSTANTS: M*voltage + B"
INPUT "ENTER M "M_PC!,"ENTER B "B_PC!
PRINT "FOR VENTILATION vs. HR, EQUATION IS LPM vs BPM"
PRINT "ENTER VENTILATION CONSTANTS: M*HR + B"
INPUT "ENTER M "M_HR!,"ENTER B "B_HR!
INPUT "ENTER SLOPE OF SELECTED ISOVOLUME CURVE "mISO!
PRINT "FOR VENTILATION vs. SUM of RIP DERIVATIVES, EQUATION IS LPM vs SUM OF
    ABD and RC DERIVATIVES"
INPUT "ENTER RIP SLOPE "mRIP!,"ENTER RIP INTERCEPT "bRIP!
INPUT "ENTER SUBJECT AGE "AGE!
INPUT "ENTER SUBJECT HEIGHT (in cm) "HTcm!
INPUT "ENTER SUBJECT WEIGHT (in lbs) "WTlb!
INPUT "ENTER SUBJECT RESTING HEART RATE "RESTHR!
PRINT "THE PUMP WILL PAUSE FOR 2 MINUTES BEFORE STARTING TO SAMPLE"
INPUT "HIT ENTER TO BEGIN "DUMMY!

WTkg!=WTlb/2.2
    //CALC WT_kg

ABcoef!=-mISO
    //CALC AB and RC COEFFICIENTS
RCcoef!=1

PSET 0,6:SDO INT(HR_OUT*255/250),8:PCLR 14:PSET 14
    //SEND DESIRED HR PUMP VOLTAGE TO A/D BUFFER
PCLR 0,6:SDO INT(PNE_OUT*255/250),8:PCLR 14:PSET 14
    //SEND DESIRED PNEUM PUMP VOLTAGE TO A/D BUFFER
PCLR 6:PSET 0:SDO INT(RIP_OUT*255/250),8:PCLR 14:PSET 14
    //SEND DESIRED RIP PUMP VOLTAGE TO A/D BUFFER
PCLR 0:PSET 6:SDO INT(PRED_OUT*255/250),8:PCLR 14:PSET 14
    //SEND DESIRED PRED. VENT. PUMP VOLTAGE TO A/D BUFFER
PCLR 15: PSET 15

FOR I=0 to 222000                    //2 MIN. PAUSE
NEXT I

STATE=PIN(13)
T_START=? : T_LST_PUL=?
WHILE (FILE_COUNT<161)

    IFF STATE<>PIN(13)
        NEW_HR!=6000/(?-T_LST_PUL)-1 :IF NEW_HR>180 NEW_HR=180
            //PULSE DETECTED
      T_LST_PUL=?: STATE=PIN(13)
      PRINT "DETECTED   ",#6.2F,NEW_HR

      //**************CALCULATE HR MOVING AVE**************
      HR_SUM!=OLD_HR_SUM+NEW_HR
      OLD_HR_SUM!=HR_SUM
      HR_COUNT!=OLD_HR_COUNT+1
      OLD_HR_COUNT!=HR_COUNT
```

```
HR_AVE!=HR_SUM/HR_COUNT

DELTAHR!=HR_AVE-RESTHR:IF DELTAHR<0 DELTAHR=0


//**********UPDATE PUMPS EVERY FIFTEEN SECONDS*********
  IFF ((?-T_START)/1500)>=FILE_COUNT
  FILE_COUNT=FILE_COUNT+1
  OLD_HR_SUM=0
  OLD_HR_COUNT=0
  OLD_PNE_SUM=0
  OLD_PNE_COUNT=0
  OLD_RIP_SUM=0
  OLD_RIP_COUNT=0

  HR_logVI!=M_HR*HR_AVE+B_HR
      //CALCULATE LOG(VI)
  HR_lnVI!=HR_logVI!*2.303
  HR_VI!=EXP(HR_lnVI)
  HR_FLOW!=0.025*1000*HR_VI
      //CALCULATE DESIRED FLOW FROM HR
  HR_OUT!=Ahr*HR_FLOW*HR_FLOW+Bhr*HR_FLOW+Chr
      //CALCULATE REQUIRED SIGNAL TO PUMPTO GET THAT FLOW

  PNE_FLOW!=0.025*1000*PNE_AVE!                    //SAME, BUT FOR PNEUMO
  PNE_OUT!=Apne*PNE_FLOW*PNE_FLOW+Bpne*PNE_FLOW+Cpne

  RIP_FLOW!=0.025*1000*RIP_AVE!                    //SAME, BUT FOR RIP
  RIP_OUT!=Arip*RIP_FLOW*RIP_FLOW+Brip*RIP_FLOW+Crip

  PRED_VI!=EXP(0.018944*DELTAHR+0.010523*WTkg+1.9008)
      //CALCULATE VI
  PRED_FLOW!=0.025*1000*PRED_VI
      //CALCULATE PRED.VENT. FROM HR
  PRED_OUT!=Apvent*PRED_FLOW*PRED_FLOW+Bpvent*PRED_FLOW+Cpvent

  SATO_logVI!=0.00938*DELTAHR+0.00422*HTcm+0.00119*WTkg+0.00222*AGE-
      0.0439                                       //CALCULATE LOG(VI)
  SATO_lnVI!=SATO_logVI!*2.303
  SATO_VI!=EXP(SATO_lnVI)          //CALCULATE VI USING EXP FUNCTION


  PSET 0,6:SDO INT(HR_OUT*255/250),8:PCLR 14:PSET 14
      //SEND DESIRED HR PUMP VOLTAGE TO A/D BUFFER

  PCLR 0,6:SDO INT(PNE_OUT*255/250),8:PCLR 14:PSET 14
      //SEND DESIRED PNEUM PUMP VOLTAGE TO A/D BUFFER

  PCLR 6:PSET 0:SDO INT(RIP_OUT*255/250),8:PCLR 14:PSET 14
      //SEND DESIRED RIP PUMP VOLTAGE  TO A/D BUFFER

  PCLR 0:PSET 6:SDO INT(PRED_OUT*255/250),8:PCLR 14:PSET 14
```

```
            //SEND DESIRED PRED. VENT. PUMP VOLTAGE TO A/D BUFFER

          PCLR 15: PSET 15
              //FLUSHES BUFFER VALUES TO PUMPS

          STORE POINTER," ",#04,PNE_AVE,",",#04,RIP_AVE,",",#04,HR_VI,",",#04,PRED_VI,
              ",",#04,SATO_VI,",",#04,HR_AVE

          PRINT "^^^^^^^ WRITING TO DATA FILE ^^^^^^^^^^"
          PRINT "LAST ADDRESS FOR DATA FILE:   ",POINTER

        ENDIF
        //*****************************************************************************

      //**********NO PULSE, SAMPLE PNEUMO**********************
       ELSE
        PNE_SAMPLE!=((4.94*CHAN(2)/65536)-OFFSET_2)/0.95
              // READ PNEUMO
              //div. by 0.95 b/c mult. orig. data by 0.95 to scale for PSCU
        PNE_FLOW!=M_PC*PNE_SAMPLE+B_PC
              // PNE IS VENTILATION IN LPM
        PNE_SUM!=OLD_PNE_SUM+PNE_FLOW
        OLD_PNE_SUM!=PNE_SUM
        PNE_COUNT!=OLD_PNE_COUNT+1
        OLD_PNE_COUNT!=PNE_COUNT
        PNE_AVE!=PNE_SUM/PNE_COUNT

        RIB_DER!=((4.94*CHAN(6)/65536)-OFFSET_6)*2
              //mult. by 2 b/c divided orig. data by 2 to scale for PSCU
        ABD_DER!=((4.94*CHAN(0)/65536)-OFFSET_0)*2
              //mult. by 2 b/c divided orig. data by 2 to scale for PSCU
        ABD_ADJ!=ABD_DER!*ABcoef!
        RIB_ADJ!=RIB_DER!*RCcoef!
        ADJ_RIP_SUM!=ABD_ADJ!+RIB_ADJ!
        RIP_VI!=ADJ_RIP_SUM!*mRIP+bRIP
          IF RIP_VI<0 RIP_VI=0
        RIP_SUM!=OLD_RIP_SUM+RIP_VI
        OLD_RIP_SUM!=RIP_SUM
        RIP_COUNT!=OLD_RIP_COUNT+1
        OLD_RIP_COUNT!=RIP_COUNT
        RIP_AVE!=RIP_SUM/RIP_COUNT
      ENDIF

WEND
    STORE POINTER," ",#04,0,",",#04,0,",",#04,0,",",#04,0,",",#04,0,",",#04,0
    STORE POINTER," ",#04,0,",",#04,0,",",#04,0,",",#04,0,",",#04,0,",",#04,0
    STORE POINTER," ",#04,0,",",#04,0,",",#04,0,",",#04,0,",",#04,0,",",#04,0
    STORE POINTER," ",#04,0,",",#04,0,",",#04,0,",",#04,0,",",#04,0,",",#04,0
    PRINT "^^^^^^^ WRITING TO DATA FILE ^^^^^^^^^^"
    PRINT "LAST ADDRESS FOR DATA FILE:   ",POINTER
    PRINT "DONE"
```

APPENDIX V: PRIMARY STANDARD CALCULATIONS

Subj.	Sess.	pneSlp	pneInt	#pts	sum	sum-offset	sum (LPM)	avg (LPM)	total (LPM)
1	1	30.091	-0.667	300001	-50612	199356.6	5798740	19.33	773.16
1	2	29.658	0.8205	300002	-57770	191822.7	5935230	19.78	791.36
2	1	30.091	-0.667	299972	-87834	160151.2	4619027	15.40	615.93
5	1	30.091	-0.667	299834	-64418	184683.4	5357318	17.87	714.70
6	1	30.091	-0.667	300001	-81992	166325.1	4804786	16.02	640.64
7	1	30.091	-0.667	299883	-77434	171023.6	4946249	16.49	659.76
8	1	30.091	-0.667	300001	-42607	207782.9	6052296	20.17	806.97
9	1	29.56	0.147	300002	-96433	151124.8	4511351	15.04	601.51
10	1	33.03	-1.0849	300001	-66359	182780.8	5711780	19.04	761.57
10	2	30.091	-0.667	300002	-73857	174889.1	5062485	16.87	674.99
11	1	33.03	-1.0849	299931	-60434	188958.7	5915912	19.72	788.97
11	2	29.71	1.2963	300001	-42392	208009.3	6568847	21.90	875.84
12	1	33.03	-1.0849	299930	-59208	190248.4	5958511	19.87	794.65
13	1	33.03	-1.0849	299800	-67309	181611.6	5673377	18.92	756.96
13	2	33.6	0.7246	300003	-71711	177149.4	6169601	20.57	822.60
14	1	33.03	-1.0849	300001	-79877	168551.4	5241781	17.47	698.90
14	2	28.89	2.1411	300001	-64250	185000.8	5987006	19.96	798.26
15	1	33.03	-1.0849	299971	-96033	151519.8	4679260	15.60	623.96
16	1	33.03	-1.0849	300002	-87119	160929.3	4990021	16.63	665.33
16	2	28.26	1.2053	300001	-56258	193413.5	5827456	19.42	776.99
17	1	33.03	-1.0849	299758	-28692	222225.7	7014907	23.40	936.08
17	2	28.62	0.0737	300001	-33124	217765.1	6254546	20.85	833.94
18	1	33.03	-1.0849	300001	-78130	170390.3	5302521	17.68	707.00
19	2	34.784	-0.442	300001	-72484	176333.5	6000983	20.003	800.12
20	1	30.832	-0.7572	300001	-92669	155086.1	4554454	15.18	607.26
20	2	30.82	-0.7572	299977	-97588	149888	4392406	14.64	585.70
21	1	32.19	-1.6821	299944	-61556	187788.6	5540380	18.47	738.86
21	2	32.151	1.2841	299999	-78754	169731.8	5842275	19.47	778.97
22	1	31.847	-2.2824	300001	-69783	179176.6	5021516	16.74	669.53
22	2	33.937	-0.263	299999	-88876	159077.1	5319698	17.73	709.30
23	1	31.854	-2.2861	299904	-85754	162283.4	4483764	14.95	598.03

Subj.	Sess.	pneSlp	pneInt	#pts	sum	sum-offset	sum (LPM)	avg (LPM)	total (LPM)
24	1	31.342	0.0158	300001	-90386	157489.3	4940769	16.47	658.77
25	1	33.895	-1.0035	300001	-77997	170530.3	5479074	18.26	730.54
25	2	34.997	-0.7864	300002	-84161	164042.7	5505082	18.35	734.00
26	1	30.938	0.6721	300002	-72048	176793.6	5671271	18.90	756.16
27	1	31.96	0.8202	300002	-89317	158615.4	5315409	17.72	708.72
28	1	34.826	-1.4477	300001	-89014	158933.5	5100706	17.00	680.09
28	2	29.043	2.2216	300001	-83109	165149.3	5462912	18.21	728.39
29	1	33.773	0.0589	300001	-91493	156324	5297201	17.66	706.29
29	2	29.81	1.1308	299923	-55690	193945.7	6120674	20.41	816.30
30	1	28.293	1.9381	299814	-52285	197438.1	6167186	20.57	822.80
30	2	29.673	-0.0962	300002	-64988	184224.9	5437647	18.13	725.01
31	1	28.85	1.7928	300001	-49336	200699.8	6328031	21.09	843.73
31	2	29.784	1.0238	299986	-60807	188612.4	5924758	19.75	790.00
32	1	29.105	1.4421	299958	-43439	206870.9	6453548	21.51	860.59

sum-offset = (sum - (#pts x -0.8)) / 0.95

sumLPM = (sum - offset) x pneSlp + pneInt x #pts

avgLPM = sumLPM / #pts

APPENDIX W: VTEST15S PSCU CODE

```
MODEL 400
DFSIZE 2048
X=0
PCLR 0,6
PCLR 14,15

T_START=?
X=0
WHILE (?-T_START) < 24000
PRINT "SENDING  ",X
SDO X*255/250,8
IFF (?-T_START) > 1500
    IFF (?-T_START) < 3000
        X=40
    ENDIF
ENDIF
IFF (?-T_START) > 3000
    IFF (?-T_START) < 4500
        X=80
    ENDIF
ENDIF
IFF (?-T_START) > 4500
    IFF (?-T_START) < 6000
        X=40
    ENDIF
ENDIF
IFF (?-T_START) > 6000
    IFF (?-T_START) < 7500
        X=120
    ENDIF
ENDIF
IFF (?-T_START) > 7500
    IFF (?-T_START) < 9000
        X=40
    ENDIF
ENDIF
IFF (?-T_START) > 9000
    IFF (?-T_START) < 10500
        X=160
    ENDIF
ENDIF
IFF (?-T_START) > 10500
    IFF (?-T_START) < 12000
        X=40
```

```
    ENDIF
ENDIF
IFF (?-T_START) > 12000
    IFF (?-T_START) < 13500
        X=200
    ENDIF
ENDIF
IFF (?-T_START) > 13500
    IFF (?-T_START) < 15000
        X=160
    ENDIF
ENDIF
IFF (?-T_START) > 15000
    IFF (?-T_START) < 16500
        X=200
    ENDIF
ENDIF
IFF (?-T_START) > 16500
    IFF (?-T_START) < 18000
        X=120
    ENDIF
ENDIF
IFF (?-T_START) > 18000
    IFF (?-T_START) < 19500
        X=200
    ENDIF
ENDIF
IFF (?-T_START) > 19500
    IFF (?-T_START) < 21000
        X=80
    ENDIF
ENDIF
IFF (?-T_START) > 21000
    IFF (?-T_START) < 22500
        X=200
    ENDIF
ENDIF
IFF (?-T_START) > 22500
    IFF (?-T_START) < 24000
        X=40
    ENDIF
ENDIF

WEND
```

APPENDIX X: VTEST60S PSCU CODE

```
MODEL 400
DFSIZE 2048
X=0
PCLR 0,6
PCLR 14,15

T_START=?
X=0
WHILE (?-T_START) < 93000
PRINT "SENDING  ",X
SDO X*255/250,8
IFF (?-T_START) > 3000
    IFF (?-T_START) < 9000
        X=40
    ENDIF
ENDIF
IFF (?-T_START) > 9000
    IFF (?-T_START) < 15000
        X=80
    ENDIF
ENDIF
IFF (?-T_START) > 15000
    IFF (?-T_START) < 21000
        X=40
    ENDIF
ENDIF
IFF (?-T_START) > 21000
    IFF (?-T_START) < 27000
        X=120
    ENDIF
ENDIF
IFF (?-T_START) > 27000
    IFF (?-T_START) < 33000
        X=40
    ENDIF
ENDIF
IFF (?-T_START) > 33000
    IFF (?-T_START) < 39000
        X=160
    ENDIF
ENDIF
IFF (?-T_START) > 39000
    IFF (?-T_START) < 45000
        X=40
```

```
      ENDIF
  ENDIF
  IFF (?-T_START) > 45000
      IFF (?-T_START) < 51000
          X=200
      ENDIF
  ENDIF
  IFF (?-T_START) > 51000
      IFF (?-T_START) < 57000
          X=160
      ENDIF
  ENDIF
  IFF (?-T_START) > 57000
      IFF (?-T_START) < 63000
          X=200
      ENDIF
  ENDIF
  IFF (?-T_START) > 63000
      IFF (?-T_START) < 69000
          X=120
      ENDIF
  ENDIF
  IFF (?-T_START) > 69000
      IFF (?-T_START) < 75000
          X=200
      ENDIF
  ENDIF
  IFF (?-T_START) > 75000
      IFF (?-T_START) < 81000
          X=80
      ENDIF
  ENDIF
  IFF (?-T_START) > 81000
      IFF (?-T_START) < 87000
          X=200
      ENDIF
  ENDIF
  IFF (?-T_START) > 87000
      IFF (?-T_START) < 93000
          X=40
      ENDIF
  ENDIF

  WEND
```

APPENDIX Y: NIOSH METHOD 1500

FORMULA: Table 1

M.W.: Table 1

HYDROCARBONS, BP 36 – 126 °C

METHOD: 1500
ISSUED: 2/15/84

OSHA, NIOSH, ACGIH: Table 2

PROPERTIES: Table 1

COMPOUNDS: (Synonyms in Table 1)	benzene	n-heptane	n-octane
	cyclohexane	n-hexane	n-pentane
	cyclohexene	methylcyclohexane	toluene

SAMPLING

SAMPLER: SOLID SORBENT TUBE
(coconut shell charcoal, 100 mg/50 mg)

FLOW RATE, VOLUME: Table 3

SHIPMENT: no special precautions

SAMPLE STABILITY: at least 2 weeks

BLANKS: 2 to 10 field blanks per set

BULK SAMPLE: desirable, 1 to 10 mL; ship in separate containers from samples

ACCURACY

RANGE STUDIED,
BIAS and OVERALL PRECISION (s_r): Table 3

MEASUREMENT

!
!TECHNIQUE: GAS CHROMATOGRAPHY, FID
!
!ANALYTES: hydrocarbons listed above
!
!DESORPTION: 1 mL CS_2; stand 30 min
!
!INJECTION VOLUME: 5 µL
!
!TEMPERATURE-INJECTION: 250 °C
! -DETECTOR: 250 °C
! -COLUMN: see step 11
!
!CARRIER GAS: N_2 or He, 25 mL/min
!
!COLUMN: glass, 3.0 m x 2 mm, 20% SP-2100 on
! 80/100 mesh Supelcoport
!
!CALIBRATION: analytes in CS_2
!
!RANGE AND PRECISION (s_r): Table 4
!
!ESTIMATED LOD: 0.001 to 0.01 mg per sample
! with capillary column [1]
!

APPLICABILITY: This method is intended for determining the OSHA-regulated hydrocarbons included within the boiling point range of n-pentane through n-octane. It may be used for simultaneous measurements; however, interactions between analytes may reduce breakthrough volumes and change desorption efficiencies.

INTERFERENCES: At high humidity, breakthrough volumes may be reduced by as much as 50%. Other volatile organic solvents, e.g., alcohols, ketones, ethers, and halogenated hydrocarbons, are likely interferences. If interference is suspected, use a more polar column or change column temperature.

OTHER METHODS: This method is based on and supercedes Methods P&CAM 127, benzene and toluene [2]; S28, cyclohexane [3]; S82, cyclohexene [3]; S89, heptane [3]; S90, hexane [3]; S94, methylcyclohexane [3]; S311, benzene [4]; S343, toluene [4]; S378, octane [4]; and S379, pentane [4]. For benzene or toluene in complex mixture of alkanes ($\leq C_{10}$), Method 1501 (aromatic hydrocarbons) is more selective.

HYDROCARBONS, BP 36-126 °C — METHOD: 1500

REAGENTS:

1. Eluent: Carbon disulfide*, chromatographic quality with (optional) suitable internal standard.
2. Analytes, reagent grade.*
3. Nitrogen or helium, purified.
4. Hydrogen, prepurified.
5. Air, filtered.

*See Special Precautions.

EQUIPMENT:

1. Sampler: glass tube, 7 cm long, 6 mm OD, 4 mm ID, flame-sealed ends, containing two sections of activated (600 °C) coconut shell charcoal (front = 100 mg, back = 50 mg) separated by a 2-mm urethane foam plug. A silylated glass wool plug precedes the front section, and a 3-mm urethane foam plug follows the back section. Pressure drop across the tube at 1 L/min airflow must be less than 3.4 kPa. Tubes are commercially available.
2. Personal sampling pump, 0.01 to 0.2 L/min, with flexible connecting tubing.
3. Gas chromatograph, FID, integrator and column (page 1500-1).
4. Vials, glass, 1-mL, with PTFE-lined caps.
5. Pipet, 1-mL, with pipet bulb.
6. Syringes, 5-, 10-, 25- and 100-µL.
7. Volumetric flasks, 10-mL

SPECIAL PRECAUTIONS: Carbon disulfide is toxic and extremely flammable (flash point = -30 °C); benzene is a suspect carcinogen. Prepare samples and standards in a well-ventilated hood.

SAMPLING:

1. Calibrate each personal sampling pump with a representative sampler in line.
2. Break the ends of the sampler immediately before sampling. Attach sampler to personal sampling pump with flexible tubing.
3. Sample at an accurately known flow rate between 0.01 and 0.2 L/min (0.01 to 0.05 L/min for n-pentane) for a total sample size as shown in Table 3.
4. Cap the samplers with plastic (not rubber) caps and pack securely for shipment.

SAMPLE PREPARATION:

5. Place the front and back sorbent sections of the sampler tube in separate vials. Discard the glass wool and foam plugs.
6. Add 1.0 mL eluent to each vial. Attach crimp cap to each vial immediately.
7. Allow to stand at least 30 min with occasional agitation.

CALIBRATION AND QUALITY CONTROL:

8. Calibrate daily with at least five working standards over the appropriate range (ca. 0.01 to 10 mg analyte per sample; see Table 4).
   a. Add known amounts of analyte to eluent in 10-mL volumetric flasks and dilute to the mark.
   b. Analyze together with samples and blanks (steps 11, 12 and 13).
   c. Prepare calibration graph (peak area of analyte vs. mg analyte).
9. Determine desorption efficiency (DE) at least once for each batch of charcoal used for sampling in the calibration range (step 8). Prepare three tubes at each of five levels plus three media blanks.
   a. Remove and discard back sorbent section of a media blank sampler.
   b. Inject a known amount of analyte directly onto front sorbent section with a microliter syringe.

c. Cap the tube. Allow to stand overnight.
d. Desorb (steps 5 through 7) and analyze together with working standards (steps 11, 12 and 13).
e. Prepare a graph of DE vs. mg analyte recovered.

10. Analyze three quality control blind spikes and three analyst spikes to insure that the calibration graph and DE graph are in control. Check for possible contamination during shipment of field samples by comparing results from field blanks and media blanks.

MEASUREMENT:

11. Set gas chromatograph according to manufacturer's recommendations and to conditions given on page 1500-1. Select appropriate column temperature:

Substance	Approximate Retention Time (min), at Indicated Column Temperature			
	40 °C	70 °C	100 °C	Programmed[a]
n-pentane	2.2	1.2		1.8
solvent (CS_2)	3.0	1.6		2.4
n-hexane	5.1	2.2		3.5
benzene[b]	7.7	3.2		4.5
cyclohexane[b]	8.4	3.4		4.7
cyclohexene	9.5	3.8		4.9
n-heptane	12	4.3		5.4
methylcyclohexane	14	5.2	2.2	5.9
toluene	17	6.5	2.6	6.5
n-octane	19	8.7	3.2	7.1

[a]Temperature program: 50 °C for 2 min, then 15 °C/min to 150 °C, 2-min final hold.
[b]Not completely resolved.

NOTE: Alternatively, column and temperature may be taken from Table 4.

12. Inject sample aliquot manually using solvent flush technique or with autosampler.
NOTE: If peak area is above the linear range of the working standards, dilute with eluent, reanalyze and apply the appropriate dilution factor in calculations.
13. Measure peak area.

CALCULATIONS:

14. Determine the mass, mg (corrected for DE) of analyte found in the sample front (W_f) and back (W_b) sorbent sections, and in the average media blank front (B_f) and back (B_b) sorbent sections.
NOTE: If $W_b > W_f/10$, report breakthrough and possible sample loss.
15. Calculate concentration, C, of analyte in the air volume sampled, V (L):

$$C = \frac{(W_f + W_b - B_f - B_b) \cdot 10^3}{V}, \text{ mg/m}^3.$$

EVALUATION OF METHOD:
Precisions and biases (Table 3) were determined by analyzing generated atmospheres containing one-half, one, and two times the OSHA standard. Generated concentrations were independently verified. Breakthrough capacities were determined in dry air. Storage stability was not assessed. Measurement precisions (Table 4) were determined by spiking sampling media with amounts corresponding to one-half, one, and two times the OSHA standard for nominal air volumes. Desorption efficiencies for spiked samplers containing only one compound exceeded 75%. Reference [12] provides more specific information.

REFERENCES:
[1] User check, UBTL, NIOSH Sequence #4213-L (unpublished, January 31, 1984).
[2] NIOSH Manual of Analytical Methods, 2nd. ed., V. 1, P&CAM 127 U.S. Department of Health, Education, and Welfare, Publ. (NIOSH) 77-157-A (1977).
[3] NIOSH Manual of Analytical Methods, 2nd. ed., V. 2, S28, S82, S89, S90, S94, U.S. Department of Health, Education, and Welfare, Publ. (NIOSH) 77-157-B (1977).
[4] NIOSH Manual of Analytical Methods, 2nd. ed., V. 3., S311, S343, S378, S379, U.S. Department of Health, Education, and Welfare, Publ. (NIOSH) 77-157-C (1977).
[5] R. D. Driesbach, "Physical Properties of Chemical Compounds"; Advances in Chemistry Series, No. 15; American Chemical Society, Washington (1955).
[6] R. D. Driesbach, "Physical Properties of Chemical Compounds - II"; Advances in Chemistry Series, No. 22; American Chemical Society, Washington (1959).
[7] Code of Federal Regulations; Title 29 (Labor), Parts 1900 to 1910; U.S. Government Printing Office, Washington, (1980); 29 CFR 1910.1000.
[8] Update Criteria and Recommendations for a Revised Benzene Standard, U.S. Department of Health, Education, and Welfare, (August 1976).
[9] Criteria for a Recommended Standard....Occupational Exposure to Alkanes (C5-C8), U.S. Department of Health, Education, and Welfare, Publ. (NIOSH) 77-151 (1977).
[10] Criteria for a Recommended Standard....Occupational Exposure to Toluene, U.S. Department of Health, Education, and Welfare, Publ. (NIOSH) 73-11023 (1973).
[11] TLVs—Threshold Limit Values for Chemical Substances and physical Agents in the Work Environment with Intended Changes for 1983-84, ACGIH, Cincinnati, OH (1983).
[12] Documentation of the NIOSH Validation Tests, S28, S82, S89, S90, S94, S311, S343, S378, S379, U.S. Department of Health, Education, and Welfare, Publ. (NIOSH) 77-185 (1977).

METHOD REVISED BY: R. Alan Lunsford, Ph.D., and Julie R. Okenfuss; based on results of NIOSH Contract CDC-99-74-45.

APPENDIX Z: PUMPCAL PSCU CODE

```
//USE CHANNEL A FOR OUTPUT AND CHANNEL 2 FOR INPUT
//RUN THE PROGRAM FOR 0 VOLTS IN AND READ THE OFFSET
//THEN PUT THAT OFFSET IN THE READING
        MODEL 400
        DFSIZE 30000
        PCLR 0,6,14,15
        PRINT "PUMP VOLTAGE vs PUMP FLOW"
        PRINT "   mV     vs   mLPM"
        PRINT "  Y = A X^2 + BX + C"
        DIM A!(6) : DIM R!(13) : DIM T!(5)
        N=0:DECISION=0
        FOR I=1 TO 5
        A(I)=0
        NEXT I
        FOR I=1 TO 12
        R(I)=0
        NEXT I
        FOR I=1 TO 4
        T(I)=0
        NEXT I
        D=2
        A(1)=N
//*************************************************************
//* THE FOLLOWING IS TO ADJUST THE OFFSET*

        PRINT "ADJUSTING FOR MASS FLOWMETER DC OFFSET..."
        SDO 0,8
        SLEEP 1000
        SLEEP 1000
        OFFSET!=(4.94*CHAN(2)/65536)-.985
//*************************************************************
PRINT "YOU CAN CHOOSE TO HAVE THE STANDARD CALIBRATION PROTOCOL"
PRINT "OR YOU CAN SPECIFY YOUR OWN"
PRINT "STANDARD=1 YOUR OWN=2"
INPUT CHOICE
IFF CHOICE=1
  LOW=30:HIGH=230:STP=40
ELSE
  INPUT"LOWER POINT "LOW
  INPUT "HIGHER POINT "HIGH
  INPUT "STEP "STP
ENDIF
FOR I=LOW TO HIGH STEP STP
                VOLTAGE!=I
                SDO INT(VOLTAGE)*255/250,8
```

```
                SLEEP 1000
                SLEEP 1000
                READING!=(4.94*CHAN(2)/65536)-0.985-OFFSET
                IF READING < 0 READING=0
                PRINT "READING ",#6.2F,READING
                FLOW!=(39.369*READING*READING+882.7*READING+51.64)
                PRINT "FOR ",#6.3F,VOLTAGE," FLOW IS ",#6.3F,FLOW!
                N=N+1
                Y!=VOLTAGE!:X!=FLOW
                FOR J=2 TO 5
                        A(J)=A(J)+EXP((J-1)*LOG(X))
                NEXT J
                FOR K=1 TO 3
                        R(9+K)=T(K)+Y*EXP((K-1)*LOG(X))
                        T(K)=R(9+K)
                NEXT K
                T(4)=T(4)+Y*Y
NEXT I
    SDO 0,8
        A(1)=N
        FOR J=1 TO 3
                FOR K=1 TO 3
                R(3*K+J-3)=A(J+K-1)
                NEXT K
        NEXT J
        FOR J=1 TO 3
                FOR K=J TO 3
                IF R(3*J+K-3)<>0 GOTO LAB
                NEXT K
                PRINT "NOT UNIQUE SOLUTION"
                STOP
LAB:            FOR I=1 TO 4
                 S!=R(3*I+J-3)
                 R(3*I+J-3)=R(3*I+K-3)
                 R(3*I+K-3)=S
                NEXT I
                Z!=1/R(3*J+J-3)
                FOR I=1 TO 4
                 R(3*I+J-3)=Z*R(3*I+J-3)
                NEXT I
                FOR K=1 TO 3
                 IF K=J GOTO LAB1
                 Z=-R(3*J+K-3)
                 FOR I=1 TO 4
                  R(3*I+K-3)=R(3*I+K-3)+Z*R(3*I+J-3)
                 NEXT I
LAB1:            NEXT K
        NEXT J
        PRINT""
        PRINT "SECOND DEGREE COEFF, A = ",R(12)
        PRINT "FIRST DEGREE COEFF,  B = ",R(11)
        PRINT "CONSTANT, C = ",R(10)
```

APPENDIX AA: FIELDPSP PSCU CODE

```
// FIELDPSP    written by Pablo Lopez, modified by Cheryl Hart 2/26/97
   MODEL 400
   DFSIZE 32000
   FILE_COUNT!=1
   OLD_HR_SUM!=0:OLD_HR_COUNT!=0
       // INITIALIZE FILTER
   PSP_OUT!=0
       // INITIALIZE OUTPUTS
   POINTER=0
       // INITIALIZE POINTER TO FILE
   SEC!=15
   STORE POINTER,"SEC, ","PSP_VI, ","HR_AVE, ","SMPLNUM

//   FOR PHYSIOLOGIC SAMPLER FLOW (mLPM) vs. VOLTAGE (mV),
       // ENTER CONSTANTS: AX^2 +BX +C
   Apvent!=0.00001934
   Bpvent!=0.03313
   Cpvent!=14.75

   INPUT "ENTER SUBJECT WEIGHT (in lbs) "WTlb!
   INPUT "ENTER SUBJECT RESTING HEART RATE "RESTHR!
   INPUT "ENTER CHARCOAL TUBE ID NUMBER "SMPLNUM!
   PRINT "THE PUMP WILL PAUSE FOR 2 MINUTES BEFORE STARTING TO SAMPLE"
   INPUT "HIT ENTER TO BEGIN "DUMMY!

   WTkg!=WTlb/2.2
       //CALC WT_kg

   PCLR 0,6:SDO INT(PSP_OUT*255/250),8:PCLR 14:PSET 14
       //SEND DESIRED PSP. VENT. PUMP VOLTAGE TO A/D BUFFER
   PCLR 15: PSET 15
       //FLUSHES BUFFER VALUES TO PUMPS

   FOR I=0 to 222000                          //2 MIN. PAUSE
   NEXT I

   STATE=PIN(13)
   T_START=? : T_LST_PUL=?
   WHILE (1)                                  //SEE PROGRAM "FLIP"
                                              //FOR A SIMPLIFIED VERSION
                                              //OF THIS ROUTINE
      IFF STATE<>PIN(13)
        NEW_HR!=6000/(?-T_LST_PUL)-1 :IF NEW_HR>180 NEW_HR=180
            //PULSE DETECTED
        T_LST_PUL=?: STATE=PIN(13)
```

```
          PRINT "DETECTED  ",#6.2F,NEW_HR

          //*************CALCULATE HR MOVING AVE*************
          HR_SUM!=OLD_HR_SUM+NEW_HR
          OLD_HR_SUM!=HR_SUM
          HR_COUNT!=OLD_HR_COUNT+1
          OLD_HR_COUNT!=HR_COUNT
          HR_AVE!=HR_SUM/HR_COUNT

          DELTAHR!=HR_AVE-RESTHR:IF DELTAHR<0 DELTAHR=0



          //**********UPDATE PUMPS EVERY FIFTEEN SECONDS*********
            IFF ((?-T_START)/1500)>=FILE_COUNT
            FILE_COUNT=FILE_COUNT+1
            OLD_HR_SUM=0
            OLD_HR_COUNT=0

            PSP_VI!=EXP(0.01959*DELTAHR+0.01094*WTkg+1.8029)
                //CALCULATE LOG(VI)
            PSP_FLOW!=0.025*1000*PSP_VI
                //CALCULATE PSP.VENT. FROM HR
            PSP_OUT!=Apvent*PSP_FLOW*PSP_FLOW+Bpvent*PSP_FLOW+Cpvent


            PCLR 0,6:SDO INT(PSP_OUT*255/250),8:PCLR 14:PSET 14
                //SEND DESIRED PSP. VENT. PUMP VOLTAGE TO A/D BUFFER

            PCLR 15: PSET 15
                //FLUSHES BUFFER VALUES TO PUMPS

            STORE POINTER," ",#04,SEC,",",#04,PSP_VI,",",#04,HR_AVE,",",#04,SMPLNUM
            PRINT "^^^^^^^ WRITING TO DATA FILE ^^^^^^^^^^"
            PRINT "LAST ADDRESS FOR DATA FILE:   ",POINTER
            SEC=SEC+15
          ENDIF
          //*************************************************************************
    ENDIF
WEND
```

VITA

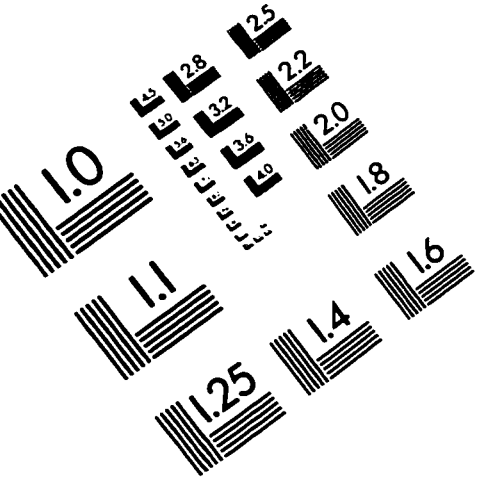
Cheryl K. Hart

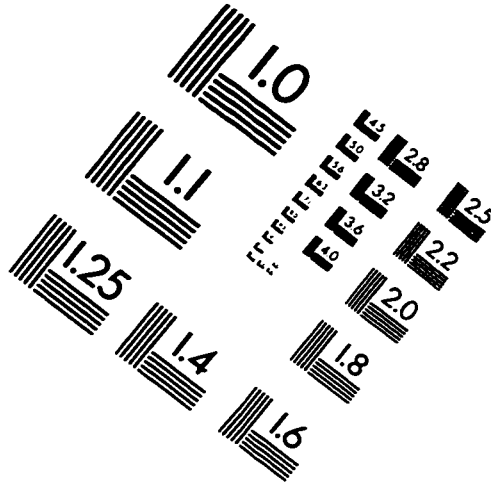
University of Washington

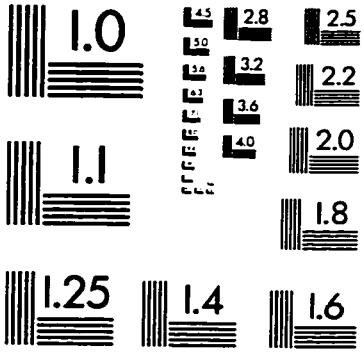
1998

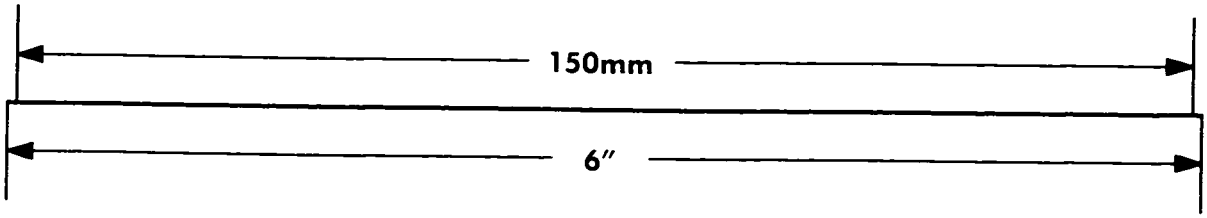
B.S., University of Washington, 1991.

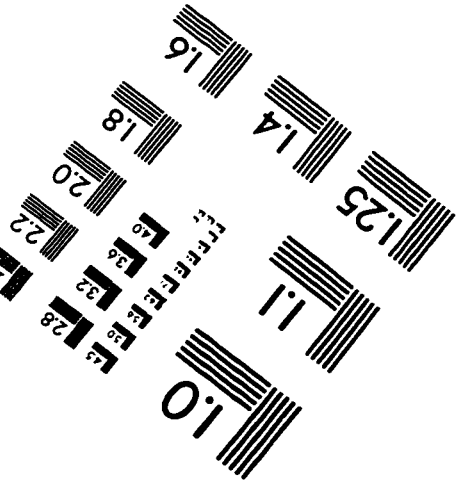
IMAGE EVALUATION TEST TARGET (QA-3)

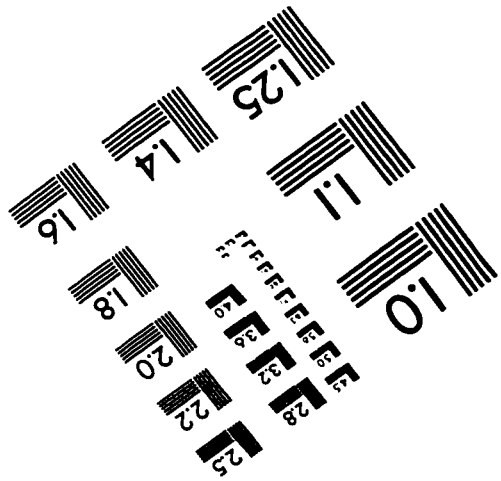












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved

99

07909

UMI

MICROFILMED 1998

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Discussing Controversial Public Issues in Secondary Social Studies Classrooms: Learning from Skilled Teachers

by

Diana Hess

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

1998

Approved by _____
Chairperson of Supervisory Committee

_____

_____

_____

Program Authorized
to Offer Degree College of Education

Date August 11, 1998

UMI Number: 9907909

Copyright 1998 by
Hess, Diana Elizabeth

All rights reserved.

UMI Microform 9907909
Copyright 1998, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized
copying under Title 17, United States Code.

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

© Copyright 1998
Diana Hess

Doctoral Dissertation

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature Diana Hess

Date August 18, 1998

University of Washington

Abstract

Discussing Controversial Public Issues in Secondary Social Studies Classrooms: Learning from Skilled Teachers

by Diana Hess

Chairperson of the Supervisory Committee:
Professor Walter Parker
College of Education

This study was about the discussion of controversial public issues (CPIs) in middle and high school social studies classes. Its purpose was to address the problem that while CPI discussions are valuable for students and for the broader society, few students are actually taught how to participate in them. I sought to better understand the instruction and conceptions undergirding the instruction of secondary social studies teachers who are skillfully teaching their students to participate more effectively in CPI discussions. To accomplish that, I studied three such teachers by interviewing them, examining discussion-related classroom artifacts, and observing videotapes and listening to audiotapes of CPI discussions in their classes. Data were analyzed in a four-step process using grounded theory methodology. During the analysis of the data, seven propositions emerged, along with six ways in which the propositions related to one another. The propositions were: (1) Teachers teach for, not just with, discussion; (2) Teachers work to make the discussions the students' forum; (3) Teachers select a discussion model and a facilitator style that is congruent with their reasons for using discussion and their definition of what constitutes effective discussion; (4) Decisions about whether and how to assess students' participation in CPI discussions are

influenced by an enduring tension between authenticity and accountability; (5) Teachers' personal views on CPI topics do not play a substantial, visible role in classroom discussion itself – however, teachers' views strongly influence the definition and choice of CPIs for discussion; (6) Teachers engage in CPI discussion teaching practices that are informed by their conceptions of democracy; and (7) Teachers are receiving support for their CPI discussion teaching from school administrators, the overall culture of the school, and their schools' missions. Implications of these propositions for teachers, teacher educators, and researchers interested in classroom discussion are also presented.

TABLE OF CONTENTS

List of Figures iii

Introduction 1

Chapter One: Background and Explanation of Problem 4

CPI Discussions as an Outcome for Democracy's Sake 4

CPI Discussions as a Method for Achieving Other Student Outcomes ... 6

Effects of CPI Discussion 10

Why Few Teachers Use CPI Discussions 16

Studying Effective CPI Discussion Teachers 17

Chapter Summary 18

Chapter Two: Research Questions and Conceptual Framework 20

First Research Question: CPI Instructional Practice 20

Second Research Question: Roles 21

Third Research Question: Conceptions 27

Conceptual Framework 34

Chapter Summary 36

Chapter Three: Methodology 38

Selection of Teachers 39

Data Collection 42

Data Analysis 49

Chapter Summary 62

Chapter Four: Portraits of Three Teachers' Conceptions and Practice of Controversial Public Issue Discussions 63

Portrait One: Joe Park at New Horizons High School 65
Portrait Two: Elizabeth Hunt at Summit Ridge Middle School 94
Portrait Three: Ann Twain at Douglas Middle School 120
Chapter Five: Propositions and Reflections on the Study 152
The Teachers: A Summary 153
Skillful Teaching of CPI Discussion: Beginnings of a Theory 155
From Propositions to an Initial Theory 174
Contributions to the Literature 177
Implications of the Theory for Practice 185
Methodological Strengths, Limitations, and Issues 188
Conclusion 195
References 198
Appendix A – Consent Form 208
Appendix B – Interview Protocols for Semi-Structured Interviews 210

LIST OF FIGURES

Number		Page
1	Relationship Between CPI Discussions and Democracy	6
2	Relationship Between CPI Discussions and Student Outcomes	7
3	Conceptual Framework	36
4	Types of Data Correlated to the Conceptual Framework	43
5	Methodological Sequence	51
6	Rationales for CPI Discussions: Joe Park	74
7	Joe's Decisions and Reasons About Which CPIs To Have Students Discuss	80
8	Rationales for CPI Discussions: Elizabeth Hunt	103
9	Elizabeth's Decisions and Reasons About Which CPIs To Have Students Discuss	109
10	Rationales for CPI Discussions: Ann Twain	132
11	Ann's Decisions and Reasons About Which CPIs To Have Students Discuss	139
12	Revised Conceptual Framework	174

Acknowledgments

Many people helped me conceptualize, research, and write this dissertation. My advisor, Professor Walter Parker, provided the perfect mix of support, intellectual engagement, and critical challenge necessary to undertake such an endeavor. I am enormously grateful to him for all he has taught me. Other professors at the University of Washington played an important role in helping me write this dissertation, including Professors Lea Vaughn, Michael Knapp, James Banks, and Arval Morris. I thank each of them for their time, expertise, and wisdom.

A special thanks to Barbara Miller and Ron Miller for hosting me as I did the field work for this dissertation. Laurel Singleton did a wonderful job editing and I am especially appreciative of her knowledge, skill, and good humor. Thanks also to Polly Linhoss and Fred Linhoss for proofreading and manuscript production with such skill and grace, and to Jennifer Brandsberg, Kate Evans, and Barbara McKean for providing such helpful feedback as I worked on final revisions.

Without the three teachers I studied, who shared their time and expertise with me, there would be no dissertation. I am especially grateful to them for opening the doors to their minds and classrooms so I could better understand the nature of the important work they have undertaken to teach young people to discuss controversial public issues.

Introduction

Finding support for the claim that students should learn to participate effectively in small and large group discussions is easy; finding evidence that students are currently provided opportunities to engage in such discussions is not. In fact, research indicates that students rarely participate in class discussions (Goodlad, 1984; McNeil, 1986). As one high school student reported, "Talking is one of the things we are pretty deprived of at school" (Alvermann et al., 1995).

Even when students do participate in classroom discussions, it is rarely the case that discussion focuses on controversial public issues (Hahn, 1991; Shaver, Davis, & Helburn, 1980). A controversial public issue (CPI) is a matter of public concern about which there is (or was in the past) disagreement (Harris, 1996). Discussing such issues has powerful educational outcomes for students (Hahn, 1996; Harris, 1996) and for the maintenance and development of a democratic society (Barber, 1989; Parker, 1996). A problem exists that can be summarized simply: students rarely participate in class discussions of CPI, even though such discussions could be educative for them and helpful for the broader society.

Notwithstanding that few students participate in classroom discussion of CPI, there are teachers who do include such discussions in their curriculum and teach their students to participate effectively in such discussions (Miller & Singleton, 1997; Rossi, 1995). Here, I studied three such teachers in order to generate an initial theory about what constitutes effective teaching of CPI discussion skills. This initial theory helps to address the problem of little emphasis on CPI discussions by providing teachers, teacher

educators, and other researchers with a better understanding of the conceptions and practices of these skilled teachers.

This dissertation has five chapters. Chapter One explains the background of the problem. It focuses on how social studies educators have traditionally made the case for including CPI discussions in the secondary social studies curriculum and reviews what previous studies have shown about the effects of CPI discussions on a number of student and societal outcomes. Chapter One concludes with an explanation of why few secondary social studies teachers include CPI discussions in their curriculum. It also explains how studying teachers who do teach students to participate effectively in CPI discussions can enhance our understanding of what constitutes effective CPI discussion teaching practice.

Chapter Two identifies and explains the three research questions that animate the study: How do secondary social studies teachers who are skilled in the use of CPI discussions teach their students to participate effectively in such discussions? What role do instructional strategies, issues, materials, and assessments play in this teaching process? What accounts for these teachers' approaches to CPI discussions? In particular, I am interested in how the teachers' conceptions of democratic citizenship, the purposes of social studies education, what constitutes good discussion, and their rationales for CPI discussions inform and influence their CPI discussion teaching practice. The explanation of the research questions emphasizes their origins, meaning, and importance. Following this, I present the conceptual framework that guides the study.

In Chapter Three, I explain the grounded theory methodology that was used in the study. This chapter focuses on the criteria used to select the three

teachers, how data was collected and analyzed, and how the initial theory was developed. Illustrations of the data analysis and theory generation process are included, along with a discussion of methodological concerns.

The conceptions and practices of the three teachers with respect to CPI discussions is presented in Chapter Four through portraits organized in accordance with the research questions and conceptual framework of the study. Each portrait explains the teacher's CPI classroom practice, the relative roles of specific parts of his/her practice (e.g., issues selection, assessment), and his/her conceptions of factors (e.g., purpose of social studies) that may influence CPI discussion teaching practice.

Chapter Five describes and explains the initial theory about effective CPI discussion teaching that emerged from the study. Drawing on similarities and differences throughout the three portraits, the theory includes seven propositions about the use of CPI discussions in secondary social studies. Relationships among the propositions are identified and a revised conceptual framework is presented. In this chapter, I also discuss how and why the initial theory generated from this study is potentially important for secondary social studies teachers, teacher educators, and other researchers of classroom discussion. It concludes with the limitations of the study.

Chapter One

Background and Explanation of the Problem

In this chapter, I describe two rationales social studies educators have traditionally used to support the use of CPI discussions: (1) positive connections between a citizenry that knows how to discuss CPIs and a healthier democracy and (2) the influence participation in such discussions has on important student outcomes. I briefly review existing research on the effect of participation in CPI discussions and the extent to which CPI discussions are actually used in social studies classrooms. The chapter concludes with a discussion of how studying secondary social studies teachers who do teach their students to participate effectively in CPI discussions can help other teachers, teacher educators, and researchers.

Since 1916, when the National Education Association's Commission on the Re-organization of Secondary Education recommended the development of a course examining the problems of democracy, social studies reformers have repeatedly called for inclusion of CPIs in the social studies curriculum. In recent years, the enthusiasm for CPIs in the social studies has not waned. Both Social Education (1996) and The Social Studies (1989), two prominent social studies education journals, devoted full issues to issues-centered social studies curriculum. Additionally, the National Council for the Social Studies sponsors a special interest group on issues-centered social studies and published a lengthy Handbook on Teaching Social Issues (Evans & Saxe, 1996).

CPI Discussions as an Outcome for Democracy's Sake

Why have social studies educators been so interested in the teaching of CPI discussions? The connection between learning how to discuss divisive

public topics and preparing for democratic citizenship is the predominant and most compelling answer. Fred Newmann (1989) argues that the most important component of effective democratic citizenship education is teaching young people how to deliberate about the nature of the public good and how to achieve that. Walter Parker (1996) makes a similar claim:

> Curricula need to emphasize civic discourse, particularly face-to-face discussion. Discussion of the public's problems, the causes of which they are effects, and alternative courses of action need to become centerpieces of the curriculum--taught, modeled, studied, practiced, and assessed. This elevates democratic deliberation to the high point of the school curriculum. (p. 197)

By deliberation, Newmann and Parker mean a particular kind of classroom discourse, one designed to teach students how to determine both public ends and means. For example, secondary social studies students might deliberate about ends when they talk about the tension between individual rights and public safety embedded in questions regarding how much safety citizens are willing to give up to be less encumbered by governmental regulations. Deliberation about means could involve analyzing whether a city council should adopt a curfew for teenagers as a means of diminishing violence by and against youth. Thus, the ends-means connection is whether a curfew would enhance or diminish the likelihood of achieving a desired goal (various levels of public safety or individual rights). Deliberation about the public good, by definition, concerns topics that are controversial, or "at issue." What makes these topics controversial is the disagreement that exists about how they should be resolved.

From this viewpoint, the rationale for deliberation of CPIs is that democratic citizenship depends on it because healthy democracies have many citizens who are engaged in high-quality public talk (Barber, 1984, 1989). As Ruth Grant reminds us:

> Decent politics, and democratic politics particularly, is conducted through talk, and thus conversation has an impact beyond the individual development of character. Conversation is a civics education as well as a moral education because the capacity for conversation is a crucial public capacity. (1966, p. 477)

The democracy rationale for CPI discussions posits discussion competence as the primary outcome; that is, young people should be taught to discuss CPIs because good citizens in a democracy need to be able to participate effectively in such discussions (see Figure 1).

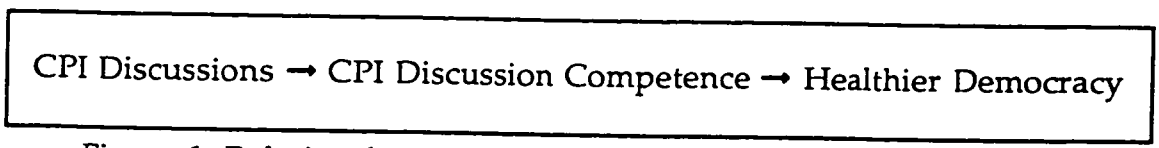


Figure 1. Relationship Between CPI Discussions and Democracy

CPI Discussions as a Method for Achieving Other Student Outcomes

The democracy rationale for CPI discussions is based on discussion competence as the primary student outcome of participation in such discussions. Another rationale for CPI discussions is that other student outcomes, such as the development of certain values or enhanced understanding of content, may be achieved through participation in CPI discussions. I am calling this rationale a method rationale because it is based on the belief that participation in CPI discussions will be a method (or

avenue) toward desired student outcomes. In other words, the discussions teach students something in addition to how to participate in discussions.

Within the method rationale are several claims about the positive student outcomes that could result from CPI discussions. The claims include the development of democratic values, increased willingness to engage in political life, enhanced content understanding, and improved critical thinking and interpersonal skills (Gall & Gall, 1990; Hahn, 1996; Harris, 1996).

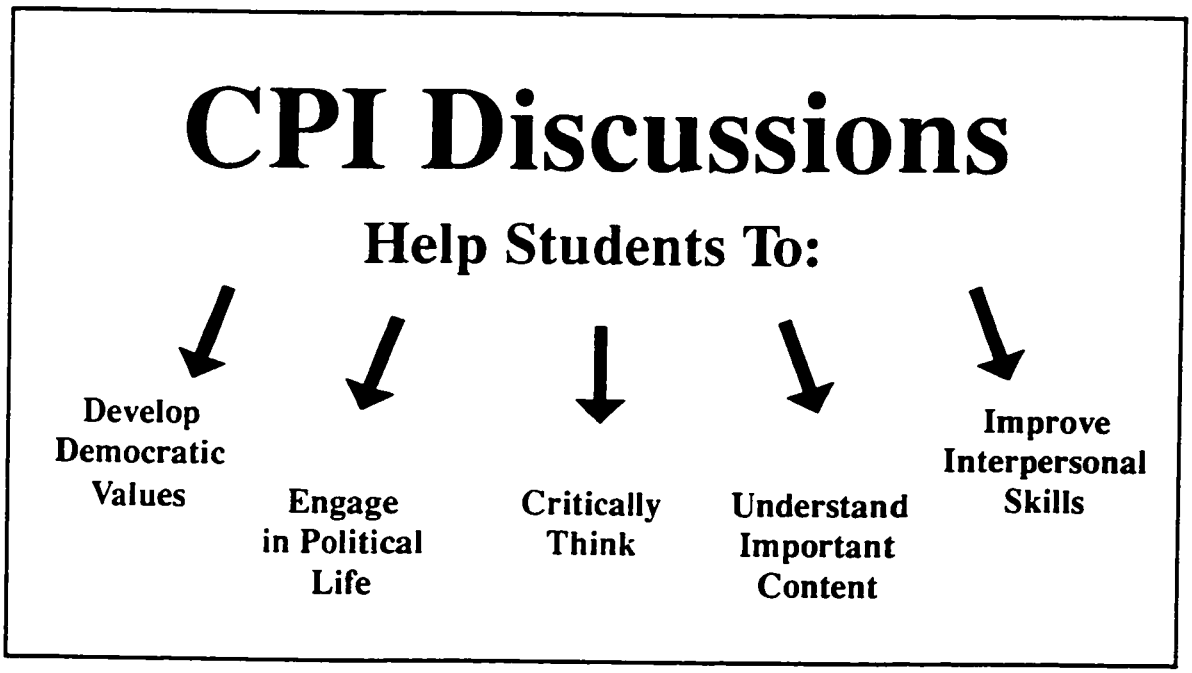


Figure 2. Relationship Between CPI Discussions and Student Outcomes

While the first two outcomes (democratic values and participation in political life) may appear to be part of the democracy rationale just discussed, I am conceptualizing them in the method category because they support types of citizenship behavior in addition to participation in public issues discussion. For example, students who develop democratic values may be more prone to behave in support of those values. A person who believes in

equality, for example, may be more likely to treat others in an equal manner. Similarly, a person can engage in political life through activities other than participation in CPI discussions, such as political protest, volunteer service, jury duty, or voting. To summarize, CPI discussions work toward the democracy rationale when the end is more effective participation in such discussions, which some theorists suggest correlates positively to a healthy democracy. CPI discussions are supported as a method when aimed at democratic values or behaviors in addition to participation in CPI discussions.

The first claim in the method rationale is that CPI discussions influence the development of democratic values, such as toleration of dissent and support for equality. This claim presumes that schooling can influence students' values and that the dynamics of effective CPI discussions, in particular, help students form values supportive of democracy (Lockwood & Harris, 1985). Informed by the research of Lawrence Kohlberg (1981), the democratic values claim posits that the cognitive dissonance created by CPI discussions, as well as the likelihood that students will hear and be attracted to moral reasoning more sophisticated than their own, will combine to shape the development of democratic values.

CPI discussions are also recommended as a way to enhance students' willingness to participate in the political world. Derived from the research on the political socialization of youth (Hahn, 1996), this claim suggests a connection between participation in discussion of CPIs and an interest in political participation. Discussing CPIs is seen as a way to help students feel more politically efficacious, an attitude correlated positively to willingness to participate in political affairs.

Participation in CPI discussions is often advocated as a means of helping students better understand important content, just as writing is recommended as a method toward enhanced understanding of content. David Harris (1996) explicated this claim when he wrote, "The effort to produce coherent language in response to a question of public policy puts knowledge in a meaningful context, making it more likely to be understood and remembered" (p. 289). For example, in a CPI discussion about physician-assisted suicide, a teacher may hope that students will form a deeper understanding of social studies content, such as the meaning of liberty in the U.S. Constitution. The idea underlying this hope is that talking with others will shape (in fact, improve) one's understanding because ideas will be challenged, broadened, and refined by the dynamics of the group discussion.

Finally, CPI discussions are advocated because it is believed they improve both students' ability to think critically and their interpersonal skills. As applied to CPIs, a critical thinker is able to think rationally by supporting and justifying arguments and conclusions (Common, 1985). By critically examining positions, and evidence supporting positions, students are able to carefully analyze their reasoning and the reasoning of their classmates. This analysis is seen as an avenue toward improving the ability to think critically. In terms of interpersonal skills, one of the overall goals of CPI discussions is to improve students' ability to work well with others, even those with whom they disagree. For example, listening attentively and disagreeing respectfully are key interpersonal skills that research suggests may be enhanced by participation in such discussions.

Effects of CPI Discussions

There are many rationales that are used to support CPI discussions. What does the research show about the effects on society and on students from participation in such discussions? In 1991, Carole Hahn wrote an extensive review of the research on controversial issues in the social studies. The review characterized the empirical evidence as "meager . . . and coming from non representative samples," yet concluded that it "consistently supports the position that positive citizenship outcomes are associated with giving students opportunities to explore controversial issues in an open, supportive classroom atmosphere" (Hahn, 1991, p. 470). Hahn was clearly unsatisfied, however, by the amount and quality of research that had been done on CPI discussions in the social studies. Her review concludes with this charge to other researchers: "There is much yet to be learned about controversial issues in social studies, and given the long-standing commitment of the profession to teaching controversial issues, this topic warrants a top priority in our collective research agenda" (1991, p. 471).

Although there is not a robust body of research on the effects of CPI discussions, the research that has been done sheds light on many of the claims made by those who advocate such discussions. In particular, two questions are addressed by the research on CPI discussions. First, what influence does participation in CPI discussions have on the building of a healthier democracy? Second, what influence does participation in CPI discussions have on students' knowledge, skills, and attitudes?

Improving Democracy

The democracy rationale for CPI discussions is based on the belief that a citizenry that can and does discuss CPIs leads to a healthier democracy. Empirical evidence suggests there is merit to this claim.

It is difficult to find examples of exemplary public talk in the United States. Barber (1989) suggests this is so because public deliberation has been reduced to "an instrument of symbolic exchange between avaricious competitors who are seen as having only private, animal interests" (p. 355). There are, however, some particularly strong exemplars of public talk. Public deliberations that are part of the National Issues Forum (NIF) network are one such example. The NIF process is designed to bring together people from all walks of life to deliberate about public issues, such as youth violence, immigration, and affirmative action. Research on the effects of NIF deliberation suggests that a positive connection exists between CPI discussions and a healthier democracy. Specifically, public deliberation "establishes and enhances communication between groups" and "improves a community's . . . ability to deal with its issues, concerns, and problems" (NIF literature, no date).

Recent focus group research on how people form relationships with public concerns (Kettering Foundation, 1993) provides additional evidence for the claim that CPI discussions influence the formation of a healthy democracy. Specifically, the researchers found that citizens want to participate in public talk, and that when they do so, they "enlarge, rather than narrow, the way they see and act on public concerns" (Kettering Foundation, 1993, p. 1). Conversation about public problems is positively linked to what people learn from other citizens and to solving important problems. The researchers

concluded that the importance of CPI discussions to citizens and to a healthy democracy shows that "talk is not cheap to people, as the axiom goes; it is the valued currency of their public life" (Kettering Foundation, 1993, p. 2). In a society where talk is often criticized in comparison to other forms of action, this research is an important reminder that public talk is a form of democratic action that appears to strengthen democracy.

In addition to the relationship between CPI discussions and a healthier democracy, researchers have also investigated what effect CPI discussions have on the students who participate in them. Recall, advocates of CPI discussions have theorized a positive relationship between CPI discussions and five categories of student outcomes: development of core democratic values, willingness to engage in political life, enhanced content understanding, improved critical thinking, and improved interpersonal skills. Empirical evidence suggests there is merit to each of the claims.

Development of Core Democratic Values

Advocates of the inclusion of controversial issues in the social studies have traditionally theorized a connection between the discussion of such issues and the development of core democratic values, such as support for civil liberties and tolerance of dissent. Research has fairly consistently shown that such a connection exists (Avery, Bird, Johnstone, Sullivan, & Thalhammer, 1992; Baughman, 1975; Brody, 1994; Goldenson, 1978; Grossman, 1975/76). For example, a study of a curriculum on free expression that included CPI discussions showed that experimental group students significantly increased their scores on a scale of Political Tolerance, and that these scores persisted four weeks after the curriculum was completed (Avery et al., 1992). Another study analyzed the influence of a nationally

disseminated curriculum on the Constitution and Bill of Rights that includes classroom CPI discussions. Participation in this curriculum, We the People (Center for Civic Education, 1987), correlated positively to students being more likely than a comparison group to support the rights of free speech, freedom of assembly, and due process for diverse groups (Brody, 1994).

Willingness to Engage in Political Life

Researchers have studied the influence of discussing CPIs on factors related to involvement in the political world, including political efficacy, political participation, and trust in the political system (Baughman, 1975; Blankenship, 1990; Ehman, 1969, 1970). The single most significant research finding focuses on the differences between CPI discussions in an open versus a closed classroom climate. An open classroom climate has three attributes: students examine multiple "sides" of an issue, students feel free to express their opinions, and teachers allow dissent. CPI discussions in an open climate have consistently been correlated positively to four effects on students:

> (1) an interest in the political world, (2) a sense that they and citizens like themselves can have some influence on political decisions in a democracy, (3) a belief that citizens have a duty to be actively engaged in politics, and (4) integration into--rather than alienation from--the school culture and the wider society. (Hahn, 1996, p. 32)

Research has also fairly consistently shown a correlation between discussion of controversial issues (even in an open climate) and increased distrust of government (Ehman, 1970; Long & Long, 1975; Zevin, 1983). These studies showed that discussion of controversial issues can enhance cynicism about politicians and government. Depending on one's political vantage point, this finding could be interpreted as either a positive outcome (i.e., good

citizens need to be cynical, it will prevent them from being "conned" by political leaders) or a negative outcome (i.e., if young people are too cynical they will not support the political structure, and ultimately might be less likely to be involved citizens).

Enhancement of Content Mastery

Several studies of CPI discussions have included a focus on the acquisition of content knowledge (Cousins, 1963; Johnston, Anderman, Milne, Klenk, & Harris, 1994; Levin, Newmann, & Oliver, 1969; Oliver & Shaver, 1974/1966). Results of these studies are mixed; some show that CPI discussions have no effect on students' content mastery (Levin et al., 1969; Oliver & Shaver, 1974/1966), while others show that CPI discussions help students learn and retain important social studies content (Cousins, 1963; Johnston et al., 1994). For example, two studies of the Harvard Social Studies Project, one with junior high students and one with high school students, showed that CPI discussions caused students to learn as much content as students in non-CPI classes (Levin et al., 1969; Oliver & Shaver, 1974/1966). In other words, CPI discussions did not cause students to learn less content, which is a concern to some teachers.

The most recent study of CPI discussions (Johnston et al., 1994) showed that the discussions were correlated positively to the mastery and retention of social studies content. In this study, experimental group students viewed Channel One broadcasts, followed by CPI discussions of issues explained on the program, while control group students viewed the broadcasts but did not participate in CPI discussions. The experimental group students scored significantly higher on a test of current events knowledge. This finding lends

support to the claim that participation in CPI discussions can enhance students' mastery of social studies content.

Improvement in Critical Thinking

More than thirty years ago, the study of the Harvard Social Studies Project showed that participation in CPI discussions enhances students' analytic competence (Oliver & Shaver, 1974/1966, pp. 262-274). This finding has recently been replicated in the Channel One study described above (Johnston et al., 1994). Using a written test adapted from the Harvard Social Studies Project study, the Channel One researchers found that students who participated in CPI discussions after viewing Channel One were better able to analyze the functions of various statements in a written dialog. As an indicator of critical thinking about CPIs, this finding is important because it illustrates the connection between the development of higher-level thinking (which is what the written test measured) and participation in CPI discussions.

Improvement in Interpersonal Skills

There is little research about the influence of participation in CPI discussions on students' interpersonal skills, such as listening, sharing the floor, and challenging others' statements without attacking. Research on various cooperative learning strategies, however, is helpful to assessing the potential of CPI discussions to improve students' interpersonal skills. Cooperative learning strategies, such as Structured Academic Controversy (Johnson & Johnson, 1979), improve students' interpersonal skills because students are provided instruction and opportunities to practice civil ways of dealing with one another. Given that much of what occurs in high-functioning cooperative learning groups is discussion, it is logical to infer that

high-quality CPI discussions can be a pathway to helping students develop interpersonal skills.

Summary of Research

In summary, research on the discussion of controversial issues in secondary social studies classes lends support to the claims of controversial issues advocates that such study can have a positive influence on important societal and student outcomes. However, there has not been much research on CPI discussions, and some of the best quality research (e.g., Levin et al., 1969; Oliver & Shaver, 1974/1966) was done three decades ago.

It is important to note that research on CPI discussions is stronger when compared to what we know about students' learning in other forms of citizenship education. For example, traditional civics classes without a focus on CPI discussions had little or no effect on student outcomes that many civic educators value. Students in such classes, for example, were not apt to be more interested in political affairs or willing to participate in the political life of their communities (Langston & Jennings, 1968; Litt, 1963). In other words, the effects of CPI discussions on a variety of student outcomes look more positive when compared to what empirical evidence suggests results from classes without an emphasis on such discussions.

Why Few Teachers Use CPI Discussions

Evidence suggests that few teachers include CPI discussions in their secondary social studies classes (Goodlad, 1984; Hahn, 1991; McNeil, 1986). Yet, survey research indicates that social studies teachers support the inclusion of CPI discussions in their courses (Engle, 1993; Hahn, 1998). Observational evidence, however, suggests that what teachers say they support on surveys does not translate into what actually happens in classrooms. For example,

John Goodlad (1984) observed that little discussion of any sort occurs in high school classes, and Fred Newmann (1988) was unable to find many social studies teachers who included discussion, even though the observations were in schools nominated for thoughtful teaching, a type of teaching that includes the use of classroom discussion.

Little is known about why teachers do or do not include CPI discussions in their courses. However, other bodies of research about why teachers teach what they do suggest some possible explanations: (1) the influence of the apprenticeship of observation (Lortie, 1975), (2) lack of attention to CPI issues in teacher education programs, (3) the difficulty of teaching CPI discussions (Dillon, 1994), (4) the impact of standardized tests, (5) lack of knowledge about how to teach students to participate effectively in CPI discussions, and (6) teachers' concerns about maintaining control of the classroom (McNeil, 1986).

Studying Effective CPI Discussion Teachers

While research suggests that many secondary social studies teachers' instruction relies on recitation and lecture (Goodlad, 1984; Hahn, 1991; McNeil, 1986), there are teachers who do teach their students to participate effectively in CPI discussions (Miller & Singleton, 1997; Rossi, 1995). What can be learned from these teachers? Lee Schulman (1983) suggests that studying "good cases" has value because it allows us to learn about the possible, instead of just the probable:

> The well-crafted case instantiates the possible, not only documenting that it can be done but also laying out at least one detailed example of how it was organized, developed, and pursued. For the practitioners concerned with process, the operational detail of case studies can be

more helpful than the more confidently generalizable virtue of quantitative analysis in many cases. (p. 495)

Examples of research on effective teachers, such as the study of exemplary secondary history teachers by Samuel Wineburg and Susan Wilson (1988) and the study of exemplary teachers of African-American children by Gloria Ladson-Billings (1994), show that an in-depth look at teachers who are doing what is difficult can contribute to the knowledge base about what constitutes effective teaching. In both of these studies, the researchers selected teachers who were atypical in order to gain a greater understanding of good teaching. Such studies are often referred to as "best practice" or "models of wisdom" studies.

In this study I sought to address the problem that few secondary social studies teachers teach their students to participate effectively in CPI discussions. I did so by studying "good cases" in order to create an initial theory of what constitutes good CPI discussion teaching practice. This, in turn, can be used to improve teachers' practice by providing examples and a theory based on them. Such a "models of wisdom" study follows the common-sense model of learning concepts based on examples. Teachers have lacked examples of what the concept "good CPI discussion teaching" entails. Lacking exemplars, it is exceptionally difficult to learn how to teach something as sophisticated as participation in CPI discussions. Moreover, the initial theory connects the examples to one another and induces from them more general ideas about the characteristics of good CPI discussion teaching.

Chapter Summary

Two rationales are commonly cited to justify teaching secondary social studies students to participate in CPI discussions: (1) positive connections

between a citizenry that knows how to discuss CPIs and a healthier democracy and (2) the influence participation in such discussions has on important student outcomes. While research on the effects of participating in CPI discussions is limited, existing research lends support to claims that such study can have a positive influence on important societal and student outcomes. Still, few teachers actually use CPI discussions in their classrooms. The reasons for their failure to do so are unknown, but this "best practice" (i.e., "model of wisdom") study should help teachers, teacher educators, and researchers understand what constitutes good CPI discussion teaching practice.

Chapter Two

Research Questions and Conceptual Framework

As documented in the previous chapter, few secondary social studies teachers teach their students how to participate effectively in CPI discussions. Neglect of CPI discussions is a problem because it deprives students, and society writ large, of opportunities to obtain the potential benefits participation in such discussions offers. This study addresses three research questions that may help us better understand how this problem can be addressed. In this chapter, I explain these questions and the conceptual framework on which the study is based.

The research questions are: How do secondary social studies teachers who are skilled in the use of CPI discussions teach their students to participate effectively in such discussions? What role do instructional strategies, issues, materials, and assessments play in this teaching process? What accounts for these teachers' approaches to CPI discussions? In particular, I am interested in how the teachers' conceptions of democratic citizenship, the purposes of social studies education, what constitutes good discussion, and their rationales for CPI discussions inform and influence their CPI discussion teaching practice. What follows is an explanation of the origin and meaning of each of the three research questions.

First Research Question: CPI Instructional Practice

The first research question is: How do secondary social studies teachers who are skilled in the use of CPI discussions teach their students to participate effectively in such discussions? This question stems from my interest in better understanding how these teachers do something that is very difficult. James Dillon (1994) reminds us that discussion is unnatural: "Discussion is

difficult. Far from coming naturally, it has to be learned" (p. 105). Moreover, discussion of CPIs is arguably more difficult than other types of discussion because the very nature of the subject matter at hand often causes fight (i.e., rancorous debate) or flight (i.e., refusal to participate). Therefore, I assume that especially skillful discussion teachers do something (most likely, many things) to engage their students in CPI discussions. That is, they do not simply provide students with the opportunity to discuss CPIs, but also teach them how to discuss these issues.

The first research question is fundamental: learning what such teachers do to teach CPI discussion skills is necessary in order to understand the role of various components of their discussion-teaching practices and the reasons for the choices they make about curriculum and instruction, which are the topics of the two other research questions.

Second Research Question: Roles

The second research question (What role do instructional strategies, issues, materials, and assessments play in this teaching process?) stems from the belief that skilled CPI teachers most likely make decisions about several basic questions intrinsic to teaching discussion. James Dillon (1994) writes that the questions include: how to conduct a discussion, how to talk in discussion, what to discuss, and how to prepare for discussion. Barbara Miller and Laurel Singleton (1997) advocate the exploration of an additional question: how to assess a discussion. In the following subsections, I examine the four components of the second research question: instructional strategies (Dillon's how to conduct a discussion and how to talk in discussion), issues for discussion, materials (the key aspect of Dillon's preparing for discussion), and assessment.

Instructional Strategy

Two of Dillon's questions (i.e., how to conduct a discussion and how to talk in discussion) imply some type of instructional strategy used by the teacher. I am defining instructional strategy to include both particular models of discussion and the ways teachers help students learn the skills necessary to participate effectively in the models. This purposely broad definition reflects the variation I expected to find in how these teachers instruct students to participate effectively in CPI discussions. For example, the instructional strategy could be selecting and orienting students to a particular model of discussion, such as Structured Academic Controversy (Johnson & Johnson, 1988) or Public Issues Discussions (Singleton & Giese, 1996), or it could be providing examples (e.g., by viewing a videotaped exemplary discussion or doing a lesson on a specific skill, such as listening) designed to help students form a more general concept of what constitutes effective discussion. Skilled CPI discussion teachers may have many instructional strategies in their pedagogical quiver, and they may use different strategies for different purposes, with different students, at different times in the school year.

Within this broad definition of instructional strategy, I am also including decisions the teacher makes about whether, when, and how to disclose her/his personal views about the CPI being discussed. Framing these decisions under the heading, "teacher's role," Tom Kelly (1986) has conceptualized four positions that a teacher can take with respect to disclosing personal views on a CPI: exclusive neutrality, exclusive partiality, neutral impartiality, and committed impartiality (pp. 114-132). Of the four, Kelly prefers committed impartiality. Committed impartiality has two components. First, it involves the teacher's stating, rather than concealing, her/his beliefs

about CPIs. Second, the teacher should encourage truth-seeking by encouraging critical discourse about competing perspectives. This means opinions other than the teacher's perspective will also get a fair hearing.

While I will be attentive to other factors related to instructional strategy (such as what the teacher does to create a safe environment for discussion), I anticipate that the three factors mentioned above will be particularly important.

Issue Selection

Of course, discussions must be about something, a topic at hand. While this study focuses only on discussions of CPIs, there are still many questions about how the teachers select them. For example, to what extent do the teachers involve students in selecting CPIs to be discussed? What specific criteria do the teachers (and/or students) use to select the issues? How do the teachers (and/or students) define what constitutes a public issue?

Many civic educators urge an in-depth approach to a relatively small number of issues instead of focusing on many issues for a short amount of time (Harris, 1996; Miller & Singleton, 1997; Rossi, 1995). Given that time is the currency of teaching, selecting issues carefully is imperative. Sound criteria must be utilized when choosing the small number of issues which most deserve students' attention.

Various criteria have been proposed for selecting issues for CPI discussions. Richard Gross (1964, pp. 3-4) advised teachers to use the following questions to select issues:

1. Is this issue beyond the maturity and experience level of the pupils?
2. Is this issue of interest to the pupils?

3. Is this issue socially significant and timely for this course and grade level?

4. Is this issue one which the teacher feels he can handle successfully from a personal standpoint?

5. Is this issue one for which adequate study materials can be obtained?

6. Is this issue one for which there is adequate time to justify its presentation?

7. Is this issue one which will clash with community customs and attitudes?

Gross's final criterion about community norms has itself been a controversial public issue for many years. Almost ten years before Gross published his list of questions, Maurice Hunt and Lawrence Metcalf (1955/1996, p. 111) urged a clash with community customs and attitudes. Unlike Gross (1964), who warned the teacher against "cutting his own throat" (p. 3) by selecting issues that would upset the community, Hunt and Metcalf recommended an examination of "closed" or taboo issues for students' own psychological well-being and for the good of democratic society at large. Only by taking issues out of the closet, so to speak, could a legitimate and productive discussion of them occur.

Other selection criteria that have been advanced include selecting issues because of their overall importance to the society and body politic or because of the possible personal significance the issues may have to particular students (Singleton & Giese, 1996). In this study I am most interested in investigating three questions related to issues selection: What issues are selected? Who selects the issues? For what reasons are the issues selected?

Materials

Once issues are selected, materials about the issue (such as films, software, written text, etc.) are gathered and studied in preparation for discussion. Without such preparation, the discussion would not be a discussion. Instead, it would be what Tom Roby (1988) calls a "bull session." Here, I will investigate how the teachers (and/or students) select materials used to prepare for discussion.

In particular, I am interested in both the depth and breadth of materials selected for discussion preparation. By depth, I mean how much students learn about an issue in preparation for the discussion. By breadth, I mean the range of materials used for discussion preparation, especially in terms of variance of perspectives on the issue. For example, if a teacher has selected affirmative action as a CPI for discussion, how many different positions on affirmative action do students study in preparation for the discussion? Do these different perspectives represent a narrow (i.e., liberal and conservative) range of views, or do they include a broader range of the political spectrum?

Finally, I will investigate whether the teacher selects the materials students use to prepare for discussion, whether students select the materials, or whether they are selected jointly by the teacher and students. This point is important because if students are finding materials about the issue they may be learning some ancillary skills to discussion, such as how to research public policy issues.

Assessment

The last component of the second research question deals with assessment. I am most interested in whether skilled CPI teachers make judgments (i.e., assessment) about how students are progressing toward the

goal of participating effectively in CPI discussions. While some experts in CPI discussions urge fairly formal assessment (Harris, 1996; Miller & Singleton, 1997), some students object to assessing discussion because of concern about whether they should be required to participate orally in class. Rahima Wade's (1994) research on preservice teachers' beliefs about oral participation in class shows that the "choice issue" (whether students should be given a choice about oral participation) is itself highly controversial. For example, 66% of preservice teachers agreed with this statement: "Participating in class discussions is a matter of personal choice. It is not essential that everyone contributes in this way" (p. 235). The Wade study suggests that teachers who do require (and assess) oral participation in class discussion may be operating at odds with their students' desires.

My primary focus for investigating the teachers' assessment of CPI discussions is whether they are formally assessing individual students' participation or assessing the class as a whole. If they are assessing individual students' participation in CPI discussions, then I would also like to know the dimensions on which they base their assessments. For example, do the teachers value some characteristics of CPI discussion participation more than others? How do the teachers determine what constitutes more or less skillful participation in CPI discussions? Have the teachers formalized their assessment of CPI discussion participation through codifying desired characteristics on a rubric? Finally, to what extent does participation in CPI discussions influence students' grades in the course?

Interactions Among Components of Teaching Practice

Embedded in the second research question is an emphasis on understanding the roles of the various components of teachers' discussion

teaching practice; that is, the function and position of the components in teaching students to participate more effectively in CPI discussions. By function and position, I mean possible relationships between components (position), such as a relationship between issues selection and instructional strategy. For example, teachers may vary their instructional strategies depending on the nature of the CPI at hand. Highly contentious issues (such as abortion or race-related controversies) may require a more structured strategy than less controversial issues.

Third Research Question: Conceptions

The third research question (What accounts for these teachers' approaches to CPI discussions?) is based on research illustrating that skillful practitioners have reasons for teaching the way they do. Samuel Wineburg and Suzanne Wilson (1988) suggest that teachers who are "wise practitioners" possess "rich and deep understandings of many things, understandings that manifest themselves in the ability to draw from a broad range of possibilities" (p. 58). I am defining these "rich and deep understandings" as conceptions, following Shavelson and Stern's (1981) thinking that conceptions include knowledge, beliefs, thoughts, and images. Together these form a mental picture--a conception--that then may inform classroom practice.

As previously mentioned, I am particularly interested in how the teachers' conceptions of four things (democratic citizenship, purposes of social studies education, what constitutes good discussion, and rationales for CPI discussion) explain and influence how they engage students in CPI discussions. I will address each in turn.

Democratic Citizenship

Preparing citizens for participation in democracy has traditionally been the mission of the public schools generally, and particularly the role of the social studies curriculum (Patrick & Hoge, 1991). This study seeks to understand how these teachers define democratic citizenship because their definitions may inform their CPI discussion teaching practice. Recall that the preparation for democratic citizenship rationale for CPI discussions contends that citizens in a democracy need to participate in decision making about public issues. The preeminence of the democratic citizenship rationale for CPI discussions suggests that how teachers conceptualize democratic citizenship will influence their CPI discussion teaching practice.

This study will investigate two questions about the teachers' conceptions of democratic citizenship. First, how do teachers conceptualize what a good citizen in a democracy should know, be able to do, and be inclined to do? Second, how do these teachers believe difference (diversity) should be dealt with in a democracy?

In an earlier study (Hess, 1997), I found that secondary social studies teachers have differing conceptions of what citizens need to know, be able to do, and be inclined to do in a democracy. One teacher believed that good citizens should be highly involved in developing answers to pressing public policy issues, while another believed that good citizens should vote wisely and participate in volunteer activities designed to alleviate human misery. Their conceptions roughly mirrored one difference found in the theoretical literature on what constitutes effective democratic citizenship: the distinction between "weak" and "strong" democracy.

Historian Paul Gagnon's (1996/1989) theory of democratic citizenship is revealed in the following:

> We seek to develop at one and the same time a taste for teamwork and a taste for critical, thorny individualism, at once the readiness to serve and the readiness to resist, for no one can foretell which way the "good" citizen ought to turn in future crises. . . . Civic education asks all this, and that citizens inform themselves on the multiple problems and choices their elected servants confront. (p. 247)

The citizen has few roles in the democracy Gagnon describes. After "serving" (for instance, willingly paying taxes and going to war) and "resisting," the citizen has only one other role--to elect the people who will do the work of deliberating about the nature of the common good and how to achieve it. This vision of democracy is "weak" because citizens are not directly involved in creating public policy. In contrast, strong democracy, as political scientist Ben Barber (1989) defines it:

> is not simply a system whereby people elect those who govern them, but a system in which every member of the community participates in self-governance. It entails not merely voting and overseeing representatives but ongoing engagement in the affairs of the civic community at the local and national levels. (p.355)

Teachers with "weak" and "strong" conceptualizations of democratic citizenship may make different choices about how to teach students to participate in CPI discussions. For example, a teacher who shares Barber's view may put greater emphasis on preparing students for public discussion of public issues, whereas a teacher who shares Gagnon's view might be more

likely to teach CPI discussions as a means of analyzing issues to prepare to vote in elections.

A second component of my investigation into teachers' conceptions of democratic citizenship is how they think about difference in a democracy. In particular, to what extent do these teachers think it is possible to have discussion across difference, especially about issues that are quite controversial? By difference, I mean the variety of ways that people define who they are relative to others, including such categories as race, gender, class, religion, and ways of looking at the world.

The theoretical literature illustrates that the possibility of discussion across difference is highly controversial. Some feminists, such as Leach (1992), suggest that power differences in contemporary American society are so firmly entrenched along lines of race, class, and gender that it is impossible to have a genuine public discussion among people who are different. Other theorists (Burbules & Rice, 1991) disagree, claiming that even though discussion across difference is exceptionally difficult, it is still possible as long as specific steps are taken to ensure that people who have historically been marginalized are allowed and encouraged to speak and be heard.

Understanding how these teachers think about difference in democracy may be important to illuminating their CPI discussion teaching practice. For example, a teacher may purposely select certain discussion models that mandate equal sharing of air time. Teachers' thinking about difference may also influence the selection of issues. For example, a teacher may select an issue that is of concern to students in the minority as a way of ensuring that minority concerns are emphasized in the classroom. Conversely, the teachers

who deny the importance of difference in democracy may make CPI discussion teaching decisions based on that assumption.

Purpose of Social Studies Education

A second conception area I am interested in investigating is how these teachers define the purposes of social studies education. The driving question undergirding this area of inquiry is: What do these teachers want their students to know and be able to do as a result of social studies? Given that CPI discussions are typically only one part of a teacher's social studies curriculum, here I seek to understand the larger whole (in terms of student outcomes) in which CPI discussions are nestled.

Numerous frameworks for delineating the purposes of social studies have been proposed, including Barr, Barth, and Shermis' (1977) typology which separates social studies into three historical traditions: social studies as citizenship transmission, as reflective inquiry, and as social science. Research since the development of this framework has suggested that teachers' conceptions of the purposes of social studies are much more complicated and variegated (White, 1982; Goodman & Adler, 1985). That is, the number of categories into which teachers' conceptions of social studies can be placed are more numerous than simply the three proposed by Barr, Barth, and Shermis (1977). For example, one of the categories that the research by Goodman and Adler (1985) added to the stew was education for social action. Given that secondary social studies teachers have tremendous power as curricular gate-keepers (Thornton, 1991), understanding how they conceptualize the purposes of social studies may be a significant step toward the larger goal of understanding why they use CPI discussions.

Good Discussion

Another area of teachers' conceptions that I will investigate is how the teachers define what constitutes good discussion. Discussion is a relatively generic term, often used to describe a range of classroom talk from recitation to more open-ended deliberation. This generality creates both a problem and an opportunity for researchers who seek to better understand the nature of classroom discussion. The problem is that social studies teachers disagree about what discussion is and about what constitutes good discussion. Recent research (Larson, 1997; Miller & Singleton, 1997) demonstrates that social studies teachers have multiple conceptions of discussion that they variously employ in the classroom. These multiple definitions of discussion can be a problem: If a concept means everything, then often it means nothing. Teachers have difficulty using discussion effectively if they have no firm concept of what constitutes discussion in the first place and, equally as significant, what makes a discussion a good discussion.

Teachers' multiple and conflicting definitions of discussion also provide an opportunity for researchers, however. Effective CPI discussion teachers may have formed an understanding of what makes a good discussion and are carefully organizing their instruction to ensure that students reach that target. In an earlier study (Hess, 1997), I found that two exemplary high school social studies teachers had strong and conflicting ideas about what constituted good CPI discussions. As an example, here I include what they said after viewing a videotape of the same discussion in a high school classroom:

> John: There's a lot of important work going on in this classroom . . . but this is not discussion because most of the questions are not only

directed at the teacher's questions, but they're compartmentalized. They almost stand as separate entities. . . . My initial definition of discussion is two or more people talking together, maybe linking their comments, spring boarding from one comment to another, give and take, response. These people might have well been talking in an empty barn.

Ian: I think they're curious to hear what other people are going to say, and it's moving along. . . . These kids are, seem to understand that they can't, you know, hold the floor forever, . . . they're reasoning, they're doing a great job . . . thinking it all the way through. They're putting pieces together and they're backing up what they're saying. They seem to have a respect for each other. . . . Reasonable people can disagree here. These kids are bringing in their factual information. I just love the way they're really, you know, interacting and following up, picking up on each other's comments.

The author of a nationally disseminated assessment rubric for CPI discussions encourages teachers to assess students' participation in CPI discussions more favorably if they use relevant background knowledge, engage others in the discussion, and constructively challenge the accuracy, clarity, relevance or logic of statements made (Harris, 1996). Other experts in CPI discussions are more interested in promoting an "open climate for discussion . . . where all points of view are valued, and all ideas merit critical examination" (Massialas & Cox, 1966). A discussion in which "all points of view are valued" may differ significantly from one in which students are encouraged to "challenge the accuracy . . . [etc.] of statements made" by other students. In the latter, accuracy and challenge receive center stage; in the

former, openness is the most significant attribute. For this study, I elicited the teachers' definitions of discussion and the specific attributes they believe a discussion must possess to be characterized as a good discussion.

Rationales for Using CPI Discussion

The last component of the third research question focuses on why secondary social studies teachers use CPI discussions. Here I am mainly interested in whether the teachers treat discussion as the outcome or as a method for achieving other outcomes. In the former, engagement in CPI discussions is designed to teach students how to participate effectively in CPI discussions. That is, this conception would identify students' ability to discuss as the primary outcome. In the latter, teachers have other reasons for using CPI discussions, such as enhanced content understanding or developing critical thinking skills.

Understanding why teachers use CPI discussions is important because rationale influences practice. For example, if teachers see enhanced discussion skills as the primary outcome, they may be more insistent that all students participate in discussions.

Conceptual Framework

The conceptual framework (see Figure 3) of the study is made up of four parts: CPI Instructional Practice, CPI Instructional Plans and Strategies, Teachers Conceptions, and Classroom, School, Community Contexts. Within each part of the framework are smaller categories, what I am calling sub-parts. I will be looking at the relationships between the sub-parts and the parts. For example, within the part of the framework on Teachers' Conceptions, I will seek to understand the teachers' conceptions about the four sub-parts (democratic citizenship, etc.), as well as how the sub-parts are interrelated,

such as a relationship between how a teacher conceptualizes democratic citizenship and the purposes he or she holds for social studies. Moreover, I will look at how elements within a particular part of the framework, such as CPI Instructional Practice, are informed and influenced by other parts of the framework, such as CPI Instructional Plans and Strategies.

Three of the four parts of the conceptual framework are directly correlated to the research questions that have already been explained. The part of the conceptual framework dealing with contexts has not been explained by the research questions because, while I think it will be important, it focuses on a more generic influence on teaching and learning than the other parts of the conceptual framework. By contexts, I mean the circumstances and settings in which the use of CPI discussions occurs, including the classroom, the school, and the community.

The classroom context includes the grouping and grade level of the students and the courses in which the teacher is including CPI discussions. This part of the conceptual framework, then, deals with ways that students and curriculum can influence how a teacher uses CPI discussions. For example, teachers may scaffold CPI discussion instruction based on how much experience they think their students have participating in such discussions. Students who have already taken courses that include a focus on CPI discussions might receive more advanced instruction on CPI discussion than students who are relatively new to this type of classroom discourse.

By school and community context, I am referring to a range of ways that school and community climate may influence how teachers use CPI discussions. For example, if administrators and other teachers value CPI discussions the teacher may be more likely to include them in the curriculum

they use in the classroom. Conversely, in some communities CPI discussions that focus on highly contentious issues are not valued, in fact, are often effectively banned. Thus, this part of the conceptual framework focuses on the need to understand the context outside of the classroom in which the teacher works.

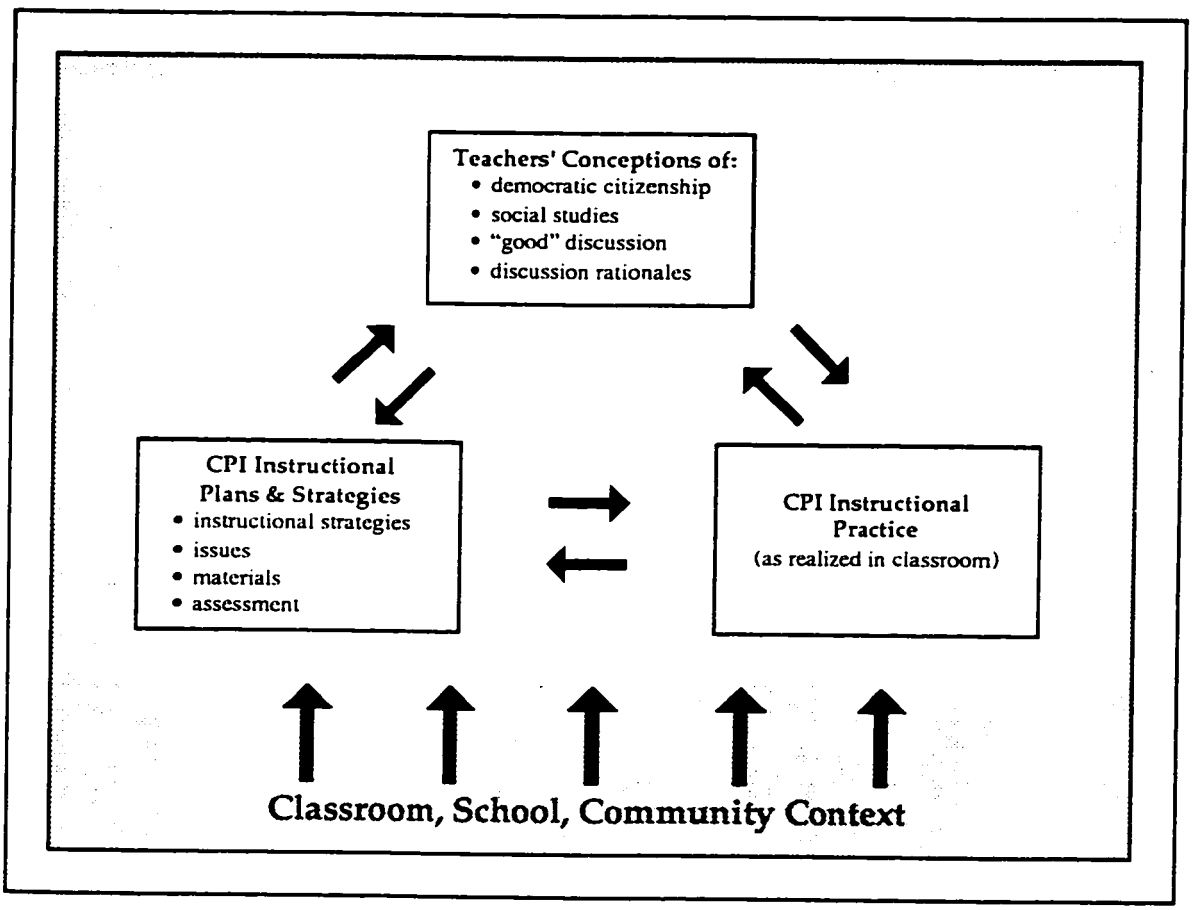


Figure 3. Conceptual Framework

Chapter Summary

Three research questions guided this study: How do secondary social studies teachers who are skilled in the use of CPI discussions teach their students to participate effectively in such discussions? What role do

instructional strategies, issues, materials, and assessments play in this teaching process? What accounts for these teachers' approaches to CPI discussions? The foci of these questions form three of the four parts of the conceptual framework on which the study is based; the fourth component of the framework is contexts--the classroom and school/community circumstances and settings in which use of CPI discussions occurs.

Chapter Three

Methodology

Models of wisdom or "best practice" studies about especially effective teachers have typically used qualitative methodology. For example, Ladson-Billings (1994), in a study of teachers who were exceptionally effective at teaching African-American students, used a combination of classroom observations, teacher interviews, and teacher focus groups to collect the data. To develop a theory about the characteristics of these effective teachers Ladson-Billings used analytic induction to make meaning from the data. Her research had the features typically associated with qualitative research, including data collected in the natural setting, data that was inductively analyzed (often likened to putting together the parts of a puzzle), and a primary emphasis on what things meant (Bogdan & Biklen, 1992). Similarly, recent research into teachers' conceptions has relied on qualitative methodology. For example, Larson's (1995) study of teachers' conceptions of classroom discussion relied on interviews, observations, and inductive analysis to develop a theory explaining teachers' purposes for using discussion.

This study, an investigation of secondary social studies teachers who are skilled at teaching their students to participate effectively in CPI discussions, is a models of wisdom or "best practice" study that focuses, in part, on the teachers' conceptions. As such, I too used qualitative methodology. Within the large family of qualitative methods, I relied on grounded theory, a qualitative method that "uses a systematic set of procedures to develop an inductively derived grounded theory about a phenomenon" (Strauss & Corbin, 1990, p. 24).

Because the grounded theory approach is appropriate for research questions that involve the understanding of complex social phenomena (Strauss & Corbin, 1990, p. 250), it was appropriate for this study. Teaching students to participate effectively in class discussion is a complex and difficult enterprise (Dillon, 1994). Understanding the conceptions and practices of teachers who do this well, by definition, involves the understanding of a complex social phenomena. The grounded theory approach is also appropriate because little is known about how teachers instruct their students to participate more effectively in CPI discussions. In other words, no existing data-based theory about how to teach CPI discussions well exists. Building the first layer of a "close to the ground" or substantive theory, defined by Gehrke and Parker (1982) as "theory that is developed for a relatively specific area of inquiry in a given context" (p. 2), then has practical utility for people interested in better understanding how to teach students to discuss controversial public issues.

In this chapter, I describe the components of the grounded theory approach as I applied them through three phases of the study: sample selection, data gathering, and data analysis. The first section of the chapter explains how I selected the teachers who participated in the study. The chapter's second section focuses on the data collection methods used in the study. The third section turns to data coding and analysis.

Selection of Teachers

Instructing young people to participate more effectively in CPI discussions can be taught by school teachers in a variety of subject areas (Engle, 1993) or by other people, such as parents, youth group leaders, or peers. This study, however, investigated only secondary social studies teachers who

were skillfully teaching their students to participate in CPI discussions. Moreover, the teachers studied were chosen purposely, instead of at random. This section addresses two questions about their purposeful selection: Why did I narrow the sample to secondary social studies teachers? How did I find and select the secondary social studies teachers who participated in the study?

Narrowing the Sample

In the grounded theory approach, a sample of groups or individuals is chosen not because it statistically represents a given population, but because it has theoretical usefulness to "discover categories and their properties, and to suggest the interrelationships into a theory" (Glaser & Strauss, 1967, p. 62). When beginning the discovery, or generation, of substantive theory, experts recommend that differences in the sample be minimized (Glaser & Strauss, 1967).

Four reasons prompted me to select secondary social studies teachers who were skilled at teaching their students to participate in CPI discussions. First, given that the purpose of the study was to generate a theory of the thinking and teaching of skilled CPI discussion teachers, a sample consisting of these skillful teachers was necessary. Second, although CPI discussions do occur in classes other than social studies, evidence suggests that social studies teachers are more likely than teachers of other subjects to include such issues in the curriculum (Engel, 1993). Third, the importance of the democracy rationale (see Chapter One) suggests that the school subject that has historically been charged with citizenship education, social studies, was a good fit for this study.

The fourth and final reason supporting my decision to study secondary social studies teachers was that secondary social studies is my area of expertise,

so selecting a sample of teachers from that category enhanced my theoretical sensitivity. As a novice researcher, I was especially concerned with theoretical sensitivity, defined by Strauss and Corbin (1990) as the researcher's "attribute of having insight, the ability to give meaning to data, the capacity to understand, and the capability to separate the pertinent from that which isn't" (p. 42). Theoretical sensitivity can come from different sources, including two that pertained to my situation: familiarity with the literature and professional experience. As a teacher of discussion classes and professional development workshops for social studies teachers, and an experienced secondary social studies teacher, my professional experience would, I reasoned, help me more readily understand the nature of what happens in CPI discussions that occur in social studies classes.

Identifying and Selecting the Teachers

Social studies educators who work with teachers in many different school districts, such as professional development providers for inservice teachers and university professors, were asked to nominate teachers for a selection pool from which I would select my sample. I sought only teachers who lived near two major cities in the West because finances precluded traveling across the United States. The social studies experts were asked to nominate secondary social studies teachers who were skilled at teaching students to participate effectively in CPI discussions. I then contacted the teachers, explained the study, and asked if they would be interested in participating. Many of the nominated teachers expressed interest in the study but did not have time to participate. Those who indicated an interest and did have time to participate were asked to identify a social studies expert who had recently seen them lead CPI discussions, either in person or on videotape. By

contacting the social studies experts they identified I received verification that the teachers were, in fact, especially skilled at teaching CPI discussions. One of the recommended teachers did not make it through this verification stage.

Once teachers were nominated and their CPI discussion-teaching prowess verified, I selected three to participate in the study on the basis of three additional criteria: the teacher was teaching a class that included discussion of CPI during the 1997-98 school year, the teacher could accommodate the interviews into his/her schedule, and necessary permission to audio- or videotape class discussions could be readily obtained from the school's principal, students, and their parents.

Data Collection

Research on especially skillful teachers and research on the conceptions of teachers typically rely on multiple data types (Ladson-Billings, 1994; Larson, 1995; Wineberg & Wilson, 1988). Researchers commonly interview teachers, observe them in the classroom, and analyze documents (e.g., lesson plans, student materials, assessments). Multiple data sources are also recommended when developing grounded theory because "it yields more information on categories than any one mode of knowing" (Glaser & Strauss, 1967, p. 66).

In this study I collected three kinds of data that correlated to the conceptual framework (see Figure 4). The three kinds of data were: (1) audiotapes and notes from interviews of the three teachers, (2) field notes from observations of three CPI discussions in each teacher's classroom (and/or notes from listening to audiotapes/viewing videotapes that the teachers made of CPI discussions in their classrooms), and (3) written artifacts related to CPI discussions. I collected all of the data from the first teacher

before moving on to teacher two. Then I collected all of the data from that teacher before moving on to teacher three.

	Semi-structured Interviews	Observations	Post-Observation Interviews	Classroom Artifacts
CPIs **Instructional Practice**		x	x	x
CPIs **Instructional Plans**				
instr. model	x	x	x	x
issues	x	x	x	x
materials		x	x	x
assessment		x	x	x
Conceptions				
democracy	x			
social studies	x			
good disc.	x	x	x	
rationales	x	x	x	x
Contexts				
school	x	x	x	x
community	x	x	x	x
classroom	x	x	x	x

Figure 4. Types of Data Correlated to the Conceptual Framework

Key to Figure 4

Note that the x means that a particular part of the conceptual framework was addressed by kind of data collected. For example, I learned about how the teachers conceptualized democracy from the semi-structured interviews.

Semi-structured interviews: Five semi-structured interviews were conducted with each teacher, many involving tasks, such as concept mapping, that the teachers performed and explained.

Observations: For the first teacher ("Joe"), I listened to audiotapes of nine CPI discussions, each lasting more than an hour. For the second teacher ("Elizabeth"), I observed two CPI discussions in her classroom and listened

to three additional discussions that had been audiotaped. For the third teacher ("Ann"), I observed three CPI discussions in her classroom and viewed three additional discussions that had been videotaped.

Post Observation/Listening/Viewing Interviews: The first teacher (Joe) and I listened to parts of two CPI discussions together; I then interviewed him about various elements of the discussions. The second teacher (Elizabeth) and I listened to parts of two audiotaped discussions together and I then interviewed her about them; I also interviewed her immediately after two CPI discussions I had observed. The third teacher (Ann) and I viewed portions of two videotapes together, immediately followed by an interview. Additionally, another interview occurred after I observed three CPI discussions in her classes.

Artifacts: I collected, for each teacher, written artifacts that pertained to CPI discussions, including a list of classroom rules for discussion, readings used to prepare students for the CPI discussions that were observed/listened to/viewed, and assessment instruments.

Interviews

Most of the data for this study came from two kinds of interviews with the teachers, semi-structured interviews and open-ended interviews. The first four interviews described below were semi-structured, meaning that I was using a detailed interview protocol for each one. Interviews five through eight were open-ended, meaning that I did not use a protocol. Below is a thumbnail sketch of each interview, describing the part of the conceptual framework the particular interview focused on and what kind of questions or tasks were part of the interview. Interview protocols for those interviews that involved specific pre-planned questions or tasks are located in the Appendices.

Interview One: Context, Purpose of Social Studies, Democratic Citizenship. In the first interview, each teacher responded to open-ended

questions about the school, community, and particular class (or classes) that included CPI discussions. They also described their backgrounds and gave short biographies of their teaching careers. Then they drew two concept maps, one illustrating what effective citizens in a democracy should know, be able to do, and be disposed to do, and another illustrating what they want their students to know, be able to do, and be disposed to do as a result of their social studies classes. After they had drawn each concept map, they explained it to me and responded to questions about its meaning. See Appendix B for the protocol for this interview.

Interview Two: Characteristics of Good Discussion. In an adaptation of the think-aloud method (Clark & Peterson, 1986), the teachers viewed two short teaching tapes that showcased class discussion and explained their reactions to the discussions. The first videotape was of a high school United States History class discussing a moral dilemma drawn from an excerpt of Richard Wright's autobiography in Reasoning with Democratic Values (Lockwood & Harris, 1985). This class discussion is on the film Minds on Social Studies (National Council for the Social Studies, 1998). The second discussion was of six students in a high school American Government class discussing whether the electoral college should be abolished or reformed. This small group discussion is included in the film series Preparing Citizens (Social Science Education Consortium, 1997).

I selected these two videos mainly because of their differences. The first discussion is a whole class discussion with students seated in rows facing the teacher. The teacher takes an active role as the discussion leader, asking many questions and occasionally correcting students' factual errors. The discussion focuses on an individual's moral decision-making about whether or not to

steal. The second discussion is not led by the teacher. Instead, a small group of discussants are seated in an inner circle surrounded by their classmates and the teacher who are evaluating the discussion using a rubric (explained in Miller & Singleton, 1997, p. 69). The discussion focuses on the public policy issue of whether the United States Constitution should be amended to change the presidential election system.

Interview Three: Rationales for CPI Discussions. Using labels drawn from the rationales for CPI discussions in the literature (see Chapter One), the teachers ranked those rationales to reflect the reasons they use CPI discussions. This kind of ranking exercise is drawn from research that connects the ability of teachers to clearly articulate their teaching aims and purposes to activities that require some kind of discrimination or ranking (Feiman-Nemser & Floden, 1986). See Appendix B for the list of rationales the teachers ranked in this interview.

Interview Four: Issues Selection. This interview involved another selection and ranking exercise. The teacher organized various CPI topics in terms of the likelihood they would be included in his/her curriculum. I selected issues drawn from a broad range of categories of public policy concerns: domestic and international topics; money and morals topics; and contemporary and historical topics. I purposely included topics that might be particularly controversial in some communities, such as gay rights, abortion, and physician-assisted suicide. See Appendix B for the list of CPI topics used in this interview.

Interview Five: Assessment. The teachers responded to open-ended questions about how they assess CPI discussions. If the teacher had developed a scoring guide for CPI discussions, then much of this interview focused on

why and how that scoring guide was developed and how it worked in practice.

Interviews Six, Seven, Eight: Enacted CPI Discussion Teaching/Post-Observation. These interviews occurred immediately after each observation or immediately after each teacher and I had listened to/viewed an audiotape/videotape of a CPI discussion. The teacher explained and evaluated CPI discussions by answering open-ended questions. The teacher also explained and analyzed the materials used for at least one of the CPI discussions I observed (or listened to) and evaluated how the materials worked with students.

Field Notes from Observations/Audiotapes/Videotapes

In addition to collecting data from interviews, I also took field notes during classroom observations and listening and viewing notes while listening to audiotapes or viewing videotapes of CPI discussions. As explained in the key to Figure 4, the way in which I learned about enacted CPI discussions in the three teachers' classes varied. For the first teacher, Joe, I listened to nine CPI discussions that had been audiotaped and did not conduct any in-person observations because the class Joe said was most suitable for this study (Important Supreme Court Cases) ended before the field work began. For the second teacher, Elizabeth, I observed two CPI discussions in person and listened to two CPI discussions on audiotape. I did not videotape these discussions because the students had not been videotaped before and I was concerned that video taping would interfere with the normal course of the discussion. For the third teacher, Ann, I observed three CPI discussions in person and viewed three on videotape. Because Ann videotapes all of the CPI discussions in her classes, I had ready access to these videotapes and was not

concerned that the additional taping would change the dynamics of the discussions. I consider the limitations caused by this variance in how I learned about each teacher's enacted CPI discussion teaching practice in the final section of this chapter.

When conducting in-person observations, I both audiotaped the discussions and took extensive field notes. The audiotapes were not transcribed as were the interviews, but I did listen to them carefully and took listening notes. The notes (both in person and from listening and viewing) focused on the quantity and content of teacher and student participation. For example, I tracked who was participating and what they were saying. When observing in person, I also noted what the teacher and students were doing (i.e., taking notes, having side conversations, looking at the speaker, etc.).

Classroom Artifacts

In addition to interviews and notes from observations (both in person and via audiotapes and videotapes), I collected a third type of data: materials used by the teacher and students to prepare for participation in CPI discussions, assessment instruments, and classroom rules that pertained to CPI discussions. Because the teachers' approaches to CPI discussions varied, these materials also varied by teacher. For example, two of the teachers formally assessed and graded CPI discussions using rubrics, but the other teacher did not. Consequently, I have assessment rubrics for only two of the three teachers.

Summary of Data Collection

Three types of data were collected for this study: semi-structured and open-ended interviews, field notes from observing and/or listening/viewing tapes of CPI discussions, and classroom artifacts related to CPI discussions,

such as assessment rubrics and preparatory readings. All three teachers were interviewed at least six times. Several of the interviews were semi-structured, followed a standardized protocol involving various tasks (such as ranking activities and concept maps) completed by the teachers. For the reasons given, both the number of CPI discussions I observed in each teacher's classes and the form (i.e., in-person observation vs. listening to an audiotape or viewing a videotape) of the observation varied. The classroom artifacts I collected also varied from teacher to teacher because of the different ways they approached CPI discussions.

Data Analysis

The grounded theory approach is distinguished from other types of qualitative research because data gathering and data analysis occur throughout the study, rather than all of the data being collected before the data analysis stage. This process allows tentative hypotheses to be checked against new data, thus assuring that verification is built into the method (Gehrke & Parker, 1982). Following this tenet of grounded theory methodology, I collected and analyzed all of the data from each teacher before moving on to the next.

In this section, I describe the four-stage process (see Figure 5) I used to analyze the data and generate the initial theory. In the first stage, I initially coded and analyzed the data by proceeding through four steps: (a) transcription and verification; (b) coding and the development of visual displays and conceptual memos; (c) portrait writing; and (d) analysis to integrate categories and their properties. After these four steps were completed for the first teacher (Joe), I moved on to the second teacher,

completing the four steps, but taking with me what I had learned from Joe. Then the entire process was repeated for the third teacher.

The second stage of the process was to compare and contrast the conceptions and practices of the three teachers by additional integration of the categories and their properties. That is, I looked for similarities and differences among the three teachers to better understand the categories, their properties, and the interconnections between categories.

In the third stage of the process, I delimited (by both limiting and demarcating) the categories and their properties, working to enhance both the parsimony and scope of the theory. This all led to the fourth and final stage, the writing of an initial theory explaining the conceptions and practices of secondary social studies teachers who are skilled at teaching their students to participate more effectively in CPI discussions.

Methodological Sequence Used to Generate the Initial Theory

Collect data from Teacher One (Interviews, Observations, Artifacts)

↓

Stage One: Code Data and Do Initial Analysis

1. Transcribe and verify
2. Code, including inter-rater coding, develop visual displays, and conceptual memos
3. Write portrait
4. Perform initial integration of categories and properties

↓

Collect data from Teacher Two

Repeat Stage One with Teacher Two's Data (see above)

↓

Collect data from Teacher Three

Repeat Stage One with Teacher Three's Data (see above)

↓

Stage Two: Integrate Categories and Their Properties Among Three Teachers' Data

↓

Stage Three: Delimit Categories and Their Properties

↓

Stage Four: Write Initial Theory

Figure 5. Methodological Sequence

Stage One: Coding and Initial Analysis

After collecting the data from the first teacher, I began to code and analyze that data. It is important to note that this coding and analysis process was initially applied to the data from teacher one, before any of the data was collected for teachers two and three. Then, the process was followed, with modifications that are explained beginning on page 58, for the data from teachers two and three. The coding and initial analysis stage included the following steps, which are described below: transcription and verification; coding, creating visual displays, and writing conceptual memos; writing portraits; and the initial integration of categories and their properties.

Transcription and Verification Opportunity. The first step in the data analysis process was to transcribe the interviews for the first teacher (Joe) and check them for accuracy. After a professional transcribed the interviews, I edited the transcriptions to correct misspellings and other inaccuracies. This editing also served to remind me of what had transpired during the interviews. After this editing, the transcripts were printed with very wide right-hand margins to allow adequate space for manual coding. I offered Joe the opportunity to read the transcripts and correct them. He did not have time to read them and indicated that he trusted the accuracy of their contents. Later, the other two teachers also declined the opportunity to read their transcripts, giving the same reasons.

Coding. Working with the transcripts, field notes, and classroom artifacts from the first teacher, I began coding. I coded each definable "chunk" of the data. Some chunks were just fragments of sentences, others were complete paragraphs. As a general rule, I defined a chunk of data as an idea

that was (1) different from what had come before and (2) important enough to stand alone.

As recommended by Glaser and Strauss (1967), I coded each chunk in the data into as many categories as possible "as categories emerge or as data emerge that fit an existing category" (p. 105). The conceptual framework provided some of the existing categories, such as the large categories of rationales for CPI discussions and characteristics of effective discussions. As I explored the data, many categories not in the conceptual framework also emerged; to illustrate, "knowledge creation" emerged early in the coding of the data from the first teacher. In many cases, a chunk of data was coded into several categories. For example, when the first teacher explained that only "good discussion" would enhance students' interpersonal skills, that "chunk" of data was coded into two categories: interpersonal skills and the defining what constitutes good discussion.

Throughout the coding process, I followed what Glaser and Strauss term "the basic, defining rule for the constant comparative method" (1967, p. 106): always compare an incident (what I am calling a "chunk") being coded into a specific category with the incidents previously coded in that category. For example, one overall category in the conceptual framework was teachers' reasons for selecting which CPI issues students would discuss. Everything in the data related to reasons for selecting issues was coded in that overall category; then sub-categories began to emerge. For example, from the first teacher's data the category of "social justice" was identified. Each time I coded an incident for "social justice" I mentally compared it to the other incidents similarly coded and to other incidents that had been coded in the overall

category of criteria for issues selection. By doing this, I began to generate theoretical properties of each category.

Inter-rater. Early in the coding process, I asked another researcher familiar with my study to code 10% of the data from the first teacher. I did this as a check on whether the categories I saw emerging from the data could be seen by another researcher. By enlisting this help, I sought to "compare the analyst's ideas [the other researcher's ideas] with [my] ideas and knowledge of the data; this comparison generates additional theoretical ideas" (Glaser & Strauss, 1967, p. 108). This process is different from the typical one used for inter-rater reliability because the other researcher was constructing categories and their properties, instead of using the categories I had identified to code the data. The other researcher and I talked about the codes she induced from the data, comparing her codes to the ones I had previously identified. This process was enormously useful because it refined my coding scheme and broadened how I was interpreting the data.

Throughout the coding process, I created visual displays and wrote conceptual memos both to reduce and to better understand the overwhelming amount of data (Miles & Huberman, 1994). For example, I created a data retrieval chart that summarized and showcased the teachers' conceptions of the rationales for using CPI discussions (see Figures 6, 8, 10). I also wrote conceptual memos explaining how I thought a particular category was emerging from the data and the properties of that category. For example, one memo explored how the first teacher used CPI discussions to help his students create new knowledge about the issue being discussed. Another memo focused on the data about the first teacher's role in facilitating discussions that emerged from the audiotaped discussions. Throughout the

process of creating visual displays and writing conceptual memos, I continually referenced the data, as is typical of the constant comparative approach of joint coding and analysis.

Writing Portrait One. The next step in the data analysis process was to write the portrait (Lawrence-Lightfoot & Davis, 1997) of the first teacher. Using the visual displays, conceptual memos, and the coded data, I wrote an extensive portrait to describe the first teacher's CPI discussion thinking and teaching. The portrait is not the theory, although it does include many of the categories and their properties that were later used for theory-creation. To organize the portrait, I relied on the conceptual framework (see Figure 3). For example, the overall categories of the portrait (such as contexts and conceptions) mirror those in the conceptual framework. Within those categories, however, are glimmers of the emerging theory. For example, within the section of the portrait describing how the first teacher defines effective discussion are categories and their properties that emerged from the coding and initial analysis (such as classroom seating to promote equality).

The portraits contain many quotes from the interviews, observations, and classroom artifacts. These quotes, however, tend to be short and are used to lend authenticity to the portrait. In keeping the quotes short, I followed Wolcott's advice (1990) to not "let informants rattle on in the written account just as they may have done during the interviews" (pp. 66-67). I developed a citing system for the quotes so readers would know who made each statement, and in which context (i.e., interview, observation, etc.). Using pseudonyms for the teachers, quotes are first cited to a specific teacher (J.P. for Joe Parks, E.H. for Elizabeth Hunt, and A.T. for Ann Twain) or to a student (ST), then to its source (IV for interview, AT for audiotape, VT for videotape,

and ART for artifact), and then, if applicable, to the page number of the transcript. For example, "J.P. IV#1, p. 4" means that the quote came from Joe Parks, during interview one, and can be found on the fourth page of the transcript of that interview.

After extensive editing, each portrait was shared with the teacher it described. This was done to obtain "member feedback" (Miles & Huberman, 1994, p. 277). Inviting the teachers' reactions to the portraits served two purposes. First, the feedback enhanced the overall reliability of the study because it ensured that factual errors would be corrected. Second, I shared the portraits and invited feedback to enhance the trust that had begun to form during the data collection process. In short, I did not want the teachers to feel this research was something being done to them, but instead with them. Notwithstanding that every effort was made to ensure that their participation in the study was confidential, the teachers' principals, other teachers, and their students knew they were participants in the study. Thus, I felt an ethical obligation to provide the teachers with an opportunity to react to my description of their conceptions and practices.

All of the teachers were pleased with the portraits that I had written describing their CPI discussion conceptions and teaching practice. In particular, they said the portraits were accurate and complete. Two of the teachers, however, caught factual errors and corrected them. One was a major error caused by how I had misinterpreted the teacher's description of how issues were selected for discussion in the curriculum. Another teacher pointed out that my use of the word "consequence" implied a negative judgment to her, which was not what I had intended. The teachers' positive

reactions to their portraits were an important indicator to me that the process I had used to collect data and write the portraits had worked well.

Initial Integration of Categories and their Properties. After the first portrait was completed, I turned to the next stage in the data analysis process, which was to further refine the categories and their properties that had emerged from the first teacher's data and to begin looking for relationships across the categories. I revisited the coded data, visual displays, conceptual memos, and the portrait. As suggested by Glaser and Strauss (1967), "this process started out in a small way; memos . . . are short" (p. 108).

At the beginning of this stage, I wrote another series of conceptual memos and created additional visual displays. I frequently returned to the coded data to be reminded of all the incidents that had been coded a particular way. Categories grew larger (becoming more integrated) during this process. For example, I created a macro category labeled "Effective Democratic Citizen Behaviors (EDCB)" that encompassed all of the previous categories about what effective citizens in a democracy should know, be able to do, and be disposed to do. By collapsing these categories into a larger one, I began to see how the first teacher's conceptions of EDCB were linked to one another. Additionally, I began to see how the larger, more encompassing categories were connected. For example, I began to identify ways in which the first teacher's conceptions of EDCB were linked to how he defined the characteristics of effective discussions. At this point, I wrote additional conceptual memos that explained my tentative hypotheses about the CPI thinking and teaching of the first teacher. While still not a theory, the theory was beginning to emerge. It was time to move to the second teacher.

Repeating the Process for Data from Teachers Two and Three. Because grounded theory requires a constant comparison between data and analysis, I next moved on to teacher two, but took what I had learned from teacher one with me. I did this in two ways. First, the interview protocol for the second interview on rationales for CPI discussions changed as I moved on to teacher two, and again as I moved to teacher three. Recall, this interview involved a card-sorting task requiring the teachers to rank order reasons for using CPI discussions drawn from the literature. Teacher one added several additional rationales. For the interview with teacher two, I added to the sorting task the rationales created by teacher one. Teacher two also added rationales that were then brought to the interview with teacher three. Allowing the list of rationales to grow facilitated theory-generation by eliciting data that would allow for comparisons across the three teachers.

The second way that the data coding and analysis process I used for teacher one informed the process for teacher two (and later, how teacher two informed the process for teacher three) was in the initial coding. Instead of starting a fresh coding system with teachers two and three, I brought the codes from teacher one (and later from teacher two) with me. Thus, as I was coding the data from teachers two and three, I constantly referred to the incidents from the previous teacher(s) that had been coded in the same categories. For example, all three teachers talked about the importance of voting, and all of these data chunks were coded into a voting category. However, they talked about voting in different ways, and comparing those differences caused this category to be understood differently. Another example can be found in the second teacher's definition of interpersonal skills. She talked about the importance of students learning how to "yield" the floor in CPI discussions.

Under the previously created category of "interpersonal skills," the property of "yielding" was added.

Except for those changes (i.e., rationale interview and coding categories), the rest of the process for coding and initially analyzing the data from teachers two and three followed the pattern established for teacher one (see Figure 5). After completing all of these steps for the three teachers, I was ready to move on to the second stage in the theory-generation process: integrating categories and their properties among the three teachers.

Stage Two: Integration of Categories and Properties Among the Three Teachers

In the second stage of integrating the categories and their properties I worked with the data, visual displays, and conceptual memos from all three of the teachers. Here I was using comparison and contrast to develop categories and properties that would apply to all three teachers. For example, the category that would eventually be titled "students' forum" emerged as I collapsed previously-coded incidents from the data of the three teachers into a general category that described ways in which the CPI discussions were controlled and influenced by the students. Several properties of this category began to emerge, such as the students' influence on issue selection, and the role of the teachers as facilitators who worked to encourage interaction among the students.

It was during this process that I realized the potential significance of what was not in the data. Based on CPI discussion literature, I had expected to find that the teachers would have to make decisions about whether and how to disclose their personal views on the CPI under discussion. As I compared and contrasted the data, I realized that I had never heard a student ask the

teacher to disclose his or her views on the issue. Moreover, none of the teachers volunteered their views. This realization illustrates the recursive nature of grounded theory, for it sent me back to the data to double-check whether I was right. I was, and the absence of teachers' disclosure of personal views on CPIs then became a property within the larger category of "students' forum."

As a result of this stage in the data analysis process, I had created many categories based on the data from the three teachers. Some of the categories were more fleshed out than others, and some were closely related to one another, while others still seemed fairly distinct. Each, however, had specific properties. At this stage in the process, the emerging theory was cumbersome and awkward, signaling to me that it was time to move to the next step in the process, one of winnowing the categories and developing relationships among them.

Stage Three: Delimiting Categories and Their Properties

In grounded theory methodology, the emerging theory is "delimited," that is, limited and demarcated, reducing the number of categories and interrelationships and constantly looking for larger and more abstract categories. This defining is done with two goals in mind: to create parsimony of variables and formulation and to enhance the scope of the theory to a wide range of situations (Glaser & Strauss, 1967, p. 111).

I progressed through this delimiting stage by systematically reducing the categories into seven propositions, and then working to develop ways in which the propositions influenced one another. As explained in Chapter Five, I sought to create propositions that described and explained the CPI discussion conceptions and practices of the three teachers. The propositions

were created by collapsing categories that were closely related to one another. For example, the proposition about how the teachers teach for and with discussion (Parker, 1996)[1] was created by combining two categories that emerged during stage two. I combined the two categories into one larger and more abstract category to enhance the explanatory power of the emerging theory. That is, what seemed important about how the teachers taught with and for discussion was that they were doing both, not just focusing on one or the other. I viewed the propositions as building blocks to the emerging theory--not the theory itself.

Once I had developed the seven propositions, I began looking for ways that they influenced one another in order to develop a theory about skillful CPI discussion teaching. For example, the teachers' conceptions of democracy were informing many of their decisions about what kind of discussion to teach their students, so a key piece of the emerging theory was that teachers' conceptions of democracy were important to their CPI discussion teaching practice.

Stage Four: Writing the Initial Theory

The final stage in grounded theory methodology is to write the initial theory by collating the memos on each category, often returning to the data for illustrations. The written theory is considered a "theory-in-progress" because "researchers expect, even welcome, refinements and extensions of the theory" (Gehrke & Parker, 1982, p. 5). I wrote the theory in two stages. First, I

---

[1] Walter Parker's definition of the difference between using discussion to teach other outcomes and teaching for discussion as an outcome in its own right (the with/for distinction) informed how I labeled the first proposition. While retaining Parker's meaning, I reversed the order of the distinction to for/with to represent the primacy these teachers place on teaching for discussion.

described and explained the seven propositions (see Chapter Five). At this stage, a number of readers offered critical feedback that was used to revise the explanations to enhance their clarity. The next step was to write the theory about how the propositions related to and influenced one another. To accomplish this, I began with the conceptual framework that had informed the design of the study. I placed each proposition on the conceptual framework, re-arranging the arrows that illustrated relationships between the propositions. Using the revised conceptual framework as an outline, I then wrote a description of the emerging theory (see Chapter Five).

Chapter Summary

I used grounded theory methodology throughout three phases of the study: sample selection, data gathering, and data analysis. Three secondary social studies teachers who are especially skilled at teaching their students to participate effectively in CPI discussions were selected to participate in the study. I gathered data by interviewing the teachers, observing their CPI discussion teaching (either in person, or by listening to audiotapes/viewing videotapes), and collecting CPI-related classroom artifacts. Data analysis followed the recommended procedures of grounded theory methodology and progressed through four stages: coding and initial analysis, integrating categories and their properties among the three teachers' data, delimiting categories and their properties, and writing the initial theory.

Chapter Four

Portraits of Three Teachers' Conceptions and Practices of CPI Discussions

At the heart of this study are the CPI discussion conceptions and practice of three teachers. Just as researchers of "wise practitioners" in other areas of teaching expertise have found (Ladson-Billings, 1994; Wineberg & Wilson, 1988), no single set of beliefs and practices defines effective CPI discussion teaching. The three teachers in this study think about and enact their CPI discussion teaching in different ways. Yet similarities can also be found among the three. This chapter illustrates these similarities and differences through portraits that describe the three teachers' CPI discussion thinking and teaching; the portraits lay a foundation for understanding the theory of what constitutes effective CPI discussion teaching that will be presented in the next chapter.

The portraits are organized using the factors described in the conceptual framework for the study (see Chapter Two). Recall, the conceptual framework was developed based on the literature about how teachers think about and practice CPI discussions in their teaching. The factors in the conceptual framework are organized into four categories: contexts, conceptions, instructional plans and strategies, and CPI teaching practice. I use the categories in the conceptual framework to organize the portraits as one would use building blocks to develop a structure. Throughout the portraits I refer to the teachers and their schools by using pseudonyms.

The first section of each portrait begins with a brief biographical sketch and description of the contexts in which the teacher is working. This introduction focuses on how and why the person became a teacher, the school and community in which she/he works, and the students he/she teaches.

Following this introduction, each portrait explains the teacher's conceptions of factors that may influence CPI discussion teaching, including beliefs about what students should know and be able to do as a result of her/his social studies teaching, and the knowledge, skills, and behaviors he/she thinks effective democratic citizens should possess and exhibit. Each portrait then turns to a specific focus on CPI discussion, explaining the teacher's rationales for including such discussions in the curriculum, conceptions of what constitutes an effective discussion, and how he/she makes decisions about the important matter of what should be discussed.

Following the explanation of the teacher's contexts and conceptions, each portrait then turns to a more specific focus on how the teacher enacts his/her CPI discussion teaching. In this third section of each portrait, I describe and explain the teacher's instructional plans and strategies when using CPI discussions. I begin with how the teacher makes decisions about what should be discussed and what materials students should use to prepare for discussions.

The centerpiece of the third section of each portrait is a "snapshot" of a CPI discussion that illustrates what CPI discussions look like in each teacher's classroom. These "snapshots" are narrative descriptions of CPI discussions that reveal the CPI instructional models used by each teacher, how students prepare to participate in a discussion, the role the teacher takes throughout the discussion, and whether and how students' participation in discussions is assessed. The "snapshot" narratives alternate between describing what actually happened in a specific discussion and the teacher's explanation of various instructional plans and strategies.

Portrait One: Joe Park at New Horizons High School

Joe Park has taught secondary social studies for 22 years, a career choice initially motivated by his interest in social studies content, especially African studies. As an undergraduate student in one of the nation's premier universities, Joe's preservice teacher education was "pretty phenomenal. . . . I learned how to teach from people who really believed that social studies teaching was a special and very, very important trust" (J.P. IV#1, pp. 2-3). Joe credits the superb preservice training he received as the reason he both likes teaching and feels successful as a teacher.

Since graduating with a B.A. in history and a certificate to teach secondary social studies in 1975, Joe has taught both middle and high school social studies at several schools, completed an M.A. degree in curriculum and instruction, and participated in a wide variety of professional development programs, in both the teacher and student roles. For example, early in his teaching career, Joe spent a year as a teacher associate for a social studies think tank, where he developed curriculum and led teacher inservice professional development programs. A few years later, he spent two years as a clinical professor in a university program, supervising student teachers and facilitating staff development programs for teachers in his school district. Since that time, Joe has continued to participate in a number of professional development activities, including a national project on developing authentic assessments in civic education, Educators for Social Responsibility (a national organization dedicated to peace education), and a locally-based study group in which teachers share their curriculum and students' work as a way to improve their teaching. As an adjunct instructor for his state's flagship university, Joe teaches a preservice class on secondary social studies methods.

Contexts

The contexts in which teachers work influence their practice (Meier, 1995; Newmann & Wehlage, 1995). Defining contexts broadly to include the community, the school, the students, and the curriculum, in this section I provide a brief orientation to the contexts in which Joe's CPI discussion teaching practice is situated.

Community. The school in which Joe teaches is located in a university community outside of a major Western city in the United States. Joe describes his community as a

> middle to upper middle class college town, high expectations, high income, lots of scientists, privileged community, that is obsessive about the success of its own children. . . . For the majority of the parents in this community, or a very vocal segment of the community, traditional schooling worked for them, they perceive that it will work for their children, and they don't want anything to get in the way of their children having the same privilege that they have. It is a community that prides itself on enlightened progressivism. (J.P. IV#1, p. 8)

Not all young people and parents in the community are drawn to schools that use traditional approaches. In fact, the school district has sponsored the development of several non-traditional schools, including the high school that Joe helped design and where he has taught for six years.

New Horizons High School. In 1992 Joe joined a design team working to develop a public high school of choice (similar to a magnet school) that opened as a "break the mold" high school in the fall of 1993. New Horizons High School was specifically designed to be "a place to experiment . . . to do

high school right. As we said, to take 25 years of research and try to implement it" (J.P. IV#1, p. 5).

Founding principles of the new school included using the community as a viable learning resource; valuing diversity, including not just race, gender, and ethnicity, but a vast spectrum of "other ways that kids are" (J.P. IV#1, p. 5); actively engaging students in their learning; teaching students to take responsibility for their learning; creating and fostering a climate of mutual respect; holding high expectations for all students; and personalizing education for students.

Many of the original plans for the school did not work; after much trial and error, the teachers "came into a very strong understanding of what it means to teach from your passions, and that authentic curriculum is different than active learning" (J.P. IV#1, p. 6). Using a shared governance structure that Joe characterizes as "real wonderful and cozy" (J.P. IV#1, p. 8), the school experienced few disputes among students, parents, and staff regarding its philosophy, in large part because students choose to attend New Horizon. Any high school student in the 540-square-mile school district is eligible to apply to attend the school.

During the 1997-98 school year, New Horizons High School had 350 students and 20 teachers. Joe, one of three social studies teachers in the school, chooses to work 70% time, teaching three courses per quarter.

Students. As a school that serves students in grades 9-12, New Horizons uses a multi-aged, mixed-ability, full-inclusion model. The school, by its very design, attracts students who want a non-traditional high school education. The promotional material for the school describes the kind of student well served by its unique approach:

> Based on our experience, New Horizons High School works really well for students who: are willing to be partners with teachers, parents and other adults; are willing to be partners with other students; work hard when they are treated with respect and given autonomy; believe a school should be a community of learners; advocate for themselves and negotiate with others to solve problems; are ready for more responsibility for their own learning. (New Horizon Promotional Literature, 1997)

Curriculum. Undergirding the philosophy of the curriculum at New Horizons High School is student choice. There are no required classes, although students must successfully complete units in such areas as science, language arts, and social studies. All courses are one quarter in length. Students also create their own "Individual Student Paths," which culminate in major projects completed for graduation in their senior year.

Joe has designed a number of nine-week elective courses for New Horizons students, including The Vietnam War, Protest and Reform, and a new course for the fall quarter of 1997 on landmark Supreme Court cases about the free speech and press clauses of the First Amendment. Class discussions in this Supreme Court course, along with Joe's conceptions of a number of factors related to his use of CPI discussions, were the focus of this study.

Summary of Contexts. Joe has a rich and varied background in social studies teaching. As both a leader and learner in multiple professional development programs, Joe views his teaching as a work in progress. The community he both lives and teaches in is supportive of education, almost obsessive, and the unique school in which Joe does his daily work is based on

principles he helped to develop and implement. The students Joe teaches are drawn to the school because of its non-traditional approach, a fact that causes a higher than usual degree of agreement about the school's philosophy and curriculum. While these contexts inform and influence Joe's teaching, it is also important to understand his conceptions of a number of other factors that may influence CPI discussion teaching.

Conceptions

Understanding the thinking that undergirds classroom practice requires identification and analysis of teachers' conceptions, defined as teachers' knowledge, beliefs, thoughts, and images (Shavelson & Stern, 1981) that may inform classroom practice. In this section, I examine Joe's conceptions of a variety of factors that research suggests may be important to CPI discussion teaching practice (see Chapter Two). These factors include the purposes of social studies; the characteristics of effective democratic citizenship; the rationales for CPI discussions; and the characteristics of effective CPI discussions.

Conceptions of the Purpose of Social Studies. At the core of Joe's teaching is his belief that social studies is "interesting, engaging, and important" (J.P. IV#1, p. 10). He tries to counteract what he considers the typical outcome of social studies instruction, which is to "teach kids that social studies is stupid, uninteresting, and has nothing to do with my life and is to be avoided at all costs" (J.P. IV#1, p. 10). Drawing a concept map of what he wants his students to know, be able to do, and be disposed to do as a result of his social studies teaching, Joe identified four central outcomes: attitudes about social studies, content and values, theory making, and citizen action.

Joe seeks to inculcate in students a positive attitude about social studies because he believes there is a strong connection between what students will learn and what they think is worth knowing and doing. Two ways Joe works to help his students develop this positive view are by providing them with alternative perspectives in his history courses and by linking what they are studying to present-day concerns. For example, Joe teaches a People's History of the United States course using Howard Zinn's book of the same title because "they've gotten plenty of the straight history in their lives, and they ought to hear the other side of that story" (J.P. IV#1, p. 11). While teaching a Civil War course, Joe engaged students in a discussion of whether it is ever right to assassinate a political leader, explicitly linking that question from the Civil War era to contemporary concerns about Saddam Hussein. By providing students with alternative views and connecting the past and present, Joe hopes to interest students in social studies and thereby enhance their ability to learn.

Joe does not believe there is a core of social studies content that all students should master. He thinks that the current standards movement, which attempts to identify core content, is misguided and unworkable, in large part because the standards sacrifice depth for breadth.

The selection of which "stuff" students will learn is driven by Joe's conception of himself as a teacher of democracy in a democracy. "It's my job to teach the concepts of citizenship and democracy. . . . In a democracy such as ours, it's my job to teach the concept of justice" (J.P. IV#1, p.12). By teaching these concepts, Joe distinguishes between working toward students' developing a general understanding of and appreciation for concepts such as justice, and specific views on controversial issues. Interchanging the words

"concepts" and "values," Joe says "those are values I will adhere to and be explicit about, though we'll argue about how they play out" (J.P. IV#1, p. 12). When discussing controversial issues that involve concepts such as democracy and justice, Joe will say "I don't have an answer, but it's my job to help you ask the questions" (J.P. IV#1, p. 12).

Joe's interest in helping students ask questions is driven by his view that one of the purposes of social studies is to involve students in the "theory-making business" (J.P. IV#1, p. 12). Drawing on constructivist learning theory, Joe believes that knowledge is created, not transmitted. Thus, when studying the First Amendment, Joe's students are encouraged to create their own theories that respond to two central questions: Why do we have the First Amendment? How absolute ought it be?

Educating young people for citizen action is another goal embedded in Joe's conception of the purposes of social studies. When asked what he would like to see one of his students doing ten years in the future, Joe responded, somewhat tongue in cheek, "burning the flag" (J.P. IV#1, p. 13). He then told a story about a recent conversation with the mother of a former student. Since graduating from college, the young woman has been a witness in Guatemala (accompanying indigenous people to protect them from government oppression) and is currently working as a political organizer. The mother attributed her daughter's citizen action to the influence of Joe and his wife, who is also a social studies teacher. Upon hearing this, both Joe and his wife "swelled up three sizes too big" (J.P. IV#1, p. 13). Joe cited numerous other examples of what he wants his students to learn about citizen action from social studies, all of them emphasizing personal agency and action.

Conceptions of Effective Democratic Citizenship. While drawing a concept map of effective citizenship, Joe included the following words and phrases: "engaged," "value this engagement," "ask questions," "know something (but not everything)," and "current events" (J.P. IV#1, ART). He used these words to describe the particular kind of democracy he values and to explain how the current democracy in the United States is lacking.

At the core of Joe's conception of democratic citizenship is engagement stemming from personal agency. "Citizenship is about action . . . an effective citizen is someone who believes that he or she has agency of some sort" (J.P. IV#1, p. 16). Joe cites numerous examples of what political engagement might look like, such as standing up for the rights of others, running for office, and voting. He believes that each of these examples is indicative of a central belief held by effective citizens: that they have a voice.

One way that Joe thinks effective citizens should use their voice is by asking questions. Characterizing the ability to ask questions before acting as "the most important thing" that citizens should be able to do, Joe defines effective citizens as people who substitute "ready, ask a question, think about it" for the more typical tendency of citizens to engage in "ready, fire, aim" (J.P. IV#1, p. 14). The high premium Joe places on thoughtful questioning is rooted in his belief that most of the decisions citizens need to make are very difficult.

The questions Joe wants effective citizens to ask are built on content knowledge related to democracy, what he calls the fundamentals of democracy, defined as certain attitudes rooted in the Bill of Rights. For example, effective citizens should believe that "free speech is better than

restricted speech and the rights of the accused are more important than kangaroo courts" (J.P. IV#1, p. 15).

Knowledge of current events is another important facet of Joe's conception of effective democratic citizenship because "citizenship is not about history, but it's about the present and the future" (J.P. IV#1, p. 15). In addition to becoming critical consumers of the news, citizens should use that knowledge to engage in voting and partisan politics because "that's the coin of the realm" (J.P. IV#1, p. 16).

Conceptions of the Rationales for CPI Discussions. Joe includes CPI discussions in virtually all of the classes he teaches. Why are CPI discussions so prominent in his social studies curriculum? To address this question, Joe engaged in a card-sorting task (see Chapter Three) designed to elicit his conceptions of the rationales for CPI discussions. Figure 6 illustrates Joe's responses to a number of rationales typically cited in the literature on CPI discussions, along with additional rationales he created.

Rationales for CPI Discussions	Do CPI discussions accomplish this goal?	Reasons why CPI discussions do/do not accomplish this goal:	Caveats and/or further explanation
Improve interpersonal skills	Yes, but only if the discussion is facilitated with ground rules and norms	Good CPI discussions teach interpersonal skills, bad discussions reinforce bad behaviors	Interpersonal skills are real, real important, but must be developed in the context of doing something
Understand important democratic values	Yes, but only if the discussion is good and appropriate	Because you bump own ideas against other people's ideas	
Participate in political life of democracy (e.g., voting, serve on jury)	No	Discussion could increase cynicism	But, discussion is model for political participation defined as participating in discussion
Learn important social studies content	Yes, if focus is substantive and students are pushed	Depending on discussion model, social studies content is learned	Purpose of Seminar model is not based on content teaching, there's no official content, least important rationale
Learn how to think critically	Yes, assuming it is not in a critical thinking way	Because you bounce your ideas off others' ideas and interact with them in a thoughtful way	Discussion is one of the best ways to get students to think critically
Improve democracy	Yes, behaviors I expect in discussions are integral to how civic discourse should work	Because you have to listen and justify what you say, learn about democratic processes	This is the broad, overall purpose of CPI discussion

Figure 6. Rationales for CPI Discussions: Joe Park
(Note: Joe's responses are from J.P. IV#2, pp. 1-8)

The primary reason Joe uses CPI discussions is that they provide the practice budding democrats need to "become part of the great conversations that take place in our society, and are taking place increasingly poorly" (J.P. IV#2, p. 8). Characterizing "great conversations" as exemplars of civic discourse, Joe explicitly links CPI discussions to preparation for effective citizenship in a democracy. He is not, however, using CPI discussions to prepare students to participate in the kind of civic discourse that currently occurs in the United States:

> One of the things that drives me crazy is that what goes for political conversation in our society, on TV especially, and talk radio, is shouting matches. And has absolutely nothing to do with thoughtful dialogue and the complexities of issues . . . and that, I think, is anti-democratic. (J.P. IV#2, p. 4)

Joe teaches his students to participate in CPI discussions to counter this anti-democratic trend, in the hope of creating a better democracy. Thus, the rationales Joe uses to support CPI discussions focus more on developing specific kinds of thinking and process skills than specific social studies content.

Describing CPI discussions as "one of the best ways to get kids to think critically" (J.P. IV#2, p.4), Joe says this goal is realized if three factors are present: (1) students are pushed to justify their thinking, (2) they are presented with alternatives, and (3) there is interaction between the ideas of various discussants. CPI discussions promote more complex thinking (i.e., more critical thinking) if the discussion is structured to help students become more comfortable "accepting a lack of closure regarding issues and questions" (J.P. IV#2, p. 5).

Joe also includes CPI discussions as a way to help students develop process skills, such as the ability to both listen and talk well. Emphasizing the importance of practice to the development of these skills, Joe exclaims, "There has to be some sort of environment in society where kids practice doing that. They practice baseball batting, for God's sakes, why can't they practice talking?" (J.P. IV#2, p. 7).

Joe recognizes that CPI discussions don't automatically achieve the outcomes he values. Throughout his discussion of rationales, Joe distinguished between outcomes achieved in good discussions and those achieved in ineffective discussions. This raises the question: What, in his view, is effective discussion?

Conceptions of Effective Classroom Discussion. Joe has developed a highly specific definition of the characteristics of effective classroom discussion. These characteristics emerged as we watched and Joe discussed two videotaped excerpts of high school students' discussions in social studies. To Joe, effective discussions are based on the following factors: setting, climate, ownership, preparation, facilitation, and knowledge construction through interaction.

1. The setting promotes equality and interaction between students. Referring to the seating arrangement in the classroom as "geography," Joe emphasizes the importance of this geography to the quality of discussion. "I actually believe you cannot, by definition, have a quality discussion when kids are looking at the back of people's heads. And where their words are going in one direction" (J.P. IV#3, p.1). Arguing for seating students in a circle, or at least in semi-circles, Joe says the physical layout of the room should communicate the message "you ought to be talking" (J.P. IV#3, p. 6).

The circle is important because it is authentic to public discourse outside of school, whereas the artificial nature of rows in a classroom "is not the way society orients itself and organizes itself" (J.P. IV#3, p.1).

2. The climate encourages students to talk and listen. Many environments are too cold or intimidating for students to participate in discussions. Creating a climate in which the opposite is the case--students want to talk and listen to one another--is a key factor for Joe when evaluating whether a discussion has been effective.

3. Students own the discussion. When evaluating class discussion, Joe asks himself the question, "Whose discussion is this?" (J.P. IV#3, p. 2). He wants students to feel they have some control over what happens in the discussion. Joe does not believe that students need to select what is being discussed to have ownership; rather, students must know they have some control over and responsibility for what direction the discussion takes.

4. Students are prepared for the discussion. Effective discussions involve student preparation of common content. "I think the quality of the discussion really is about the degree to which kids are prepared to actually do some careful thinking about what's going on. Or at least share the common content in some way" (J.P. IV#3, p. 3).

5. If the teacher is facilitating the discussion, she/he does so actively to encourage knowledge construction based on probing and critical challenge. Joe does not believe that an effective discussion must have a facilitator; when a person takes that role, however, Joe favors relatively active facilitation. Through the use of prompts and challenges, teachers who are facilitating discussions can help students construct new knowledge.

6. Students interact with one another, extending ideas, questioning, and probing to create new meaning. Whereas some teachers believe that equality of participation is more important than what students say, Joe is concerned with interaction among students, and between students and the teacher, in order to both develop the critical thinking skills of individual students and to create new meaning derived from mixing ideas. Simply sharing ideas, without any transformation of those ideas, does not constitute effective discussion. Evaluating one of the videotaped discussions we viewed, Joe used a baking metaphor to express this belief:

> But, it's almost like . . . now, on the counter, she's [referring to the teacher on the tape] got yeast, and flour, and water, and honey, and mixing bowls, and all that kind of stuff. And the question is whether or not they're going to stay ingredients. You know, whether it's going to become bread that will get leavened up and become this really, you know, a whole bunch of fermentation of those ideas will go around. (J.P. IV#3, p.4)

Summary of Joe's Conceptions. How Joe conceptualizes what he wants students to learn from his social studies classes is remarkably congruent with what he thinks effective democratic citizens should know, be able to do, and be disposed to do. In both cases, Joe puts a premium on thoughtful participation, informed by an understanding of, and appreciation for, democratic values, such as justice. His conceptions of CPI discussions highlight an important link between teaching students how to participate effectively in discussions of important public issues and the necessity for thoughtful discussion of such issues in a functioning democracy. In defining the important characteristics of effective classroom discussion of CPIs, Joe is

most interested in interactions between participants that promote the development of new understandings and meanings.

Selecting Content for Discussion: Issues and Materials

Although some educators view discussion as primarily a content-neutral process, others (Miller & Singleton, 1997; Parker, 1995) emphasize the importance of carefully selecting the content to be discussed. Emphasis on content selection raises two questions: Who decides what content (i.e., issues) will be discussed? What criteria are used to select the content? Once an issue is selected, a third question is introduced: What materials will students interact with to prepare for CPI discussions?

Joe includes controversial public issues in many of the courses he teaches. Students are infrequently given a choice about the issue to be discussed because, "One of the responsibilities that an adult has in a classroom is to determine, amongst all the various contents and things, which ones are the most important to use." (J.P. IV#6, p. 3). Given that Joe is the adult in the classroom who selects issues for discussion, what criteria does he use when making those decisions?

These criteria emerged as Joe explained how he selected the nine Supreme Court cases students discussed in the Important Supreme Court Cases class and as he sorted a list of controversial public topics into three piles: those he would include in his curriculum, those he wouldn't, and those he was unsure about (see Chapter Three and Figure 7). Joe selects issues based on the extent to which they are about social justice, are relevant to students' lives, focus on important content, and fit into his curriculum.

CPI Topics	Joe's decision	Joe's reasons
Abortion	rarely include	issue is too complicated and sophisticated
Affirmative Action	yes	connected to curriculum, human rights connection
Balanced Budget	no	phony issue, it's about manipulation of numbers
Gay Rights	yes	teach as a civil rights issue, not a CPI
Immigration	yes	connected to curriculum, human rights connection
Legalizing Drugs	no/maybe	plays into students' culture too much
Physician-assisted suicide	no	doesn't fit into his curriculum
Trade policy	yes	connected to curriculum, human rights connection
Vouchers for private schools	yes	connected to curriculum
Welfare Reform	yes	connected to curriculum, human rights connection

Figure 7. Joe's Decisions and Reasons About Which CPIs To Have Students Discuss

Social Justice. In selecting issues for discussion, Joe favors issues that call for the analysis of how the "words of our democracy play out against our . . . actions. You know what we say we believe versus what everybody is so proud of on the Statue of Liberty, versus what's in Proposition 209" (J.P. IV#6, p. 9). Defining social justice largely as a quest for equal rights, Joe distinguishes between the goal of equal rights, which is not a controversial issue in his schema, and specific questions about the means to achieve that goal, which are controversial issues. For example, whether minorities should have equal rights is not a controversy, but whether affirmative action is an appropriate way to achieve that goal is a controversial issue.

Relevance to Students' Lives. A second selection criterion Joe employs is relevance to the lives of his students. In the Important Supreme Court Cases course, Joe included three cases that dealt with students' rights in schools. Experience with these cases had taught Joe that students found them particularly interesting because they were "related to their lives; the differences of being kids in school and what those rights are, versus being adults out, or even kids, out in the community and what those are" (J.P. IV#6, p. 2). Joe's interest in selecting issues that are relevant to his students is tempered, however, by concerns about "playing into their culture too much" (J.P. IV#6, p. 9). Explaining why he placed legalizing drugs in the "maybe" pile, Joe said "It's too high school . . . I've got limited amounts of time to work with kids. It's not the issue that I want to deal with them. It's sort of playing into their culture too much, in some ways" (J.P. IV#6, p. 9).

Focus on Important Content. Although Joe opposes the notion of a core curriculum, he does think that some issues involve content that is especially important, and selects issues with that criterion in mind. For example, in the Supreme Court course students studied a famous 1948 case, Terminello v. Chicago, that broadened the First Amendment rights of speakers whose words spark an angry crowd. Joe selected this case because it is a "pretty important foundational free speech case" (J.P. IV#6, p. 2). Joe refers to others cases as "classics" by way of explaining why they were selected: "I don't think a kid should walk out of school without knowing something about Tinker v. Des Moines School District. You know, in terms of civic education" (J.P. IV#6, p. 3). Joe's concern with externally imposed content standards influences how he explains what makes content important. Consider the following exchange, which followed Joe's explanation of why

the Terminello and Tinker cases were important for students to learn: "Hess: So, they're kind of core content? Joe: You wouldn't get me to say that. [Both laugh]" (J.P. IV#6, p. 3).

Another determinant of whether issues are important is whether they are currently matters of public debate. For example, Joe selected the flag burning case Texas v. Johnson because Congress is currently considering an amendment to the United States Constitution that would effectively overturn the Court's decision that flag burning was a protected form of free speech. Labeling this "a good hot case," and another on censoring the Internet (Reno v. ACLU) a "good recent case" that had "been in the press" (J.P. IV#6, p. 3), Joe communicated that issues can be judged important because they are currently matters of public concern.

Curricular Fit. Joe selects issues for his students to discuss with an eye toward curricular fit. All things being equal, Joe seeks issues that are directly related to the topics of the courses he teaches. Given that Joe's high school has no required social studies classes and that he has tremendous flexibility in creating courses to offer as electives, this selection criterion is probably less constraining for Joe than for teachers in other contexts.

Selecting Materials. In the Supreme Court cases course that was the primary focus of this study, Joe assigned the texts of entire Supreme Court cases as the materials for classroom discussion. Even though Joe believes more than half of the students in the class are reading below grade level, and Supreme Court decisions are written in complex language, he says, "If you give kids real stuff, and treat them as real learners, they will do amazing things" (J.P. IV#6, p. 4). By scaffolding instruction on the difficult materials, Joe seeks to provide all students with access to the "high stats [status]

curriculum that's given to AP [Advanced Placement] and IB [International Baccalaureate] students because [mockingly] 'they can handle it' . . . everybody can handle that if they're led through it and there's support for it" (J.P. IV#6, p. 5).

Materials selection is informed by Joe's conception of democracy and the connection between democracy and schooling. He is most interested in providing difficult and authentic materials to students to enhance equality.

> I really want to emphasize that, if we live in a democracy, not only the brightest three percent should be reading Supreme Court cases and understanding what the nature of the Supreme Court is, it seems inherently anti-democratic to me. And what do we do to those other kids? We give them worksheets on, there are a blank number of justices on the Supreme Court. (J.P. IV#6, p. 6)

Summary of Joe's Approach to Selecting Content. Joe believes that the teacher is responsible for selecting which issues should be discussed in the classroom. He selects issues based on the extent to which they are about social justice, are relevant to students' lives, focus on important content, and fit into his curriculum. Joe selects materials for student use that are difficult and authentic.

Snapshot of a Seminar

During the 1997-98 school year, Joe taught a nine-week course focusing on historically significant controversial public issues related to freedom of speech and press. Joe designed the course to rely on Supreme Court cases because "you would be hard pressed to find a more authentic text than a Supreme Court decision" (J.P. IV#4, p.3). Although all nine of the cases the students read are about the First Amendment's speech and press clauses, Joe

also hopes his students will gain a general understanding of content that extends beyond this amendment. On the first day of class he said to his students: "I want to grow old in a society that has many people understanding the way the Constitution and the Supreme Court works" (AT#1).

The 24 students enrolled in the course met three times per week, for a weekly total of four hours. During most weeks the grade 9-12 students read one First Amendment Supreme Court case, prepared to participate in a seminar discussion on the case by completing a pre-discussion assignment called a "ticket," worked in small groups to review the facts of the case, participated in a seminar, and wrote an issues-analysis paper.

The seminar model of discussion that Joe used in the Supreme Court course is pervasive throughout New Horizons High School. Joe learned the model from the school's principal (who participates in the seminars as a model participant and, on occasion, as the facilitator) and has been using it for several years. The model, labeled simply "seminars" at the school, is text-based large group discussion designed to help participants develop a deeper understanding of the issues, ideas, and values in the text (Gray, 1989). Joe favors the model because of its potential to enhance critical thinking and the generation of new ideas.

Preparing for the Seminar. This day's seminar focuses on the Supreme Court's decision in New York Times Co. v. United States, the famous "Pentagon Papers" case decided in 1971. This case focuses on the tension between freedom of the press and national security. Joe's students read the 50 pages of the case and completed a ticket (a pre-seminar assignment) to participate in the seminar discussion. The ticket for this case required students to create and complete a data retrieval chart that identified the basic

arguments made by each of the justices in the nine separate opinions issued in the case. Joe's ticket assignments require students to read and interact with the text. Joe does not expect, however, that the ticket will cause students to understand the text: that is the purpose of the seminar discussion. The day before the seminar, the students work in small groups to figure out the basic facts of the case and how it moved through various courts to be heard by the United States Supreme Court.

As the students enter the classroom on seminar day, Joe checks whether their tickets are completed. Graded as a pass if completed and fail if not completed, the tickets also determine who may participate in the discussion. Students without a ticket are not allowed to sit in the circle, even if they say they have done the reading. Instead, they are assigned an observer role and must sit outside the circle and take notes on who is participating. Joe encourages students to complete the ticket by saying, "It's preferable to sit in the circle and choose not to participate, rather than be on the outside and not be able to participate" (J.P. IV#4, p. 6). Requiring students to complete the ticket is one way that Joe attempts to deal with the difficulties involved with talking across difference:

> The only thing that we know we have in common in a seminar is the text that we share in common. We've been raised differently. We have studied different materials in this class. We may have had U.S. History classes, others have not had U.S. History classes. All sorts of things. But what we do know is that we all have the text in common. A good discussion, a good seminar, begins from the premise that we are talking about a shared text. (J.P. IV#4, p.7)

Setting and Maintaining Guidelines for the Seminar. As the two-hour class period begins, Joe, 19 students, and the school's principal (acting as a model seminar participant) are seated in the seminar circle. One student who didn't complete the ticket is creating a list of participants, with check marks for each time they talk. Before the seminar begins, Joe reminds the students to work hard and to "do the work of the seminar" (AT#9). By that Joe means adhering to the guidelines created by the students at the beginning of the course; these guidelines are now posted on butcher paper on the classroom wall. Some of the guidelines written on the poster are "listen, respond to ideas out there, make the agenda yours, and refer to the text" (ART#1). In a later interview, Joe explained what he meant by that exhortation to his students: "I think it was a phrase they all understood . . . that doing the work of the seminar is using the behaviors [required in seminars]; is working hard with the text; it's living with ambiguity" (J.P. IV#5, p.1).

Focus Questions to Begin the Seminar. Joe begins the discussion with a focus question: "What was the most compelling argument in the case?" (AT IV#9). Joe has developed this focus question using specific criteria: it cannot be answered without using the text; it is open-ended in that there is no right or wrong answer; and it is a question about which he, as the seminar facilitator, has some genuine curiosity.

A student immediately responds to Joe's focus question by changing the question. "Well, I can tell you the least compelling argument" (AT#9). The student then points the class to a part of Chief Justice Burger's dissenting opinion that laments the short amount of time the Court had to spend on the case and says, "he is just whining here" (AT#9). Later, I asked Joe why he didn't direct the student to stick with the question that was asked. Joe

responded, "That's a no brainer. Just because I asked a question, doesn't mean that I asked the right question. . . . Just because I was fishing for trout doesn't mean that I'm going to ignore the bass that bites" (J.P. IV#5, p.3). Moreover, Joe believed the student's response accomplished the primary purpose of his focus question, to open a door to the text in a way that will focus students on the reasoning of the justices.

Referring to the Text. None of the other students comment on Burger's reasoning; after a short pause, several chime in and say that Justices Douglas and Black had particularly compelling reasons supporting their opinions. Joe asks the students to find where the Douglas opinion begins, and they turn to a specific page in the text. Joe immediately probes with a question to one of the students who liked the reasons of Justice Douglas, "Betty, what was your sense of what Douglas was arguing?" She responds by paraphrasing the position Douglas takes in his opinion. Joe follows up by labeling Douglas's reasoning, "So he was a First Amendment absolutist?" Students agree and Joe follows up again, "Talk to us more about Douglas's arguments." Another student responds with an elaborated description of why he finds the arguments compelling. This type of interchange continues for several minutes. Students refer to the text and talk about the basic tenets of the two First Amendment absolutists. During the opening several minutes of this seminar, Joe asks quite a few questions, continually reminding students to find a specific part of the text they are talking about.

Taking Minority Views Seriously. During the seminar, it becomes apparent that most of the students support the opinions of the court majority, which held that publishing the Pentagon Papers was protected by the First Amendment. One student, however, takes the contrary position. Joe then

says to the class, "There's our lone conservative, this time. We actually don't have to support Logan, but let's . . . pretend to do so for a minute. Okay? Let's try to construct and give credence to the argument of the government in this case" (AT#9). For several minutes the seminar continues with students identifying parts of the dissenting opinions that represent the view of the government and explaining what they think those arguments mean. Later, Joe explained why he refocused the seminar on the arguments that did not have the support of most of the students.

> I think a real important critical thinking skill is the ability to take a different position and to argue it with credence and credibility. I think it's an incredible skill for citizens, for enlightened citizens in a democracy, because it's rare that issues are completely black and white. It's important to give minority voices a really serious airing in a classroom. Because then people will give their true opinion. I think it's also real important to have kids take on different viewpoints as a way of better understanding their own viewpoints. So related, back to the earlier question, that's about doing the work of seminars. Doing the work of seminars is trying on ideas. (J.P. IV#5, p. 5)

Investigating the Meaning of Words. Several times in the seminar students do not understand the meaning of words in the text. The first time this occurs, Joe says, "Let's look it up; here's the dictionary" (AT#9). A student then looks up the word and reads the definition to the class. Later, when another word is not understood by several students, a student says, "I already looked this up last night; it means . . ." (AT#9). Throughout the seminar, it is apparent that students believe that the meanings of words matter; they are willing to stop the seminar to make sure they understand a word.

Moving to Moral Issues. The Pentagon Papers were stolen government documents, a fact that becomes the focus of conversation toward the end of the seminar. Joe asks the students, "So, what should the New York Times have done when Daniel Ellsberg came to them with boxes of stolen government documents? If Logan steals a TV and gives it to me, and I know that he stole the TV, have I done something wrong?" Several students exclaim, "Yes." Joe asks, "Is that the same thing as the New York Times did with the documents?" A student replies, "They didn't know." Another counters, "Oh yes, they knew." A third says, "But they thought the public had a right to know." Joe now takes them back to the text: "Doesn't one of the justices say something to the effect that there is this right to know right now and the New York Times feels a responsibility to provide that information? Who said that?" After a few seconds of looking, someone shouts, "page 749," and Joe then reads an excerpt from that page. Joe then says, "You guys, most of you believe, that what the Supreme Court did was right in this case." Several students say, "Yes." Joe continues, "Did the NYT do the right thing?" A student responds, "In my opinion, it's just a matter of your opinion, more important that the public know – they did what they needed to do and I agree with them." Another student says, "I agree, it's like this pull--they were publishing stolen documents which was basically not the right thing to do, but yet it was important to let the public know what the government was doing. I have a question, did anything happen to the New York Times as a result of this?" Joe answers, "The New York Times was fine, Daniel Ellsberg was tried for taking the Pentagon Papers--do you want to know now or later what happened to Daniel Ellsberg?" One student says, "Now, right now, Joe." Another jokingly adds, "We have a right to know."

This excerpt from the seminar illustrates a move frequently made by Joe to use the text to spark discussion of larger moral questions, in this instance, Is it ever right to steal? The text still reigns supreme, however, as can be seen in Joe's reference back to the text.

Seminar Participation. During the hour-long seminar, the student observer counts 150 different contributions, 104 (or 70%) made by seminar participants and 46 (or 30%) by Joe. Of the 19 students in the seminar circle, 13 verbally participated. However, 20 of the 104 statements were from the principal. Comparing this seminar to the eight others in this class, the overall participation numbers are fairly constant. Joe talks quite a bit, although most of his participation is in the form of questions to the students.

Student Critiques of the Seminar. While students are not required to participate orally in the seminar discussion, they are required to share their critique of the discussion during a debriefing period held immediately after the seminar ends. This particular seminar was the final one in the nine-week class. In a celebratory manner, Joe begins the debriefing session with the statement, "Give yourselves a round of applause, you guys got this thing" (AT#9). Following enthusiastic applause, one student exclaims, "I was terrified when I first saw it" (AT#9). Joe then says, "Regarding this seminar, I would like to know what your sense of this seminar was as it compared to others and on its own merits" (AT#9). A student volunteers:

> I'll start. . . . I just thought this was a really comfortable seminar, not a lot of people talked, but those people who did really knew what their ideas were about the case, and that helped me, a person who didn't understand it a whole lot, to get a better sense of it all. I enjoyed the

> relaxed energy of it because it made it a lot more easy to get into. (AT#9)

Although many students agree the seminar had a relaxed pace, views about the text differ. Some students liked the text, but a few others said it was confusing or worse. One student plainly states, "This text sucks" (AT#9). Another student critiques her own participation in the seminar:

> I finally completed my goal, which was to not talk during the seminar. I kept wanting to talk because I think this case was very confusing, but the seminar cleared it up. But I thought it was pretty good, but it is kind of weird trying not to talk, I think I listen more when I am talking because I listen in order to respond. (AT#9)

Throughout the critique, Joe says very little. But, when his turn comes as the critique moves around the circle, he says: ". . . the coolest thing about this seminar was the opportunity to read this case because I have known about this case for a very long time. I found the ticket helped me a whole lot in terms of organizing nine separate opinions" (AT#9).

Joe described the purpose of the seminar critique as "feedback loop for all of us"(J.P. IV#5, p. 14) and a way to enhance the students' abilities to be reflective:

> It's a whole metacognitive process that I think is incredibly powerful for kids. It also, again, goes to that notion of power and voice and the politics of the classroom that I value by holding a space for them to reflect on what's going on--which says . . . whether or not it's my classroom or their classroom. My conversation or their conversation. And this is very, very concrete evidence that it's our conversation. (J.P. IV#5, p. 15)

Joe's Views on Assessment of Seminar Participation. Joe distinguishes between informally assessing the quality of students' participation in seminar discussions and formally assessing, or grading, that participation. He supports the former and is resolute about not doing the later. While Joe does provide each student with oral and written feedback on their seminar discussion skills, he does not factor in their participation in seminars into their course grades. Joe believes the authenticity of the seminar would be harmed if students were being graded on their verbal participation. The unique nature of seminars as discussions aimed to collaboratively create meaning makes the grading of individual's verbal participation problematic for Joe:

> Seminars are about public performance[,] . . . are explicitly about collaborative work. Seminars are about coming together in a public space to interact with people who are and are not like you. And that is freighted with different things, not less important, but different things than individual writing assignments [which Joe does grade]. Not the least of which is the public performance aspect of it. And the fact that we're trying to make meaning together. If I want to make meaning together, I want only the contribution of authentic ingredients. (J.P. IV#5, p. 4)

Another reason for Joe's refusal to grade seminars is his belief that students participate in seminars in various ways. "You see, I like it when kids speak in seminars. I've learned to get just as big a kick out of kids who say, 'I really, God I was there the whole time. I just didn't have anything to say because I was listening so intently and trying to figure stuff out.' So I value being in seminars in a variety of ways" (J.P. IV#5, p. 7). Joe believes it is impossible to

create an assessment rubric for seminar participation that would honor the various ways that students participate in the discussions.

Summary of the Seminar. The seminar discussion on New York Times Co. v. United States typifies how seminar discussions work in Joe's classes. Before a seminar, students read a text, complete a ticket, and work in small groups to become more familiar with the basic facts of the case. On seminar day, guidelines for seminar behavior are reviewed, and Joe then begins the seminar with a focus question. Although there is often disagreement about what the text means and what its implications are, the conversation is rarely heated or adversarial. Instead, the pace is relaxed and the tone is highly civil. Frequent references to the meaning of specific words and to the text pepper the conversation, although both Joe and the students use non-text metaphors and analogies. Each seminar ends with a critique in which the verbal participation of all students is required. The critique focuses not on the issues or ideas in the text, but on students' reactions to the text and their opinions about the quality of the seminar. While Joe informally assesses his students' participation in the seminars, he believes that grading students' verbal participation would decrease the authenticity of the seminars. After most seminars (but not this one because it was the end of the class), students are required to write a paper about the issues in the text.

The next chapter of the dissertation interprets the conceptions and practices of Joe and the other two teachers who participated in the study by presenting an initial theory that describes and explains seven propositions. Before this theory is presented, however, portraits of the two other teachers' CPI discussion conceptions and practices are described and explained.

Portrait Two: Elizabeth Hunt at Summit Ridge Middle School

Elizabeth Hunt always knew she wanted to be a teacher: "I can remember even when I was in third grade . . . having an excellent teacher and thinking, 'Oh, what fun this would be'" (E.H. IV#1, p. 4). Elizabeth now teaches social studies to eighth-graders at Summit Ridge Middle School, located in a suburb of a major city in the West. Elizabeth has taught for seventeen years. She taught language arts for twelve years but switched to social studies five years ago for the challenge of teaching different content.

Elizabeth earned a B.A. in Middle School Education at a university that has one of the nation's largest teacher certification programs. At the time Elizabeth attended the university, it was one of only two in the country that offered a special certification in middle school education. Within that program, Elizabeth specialized in social studies and language arts.

Since receiving her B.A., Elizabeth has taken numerous graduate-level courses in language arts and social studies. The past several summers she has attended institutes focusing on civic education. As part of the 1996 institute, she took a course about teaching with discussion that I taught.

Since the 1996 institute, Elizabeth has participated in a study group with other middle school social studies teachers. The study group gathered once a month to discuss how to improve social studies teaching and learning. Facilitated by social studies experts from a nationally recognized social studies think tank, the study group meetings initially focused on improving classroom discussion. For each study group meeting the teachers brought representations of their students' experiences in classroom discussions (such as videotapes) and using a "show and ask" format, solicited feedback from their peers. During the past year, the focus has broadened beyond classroom

discussion to encompass a range of issues related to teaching social studies. For example, several recent meetings centered on how to teach such important social studies concepts as federalism and executive privilege.

Contexts

Similar to the CPI discussion conceptions and practice of Joe Park, Elizabeth Hunt's work is informed by the various contexts in which it is situated. Her community, the school, her students, and the curriculum all shape how Elizabeth approaches teaching her students to participate more effectively in CPI discussions.

Community. The community in which Elizabeth lives and teaches is typical of the suburban sprawl found outside many major cities in the United States. Primarily residential, it is a "middle to upper middle class, predominantly white community" (E.H. IV#1, p. 5). Many adults in the community work in a nearby technology center, which is often labeled the "new downtown" (E.H. IV#1, p. 5) to distinguish it from the downtown of the major city it borders. Elizabeth's community is a new one--still working to create its traditions and institutions. One institution that both reflects and creates the community is the schools. Many people, in fact, moved to this community because of the excellent reputation of the school district in which it resides. Parents in the community are supportive of education, and want to be involved in the schools. One way adults show their support is to vote for increased funding to build new schools. Summit Ridge Middle School is only six years old, and new middle and high schools are under construction.

Summit Ridge Middle School. Elizabeth's school is huge, with more than 1600 students in grades 6-8. The wealth of the community is immediately apparent when touring the school. Its library, for example,

overflows with computers, expensive research materials, and new books. Students are placed in three separate "communities" organized around the classic middle school team concept. The "communities" operate like schools within a school, each with a separate administrative team and teaching staff. Students stay in the same community from sixth through eight grade. Teachers work in teams to facilitate communication about individual students and to coordinate the curriculum. Additionally, subject matter teachers communicate across the teams to ensure some commonality in the curriculum.

Students. Summit Ridge Middle School is a neighborhood school and, as such, is filled with students who run the gamut in terms of interests, capabilities, and goals for the future. While the overall student population is not very diverse in terms of race (approximately 90% of the students are white), it is diverse in other ways. For example, as the school district's designated school for students with multiple physical disabilities, some students are dealing with serious physical difficulties. Other students have learning and behavioral challenges. For example, Elizabeth has students in her classes who read significantly below grade level, in some instances as low as the first or second grade levels. In this school that follows a special education inclusion model, some of the students in Elizabeth's classes have individualized education plans (IEPs) and receive supplemental services from special educators.

Most of the students in Elizabeth's classes, however, are not dealing with serious physical or learning challenges. Quite the contrary. They are reading at or above grade level and behave in age-appropriate ways. Moreover, the students seem to get along with one another. Walking

through the hallways, one hears few taunts or bursts of anger. Instead, the warm atmosphere seems to reflect the many posters on the hall walls encouraging students to respect one another, take learning seriously, and celebrate diversity.

Curriculum. The required course that all eighth-graders at Summit Ridge Middle School take is officially labeled "American Studies," and focuses on United States history and civics. Using the state's civics standards for direction, teachers from the five middle schools in the district recently developed a common curriculum for the eighth-grade course that mandates beginning with a unit on the American Revolution, then spending more than half of the school year on an in-depth study of the Constitution and civics. Within each school, the curriculum varies after the Constitution and civics unit is completed. At Elizabeth's school, the eighth-grade teachers in the three communities fill out the curriculum with the study of political parties, Westward expansion, and the beginning of the Civil War. Thus, unlike Joe, who has total control over the content of the courses he teaches, Elizabeth teaches a common curriculum that she has developed with other teachers in her school and district.

Summary of Contexts. Elizabeth is an experienced middle school teacher who works hard to continually improve her practice. Through taking courses and participating in a unique study group, she demonstrates a commitment to reflective practice. The young and growing community Elizabeth teaches in values education, financing the building of new schools and purchasing expensive educational resources. Elizabeth teaches in a new school, a large middle school organized into three separate communities. As the eighth-grade social studies teacher in one of the school's communities,

Elizabeth teaches a district-wide curriculum she helped to develop that focuses on 18th- and 19th-century United States history and civics. While these various contexts influence how Elizabeth teaches her students to participate effectively in CPI discussions, her conceptions of a number of factors also inform that practice.

Conceptions

Recall, in this study I am defining conceptions to include a teacher's knowledge, beliefs, thoughts, and images that may inform classroom practice (Shavelson & Stern, 1981). In this section, I focus on Elizabeth's conceptions of four factors: the purposes of social studies; the characteristics of effective democratic citizenship; the rationales for CPI discussions; and the characteristics of effective CPI discussions.

Conceptions of the Purposes of Social Studies. When drawing a concept map to illustrate her thinking about the purposes of social studies, Elizabeth created a wheel, with "participatory citizens" (E.H., ART) in the center. This visual display represented the primacy she attaches to the citizenship goal: "I'm hoping that I'm teaching students to be effective citizens" (E.H. IV#1, p. 18). Toward that goal, Elizabeth identified six outcomes of social studies: knowledge of the development, structure, and function of the United States government; understanding of how historic events and policies have shaped the United States; informed about domestic and foreign events and relations through current events; civic virtue; civic discourse; and preparation for future voting. In combination, Elizabeth conceptualizes these six outcomes as leading to a greater likelihood that her students will be effective citizens.

The first two are knowledge goals, one focusing on civics (structure and function of the government), the other on history. Elizabeth sees a

connection between understanding the development of constitutionalism in the United States and future citizenship participation: "I think if they have good background knowledge about how our government works they may be more likely to participate in it" (E.H. IV#1, p. 12). In particular, she wants her students to know how to effect change and that "they do count" (E.H. IV#1, p. 13).

Important historical events and policies are also crucial for students to learn, especially if "it's something that is still ongoing" (E.H. IV#1, p. 13). For example, Elizabeth teaches a unit on "Turning Points of Equality" that focuses on significant milestones related to equality in United States history. Defining equality broadly, Elizabeth explains the rationale for the unit: "I think when we talk about civil rights, a lot of people think we talk just about the black minority. But I think that there are other minorities in our country that are still striving for equality. . . . I put a big emphasis on that because I want them to be aware . . . that we're all struggling for civil rights" (E.H. IV#1, p. 13).

Elizabeth's emphasis on teaching about contemporary domestic and foreign relations issues through current events is aimed at helping her students develop a habit that she hopes they will carry into adult life: to "want to read the newspaper and watch CNN, other news programs--to be informed" (E.H. IV#1, p.11).

After arming her students with knowledge about government structure, and historic and contemporary events, Elizabeth wants them to demonstrate civic virtue, defined as "putting the good of the community ahead of their own individual needs" (E.H. IV#1, p. 12). Demonstrating civic virtue occurs through volunteerism; "It can be at a very simple, basic level, where you are involved in a neighborhood association in your community to

keep it a nice place or be involved in your church" (E.H. IV#1, p.12). Regardless of venue, the purpose of exercising civic virtue is the same: to "help our nation be a better place" (E.H. IV#1, p.12).

Two additional purposes that Elizabeth holds for teaching social studies are to teach students the skills to engage in civic discourse and to prepare them for their future roles as voters, both of which she detailed when describing what she thinks effective citizens in a democracy should know, be able to do, and be disposed to do.

Conceptions of Effective Democratic Citizenship. Elizabeth's thinking about effective democratic citizenship is remarkably similar to how she conceptualizes the purposes of social studies. Effective citizens must possess background knowledge about how the government works and believe that a common good should take precedence over their individual needs. Moreover, such citizens should engage in certain activities, including public discourse, voting, local political activity, and volunteerism in the civil sector.

Elizabeth cited many reasons for supporting the inclusion of public discourse in her definition of effective democratic citizenship. Engaging in public talk with others is a way to become informed and to communicate that you are informed. It is a way to both learn about issues and influence the opinions of other people. Public discourse also has social utility: "I think in social situations, lots of people talk about current events, about issues that matter to them, and it's fun to be able to participate in those kinds of discussions" (E.H. IV#1, p. 16). While many of the other behaviors embodied in Elizabeth's conception of effective democratic citizenship are done occasionally, public discourse is a daily activity. After reading the newspaper

or listening to the news, effective citizens, Elizabeth believes, should talk about what they have learned:

> When things come up on the news or in the newspaper that you're really curious about, are really interested in, or that make you really angry, you talk about it . . . and when you have those discussions, I think it helps you to flesh out your opinions. And if you have to defend your opinion with things that you've read or things that you've heard, I think it helps you to understand the issue better yourself. (E.H. IV#1, p. 17)

Although Elizabeth urges daily public discourse, she also thinks that effective citizens should engage in activities that communicate their views to a broader audience, including policy-makers. Citing a recent example from her own life, Elizabeth explained how postcards were distributed at her church so the congregants could write messages to legislators urging a ban on partial birth abortions. Although the proposed abortion legislation was federal, Elizabeth believes citizens have more influence at the local level because the smaller numbers make it more likely that citizens' concerns will be heard. Regardless of the level of government that one attempts to influence, Elizabeth places a primacy on voting. Responding to my question about the oft-stated view that voting does not matter, Elizabeth said, "Well, there have been elections in our nation's history where the votes were very close. And I think that if everybody had that attitude, that one vote doesn't matter, then it would make a huge difference. I think that voting does matter" (E.H. IV#1, p. 19).

Effective citizens, however, should not limit their activities to influencing political issues. Echoing political theorists who argue for the need

to re-establish a stronger civil society in the United States, Elizabeth said that people need to participate in their community in order to feel a connection to the community. By joining a local homeowner's association, involvement in local church activities, or similar activities, effective citizens work to build a healthier civil society.

Conceptions of the Rationales for CPI Discussions. Throughout the school year, Elizabeth's students participate in at least one full-blown CPI discussion each month. Given that preparation for the discussions can take a few class periods, and the discussion and debrief a few more, this is a major commitment of instructional time. Elizabeth has numerous rationales to support this emphasis on CPI discussions, including democratic citizenship goals, critical thinking, interpersonal skills, and better understanding of issues (see Figure 8).

Rationales for CPI Discussions	Do CPI discussions accomplish this goal?	Reasons why CPI discussions do/do not accomplish this goal:	Caveats and/or further explanation
Improve interpersonal skills	Yes	Students get practice exercising key skills, such as listening and yielding	This is one of the most important reasons Elizabeth uses CPI discussions
Understand important democratic values	To some extent	If a CPI is value-laden, then students will learn about it through discussion	Some CPIs (but not all) emphasize conflicts between values
Participate in political life of democracy (e.g., voting, serve on jury)	Perhaps in the long term	Because students want to participate in discussion, they watch the news and read the paper	Broaden the definition of participation to include being informed
Learn important social studies content	Yes	Issues are chosen that link to important content	Especially helpful to students with learning difficulties
Learn how to think critically	Yes	Students are exposed to the ideas of others; this stretches their brains	Improving critical thinking skills is a major rationale for Elizabeth
Improve democracy	Yes	Talking helps democracy--exposes people to multiple perspectives	Discussion can cause action, also function as a democratic safety valve

Figure 8. Rationales for CPI Discussions: Elizabeth Hunt
(Note: Elizabeth's responses are from E.H. IV#3, pp. 1-14.)

When asked to rank rationales for including CPI discussions in her curriculum (see Chapter Three), Elizabeth placed the connection between such discussions and a healthier democracy at the top of the list. Her primary explanation of this choice is rooted in the First Amendment: "The First

Amendment is what allows us to have, or helps us maintain a healthy democracy; that people are able to talk about it, are able to criticize the government, are able to get their feelings out" (E.H. IV#3, p. 8). By providing her students instruction and practice in how to participate in public discourse about CPIs, Elizabeth believes she is helping them learn how to exercise their First Amendment right to speak. While Elizabeth distinguishes speech from action, she does think that speech can cause action: "When people can talk about these things [CPIs], it can initiate action" (E.H. IV#3, p.8).

The other rationales Elizabeth holds for CPI discussions are more clearly rooted in the unique needs of middle school students. Because students of this age are beginning to think on a more abstract level, Elizabeth believes CPI discussions can improve their critical thinking skills. Asked to define critical thinking skills, Elizabeth responded: "Stretching their brain beyond where they would if it was just left to them to do themselves, looking at Bloom's Taxonomy and reaching for the higher levels like synthesizing and evaluating. I think that public issues discussions really foster those skills" (E.H. IV#3, p.9). The key element in CPI discussions that enhances critical thinking skills is that students hear and are challenged by viewpoints different from their own.

Another rationale Elizabeth holds for CPI discussions, improving interpersonal skills, also stems from the needs of middle school-age students. Heading a long list of specific interpersonal skills that students need to develop is "simple manners," defined, at the minimum, as calling other discussants by their names. The other skills are more complex and difficult to develop, including teaching students to yield the floor, not dominate the discussion, and listen carefully to one another.

Finally, Elizabeth includes CPI discussions in her curriculum because she believes they are a useful tool for helping students, especially students with learning difficulties, better understand social studies content. Even if students with learning difficulties only participate by listening, Elizabeth thinks the discussions help them learn more: "A lot of times they struggle with the content of what we're studying. But the discussions really help them to understand it better" (E.H. IV#3, p. 12).

Conceptions of Effective Classroom Discussion. As Elizabeth and I viewed two videotaped excerpts of classroom discussions (see Chapter Three), it was immediately apparent that she, like Joe Park, had well-formed ideas about the characteristics of effective classroom discussion. These characteristics are normative beliefs about what Elizabeth thinks should be present in the following categories: seating, interpersonal skills, facilitator's role, critical thinking, interaction, preparation, and the role of prior knowledge. Through brief descriptions of Elizabeth's conceptions of each of these categories, an overall picture of what she values in class discussions will emerge.

1. Seating. Where students are seated in relation to one another and to the teacher is a cue to Elizabeth about the kind of discussion one can expect to hear. She recommends seating discussants in a circle. When asked about what a circle produces that rows would not, Elizabeth replied, "Eye contact is the biggest thing. And the feeling among the students that they're speaking to each other and not to the teacher" (E.H. IV#2, p. 14).

2. Interpersonal Skills. Effective discussions are characterized by discussants demonstrating interpersonal skills. Elizabeth is interested in a highly civil atmosphere in which discussants listen carefully, yield the floor,

make eye contact with one another, exhibit patience, and make one another feel a part of the discussion. One way that students who are reluctant to participate orally can be encouraged is by other students consciously trying to draw them into the discussion, either with a broad invitation, "What do the rest of you think about . . .?" or a more directed overture, "John, what is your opinion about . . .?"

3. Facilitator's Role. Elizabeth distinguishes between teachers who are discussion leaders and those who are discussion facilitators, and favors the latter role. Discussion leaders direct the flow, pace, and agenda of the discussion, whereas discussion facilitators open the discussion up to where students want it to go. If a discussion gets off track, a facilitator will intervene, but the students generally have more responsibility for the content of the discussion than in one with a discussion leader. Elizabeth favors discussions with facilitators because leaders so often preempt what would come up naturally if left to germinate.

4. Critical Thinking. Effective discussions are not merely sharing information, but places where critical thinking about the information being shared is exhibited. Elizabeth wants students to challenge one another, use analogical reasoning, direct questions to one another, and use statements like "Wait, Let me see if I get this right" (E.H. IV#2, p. 18) to seek clarity.

5. Interaction. Not surprisingly, given Elizabeth's preference for the teacher's role as discussion facilitator vs. discussion leader, she places a premium on interaction between students and is wary of teachers who ask too many questions. Instead, Elizabeth wants to hear students talking directly to one another.

6. Preparation. Students who have prepared for a class discussion (through reading or working through pre-discussion questions) are more likely to produce one that is of high quality. Thus, Elizabeth's conception of effective discussion is not talk that occurs "off the cuff," but talk that is informed by preparatory work by the discussants.

7. Role of Prior Knowledge. Elizabeth also views favorably evidence that the students are making connections between what they have learned in the past and the topic under discussion. Commenting on a student's reference to Plessy v. Ferguson to substantiate a point about institutional racism during the discussion of Richard Wright's autobiography, Elizabeth said, "They're doing a great job of making connections between history and literature" (E.H. IV#2, p. 4).

Summary of Elizabeth's Conceptions. Elizabeth's conceptions of the purposes of social studies, the rationales for using CPI discussions, and the characteristics of effective discussion share common themes. Similar to Joe Park, she cites a classic reason--preparation for democratic citizenship--to support what she is teaching her students in their eighth-grade social studies class. The democratic citizenship theme also trumps the many rationales she has for teaching students to participate more effectively in CPI discussions. Another reason Elizabeth has for using CPI discussions is to teach interpersonal skills, which she links to the development of an atmosphere of overall civility. When defining what makes a discussion particularly effective, Elizabeth draws on her rationales for CPI discussions and for teaching social studies. She emphasizes students talking directly to one another, instead of through the teacher, and the use of interpersonal and critical thinking skills in discussion.

Selecting Content for Discussion: Issues and Materials

To this point, the portrait of Elizabeth has focused on her background, contexts, and conceptions of a variety of factors that may influence her conceptions and practice about CPI discussions. Now I turn to a description of how Elizabeth selects issues and materials for CPI discussions. This section addresses three questions: Who decides what content (i.e., issues) will be discussed? What criteria are used to select the content? What materials will students interact with to prepare for CPIs discussions?

Who Selects Issues. In most instances, Elizabeth decides which CPIs will be included in the curriculum of her eighth-grade social studies classes. If students are particularly interested in a current events issue, Elizabeth will sometimes include that in the course. For example, in February 1998, when it appeared that the United States might go to war against Iraq, Elizabeth responded to her students' desire to discuss the issue by including it in the curriculum. Periodically Elizabeth teaches an elective mini-course, Public Issues Discussions, that has as its primary curricular outcome the ability to participate more effectively in CPI discussions. In that course, students select the issues to be discussed.

Criteria for Selecting Issues. Elizabeth uses four criteria to determine which issues should be included in the required social studies course: curricular connection, student interest, anticipated community reaction/student age, and teacher interest/knowledge. These criteria emerged as Elizabeth engaged in an issues-selection task. Recall, this task required Elizabeth to sort ten CPI topics into three piles to reflect what she would include in her curriculum, what she might include, and what she would not

include (see Chapter Three). Figure 9 illustrates Elizabeth's decision-making on issues selection.

CPI Topics	Elizabeth's decision	Elizabeth's reasons
Abortion	maybe, focus on structural issues (i.e., which govt. should regulate?)	community disapproval; generate more heat than light
Affirmative Action	yes	connected to curriculum, high teacher knowledge
Balanced Budget	no	low teacher knowledge; not tied to curriculum
Gay Rights	maybe, focus on structural issues	community disapproval
Immigration	yes	connected to curriculum, ties to current events
Legalizing Drugs	no	students too young, community disapproval
Physician-assisted suicide	maybe, focus on structural issues, more inclined to include if on ballot	community disapproval
Trade policy	no	low teacher knowledge
Vouchers for private schools	no	too complex; not tied to curriculum
Welfare Reform	maybe	not connected to curriculum

Figure 9. Elizabeth's Decisions and Reasons About Which CPIs To Have Students Discuss

1. Curricular Connection. The criterion that most influences which issues Elizabeth selects is the extent to which the issues are connected to the eighth-grade curriculum. This is not surprising, given that Elizabeth believes that CPI discussions can help her students better understand important social

studies content. For example, the curriculum includes a unit on the United States Civil War, which causes Elizabeth to favorably view the inclusion of affirmative action:

> It's [affirmative action] a topic that I would do with students, and I would do it as an outgrowth of slavery, to talk about what are the effects of slavery in our society today. We study slavery during the Civil War and I think that's a good tie-in of a current event. (E.H. IV#4, p. 2)

The way Elizabeth employs this criterion can be illustrated by her thinking about whether she would include welfare reform as a CPI in the eighth-grade course. She indicated that she probably would not include welfare reform because it has no direct link to the curriculum. If her curriculum extended beyond the United States Civil War (to include the New Deal), then welfare reform would be more likely to be included because its historical antecedents can be found in the New Deal period.

2. Student Interest. Another criterion Elizabeth uses when selecting issues is her assessment of whether the issue would be interesting to students. Because the study of current events is such a mainstay of her classes, her students tend to be particularly interested in CPIs that are matters of current public discussion. Student interest as a CPI selection criterion often conflicts with other criteria Elizabeth uses to select issues. For example, Elizabeth knows that her students are interested in discussing many issues that she thinks community members don't want them talking about, either because they challenge community values or because the students are perceived as being too young.

3. Anticipated Community Reaction/Student Age. When selecting issues for class discussion, Elizabeth is enormously attentive to whether the community would support having students discuss the issue. As Figure 9 illustrates, gay rights, legalizing drugs, abortion, and physician-assisted suicide are all issues that Elizabeth would be hesitant to include in the curriculum because she fears community members would disapprove. Her fear of community disapproval is not just paranoia. A teacher in her school who taught a unit on the Salem Witch Trials was castigated by some community members who thought she was encouraging the practice of witchcraft. Elizabeth admits, however, that her fear of community disapproval is entangled with her concerns about whether some issues are inappropriate for her students because of their young age.

4. Teacher Knowledge. Elizabeth is more likely to select issues about which she possesses strong background knowledge. Given the numerous issues that could be selected for her curriculum, she is candid about how her own knowledge influences the selection process. For example, when describing why she probably would not include trade policy issues, Elizabeth said, "I guess I wouldn't include it because I don't know enough about it. It's something that I probably could research and figure out a way to connect [to the curriculum], but I haven't done that yet" (E.H. IV#4, p. 2).

Materials for CPI Discussions. Most of the materials Elizabeth's students use to prepare for CPI discussions are readings, although she occasionally shows a video for background information. Because her students are reading on such different grade levels, the materials Elizabeth selects are carefully targeted for various students. For example, in preparation for a discussion on gun control, Elizabeth found more than ten different articles,

some written on a very elementary level, others at a much higher level. Finding materials to meet the needs of students reading on varying grade levels is an extremely difficult task.

> This year I have the lowest students that I've ever had. I have kids who are reading at the first, second, and third grade levels. And the materials, it's been really tough on me--just with the materials I provide them to prepare for a discussion . . . so, I've tried to differentiate for those low end, and for the high end kids too. But it's a phenomenal task to be able to find materials that are appropriate at every level. (E.H. IV#5, p.1)

The need to find articles about the same CPI written at various grade levels has an effect in addition to the considerable work for Elizabeth. Because different articles have different information and perspectives, her students come to the discussions with different knowledge about the issue under discussion.

Summary of Issues Selection and Materials. Although Elizabeth's students sometimes get the opportunity to select which CPIs will be discussed, Elizabeth typically makes those decisions. When doing so, she employs a number of criteria, including connection to the curriculum, student interest, anticipated community reaction/students' age, and her own background knowledge about the issue. Once she selects an issue, Elizabeth searches for background reading materials written at various grade levels to accommodate the diverse needs of her students. As an effect, students often bring different background knowledge to the CPI discussion.

Snapshot of a Public Issues Discussion

During the 1997-98 school year, Elizabeth's eighth-grade social studies students participated in nine CPI discussions. Interspersed throughout other instructional activities, the discussions usually focused on issues that were directly connected to the unit being studied. On occasion, however, a CPI was inserted because it was related to an important current event. The CPI discussion showcased in this snapshot was such an occasion. Shortly after two boys in Jonesboro, Arkansas, killed four of their classmates and a teacher with hunting rifles, Elizabeth prepared her students to participate in a discussion of the question, "Should the United States place more limits on guns?"

This snapshot of the gun control discussion describes the model Elizabeth uses for CPI discussions, how her students prepared to participate in the discussion, and the process used by the students to develop guidelines for the discussion. Then, I turn to the actual discussion and describe what students say in the discussion, Elizabeth's role as a facilitator, and how Elizabeth assesses their participation.

The Public Issues Model. As previously mentioned, Elizabeth was a student in a class I taught in the summer of 1996 that focused on various models of classroom discussion. One of the discussion models taught in the class was Public Issues, a model developed as part of the Harvard Social Studies Project in the 1960s (Oliver & Shaver, 1974/1966). The Public Issues model involves selecting issues that bring to the fore tensions between core democratic principles (such as liberty vs. property), and the use of three different types of sub-issues within the discussion: definitional, ethical, and factual. Although some teachers use the model in small-group discussions (Miller & Singleton, 1997), Elizabeth learned the model in a large group and

uses it in that manner. Elizabeth taught the Public Issues model to her students at the beginning of the school year by explaining the different types of sub-issues and working as a large group to practice identifying the sub-issues in an article they read together. Additionally, each of her classes developed a list of discussion guidelines that she then compiled into a master list that is posted in her classroom.

Preparing for the Discussion. Three days before the discussion occurred, Elizabeth distributed various articles on the gun control issue to her students. As previously mentioned, the articles were selected because of their varying difficulty. Some articles were chosen for students reading well below grade level, others were more challenging. Students were instructed to read the articles and create a chart that listed arguments in favor of and against placing more limits on guns, and to identify ethical, definitional, and factual issues undergirding the larger gun control issue. Almost all of the students completed this assignment, which is typical; without it, they cannot participate in the discussion. Students "want to be a part of it. . . . More kids do their work for discussions than on an average basis because they really enjoy them" (E.H. IV#5, pp. 12-13).

Classroom Arrangement. As Elizabeth's 30 students enter the room, they take a seat in either an inner or outer circle. The two circles represent a change in how the classroom furniture is typically arranged for CPI discussions, which is one large circle. Only students in the inner circle are allowed to participate orally in the discussion. When students in the outer circle want to participate orally, they must tap the person in front of them on the shoulder and change seats. Because Elizabeth's students have difficulty inviting one another into the discussion, she hopes that the physical

movement will cue them to yield the floor and invite others to participate. Moreover, she thinks the new seating arrangement will cut down on the airtime of students who often monopolize the discussions. She says to the class:

> The people who have typically monopolized the discussion, this cuts their airtime almost in half. You are lined up with somebody who talks about as much as you do, and if you talk a lot, you probably have a tendency a lot of times to control the discussion. This going to give more airtime to people who don't talk as much. (E.H. AT#2)

Reminding Students of Assessment Rubric and Discussion Guidelines. Elizabeth begins the discussion by reminding students about what is on the rubric that she is using to assess their participation in the discussion:

> Please remember that I am still observing for all of the things that I have before on discussions: that you bring in your background knowledge, that you state some of the issues you have read about this week--the definitional, factual, ethical issues, that you build on or challenge someone else's comments, that you question when you want more clarity, and inviting others in is a big one this time. (E.H. AT#2)

Next, Elizabeth directs her students' attention to the board, where she has written the discussion guidelines previously developed by the class. The guidelines state: listen, participate, invite others in, be responsible, be open-minded, and respect (field notes). She then encourages her students to read through them as a reminder of the types of behaviors they should demonstrate in the discussion.

Posing the Discussion Question. Elizabeth then states the discussion topic, "The question that we pose is: Should the United States place more

controls on guns? First period realized when they got part way through the discussion that they needed to clarify what kinds of guns they were talking about. Do you think we should do that in advance?" (E.H. AT#2). A student responds, "No, guns are guns." Several others murmur agreement with his view. Elizabeth then says, "But, one thing I wanted you to remember is that when we say more controls, that some weapons have already been regulated. Who knows about this?" (E.H. AT#2). Two students explain that assault and fully automatic weapons are banned. Elizabeth then focuses the entire class on the issue once again: "So, let's go ahead and begin with the people in the inner circle, Should the United States place more controls on guns?" (E.H. AT#2).

The Discussion. For the next forty minutes, students have a wide-ranging discussion on gun control that focuses on various proposals to regulate guns, such as requiring stricter background checks, requiring that people who have guns take training courses, requiring secure storage of guns, and a total ban on all guns except those used by the police and the military. Throughout the discussion of these proposals, students use various kinds of evidence to support or challenge other students' views, such as factual information from their readings and personal experience.

Referencing What They Have Read. The first several statements in the discussion include direct references to the articles the students have read. For example, one student says, "In one of those articles I read it said it is better to protect your life than your property" (ST, AT#2), thereby immediately surfacing the value conflict between property vs. life that becomes a major focus of the discussion. Although students occasionally talk about using guns for hunting, most of them think people have guns to protect themselves and

their property. After several minutes, a student summarizes the conflict between values by saying, "Wait, Trina said something about wanting to protect your property, but someone else said that you should put your life above your property, because if you try to get your stuff back you could end up shooting yourself or the robber" (ST, AT#2).

The factual issue of whether people are safer if they have guns in their homes becomes the next point of dispute when a student says, "I read an article that said that gun owners are more likely to kill themselves than the attacker" (ST, AT#2). For the next few minutes, the students talk about whether that is accurate and, if so, what is the cause. For example, a student says, "To add on to what was said, many people who shoot themselves are not trained to use a gun, and I read an article that with training anyone can learn to use a handgun, but I think they should still be banned or regulated, but hunting rifles should not be" (ST, AT#2).

Using Personal Experiences. In addition to drawing on the articles they have read, the students also make liberal use of personal experience. When talking about whether people should be required to participate in training to use a gun, a student comments, "I know how to use a gun because my Dad taught me, we have a bunch of them in our house, but if a kid, like, if kids aren't taught how to use them, then we'll have what happened in Arkansas" (ST, AT#2). Another student responds to this by saying, "Kids do know how to use guns; the kids in Arkansas sure did" (ST, AT#2). Referencing personal experiences again, many students state that there are guns in their homes, even going so far as to explain where they are stored.

Challenging the Views of Other Students. Throughout the discussion, students challenge the views of their classmates. During one interchange

comparing gun crime in the United States to that of other nations, a student states that crime in Great Britain has increased on a percentage basis more than crime in the United States, even though Great Britain has stricter gun control laws. A student immediately corrects that statement by pointing out that, overall, gun crime is still much more prevalent in the United States. On occasion, the students use gentle sarcasm to respond to others' views. For example, one student references an article that said that criminals interviewed feared an armed citizenry more than the police, so having guns causes crime rates to go down. A student immediately quips, "Well, aren't we just one happy society" (ST, AT#2).

Seeking a Compromise. A few students take on the role of compromisers, attempting to synthesize the views of others and forge a middle ground. For example, one student says, "We need to compromise and figure out a solution; guns are good for protection but only in the right hands, but a lot of the problems in society today are young teens using guns in the heat of the moment--don't use your best judgment so there needs to be some kind of control" (ST, AT#2).

Elizabeth's Stance as a Facilitator. During the discussion, Elizabeth rarely intervenes; when she does, it is almost always to encourage equal participation. For example, about midway through the discussion, Elizabeth says, "Okay, I am going to interrupt for a minute because some of you guys are monopolizing again. Bob just opened his mouth to talk and you talked right over him, so Bob what did you want to say?" (E.H. AT#2). A few minutes later, she encourages students who have not yet talked to do so by saying, "I am going to ask that the people who have not yet been in the inner circle move to the inner circle" (E.H. AT#2). Later, she encourages careful listening

by saying, "He's agreeing with you," and reinforces the goal of inviting others to participation by pointing out, "I want you to notice that Deanna has just moved into the circle" (E.H. AT#2). Only once during the discussion does Elizabeth say anything about the content. In response to a student's direct question about the Brady Bill, Elizabeth directs the students' attention to the chalkboard, where the Brady Bill has been summarized, and briefly explains its provisions.

Closing the Discussion. As the end of the period nears, Elizabeth points out that time permits only a few more comments. Many students groan, indicating they don't want the discussion to end. With just a few minutes left in the period, Elizabeth ends the discussion and quickly directs the students to assess their participation in the discussion by filling in a copy of the discussion rubric. Additionally, she asks them to write down anything they had wanted to say but didn't, to comment on the inner/outer circle seating, and to identify a goal for the group for future discussions. The students then hand in the rubric and the notes they wrote to prepare for the discussion.

Debriefing and Assessing the Discussion. Because the class periods are so short (forty-eight minutes), students rarely have time to debrief the discussion until the next day of class. The debriefings discussions are usually 10-15 minutes long and focus on what went well and what didn't go so well. Because Elizabeth has assessed each individual student's participation in the discussion, students also receive feedback on where they stand in relation to developing the discussion skills that Elizabeth values. Elizabeth has developed a new rubric for assessing discussions that draws on the work of other discussion experts. She completes the rubric for each student after each discussion, using notes taken on a specially designed class roster. While it

may seem difficult to both facilitate and assess the discussions, Elizabeth says that is not the case. Students receive points for their participation in the discussion and for the assignment they completed to prepare for the discussion. Students who don't participate orally in the discussion but did complete the assignment receive some points, but they are penalized for not talking. Thus, Elizabeth is communicating the value she places on participating orally in discussion through her assessment and grading policies.

Summary of Gun Control Discussion. Although the quality of discussions varies from class to class, the gun control discussion that took place in Elizabeth's third period social studies class typified her approach to Public Issues discussion. Students prepared for this discussion by reading several articles and completing an assignment. When the class period began, Elizabeth reminded the students of what she was assessing and of the discussion behaviors they had previously developed as a class. Elizabeth rarely intervened in the discussion; when she did, it was usually to encourage equal participation. Students used evidence from their readings and personal experiences to support and challenge one another's views. The discussion ranged over a number of specific proposals to regulate guns with no expectation that students reach consensus. The discussion was debriefed during the next meeting of the class, when students also found out how they scored on the discussion rubric.

Portrait Three: Ann Twain at Douglas Middle School

Unlike Joe and Elizabeth, who started teaching immediately after being awarded their B.A. degrees, Ann Twain's teaching career began in her early 30s, after she earned B.A. degrees in film and dance at a major university and

spent several years as a video producer in Washington, D.C. Working as a video producer exposed her to many impressive young people in the nation's capital who were political activists. As a result, she began thinking about the "notion of kids as activists" (A.T. IV#1, p. 1) and realized she could help more young people become politically active if she taught social studies.

After Ann moved to the Pacific Northwest, she produced an educational film about service learning and the social studies and gave educational lectures to high school students. These experiences caused her to think, "This is something I really enjoy" (A.T. IV#1, p.3), and propelled her to enter a teacher education certification program in her new state's flagship university. She enrolled in a special Middle School certification program, "which included a middle school seminar, twice weekly, for two hours, with four professors. I felt totally spoiled" (A.T. IV#1, p.4). After receiving certificates in both K-8 education and 4-12 language arts, Ann landed a position as a social studies teacher for seventh- and eighth-graders at an alternative school in one of the state's largest school districts.

During her five years of teaching, she has participated in numerous professional development programs, including a national program on authentic assessment in civic education and a yearlong district-sponsored program on linking curriculum, instruction, and assessment to the district's new curriculum standards. She regularly attends social studies conferences and has presented workshops for other teachers on assessing classroom discussion. Now, after five years of teaching, Ann is thinking about earning a graduate degree in social studies.

Contexts

As was the case for Joe and Elizabeth, a variety of contexts influence how Ann thinks about and enacts her CPI discussion teaching practice. In particular, the community in which she teaches, her school, its students, and the curriculum inform Ann's approach to CPI discussions.

Community. Ann's school is located in a large school district outside a major city in the Pacific Northwest. The community served by the school district is established, growing rapidly, and economically diverse. Students at Ann's school are bused in from all over the 36-mile school district. Ann defines the community in very broad terms: "It's a nice community, supportive community of the schools. Levies pass" (A.T. IV#1, p. 11). The parents who send their children to Ann's school tend to be more affluent than is typical in the community, and less likely to move, causing Ann to describe the sub-community made up of the school's parents and their children as "established and pretty stable. . . . Kids tend to be with us from the first grade through eighth grade" (A.T. IV#1, p. 10).

Douglas Middle School. Ann's K-8 school is eleven years old and has 700 students, 150 of whom are in the middle school. There is a waiting list to get into the school. The two terms most frequently used to describe the school are "multi-age" and "non-graded," which reflect how the school is an alternative to more traditional schools in the district. "Multi-age" means that students are in classes with children either younger or older than they are. For example, all of Ann's classes combine students who in a traditional school would be in the seventh- and eighth-grades. "Non-graded" means that traditional letter grades are not used to report students' progress to their parents. Instead, each teacher develops an elaborate report card based on the

student outcomes for her/his subject. Ann's report card, for example, includes 33 specific outcomes organized into six categories. For each outcome, students are rated on a scale ranging from "unacceptable progress toward standard" to "exceeds standard" (A.T., ART).

Parents tend to be quite involved in school activities. Ann explains this generalization with a specific example: "I am able to have a service learning program that requires kids to be driven to service learning sites off campus twice a month. And I have close to 25 parents that come and take these kids. So I think that's pretty phenomenal" (A.T. IV#1, p. 5). Parents also participate in the curriculum within the school walls, by serving as jurors in mock trials or facilitators for a simulation of the United Nations, for example.

Students. Ann teaches 150 students each day. She categorizes her students as those who are "really self-directed and capable" and those who are "very needy and their parents thought that a program that is not textbook-based, that is more individualized, would be better for them" (A.T. IV#1, p. 5).

The school district's hearing impaired program is housed at Douglas Middle School, which contributes to diversity among the students, but in other categories, especially race and ethnicity, the student body is extremely homogenous. Ann estimates that 85% of the middle school students are white. Moreover, the students come from families with economic resources. Ann estimates that fewer than 10% qualify for the free or reduced-price lunch program. Ann is concerned about the homogeneity of her school's student population: "What we're now looking at is that our population is not diverse enough . . . so we really need to up the diversity" (A.T. IV#1, pp. 6-7).

Curriculum. Each day, Ann teaches four sections of multi-age social studies and one section of drama. Because Ann has both seventh- and eighth-

graders in the social studies classes, and they stay with her for two years, her curriculum must change each year to avoid repetition. To accommodate this unusual circumstance, Ann has designed a curriculum that spans a two-year period. The fact that Ann teaches both seventh- and eighth-graders in the same class creates two additional challenges: (1) it heightens the academic differences within each class ("there is a big difference between a low seventh and a high eighth") (A.T. IV#1, p.9), and (2) it broadens the curricular focus.

During any one school year, Ann's course focuses on an amalgam of United States history, government, world geography, and current events. The specific eras in history change from year to year, but overall Ann's students study more 20th-century history than would typically be the case in other middle school social studies courses in the district. The government component of the course centers on the Constitution and other founding documents, the form and structure of the federal government, with an emphasis on the Supreme Court, and controversial public issues, especially those that are scheduled to be on the state ballot as initiatives. The world geography units in her course emphasize themes of geography (such as the interaction between humans and their environment) and the role played by world organizations, such as the United Nations. The study of current events is carefully woven into the curriculum, both to enhance its relevancy to students and to help them better understand the world in which they live.

Teachers in the school rarely use textbooks, and Ann is no exception to this rule. Instead, she creates "original curriculum" that involves "hands on" activities, such as historical re-creations, other simulations, and inquiry learning (A.T. IV#1, p. 7). Because her students have been steeped in this type of learning for many years, they have developed the skills necessary to

participate effectively in them. Ann says her students are "very used to getting up and presenting orally. That's not something that I had to introduce at this level. It's something that they feel pretty comfortable with" (A.T. IV#1, p. 8).

Summary of Contexts. Ann has been teaching middle school social studies for five years in a suburb of a major city in the Pacific Northwest. She is actively involved in a number of professional development programs, both as a student and a leader. The community in which Ann's alternative school is located supports education, and the parents who elect to enroll their children in Ann's school are extraordinarily involved in school activities. The K-8 school is "multi-age" and "non-graded," reflecting its non-traditional approach. The social studies classes Ann has created focus on United States history, government, world geography, and current events. Because she teaches seventh- and eighth-graders in each class, the curriculum must change each year. All of these contexts (community, school, students) influence how Ann thinks about and practices her teaching. Additionally, her conceptions of a number of important factors also inform one aspect of her teaching, which is her quest to teach students to participate effectively in CPI discussions.

Conceptions

Recall, in this study I am defining conceptions to include a teacher's knowledge, beliefs, thoughts, and images that may inform classroom practice (Shavelson & Stern, 1981). In this section, I focus on Ann's conceptions of four factors: the purposes of social studies; the characteristics of effective democratic citizenship; the rationales for CPI discussions; and the characteristics of effective CPI discussions.

Conceptions of the Purposes of Social Studies. I learned about Ann's conceptions of the purposes of social studies by looking at a concept map she drew on the topic (see Chapter Three) and by listening to her describe and explain the map. In the center of the map was a stick figure representing a student. Lines connected the student's head to four large circles labeled to represent what Ann wants her students to know, be able to do, and be disposed to do as a result of her social studies curriculum. The labels said: issues I'll face in my future; citizenship; geography; and United States history (A.T., ART).

Unlike Joe and Elizabeth, Ann began her description of the concept map by describing the origins of her curriculum:

> When I started my first year of teaching and I knew I would be making up my own curriculum, I didn't start with these other categories of United States history, geography, and citizenship . . . I started by sitting down and saying to myself, "Okay. I have these kids and what are the most important things that I can equip them to deal with?" (A.T. IV#2, p. 1)

Ann answered this question by identifying four overarching issues she felt would be present throughout her students' lives: immigration, population pressures, environment, and technology. These issues became conceptual drivers as she created a curricular path for her students to follow in achieving the outcomes she valued. The issues were also curricular topics in and of themselves. For each one, she had three specific outcomes. First, she wanted her students to understand that these were issues they should care about. Second, she wanted them to be informed about multiple perspectives on the issues. Third, she wanted to influence their behavior: for example, "to

be able to think about and make smart choices in terms of resource use" (A.T. IV#2, p. 2). Ann was careful to point out that she recognizes the controversial nature of these issues and does not want her students to form a particular point of view on specific policy proposals. Instead, she wants them to recognize and care about each issue writ large. The following excerpt from the concept map interview explains this distinction:

> Hess: I want to come back to these issues. Are these issues that you want kids to have a particular point of view on?
>
> Twain: No, Um, hmmm. Now that I've said that, let me think about that. In a couple of cases, I'll have to say "yes." Environment. I want them to feel they have a role as a custodian of the environment. I don't necessarily want them to feel that, you know, the Endangered Species Act is the way to go. . . . My hope is that in each of these cases [issues], with more knowledge, they will take an enlightened viewpoint, which is a responsible one.
>
> Hess: What do you mean by "enlightened?"
>
> Twain: In looking at the environment, [etc.], that they not be self-centered in their views and in their approaches. That they're always looking at the whole. The collective whole, and not what's good for them, but what's good for people. What's good for the world. What's good for the community they live in. (A.T. IV#2, p. 6).

Although Ann's identification of four big issues that drive her curriculum certainly has strong connections to citizenship education, on her concept map she created a special category labeled "citizenship" that had several components, including multiple perspectives, giving to the community, appreciating diversity, activism, and ethics. The citizenship

outcome is motivated by her opinion that "we want people who can be trusted with a really important role in society, which is to be a good citizen" (A.T. IV#2, p. 3).

Within the citizenship category, Ann distinguishes between giving to the community and activism, although both are purposes of her social studies teaching. Her definition of giving to the community is represented by a service learning program she has created for her students. Each year, her students are released from eight half-days of school to volunteer at various sites where their service is needed, such as nursing homes and elementary schools. By contrast, activism is taught to her students as a form of proactive problem solving. Ann wants her students to learn how to look at a "problem from the faucet where it starts and not the puddle on the floor that you have to mop up after it's already flowed over" (A.T. IV#2, p. 3). An example of activism that Ann cited was, "my kids going to the city council and proposing a day of concern for youth violence, and proposing a recreation center for kids during the hours when there's the most juvenile crime in our community" (A.T. IV#2, p. 3).

In addition to conceptualizing the purposes of social studies as knowing and caring about the four overarching issues and other citizenship outcomes, Ann also wants her students to learn content drawn from the disciplines of history, geography, and political science "so we're all on the same field, and know we have the same equipment" (A.T. IV#2, p. 6). "Knowing" takes many forms in Ann's conception of social studies: "I think that knowing can also have a human side . . . knowing what certain experiences are like. It may be knowing there is human knowing and not just fact-based knowing" (A.T. IV#2, p. 6).

Finally, Ann is concerned that social studies prepare her students to understand how to interpret and think about persisting issues of history that are present in today's world. Toward this end, she focuses on current events and selecting issues drawn from history and geography that require students to discuss tensions between such core values as individual rights and the common good.

Conception of Effective Democratic Citizenship. When describing and explaining how she thinks about citizenship, Ann distinguished between knowledge that effective democratic citizens should possess, skills they should be able to demonstrate, and actions in which they should be engaged. The knowledge category includes four components: understanding founding principles of the government in the United States (such as rule of law); being aware of the history related to national identity; understanding significant historical milestones, especially those related to past inequalities; and understanding historical examples of good and bad political leadership.

Ann identifies founding documents (such as the Constitution) as the primary source for the founding principles of government in the United States and believes that an understanding of those principles correlates positively with effective democratic citizenship. She links knowledge of founding principles to behavior that effective citizens would exhibit. For example, if people understand that "No one is above the law" (AT. IV#1, p. 20), they might be more likely to respect the law.

In addition to knowledge of founding principles, Ann also hopes that citizens will have historical knowledge about who lives in the United States and how different groups have influenced national identity. This knowledge can be used by people to address a key question: "Do we still share a vision

[of/for the United States] ?" (A.T. IV#1, p. 20). Historical background knowledge can also help prevent repeating the mistakes of the past. In particular, Ann believes that citizens should know how past inequalities have shaped the issues facing the United States today. To support this point, Ann asked a rhetorical question: "How can we understand something like affirmative action, or the Civil Rights Act, if we don't understand what wasn't right about what was? And what issues are we still dealing with today because of what wasn't right?" (A.T. IV#1, p. 21).

A final knowledge component in Ann's conception of effective democratic citizenship deals with political leadership. Ann hopes that citizens will know examples of good and bad leadership over time so "we know what we should be striving for" (A.T. UV #1, p. 21).

Ann also expects that effective democratic citizens will have developed a number of specific skills they can employ in their citizenship role. Recognizing that we live in the information age, Ann wants citizens to be able to acquire knowledge drawn from multiple perspectives that is learned from multiple stake holders and represents multiple value orientations. Believing that not all information is equally valid, Ann also sees the need for citizens to have a "BS detector so they can sniff out bias and falsehoods" (A.T. IV#1, p 21). She also places considerable stock in individual agency and wants citizens to make up their own minds about the issues that confront them. Once these ideas have been formed, Ann adds another skill citizens need: "the ability to articulate your ideas orally and in writing so that other people can understand them" (A.T. IV#1, p. 21).

Ann further believes that effective citizens in a democracy, armed with citizenship knowledge and skills, should want to engage in numerous citizen

actions, including voting, volunteering, attending public meetings, and standing up for the rights of others, especially people who are oppressed. She places special importance on the last activity: "The one that I think is really important is to stand up for others and stand up to others and not to be silent. I think there's a real problem with bystander behavior in our society" (A.T. IV#1, p. 22). Linking silence to compliance, Ann continues, "There are times when silence is a really negative thing and something to be avoided. That's part of being a good citizen" (A.T. IV#1, p. 22). I asked her, "So what do you mean by that? What does it look like to stand up for and to others?" (Hess, IV#1, p. 22). Ann responded:

> If someone is having their rights violated or someone who lacks a voice in our community, a person who is mentally ill, a senior citizen, a child who can't yet vote. Being able to help be a voice for people who maybe are oppressed or who don't have the skills to have a voice. And then standing up to others. . . . What are the ways that we, as citizens, can be heard? If we aren't the power holders, what can we do to have some power? (A.T. IV#1, p. 22).

Conception of the Rationales for CPI Discussions. Ann holds multiple rationales for teaching her students to participate more effectively in CPI discussions. Ann described and explained these rationales as she engaged in a card-sorting task that required her to rate and rank various rationales for CPI discussions that were drawn from the literature (see Chapter Three). Her conceptions of the rationales are summarized in Figure 10, but here I will explain in greater detail how Ann thinks CPI discussions improve democracy, activate youth's passion for issues, improve critical thinking skills, enhance

appreciation for diversity and perspective-taking, and inform personal decision-making.

Rationales for CPI Discussions	Do CPI discussions accomplish this goal?	Reasons why CPI discussions do/do not accomplish this goal:	Caveats and/or further explanation
Improve interpersonal skills	yes	Four important skills: listening, understanding others, inviting participation, not monopolizing	
Understand important democratic values	Maybe	Depends on the CPI, but some CPIs illuminate democratic values	Help students see democratic values as "living"
Participate in political life of democracy (e.g., voting, serve on jury)	yes	Discussions won't cause people to vote or serve on a jury, but will make them better at doing those things	Discussions help people be informed voters and jurors
Learn important social studies content	yes	Brings content to life, helps students see content in 3-D	You "own" ideas when you articulate them
Learn how to think critically	yes	Helps students think critically, quickly, and thoughtfully	Causes discussants to learn how to use analogies, create relationships among ideas
Improve democracy	yes	Improves the marketplace of ideas that is fundamental to democracy	Truth and logic emerge from CPI discussions, divergent ideas are sometimes best

Figure 10. Rationales for CPI Discussions: Ann Twain
(Note: Ann's responses are from A.T. IV#3, pp. 1-15).

Identical to Joe and Elizabeth, the rationale that has the most power for Ann is the positive connection between CPI discussions and a healthy democracy. Linking the health of a democracy to the vibrancy of the "marketplace of ideas" within it, Ann said, "we progress as a society because new ideas are constantly getting aired and we are not stagnant" (A.T. IV#3, p. 4). Moreover, the marketplace of ideas has the potential to prevent or redress mistakes. Ann used a negative example to support this point:

> We're constantly re-evaluating what we're all about. And as new issues come out, I mean it was Einstein, wasn't it, with the atomic bomb technology, who said immediately, "This needs to be taken into a public forum and discussed by people." (A.T. IV#3, p. 4)

CPI discussions help improve democracy because within the marketplace of ideas, the truth will emerge. Representing the value she places on multiple perspectives on CPIs, Ann says, "what is true, what is real, what is meaningful rises . . . and I think that's crucial in a democracy. That's what democracy is all about" (A.T. IV#3, p. 9). Ideas that are outside of the mainstream are necessary because progress depends on fresh ideas and "because occasionally it's that divergent idea that is the right idea" (A.T. IV#3, p. 9).

Ann reaches into her childhood memories to explain another rationale that she holds for CPI discussions: activating youth's passion for issues. "I felt clueless when I was an adolescent. Adults would be discussing things [issues] around me and it made me very passive about what was going on. I think that by inviting them into discussions of real issues, it's very empowering to them" (A.T. IV#3, p. 7). Ann uses CPI discussions to make students passionate about issues in the world around them. Young people "can get so

darn apathetic about school, about the world. This [CPI discussions] lights a fire under them" (A.T. IV#3, p. 10).

Improving her students' critical thinking skills is another rationale Ann uses to support her emphasis on CPI discussions. Unlike Joe and Elizabeth, who often mentioned slowing down the pace of discussions to help students think better, Ann wants her students to think quickly and thoughtfully and uses CPI discussions to that end. Defining critical thinking skills to include questioning, applying information to new situations, creating analogies, and seeing relationships, Ann uses CPI discussions because she thinks they help students develop these skills and give them a forum for demonstrating their critical thinking prowess. She also recognizes that talking in a CPI discussion does not necessarily correlate positively to critical thinking. Explaining this point, she says, "I would argue that I have some kids who talk in town meetings [the model of CPI discussions she uses in class] and aren't really thinking; they like to hear themselves talk" (A.T. IV#3, p. 12).

The CPI discussion model Ann uses requires the explanation of multiple perspectives on issues from the roles of multiple stake holders in the issue (see pages 144-145 for explanation of the model). This model reflects Ann's emphasis on developing appreciation for diversity. Ann defines what this means in three ways: students will (1) become more appreciative of multiple perspectives on issues, (2) appreciate the diverse groups in the United States, and (3) develop greater empathy for other people and their ideas. She thinks that CPI discussions can help students step outside of their own situations and backgrounds and appreciate that others have different experiences that inform their views on CPIs.

Finally, Ann sees a positive correlation between students' participation in CPI discussions and the personal decisions they make about how to lead their lives. For example, after participating in a CPI discussion on abortion policy, "undoubtedly kids went away and thought, 'How do I feel about this, personally?'" (A.T. IV#3, p. 6). The decisions students make as a result of CPI discussions could be either political or personal, both of which would be informed by a clarification of their values.

Conception of Effective Classroom Discussion. Ann's ideas about the characteristics of effective classroom discussion emerged as she explained her reactions to two videotaped excerpts of classroom discussions (see Chapter Three). Like Joe and Elizabeth, Ann has strong opinions about the elements of a good discussion. In an effective discussion, discussants are seated in a circle; if there is a facilitator, that person should say little. Participants should use evidence and logic and frame the issue at hand as an example of larger, more transcendent value conflicts. During the discussion, students should exhibit interpersonal skills, such as listening attentively to one another and not monopolizing.

1. Setting. Both Joe and Elizabeth recommended seating discussants in a circle. Ann shares that view, linking where discussants sit in relation to one another to the kind of interaction that occurs among them. When students are seated in rows facing the teacher, too much of the interaction is like a tennis match, with the teacher volleying back to the students. Conversely, if the students are seated in a circle so they can all see one another, interaction is more likely to occur among the students--which is what Ann desires.

2. Facilitator's Role. Ann reacts negatively to talkative facilitators. Because she places such a premium on interaction between students, a

facilitator who takes up too much airtime, by definition, diminishes the quality of the discussion. The facilitator still has an important role, however. Because Ann uses a formal discussion assessment process, she sees one role of the facilitator as clearly stating standards of what constitutes effective discussion. Explaining why she thinks it is important to formally assess discussion, Ann says, "Share with the kids what you're assessing, so they know what that looks like. And then they can be successful if the criteria are unmasked and they'll know that participation is really valued" (A.T. IV#2, p. 15).

3. Interaction Among Students. In an effective discussion, students should be interacting with one another, working to build and challenge previous statements. By doing this, the discussion is "something they are all creating" (A.T. IV#2, p. 20). Ann attends to how many students are participating and favors discussions that have a high rate of oral participation. Moreover, she wants the responsibility for the content, pace, and flow of the discussion to rest with the students--not the facilitator. Explaining why the Electoral College discussion was so impressive, Ann said, "It's moving somewhere. It's moved in a lot of directions. And, they're moving it, there's no teacher moving it, they're moving it" (A.T. IV#2, p. 18).

4. Interpersonal Skills. Just like Elizabeth, Ann uses the word "civility" to summarize the goal of various interpersonal skills. Attentive to whether students are listening to one another, inviting others to participate, and taking care not to monopolize, Ann connects the exhibition of these skills to the overall quality of the discussion.

5. Ideas in Discussion. Ann wants to hear ideas talked about in several ways in discussion. She favors the use of historical examples and connections

to modern life, along with a liberal dose of evidence and logic. Moreover, the range of ideas that students discuss should be broad, illustrating the value she places on multiple perspectives. Reacting negatively to a forced-choice exercise the teacher in the Richard Wright discussion used, Ann said, "They [students] didn't have a choice other than 'yes' or 'no,' I was wondering if she was going to say, 'Or are you torn?' "Or are you somewhere in between?'" (A.T. IV#2, p. 4). Finally, Ann responds positively to discussions in which the talk focuses on the larger value conflicts represented by specific issues. For example, while viewing the section of the videotape of the Richard Wright discussion in which a student talks about the conflicts between individual morals and the moral code implicit in democratic governance, Ann says: "Excellent comment. Because he is getting at the whole notion of individual rights vs. the common good" (A.T. IV#2, p. 6).

Summary of Conceptions. Ann has multiple and complex conceptions of the purposes of social studies, the characteristics of effective democratic citizens, the rationales for using CPI discussions, and the characteristics of effective discussion. Like those of Joe and Elizabeth, Ann's conceptualization of the many purposes of social studies is congruent with what she believes citizens should know, be able to do, and want to do. Coming from an activist bent, she does not see citizenship preparation only in the future tense. Instead, she envisions "kids as activists" who learn how to be good citizens in social studies classes through participation in citizenship activities. Consequently, her primary rationale for using CPI discussions is that they mirror the marketplace of ideas that exists in the world outside of school and are fundamental to the health of a democracy. Classroom discussions that are

especially effective involve interaction among the students across a broad range of ideas.

Selecting Content for Discussion: Issues and Materials

Thus far, the portrait of Ann has centered on her background, the contexts in which her teaching is situated, and her conceptions of factors that may influence how she teaches students to participate in CPI discussions. Now I begin describing and explaining Ann's CPI discussion teaching practice by focusing on how she selects the issues and materials for social studies curriculum. As in the other two portraits, this section addresses three questions: Who decides what content (i.e., issues) will be discussed? What criteria are used to select the content? Once an issue is selected, a third question is introduced: What materials will students interact with to prepare for CPIs discussions?

CPI discussions in Ann's social studies classes are structured using a model she calls Town Meetings. The topics of the Town Meetings are determined by Ann, using the criteria described in this section. While some of the Town Meeting topics are historical CPIs (such as, "Was it right to drop the bomb?"), most are contemporary CPIs.

Criteria for Selecting Issues. Ann and her students use five criteria when selecting the topics for the Town Meetings: the extent to which the CPI lends itself to multiple perspectives from multiple stakeholders; the availability of resources; relevance to students' lives and interests; connection to democratic values; and whether the CPI is currently a matter of public deliberation. These criteria emerged as Ann engaged in an issues-selection task. Recall, this task required Ann to sort ten CPI topics into three piles to reflect what she would include in her curriculum, what she might include,

and what she would not include (see Chapter Three). Figure 11 illustrates Ann's decision-making on issues selection.

CPI Topics	Ann's decision	Ann's reasons
Abortion	Yes	tough issue for students to depersonalize
Affirmative Action	Yes	historical connections, matter of current public discussion
Balanced Budget	No	issue is too flat, students would not be engaged
Gay Rights	Maybe	students too immature, concerned about gay students in class
Immigration	Yes	huge issue facing the United States
Legalizing Drugs	Yes, but focus on medical use question	red flag with parents, Web site resources are questionable
Physician-assisted suicide	Yes	human element, matter of current public discussion
Trade policy	Yes, if focus is on human rights issues	human rights
Vouchers for private schools	Probably not	hot potato for parents
Welfare Reform	Yes	human dimension, personally affects some of their lives

Figure 11. Ann's Decisions and Reasons About Which CPIs To Have Students Discuss

Multiple Perspectives from Multiple Stakeholders. CPIs Ann selects for her curriculum must correspond to the unique demands of the Town Meeting model; that is, the issues must be ones for which there are multiple perspectives held by different stakeholders. By multiple perspectives, Ann means a variety of different ways that people can analyze an issue and a variety of different views that people might hold about an issue. Although, in theory, multiple perspectives can be identified for all CPIs (because there is something "at issue"), Ann believes that some CPIs are more perspective-laden than others. For example, CPIs about immigration topics meet this criterion for Ann because such issues are loaded with economic, political, and social perspectives.

Ann defines stakeholders as "people who have an interest in the issue because it affects them" (A.T. IV#4, p. 7). As a general rule, the more people who are directly affected by an issue, the more likely Ann is to select it for class discussion. She mentioned physician-assisted suicide as an issue that potentially affects the lives of many people; she therefore included it as the topic of a Town Meeting in the school year before this study took place.

Availability of Resources. Because Ann's students read and view many different resources to prepare for Town Meetings, she selects CPIs for which she knows many high quality resources representing a range of perspectives are available. Ann explained this criterion by linking it to the previous one: "I'm looking for topics that lend themselves to multiple perspectives and perspectives that I can find resources on, such as guest speakers, Web sites, printed material, and video clips. I want topics that are in the news enough for kids to find information" (A.T. IV#4, p. 6).

Relevance to Her Students' Lives. A third criterion Ann uses to select CPIs for Town Meetings is whether they have a "kid hook" (A.T. IV#4, p. 7). Issues that "hook kids" do so because of their relevancy. Summarizing this criterion, Ann said, "I select issues that involve something that a middle school kid would care about. That has relevance in their lives or will have relevance in their lives in the future" (A.T. IV#4, p. 7). She believes her students are particularly attracted to issues that have a "human element" and cites the balanced budget issue as an issue that lacks in that element and would therefore not be selected. Conversely, Ann identifies immigration, abortion, and physician-assisted suicide as issues that possess the "human element" because they are fundamentally about people.

The "human element," however, sometimes makes an issue too difficult for her students to discuss. Explaining her reluctance to include gay rights as a Town Meeting topic, Ann said, "I am worried about opening the flood gates and having really intolerant comments come out. I don't want anybody to be exposed to that kind of ugliness from other kids. Because I've got some kids who, you know, might be gay. And I just don't want to go there" (A.T. IV#4, p. 12).

Connection to Democratic Values. Because Ann is interested in shaping the democratic values held by her students, she selects issues that are especially value-laden. Through CPI discussions, she hopes that students' "values, their democratic values . . . are being strengthened" (A.T. IV#4, p. 8). Distinguishing between causing her students to hold a particular view on a CPI and enhancing their overall democratic values, Ann said, "I think a person could still value equality in this country, but be opposed to current legislation on affirmative action" (A.T. IV#4, p. 9). She does, however, want

students to develop an understanding and appreciation for democratic values, such as equality, the absence of tyranny, and the freedom of thought and belief.

Currently a Matter of Public Deliberation. The final criterion Ann uses to select issues is whether they are currently matters of public debate in the world outside of school. In the state where Ann lives, CPIs are often on the ballot as initiatives. Each year, Ann selects at least one initiative issue for students to discuss because it provides them an opportunity to practice preparing to be informed voters. Additionally, these issues provide students a good opportunity to educate their parents. Ann explained how this worked when she said, "I think that [discussing ballot initiatives] also gets kids talking to their parents about the importance of being informed on an issue that you're going to have to vote on" (A.T. IV#4, p. 10). Recalling how this has worked in the past, Ann described how her students carried their new understanding formed by a Town Meeting on property rights to their homes: "I truly believe that kids in my room were more informed than their parents and were going home telling their parents things. That models discussing and really getting into the meat of these issues and I don't think we can do that enough" (A.T. IV#4, p. 11).

Materials for CPI Discussions. Of the three teachers in the study, Ann uses the widest range of materials to prepare students for CPI discussions. In preparation for each Town Meeting, students read a packet of articles that provide common background, and specialized articles selected because they illustrate the particular perspective each student will represent in the discussion. Additionally, the students search the World Wide Web for additional information, view videotaped programs, listen to guest speakers

who have specialized knowledge about the issue, and sometimes have a conference call with experts on the issue. When the Town Meeting is on a ballot initiative, Ann's students also read the exact wording of the initiative and state-prepared materials summarizing various arguments.

Similar to the demands placed on Elizabeth, the varying reading levels of students in Ann's classes cause her to seek out materials that are written in simple language so students reading dramatically below grade level can access enough information to participate in the Town Meeting. One of the reasons Ann goes to the trouble of finding guest speakers and videotaped programs about an issue is to help students who read below grade level get a better grasp of the background and various perspectives on the issue.

Summary of Issues Selection and Materials. Ann and her students select the CPIs for Town Meeting discussions. The criteria they use when selecting issues are the extent to which the CPI lends itself to multiple perspectives from multiple stakeholders; the availability of resources; relevance to her students' lives and interests; connection to democratic values; and whether the CPI is currently a matter of public deliberation. To prepare her students for the Town Meeting discussions, Ann recruits guest speakers, shows videotapes about the issue, and prepares reading packets. Her students also search the World Wide Web for additional information. All students read some core articles explaining the background of the issue and then read specialized packets to prepare for a specific role in the Town Meeting.

Snapshot of a Town Meeting

During the 1997-98 school year students participated in eight Town Meetings on issues ranging from whether it was right to drop the atomic

bombs to gun control. As previously mentioned, one criterion Ann uses to select issues is whether an issue is currently a matter of public deliberation. In the spring of 1998, it became apparent that an initiative banning local and state government affirmative action programs based on race and gender would be on the ballot in the fall. Ann decided it would be an especially good topic for a Town Meeting.

The affirmative action Town Meeting is showcased in this section of Ann's portrait. First, I explain the unique Town Meeting discussion model that Ann has created and how she teaches the model to her students. Then I describe how her students prepare to participate in Town Meetings, and how Ann uses traits on an assessment rubric to remind them of the types of oral participation that contribute to successful Town Meetings. Next, I turn to what actually occurred during the Town Meeting, focusing on what the students said, Ann's role as a facilitator, and how Ann assessed her students' participation.

The Town Meeting Model. While a first-year teacher Ann developed the Town Meeting discussion model, which she defines as "a public forum where participants air their views on an important controversial issue as a way to either affect public policy, or educate others, or persuade others to come around to their point of view" (A.T. IV#5, p. 6). The primary reason she uses Town Meetings is her belief that it helps her students better understand multiple perspectives about the issue being discussed.

The Town Meeting model is a large group discussion in which each participant assumes the role of a person who has a particular perspective on the issue. The roles are crafted by Ann and her students to cover a broad spectrum of views on the issue. Students select their roles; Ann encourages

them to pick a role that represents a position other than the one they currently have on the issue. Additionally, Ann makes sure there is a relatively equal distribution of roles between the various points of view on the issue. That is, the students are not allowed to all select roles in favor of one position.

Before the first Town Meeting in the fall, Ann taught the model to her students by explaining the assessment rubric for the Town Meeting and showing them a videotape of an especially good Town Meeting from the previous year. Ann occasionally stopped the videotape and pointed out students' contributions that met the exemplary standards of the rubric. Thus, students first learned the model by viewing a positive example. Ann followed up with a negative example, showing a videotape of adults participating in a Town Meeting that was not effective. She had her students identify what the adults were doing wrong, such as monopolizing, not using evidence to support their opinions, and talking over one another.

Preparing for the Town Meeting. One week before the Town Meeting, Ann's students received a packet of background material on the affirmative action initiative to read. After one class period of didactic instruction on the issue, Ann and her students created the roles for the Town Meeting. Some of the roles were the Governor of the state, a university admissions officer, a newspaper reporter, a white business owner, a minority student, and representatives of education and advocacy organizations that had taken positions on the initiative. After selecting which role they wanted to adopt, each student was given a specialized reading packet that focused on the particular position of her/his role and a role sheet that required him/her to state the position and identify pro and con arguments. For the next three class

days, the students worked individually and in pairs reading the articles and preparing their role sheet. During this time, Ann also showed a video about affirmative action, opened her classroom to a guest speaker from an educational organization that had studied and taken a position on affirmative action, arranged for several students to call an advocacy organization that had funded the initiative drive, and gave students time to access the World Wide Web to search for additional information.

Classroom Arrangement. As Ann's 29 seventh- and eighth-grade students enter the classroom, they immediately notice that the furniture has been reconfigured for the Town Meeting. The tables are arranged in a large circle and Ann has placed name tents listing the various roles on the tables. The students take their places. Ann intersperses the roles to ensure that a debate-like atmosphere is prevented. For example, at one table is the Governor (who opposed the initiative), while sitting at the next table are representatives of an advocacy organization that is supporting the initiative.

Explaining the Assessment Rubric. Because Ann assesses each student's participation in the Town Meeting, she begins the class period by reminding students of what is on the assessment rubric. The three categories of traits on the rubric are: knowledge of subject matter, portrayal of role, and effectiveness as a participant. Ann briefly explains each trait while holding up the tally sheet she will use to mark throughout the Town Meeting.

Identifying Their Characters. The Town Meeting officially begins with each student stating her/his role and the position with which it corresponds. Ann does this as a verbal warm-up, as well as so that students will be reminded of the many roles. One student does not have a role because he has been absent.

Stating the Purpose of the Town Meeting. Next, Ann tells the students to "stand behind that character; give him the benefit of your voice" and states the purpose of the Town Meeting: "We're here to get the facts on how you feel about the initiative (A.T., FN). She then asks, "What does the initiative say? (A.T., FN). Several students quickly respond, explaining the major points of the initiative.

The Town Meeting. For the next ninety minutes all but two of the 29 students participate orally in the discussion about the affirmative action initiative. This is an unusually long Town Meeting; more typically they last fifty minutes. A class will occasionally have a discussion this lengthy due to a changing block period schedule.

Throughout the Town Meeting students raise their hands when they want to speak, and Ann calls on them, going back and forth between students who support and those who oppose the initiative. Because the initiative addresses affirmative action based on race and gender, the discussion alternates between statements specifically focused on each. For example, in the beginning of the discussion a student says, "The initiative says you can't discriminate [by preferring racial minorities and women]," and another student responds, "But sometimes people are naturally racist, really just trying to even out [the playing field] (FN)." A student immediately shifts the focus to gender when she says, "To add to that, 95% of management jobs in this state and in the nation are taken by men" (FN).

Bringing Value Tensions to the Fore. This shifting between race and gender continues throughout the Town Meeting, although several students contribute statements that contextualize the initiative in broader tensions between competing goods and values, such as equality vs. merit, and equality

vs. safety. One lengthy interchange about firefighters represents the latter tension. A student focuses attention on this tension by saying, "I don't think we should risk people's lives. Fire departments are forced to hire women because of affirmative action and they can't do the job" (FN). Another student agrees and adds, "Many women couldn't pass the physical tests that men had to pass to become firefighters so they changed the tests. Again, that is risking people's lives" (FN). A few statements later, another student challenges this view by telling a long story about how women are discriminated against in the local fire department even when they score the highest on the tests. With this story she is shifting the discussion to the issue of merit, which is immediately reinforced by another students who says, "They [the fire department] should hire the person who is the most qualified" (FN).

Citing Factual Evidence. Ann's students' conflicting opinions on the affirmative action initiative parallels the public debate occurring outside of school. Some students think there is still a lot of discrimination against racial minorities and women, while others disagree. A student raises this issue by quoting from a study he has read: "A women or minority has only a 2% chance of being hired by a company that is run by white men" (FN). Another student immediately challenges that statistic and asks for where he read it. The first student points to an article he has in front of him. Another student uses stipulating language when he says, "Well, if what he said is true [the 2% statistic], that's why we need to keep affirmative action" (FN). Several students stick to the topic, which finally causes one boy to ask, "Why do people discriminate when we're all one race--the human race?" (FN) The simple wisdom implied in this rhetorical question momentarily silences the

entire class. A student then answers in a quiet voice, "I think people are afraid of what they don't know" (FN).

The Influence of Their Roles. Throughout the Town Meeting, it is clear that the assumed roles are influencing the content of students' comments in different ways. The student playing the Governor accurately and consistently represents what he has read about why the Governor of the state opposes the initiative. Conversely, some students quickly stray from their roles. For example, a student playing another Governor seems to change his mind midway through the Town Meeting and begins to advocate continuing affirmative action, which is not the position the real Governor supports. It appears that the roles played are not viewed as rigidly binding ideological directives by the students.

Ann's Stance as a Facilitator. Ann is incredibly busy throughout the discussion. She directs discussion by calling on students who have their hands raised, assesses by marking on the tally sheet, and, on occasion, redirects the content of the discussion by asking clarifying questions and raising issues the students haven't yet considered. When there are factual disputes that need to be clarified, Ann often inserts very short questions, such as "Are quotas legal?" and "Is the playing field level?" (FN). At other times, she helps students who are having difficulty clearly stating their comment or question by rephrasing it for the entire class. For example, after a student's very murky explanation of the predictive value of test scores, Ann says, "So you are saying that test scores don't predict job success" (FN). Midway through the Town Meeting, she stops the discussion and directs students to look at their role sheets. She then says, "Look for points that have not been made yet or that could be strengthened" (FN).

Only once during the ninety minute discussion does Ann need to intervene on a behavioral issue, and she does so without saying a word. About thirty minutes into the discussion, it becomes clear that three boys are not paying attention and are distracting others with whispered side conversations. She makes eye contact with them, gives them a classic "teacher look," and points to the couch that is outside the circle, indicating that if they don't start paying attention she will remove them from the Town Meeting. It works. They stop the whispering.

With a few minutes left in the period, Ann ends the Town Meeting by directing the students to turn to the person next to her/him and say anything about the initiative that he/she has not had time to contribute during the discussion. They do this, then hand in their role sheets to Ann and exit the room.

Debriefing and Assessing the Discussion. Similar to Elizabeth, Ann's discussions often occur on Fridays, so the debriefing of the discussion does not occur until the following Monday. I did not attend the debriefing but Ann reported to me that two things typically occur after the Town Meetings. First, the students talk about what went well and what didn't go so well. Second, Ann gives them her assessment of the Town Meeting, focusing on the traits listed on the rubric. Students also have their role sheets returned, with comments from her, find out how she assessed their participation in the Town Meeting, and sometimes get individual feedback on their participation.

Summary of the Town Meeting. Although the Town Meeting described in this snapshot was lengthier than is usual, in other ways it typified how Town Meeting discussions work in Ann's classes. Ann usually solicits her students' input when selecting the issue that will be the focus of

the Town Meeting and always co-creates the various roles with her students. Students select which role they want to adopt for the Town Meeting and spend several class periods learning about the issue, and the perspective represented by their role, by reading articles, searching Web sites, listening to guest speakers, viewing videotaped programs, and talking on the phone with experts on the issue. On Town Meeting day the students sit in a circle behind a name tag that states their role. Ann begins the discussion by reminding students of the traits on an assessment rubric and asking them to go around the circle stating their role and position. Students control the content of the discussion, although Ann does not let students speak unless until she calls on them. Throughout the discussion, most students participate orally, in part because Ann has signaled the value of talking in discussion through her assessment rubric. The next class session includes a short debrief of the discussion. Students also receive their role sheets back with feedback from Ann and find out how she assessed their participation in the Town Meeting.

Chapter Five

Propositions and Reflections on the Study

In this chapter, I state and reflect on propositions that capture theoretically what skillful teaching of CPI discussion entails. These propositions, and relationships among them, were induced from the CPI discussion conceptions and practice of Joe, Elizabeth, and Ann. Thus, understanding this chapter depends on reading the portraits of the three teachers in Chapter Four and the data on which the propositions and their relationships are based. As a reminder, however, I begin with a short summary of each teacher's conceptions and practice. These are followed by four sections: (1) propositions that emerged from the study that, when related to one another, constitute an initial theory, (2) what this initial theory contributes to the literature on classroom discussion in the social studies and suggests for future research, (3) the implications of the initial theory for teacher educators, and teachers, and (4) limitations of the study. The chapter ends with a summary and some concluding thoughts about what I learned from doing the study that will inform my teaching and research about classroom discussion in the social studies.

The propositions cross the three research questions. As a reminder, however, the research questions undergirding the study were:

1. How do secondary social studies teachers who are skilled in the use of CPI discussions teach their students to participate effectively in such discussions?

2. What role do instructional strategies, issues, materials, and assessments play in this teaching process?

3. What conceptions account for these teachers' approaches to CPI discussions? I investigated the teachers' conceptions of four things: (a) democratic citizenship, (b) the purposes of social studies education, (c) what constitutes good discussion, and (d) their rationales for CPI discussions.

The Teachers: A Summary

Joe Park has taught middle and high school social studies for 22 years, the last five at a "break the mold" high school in a university community close to a major city in the western United States. Because his school has no set curriculum, Joe is able to create courses that closely mirror his conceptions of what knowledge and skills in social studies are most important for students to develop. One such course, Important Supreme Court Decisions, was taught for nine weeks in the fall and winter of 1997 and was the basis of my learning about how Joe uses CPI discussions. This course focused on landmark free speech decisions of the United States Supreme Court. Throughout the course, Joe's 9th- through 12th-grade students each week read the decisions in one case, completed pre-discussion preparation activities, participated in a lengthy seminar discussion, debriefed the discussion, and wrote papers about their opinions on the Court's decision. During the seminars, Joe is an active facilitator, asking many questions and working to keep his students focused on major ideas communicated in the Supreme Court decisions. Students are not formally graded for their oral participation in seminars, although they assess the seminar as a whole, and their individual participation, after each seminar has been completed. Joe also comments on their oral participation in the seminar on the narrative assessment he writes for each student at the end of the course.

Elizabeth Hunt has been teaching middle school language arts and social studies for 17 years. Her current middle school is located in a suburb outside of a large city in the West. Because her school is so large, students are divided into communities, each taught by a team of teachers. As the eighth-grade social studies teacher in one of the school's "communities," Elizabeth teaches a year-long course in American Studies that focuses primarily on civics and United States history. Throughout the course, Elizabeth selects CPIs for her students to discuss using a model, Public Issues Discussions, developed in the 1960s. Reflecting her interest in teaching her students the interpersonal skills necessary to participate effectively in such discussions, Elizabeth stresses equal participation and facilitates primarily to ensure that all students have the opportunity to participate orally. Using a formal assessment instrument, Elizabeth provides her students with feedback on their discussion skills and factors their preparation for and participation in discussion into their grades.

Ann Twain has been teaching social studies for five years at a magnet middle school in a suburb outside a large city in the Pacific Northwest. Because Ann's school has multi-age classes, she teaches social studies classes comprised of both seventh- and eighth-graders. The social studies class Ann has created combines United States history, civics, and world geography, laced with an extensive service learning program and whole-class discussions of CPIs in a format she has labeled Town Meetings. Ann and her students select the CPIs that will be the basis for the Town Meetings and then create a variety of roles representing people who would be interested in or affected by the CPI. After the students select roles to represent in the Town Meeting, they spend several days preparing by reading background articles, calling experts, and

viewing videotapes. During the Town Meetings, Ann facilitates with an eye toward the representation of multiple perspectives on the issue under discussion. Like Elizabeth, she uses a rubric to assess each student's oral participation in the discussion and counts both their preparation and participation as part of their formal grade.

Skillful Teaching of CPI Discussion: Beginnings of a Theory

From the similarities and differences among these teachers' CPI discussion conceptions and practice, one can induce statements that capture theoretically what skillful teaching of CPI discussion entails. The following propositions synthesize what can be learned from these cases. A proposition is something put forward for discussion; given the initial nature of this theory, this definition accurately represents how I want the propositions that make up this theory to be viewed. The use of the word "proposition" (instead of, for example, "hypothesis") is not meant to connote a lack of grounding for the ideas represented in each proposition. As discussed in Chapter Three, each proposition was induced from the data by following the tenets of grounded theory methodology.

Each proposition is described and explained in detail. I begin with an outline of the seven propositions. Please note that the word "teachers" in the wording of the propositions refers to skilled CPI discussion teachers. Following the outline is a description and explanation of each proposition.

Propositions

I. Teachers teach for, not just with, discussion. Discussion is both a method (of teaching students to create new knowledge, critical thinking skills, social studies content, interpersonal skills, and the like) and a desired outcome.

A. Teachers believe that discussion is a difficult outcome, so instructional time is devoted to preparing for, enacting, and debriefing discussions. Discussion is a priority in the teachers' curriculum.

B. Teachers teach their students discussion skills instead of presuming they already possess them.

II. Teachers work to make the discussions the students' forum.

III. Teachers select a discussion model and a facilitator's style that is congruent with their reasons for using discussion and their definition of what constitutes effective discussion. Thus, the selection and use of a discussion model is conception and rationale-driven. The selection and use of a particular discussion model creates tensions and tradeoffs that influence the type and quality of discussion in teachers' classes.

IV. Decisions about whether and how to assess students' participation in CPI discussions are influenced by an enduring tension between authenticity and accountability.

V. Teachers' personal views on CPI topics do not play a substantial, visible role in classroom discussion itself. However, teachers' views strongly influence the definition and choice of CPIs for discussion.

VI. Teachers engage in CPI discussion teaching practices that are informed by their conceptions of democracy.

VII. Teachers are receiving support for their CPI discussion teaching from school administrators, the overall culture of the school, and the school's mission. Thus, their CPI discussion teaching is aligned with, not in opposition to, what is expected in the school.

Description and Explanation of the Propositions

Teachers teach for, not just with, discussion. Discussion is both a method (of teaching students to create new knowledge, critical thinking skills, social studies content, interpersonal skills, and the like) and a desired outcome.

Across the conceptions and practices of the three teachers, there is evidence that an important reason they include discussion in their curriculum is to teach students how to participate effectively in discussions in other situations. Joe, for example, talks about the importance of scaffolding discussion instruction so students can participate in the "great conversations" of democratic society. Elizabeth's emphasis on interpersonal skills, such as listening and yielding, is aimed at an outcome larger than the development of such skills. Recognizing that speaking and listening to others is critical to the success of a society in which CPIs are the norm, not the exception, she aims to inculcate in her students habits of communication that will allow them to participate effectively as citizens in a democratic society. Ann's emphasis on preparing her students to participate actively in solving problems in their community drives her use of Town Meetings. Because she remembers what it felt like to be excluded from adult conversation about CPIs, she wants her students to feel they are citizens now--not just future citizens in training. Thus, an outcome of her use of CPI discussions is that students are familiar with and skilled at engaging in public discourse.

The primacy that these three teachers place on discussion as an outcome in its own right does not mean they are unconcerned about other outcomes that might result from students' participation in CPI discussions. Instead, it adds to the complexity of the rationales they hold for using discussion in the courses they teach. The breadth of their rationales for using

discussion and of the outcomes they hope their students will achieve as a result of participating in discussion also accounts for another phenomenon that extends across the conceptions and practices of the three teachers: their devotion of a large amount of class time to preparing for, participating in, and debriefing discussions.

If skilled participation in discussion is an outcome, it cannot be achieved through the use of other instructional strategies. For example, students cannot learn to participate more effectively in discussions by writing papers. Recall Joe's statement about teaching students to talk and listen in discussion: "There has to be some sort of environment in society where kids practice doing that. They practice baseball batting, for God's sakes, why can't they practice talking?" (J.P. IV#2, p. 7). A result of viewing discussion as an outcome is that these teachers devote a generous amount of classroom time to teaching students how to prepare for discussions, how to participate in them, and how to debrief them so they will be improved. These teachers have elaborate lesson plans and direct the full battalion of their pedagogical content knowledge to the lesson planning process. Everyone (teachers and their students) does a lot of work to prepare for discussion. Clearly, there is an opportunity cost associated with devoting so much classroom time to discussion. Simply stated, time spent on this instructional goal is time lost from other educational activities that have value for students. For these three teachers, however, the value they place on discussion as both a method and an outcome outweighs the opportunity costs involved.

A further elaboration of the proposition that teachers teach for, not just with, discussion is seen in the ways that instruction is provided to students on the skills needed to participate effectively in discussion. All of the teachers

involve students in setting guidelines for what constitutes effective CPI discussion. Their practice is also similar in the way they use the debriefing that follows each discussion to focus students on what could be done as a group and as individuals to improve the quality of the next discussion.

There are also differences in how they approach the teaching of discussion skills. Not surprisingly, the teaching is more direct with the younger students. Elizabeth, for example, provides direct didactic instruction on various discussion skills, such as recognizing that other participants want to join the discussion. Ann has developed the most elaborate process for helping her students form an understanding of the differences between high and low quality discussions. By showing her students videotapes of effective and less effective discussions at the beginning of the school year, she aims to help her students create a common vision of what good discussion looks and sounds like. Joe provides more implicit instruction through including an "expert discussant" in most of the seminars and modeling the use of questions as the facilitator. Notwithstanding the different approaches the three teachers take to providing instruction on discussion skills, the three teachers do not assume that students already possess these skills. Quite simply, they think that discussion skills must be taught continually--it's not a one-time lesson.

In the process of teaching for and with discussion, the teachers are making many instructional decisions that bring to the forefront a key issue about the use of classroom discussion. The issue is one of ownership and control, and can be phrased as: Whose discussion is this? Addressed in the second proposition, evidence from this study suggests that the teachers are

concerned about structuring the use of discussion to create a forum for their students.

Teachers work to make the discussions the students' forum.

This proposition means that key decisions about CPI discussions are jointly made by the teachers and the students. For example, as previously mentioned, all three teachers involve students in creating guidelines that will be followed throughout the discussions. These guidelines are made public and referenced periodically as a way to both hold students accountable for following them and to remind students that the guidelines represent the group's will.

A second way that the three teachers work to make the discussions the students' forum is through involving them in selecting some, but not all, the issues to be discussed. The two middle school teachers were particularly responsive to issues that their students identified as interesting. For example, at the beginning of the school year Ann asked all her classes to brainstorm issues that would be the focus of future Town Meetings. While she did not select all of the issues for inclusion in the curriculum, some were selected, signifying to students that their concerns mattered to her. Similarly, Elizabeth included issues that were sparked by her students' knowledge of current events, such as a recent school shooting in Arkansas, and tensions between the United States and Iraq. While Joe's curriculum was more set prior to the beginning of the course, his students did have a choice between two Supreme Court cases as the topic for their final seminar.

The teachers' roles as facilitators of the discussions most clearly demonstrate how the three teachers work to enhance the likelihood that students will view the discussions as their own forum. Even though the

teachers' facilitation styles differ, the unifying emphasis is on encouraging the students to speak to one another, and not go through the teacher. Furthermore, while students are encouraged to hold and state opinions on the issues, the teachers' opinions are not explicitly stated. This finding will be elaborated in greater detail under the fifth proposition, but it is important evidence for this proposition as well. In all of the discussions I observed, viewed, or listened to, never did I hear a student ask the teacher his or her opinion on the issue, nor did I hear the teacher volunteer a position. More than any other single piece of evidence, this suggests to me that the teachers viewed CPI discussions not as a classroom soapbox on which they stood, but as a forum for their students.

The power of the students is limited in all three teachers' classrooms. The discussion model that the students learn to use (and use to learn) is selected by the teacher, without input from the students. Although all three teachers asked for students' feedback on how the model was working for them, they did not suggest that a different model for CPI discussions could be selected. One can imagine a curriculum in which students are introduced to various models for discussing CPIs and then given some choice about which model would be used for various issues. To take that route, however, would negate the strong linkage between the models selected by the teachers and what they are trying to teach. As explained in the first proposition, because all three teachers are teaching discussion as both a method and an outcome, it appears they are retaining the power to select a discussion model as a way to ensure their multiple outcomes are being met. Thus, there is a tension between the first two propositions. If one has multiple outcomes for using

discussion in the classroom, then there are limits to how much power students are given to influence the discussion model that will be used.

Given that the teachers did retain control to select the discussion model taught to their students, it is important to explain on what basis discussion models were selected.

Teachers select a discussion model and a facilitator's style that is congruent with their reasons for using discussion and their definition of what constitutes effective discussion. Thus, the selection and use of a discussion model is conception and rationale-driven. The selection and use of a particular discussion model creates tensions and tradeoffs that influence the type and quality of discussion in teachers' classes.

This proposition suggests that a particular discussion model is selected because it concretizes and exemplifies how the teacher defines effective discussion and the rationales she/he holds for teaching for and with discussion. Moreover, once a model is selected, the style of facilitation that each teacher has created is also closely aligned with her/his definitions of and goals for discussion.

To illustrate this proposition, I will briefly restate each teacher's thinking about discussion and then link it to the model selected. First, recall that Joe believes that effective discussions involve a setting that promotes equality among participants, a sense that the participants "own" the content of the discussion, intense preparation on common content, active facilitation, and the creation of new ideas. Although he has many reasons for including discussion in his courses, he is most interested in promoting critical thinking and teaching students to participate in democratic discourse. His selection of the seminar model is so closely aligned to these conceptions of discussion to

be almost tautological. In seminar discussions, the creation of new ideas and understandings is paramount. But, unlike the models used by the other two teachers, the precise content that is being examined, challenged, analyzed, and extended must be shared in common. Instead of providing students with many different articles on the various free speech CPIs that formed the basis for the course we learned about in Chapter Four, Joe selected Supreme Court decisions that all students read. This common text is one of the primary attributes of the seminar model and is intrinsic to how Joe thinks about the primary goal of discussions, which is to teach students how to think. Without a common text, Joe fears that the students would not have anything in common about which to think.

Joe's emphasis on improving students' critical thinking skills is also linked to the active facilitation role he takes during seminars. Recall that Joe's facilitation, primarily through questions, took up approximately 30% of the seminar air time. An apparent contradiction is indicated between Joe's desire to have students view the discussion as their own and the fact that such a high percentage of the available time is taken by his questions. Joe, however, does not recognize any such contradiction because his active facilitation is designed to scaffold instruction on critical thinking. Joe sees his responsibility as the facilitator as both acting as a traffic cop and leading students in critically examining the ideas in the text.

Elizabeth and Ann also have selected discussion models and created facilitation styles that are congruent with how they define effective discussion and conceptualize the purposes for teaching for and with discussion. Elizabeth wants her students to develop interpersonal skills, an understanding of social studies content, critical thinking skills, and the ability

to participate in democratic discussion of difficult issues. The Public Issues Model she has selected is aligned to those goals. For example, as we learned in Chapter Four, this model separates issues into three types: factual, definitional, and ethical. Understanding the three types of issues is a significant part of the preparation that students engage in prior to the actual discussions. Her attraction to this model illustrates her interest in teaching the content of social studies, for the separation of issues into these sub-types reflects core social science content.

Through analyzing the choices Elizabeth makes about the role she will take as the facilitator, we see a clear intersection between what she values and what she does. Of the attributes that Elizabeth thinks effective discussion should have, two stand out as particularly important: the demonstration of interpersonal skills and equal participation among the discussion participants. Highly concerned with teaching her students to be civil discussants, virtually all of Elizabeth's comments as a facilitator are directed toward that goal. When she does speak as a facilitator (which is rare), her comments typically focus on the need for students to listen, not monopolize, and to draw other participants into the conversation. This facilitation style creates a tension between deep analysis of particular social studies content versus widespread and relatively even participation in the discussion. Elizabeth is not aware of this tension because the twin goals of interpersonal skills and equal participation loom so large in her conceptions of discussion.

Finally, Ann's conceptions and practice also support the proposition that the selection of a discussion model and a facilitation style is rationale-driven. Given that Ann specifically created a discussion model to implement her ideals for discussion, it is not surprising that such a tight link exists

between what she values and what she does. Particularly concerned about how multiple perspectives on issues are raised in discussions of CPIs, Ann has created a model that forces different perspectives to come into play in the different roles her students assume. While the students do not always stay "in role," they at least start that way and, as a result, the various ways that different people and groups would be influenced and affected by CPIs form the basis of CPI discussions in her classroom. When facilitating CPI discussions, Ann makes an effort to call on roles that have not yet been represented, again indicating a link between how she behaves and what she values. Ann's emphasis on multiple perspectives creates a tension not dissimilar to what is often seen in democratic discourse outside of the school. When multiple perspectives on an issue is the primary goal, the depth of exploration on particular parts of the issue may be sacrificed. Ann appears not to be aware of this tension.

In addition to those created by the discussion model and facilitator's role selected by each teacher, there are ways in which tensions between competing goods influence their use of CPI discussions. This was most evident in the decisions made by the teachers about the assessment of their students' participation in CPI discussions.

Decisions about whether and how to assess students' participation in CPI discussions are influenced by an enduring tension between authenticity and accountability.

One of the clearest differences in how the teachers approach CPI discussion teaching is seen in the decisions they make about assessment. I am defining assessment broadly to include how teachers and students find out and make judgments about students' progress toward desired educational

outcomes. This definition collapses the distinction that is often made between assessment as finding out and evaluation as judging (Parker & Jarolimek, 1997). All three of the teachers assess their students' participation in discussion. Both of the middle school teachers assess their students' preparation for and participation in CPI discussions through the use of codified rubrics. Both of them also count how their students prepare for and perform in discussion as a formal part of their grades. Holding students accountable for their preparation and performance in CPI discussions and rewarding oral participation in CPI discussions reflect the middle school teachers' concern about aligning their assessment procedures to what is valued in their classrooms. Conversely, while Joe does assess his students' seminar participation, he is adamantly opposed to the grading of seminar participation, because he believes that "paying kids to talk" is inauthentic. That is, it does not represent the way public discourse operates in the world outside of school. Moreover, Joe believes that grading oral participation would be at odds with the creation of effective seminars, in which participants should talk because they have something to say, not because they are being rewarded by an authority figure.

Thus, comparing and contrasting the assessment practices of the three teachers makes apparent a tension between accountability and authenticity. Ann and Elizabeth have chosen to privilege accountability because they believe that if they value discussion, assessing it in a fairly formal way with a rubric delivers the message of its importance. Additionally, formal assessment gives them the opportunity to provide specific feedback to students about what they do well and what they still need to improve. Explicitly then, both Ann and Elizabeth have made a choice about the issue of

whether students should be required to participate orally in CPI discussions. They have decided that requiring such participation (through formal assessment) is important because it communicates a message that democratic discourse is a critical outcome of their curriculum.

Joe, on the other hand, has chosen to privilege authenticity. Because he believes that "paying kids to talk" will destroy the genuineness of a seminar, he is willing to allow some students to remain silent. While students in Ann's and Elizabeth's classes can also make that choice, there is a cost involved. No such cost exists for students in Joe's class.

Just as the teachers' views on whether authenticity or accountability should be given more weight when making decisions about assessing CPI discussions, there are other ways in which teachers' views influence their CPI discussion teaching practices. In particular, the teachers' personal views on CPI issues mattered--but not in the way I expected they would.

Teachers' personal views on CPI topics do not play a substantial, visible role in classroom discussion itself. However, teachers' views strongly influence the definition and choice of CPIs for discussion.

The literature on CPI discussions suggests that what is important about teachers' personal views on issues is the way and extent to which they are directly communicated to their students in the classroom (Lockwood, 1995; Kelly, 1986). This proposition suggests that the disclosure of teachers' personal views on issues is not what is significant to the practice of these skilled discussion teachers. As mentioned in the explanation of proposition two, there was <u>not a single example</u> in the many discussions I observed, listened to, or viewed of a teacher being asked his or her personal opinion on an issue by a student. Moreover, none of the teachers <u>ever</u> volunteered their personal

opinion on an issue. While this lack of disclosure is evidence that the discussions are the students' forums, it does not mean that teachers' personal views on issues do not matter in their CPI discussion thinking and teaching. The teachers' personal views on what is a CPI in the first instance, and on specific CPIs, clearly influences what issues students are allowed to discuss and the materials they are exposed to when preparing for discussion.

As an example of how teachers' personal views about CPIs influence their discussion teaching, Figures 7, 9, and 11 illustrate the different reasons the three teachers gave for not including gay rights as a CPI in their curriculum. Note that Joe does not believe that gay rights issues are CPIs. Instead, he likens such issues to human rights issues on which there are no legitimate differing views. Thus, he does not select gay rights as CPIs because his personal value system directs him to treat such issues as moral issues with one clearly right position. About gay rights, Joe stated, "The correct answer is that people should not be discriminated against on the basis of race, gender, ethnicity, sexual preference, physical disability" (J.P. IV #6, p. 10). Likening the denial of equal rights for gays to historical abuses of human rights, such as slavery and Nazism, Joe advocates including gay rights in the curriculum, but as an example of the denial of civil rights, not as a CPI.

Elizabeth and Ann's decision not to include gay rights as a CPI discussion topic also reflects personal views, but their views are different from Joe's. Elizabeth's personal discomfort with gay rights issues keeps her from including them as CPIs open for discussion in her classroom. Additionally, teaching in a conservative community, she worries that including gay rights issues would spark too much controversy and community disapproval. Ann does not include gay rights issues because she

worries her students would not discuss them with sensitivity and that gay students in her classes would feel uncomfortable.

Another way in which the teachers' views influence their discussion teaching practice is found when exploring their conceptions of democracy and social studies.

Teachers engage in CPI discussion teaching practices that are informed by their conceptions of democracy.

All of the teachers in this study are teaching for and with discussion, at least in part, because they see a connection between a healthy democracy and a citizenry that can participate skillfully in discussions of CPIs. While there are many similarities in the teachers' discussion conceptions and practice, there are key differences as well, especially evident in the model of discussion used in the classroom and the role the teachers take when facilitating and leading discussions. One reason for these differences is that they represent differences in how the teachers view what should happen to ideas in a democracy.

Joe believes that ideas should be collaboratively created by discussants. Recall the bread-baking metaphor he used to explain what constitutes effective discussion. Various ideas put on the discussion table were like ingredients in the making of bread, and only through combining them would true democratic discussion occur. In the process of this combination, ideas would be challenged and new understandings would emerge. This conception places a premium on the ability of discussants to create, which represents his view that personal agency is what matters in a democracy. People must feel they have the ability to create new understandings and new solutions to problems. His selection of the seminar discussion model is closely aligned with this understanding of what happens to ideas in a

discussion. That model demands the exploration and creation of ideas through a tightly-focused analysis of specific text. Moreover, his active role as a discussion leader is designed to facilitate his students' exploration of and creation of ideas. Recall that virtually all of his oral participation as a discussion facilitator involves asking questions and encouraging his students to challenge the ideas in the text and those created by their classmates. By taking this role, Joe believes he is scaffolding and modeling what citizens should do in a democracy--create new ideas and understandings through the analysis and challenging of ideas.

Elizabeth's discussion teaching is also informed by her conceptions of what should happen to ideas in a democracy. Foregrounding equality as a democratic principle, she is most interested in teaching her students that all discussants have both a right and a responsibility to share ideas about the CPI that is being discussed. Unlike Joe, who insists that students read and discuss a common text, Elizabeth carefully selects different materials geared to the reading ability of various students. In this way, she is hoping that all students will be able to contribute to the discussion. Thus, she is differentiating background information in order to achieve equal participation in the discussion. The discussion model she has selected represents her belief about the importance of sharing ideas in discussion. Because the model focuses on an issue instead of a specific text, all of her students are able to bring something to the discussion. Her primary concerns are that all students participate orally in the discussion and that they learn to do so in a civil manner. Her facilitation style further exemplifies the importance she places on the equal sharing of ideas in discussion. Recall that she intervenes rarely,

but when she does say something the goal is to encourage civil behavior and equal participation.

Like Elizabeth, Ann also believes that ideas should be shared in a democracy. Her concern with the representation of multiple perspectives, however, illustrates her interest in ensuring that the ideas that are shared reflect the interests and concerns of a broad range of people. Although Ann differentiates the materials students use to prepare for discussion, the materials are selected for both accessibility in terms of reading level and the extent to which they represent the views of various groups in society. In her creation of a discussion model, Ann's interest in sharing ideas that represent multiple perspectives is also evident. The Town Meeting model that Ann has created mandates the representation of different views because students select roles designed to represent a wide range of opinions about a particular CPI. In her facilitation of the Town Meetings, Ann works to get a variety of perspectives represented, which results in fairly equal participation because the students are in different roles.

The final proposition shifts the focus away from how the teachers' conceptions inform their practice to how the school context in which the teachers are working supports and influences their ability to teach CPI discussions.

Teachers are receiving support for their CPI discussion teaching from school administrators, the overall culture of the school, and the school's mission. Thus, their CPI discussion teaching is aligned with, not in opposition to, what is expected in the school.

Representations of skilled teachers in popular culture, especially in film, often portray them as remarkable because they differ from what is

valued in their schools. This proposition suggests the opposite--these skilled discussion teachers are supported by the larger school environment and are teaching in alignment with what is valued by others in the school community.

Joe's school was created to be an alternative to the traditional high school. Formed on a foundation of constructivist learning theory, the school day and school week are purposely structured to encourage seminar discussions. While not all teachers use seminar discussions, they are not unusual. Consequently, some of Joe's students come into his classes with experience in participating in discussions and can act as role models for others without such experience. Moreover, the principal of Joe's school participates frequently and actively in seminar discussions in Joe's classes. This level of support from a principal is unusual and undoubtedly contributes to Joe's sense that what he is trying to teach his students is considered important by others in the school. Finally, Joe is able to create a unique curriculum that is particularly well-suited to the discussion of CPIs. The Important Supreme Court Cases class focuses just on First Amendment cases, allowing his students to analyze and deliberate about CPIs embedded in the First Amendment with more depth than would be typical in a high school curriculum. Joe has found a school that values what he thinks is important and, as a consequence, does not need to spend his time convincing others to let him teach in a particular manner. His time can be devoted to the difficult task of planning instruction on the discussion of CPIs.

Like Joe, Elizabeth has also found a school environment that matches what she thinks is important. Although her school is much more traditional than the one in which Joe teaches, it matches her ideas about the kind of

school experiences that are appropriate for middle school-age students. The curriculum that Elizabeth teaches is one that she helped to create and it allows her enough flexibility to select issues for discussion that are relevant to her students because they are current matters of public deliberation. Recognizing that the school and community are fairly conservative, Elizabeth is careful when selecting issues and shies away from those that would directly challenge community norms. Like Joe, Elizabeth is comfortable in the school and feels like she is a valued team member.

Ann has also found a school that is particularly well-suited to her conceptions of the type of education that young people should experience. Like Joe's school, hers was formed as an alternative to the traditional schools that are more common in her school district. Her school values innovation and provides support for the curriculum she has created. For example, the school's daily schedule can be rearranged to allow a longer period for Town Meetings. While her principal does not participate in Town Meetings as frequently as Joe's principal participates in seminars, she does participate in at least one per year. In their other classes, Ann's students are frequently asked to participate orally in various kinds of discussions, which results in a school-wide norm that supports the kinds of behaviors that Ann is teaching her students in the Town Meetings.

All of the teachers in the study are in the fortunate position of working in a setting that reinforces and supports their teaching practice. As is the case when buying real estate, the lesson here may be that one of the factors that is important to the skilled teaching of CPI teaching is location, location, location.

From Propositions to an Initial Theory

The seven propositions are the building blocks of an initial theory about skilled CPI discussion teaching. Here I will identify and describe ways in which the building blocks connect to one another toward the goal of building an initial theory. As recommended by Glaser and Strauss, this initial theory is based on the idea of "theory as process; that is, theory as an ever-developing entity, not a perfected product" (1967, p. 32).

Recall, in Chapter Two I presented the conceptual framework on which this study was based. Figure 12 is a revised conceptual framework that illustrates ways in which the propositions are linked to one another. These linkages form an initial theory because they suggest relationships among the propositions. The relationships are broader and more abstract than the propositions, giving them more explanatory power.

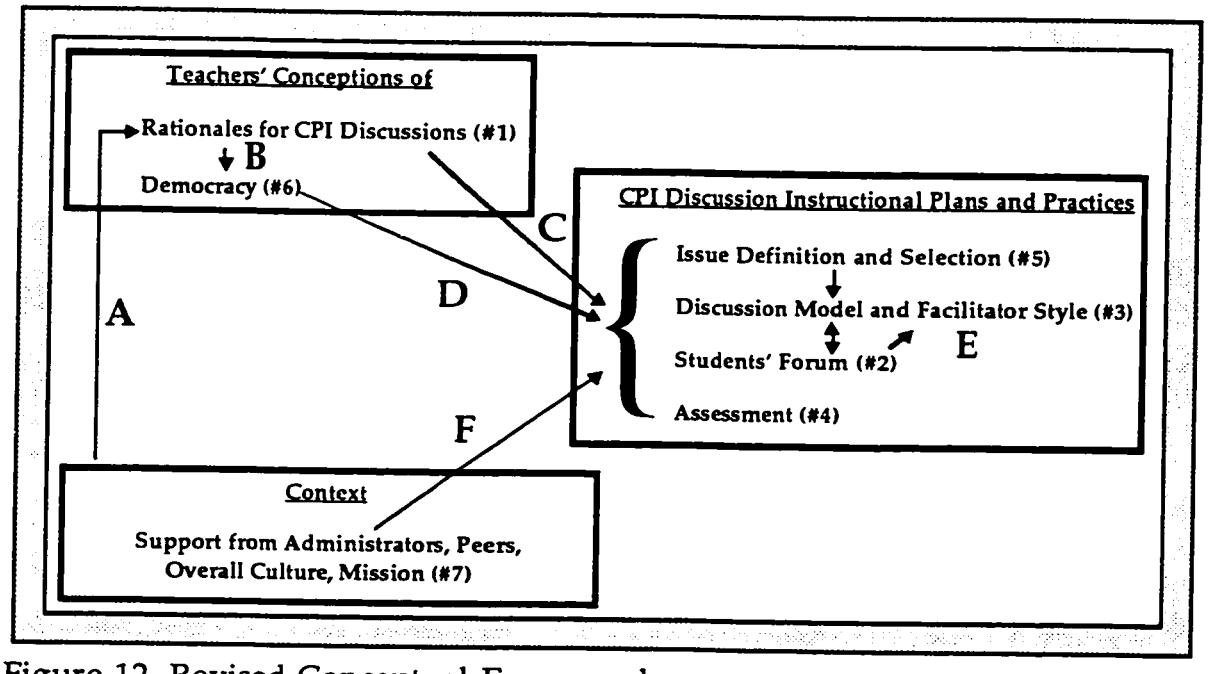


Figure 12. Revised Conceptual Framework

While many of the relationships are suggested in the explanations of the propositions, here I briefly describe each of the six major ways in which the propositions influence one another. The first relationship is marked "A" in the figure. It shows that propositions one and seven are linked. Recall, that the first proposition explains the multiple rationales that teachers have for teaching CPI discussions and further states that discussion is a priority for the teachers because there are so many rationales they have for teaching it to their students. Linkage "A" illustrates that the teachers have support from the school community and school mission for making discussion a priority in their curriculum.

The teachers' multiple rationales for CPI discussion are also linked to their conceptions of democracy, a relationship that is shown in "B" on the figure. Teachers believe that a healthy democracy depends on a citizenry that can effectively participate in discussions of CPIs. That is exactly what "teaching for discussion" is designed to create. Skillful discussion participation is one of the desired outcomes because the teachers believe it will lead to a healthier democracy.

A third way in which the teachers' rationales for CPI discussion are important is seen in the relationship marked "C" which illustrates how the teachers' instructional plans and practices are informed by their rationales for teaching discussion. The most specific relationship in this category was previously explained in proposition three. Recall, that proposition states that teachers select a discussion and a style of facilitation based on their rationales for using discussion. There are, however, other ways in which rationales influence plans and practice. For example, if a teacher is especially concerned

with using CPI discussions to teach interpersonal skills, the assessment of a students' discussion ability would focus on those skills.

Teachers' conceptions of democracy influence their CPI discussion teaching plans and practices, labeled relationship "D" in the figure. As previously explained in proposition six, what the teachers think should happen to ideas in a democracy affects the selection of a discussion model and the style of facilitation used during discussions. What the teachers think about democracy also influences their interest in working to make the discussion the students' forum--seen at its clearest in the unexpected finding that the teachers never disclosed their personal views on a CPI to their students.

Within the category of instructional plans and practices are relationships between students' forum and facilitation style, students' forum and issue selection, and students' forum and model selection. Marked as "E" in the figure, these relationships have already been explained within the propositions. It is important to reiterate, however, that the tension between creating a students' forum and selecting a discussion model represents yet another way in which the teachers' multiple rationales for discussion are important. That is, teachers select a discussion model that is best suited for the many rationales they hold for CPI discussions. The teachers' "voice" is strongest with this decision. Perhaps to balance the power somewhat, students are given more of an influence when selecting issues.

The final relationship illustrated in the revised conceptual framework figure is marked "F" and connects the larger school environment to the teachers' CPI instructional plans and practices. Beginning with the big picture, the very fact that the teachers can devote so much time to teaching CPI

discussions without getting any administrative or peer disapproval is important, for these are schools in which the teachers and administrators talk about the curriculum. That is, these schools are not like some where the twin norms of privacy and autonomy result in teachers and administrators not knowing what others are doing. In more specific ways, the teachers are also receiving support from others in the school community for their CPI discussion teaching. For example, in two of the schools, participation in CPI discussions is represented on the social studies portion of the report card sent home to parents. Another example is the involvement of two of the schools' principals as participants in CPI discussions.

Taken together, the relationships between the propositions form an initial theory that explains how skilled CPI discussion teachers conceptualize and practice with respect to CPI discussions. What this theory contributes to the relevant literature, and how it could inform the work of teachers, teacher educators, and researchers is discussed in the next section.

Contributions to the Literature and Implications for Research

This study elaborates, specifies, and challenges existing literature related to classroom discussion in social studies in three areas: teachers' conceptions of discussion, assessment of discussion, and the influence of teachers' personal views on CPIs. First, I analyze the relationship between the initial theory presented in this study and what previous research has shown about teachers' conceptions of discussion and what influences those conceptions. Second, I discuss the tension between authenticity and accountability in light of recent literature on authentic assessment. Third, in the area of curriculum content, I explain how this study challenges what previous literature states is important about teachers' personal views on CPIs.

Following each explanation of the contribution this study makes to the literature, I suggest further research related to each of the three areas.

Teachers' Conceptions of Discussion

Recent research on teachers' conceptions of classroom discussion has shown that high school social studies teachers have multiple conceptions of discussion and that these conceptions are variously implemented based on the objectives of a lesson (Larson, 1995). Moreover, teachers use discussion to accomplish varied goals, which can be separated broadly into process and product (or outcome) categories (Larson, 1997). This study confirms, challenges, and adds to these important empirical findings on classroom discussion. As explained in the first proposition of the theory, teachers use CPI discussions as both an avenue to accomplish multiple objectives (such as critical thinking and interpersonal skills) and an outcome. This is exactly what Larson found in his study.

Larson's (1995) finding suggests that teachers use a variety of types of discussion, selected to achieve different goals. This study found the opposite--that each teacher uses just one model of CPI discussion, selected based on the teacher's rationales for discussion and definitions of what constitutes effective discussion. There are at least two possible explanations for the differences in Larson's findings and my own. One is that this study concentrated on discussions of CPIs, whereas when Larson selected his sample of six high school social studies teachers, he sought those who used classroom discussion of various topics, not just CPIs. This narrowing of discussion topics to just CPIs, by definition, may have resulted in the selection of teachers who were more prone to concentrate on just one type of discussion. Another explanation is that teachers who are selected because they are considered

skilled discussion teachers may be more likely to use just one model because it allows multiple practice opportunities for their students. If students use the same model time and time again, it is more likely they will become fluent in participation in that type of discussion. Larson's sample was selected based on their principal's nomination of them as thoughtful and effective teachers, which does not necessarily mean they were especially skilled at teaching discussion.

What this study adds to Larson's (1995) findings is another explanation of what accounts for teachers' various conceptions of discussion. As explained in proposition five, teachers' conceptions of discussion (and, by extension, what they consider effective discussion) are informed by their conceptions of democracy. Given that the teachers in this study are using CPI discussions, at least in part, because they see a connection between participation in such discussions and a healthy democracy, it logically follows that their ideas about democracy will inform how they conceptualize discussion. The teachers' differing notions about democracy, then, help to explain why each selected a particular discussion model, and what kind of communication is valued in discussion. This finding adds to Larson's (1995) explanation of how lesson objectives influence teachers' conceptions of discussion. In his study, teachers used various types of discussion based on their particular objectives for a lesson. In this study, teachers' lesson objectives for discussion were informed by broader, overall objectives based on their conception of democracy.

Further research could broaden and deepen the emerging theoretical framework on teachers' conceptions of discussion. Working with the distinction between teaching for and with discussion, two questions are important for further research. First, how does conceptualizing discussion as

both a democratic outcome and a method relate to skilled discussion teaching? Given that the teachers in Larson's study were not selected because of their discussion teaching prowess, and that only three teachers participated in this study, it is necessary to explore further whether the connection that I am suggesting between teachers' conceptions and skilled discussion teaching holds up or is repudiated or modified when additional skilled teachers are studied.

In addition to increasing the number of teachers who are studied, it is also important to investigate whether the hypothesized connection between skilled discussion teaching and for/with conceptions applies to teachers working with a more racially diverse student population. This initial theory was formed based on the thinking and teaching of three white middle-class teachers who teach, for the most part, white middle-class students. Thus, the initial theory begs to be "tested" in other contexts that are more representative of the socioeconomic, racial, and ethnic diversity that exists in the United States, and in which CPI discussion teaching may take on new dimensions.

Assessment of Discussion

Literature on classroom assessment suggests that the most powerful assessments of students' learning are classroom-based (as opposed to district or state level) and tightly aligned to curriculum and instruction (Miller & Singleton, 1997; Stiggins, 1997). Additionally, assessment experts (Martin-Kniep, 1998; Newmann & Wehlage, 1995) recommend that teachers assess students' progress toward goals that are valued in the world beyond school, which is often called authentic assessment. Educators who specialize in the assessment of CPI discussions (Harris, 1996) recommend formal assessment of students' participation in discussion as a way to communicate to students that

discussion is valued and to provide students with the specific feedback they need to improve their discussion skills. As explained in proposition four, skilled discussion teachers vary in how they approach the assessment of students' participation in CPI discussions. Framed as a tension between accountability and authenticity, this proposition both reinforces and challenges the literature on classroom assessment.

The teachers in this study who gave precedence to accountability over authenticity did so because they felt formal assessment (i.e., using rubrics and grading) of discussion participation communicates the importance they place on discussion to their students. It also provides their students with a sense of how they are progressing toward the discussion goals the teachers had identified and codified in the discussion rubric. Equally important, however, is the explicit decision the teachers who are formally assessing have made about whether oral participation is required of all students. In short, all students must talk or pay a price for their silence. The advantage of this stance toward the "choice" issue (i.e., whether students may choose to be silent without penalty) is that it reinforces high and common standards. Unlike some classrooms where only the already-verbally proficient students participate orally in classroom discussion, these teachers recognize the connection between practice and progress. Grading students' participation orally in discussions is, therefore, an example of the connection assessment experts see between what is assessed and what is communicated about the importance of all students' learning.

But, there is a downside to formal assessment of discussion participation, which is captured in the phrasing of proposition four as a tension between authenticity and accountability. Recall that Joe refuses to

formally assess students' oral participation in discussion because to do so could jeopardize the authenticity of the discussion. This reasoning directly challenges the literature on authentic assessment because it underscores the problems associated with common standards for students. Common standards only work if there is agreement about what good performance looks and sounds like, and it may be that discussions, especially those that occur in a large group, work best if participants are behaving in different ways. Authentic examples of CPI discussions that occur in the world beyond school do not demand that all participate in the same manner. Think of a particularly good discussion among community members about how to solve a public problem. We would expect that some people would talk more, and some less. We would expect that some people would use analogies to explore the problem, while others would use statistical evidence. We would expect that some people would ask many questions, while others would use examples from their personal history to explore the problem or suggest solutions. In short, we would expect difference. Yet, discussion rubrics that are specific enough to be helpful to students do not allow for difference. They explicitly identify common ways that people should behave in a discussion. Thus, proposition four, which identifies a tension between accountability and authenticity, can be viewed as a direct challenge to the advantage of high and common standards that are assessed in a meaningful manner.

Assessment experts (Martin-Kniep, 1998; Newmann & Wehlage, 1995) are clear that two of the purposes of classroom assessment are to improve teaching and learning. With respect to the assessment of students' participation in CPI discussions, further research is needed to examine whether the tension between accountability and authenticity is serious

enough to inform decision-making about whether and how to assess classroom discussion. A study of what influence various kinds of assessment practices have on students' abilities to participate effectively in classroom discussion of CPIs is needed. It may be that a student study could shed light on whether losing some authenticity is an acceptable trade-off if the result is enhanced student learning.

Teachers' Views of CPI Issues

Little is known about how teachers select CPIs for classroom discussion: accordingly another contribution this study makes to the literature is to demonstrate/identify that, at least for some teachers, their personal views on CPIs inform their selection of issues. As was explained in proposition five, these skilled discussion teachers do not believe that all CPIs have equal curricular value. In fact, just the opposite was found. These teachers select CPIs based on a variety of factors--one of which is their own view about whether the issue is really a CPI, and, if so, whether it meets enough of the criteria they have created for content selection to warrant inclusion in the curriculum. Recall, that some issues were also not selected because of the teachers' personal discomfort with the issue, concerns about whether students could discuss particular issues in a sensitive manner, or worries of community disapproval.

This is a particularly important finding because of its potential to influence the focus of scholarly discussion on how teachers' personal views on CPIs influence their discussion practice. Previously, most of the literature on teachers' personal views on CPIs focused on whether they should be shared with their students (Lockwood, 1995; Kelly, 1986). But, that question presumes that a CPI has already been selected for discussion. This study

suggests that an equally important question for teachers and researchers is how teachers' personal views inform the decisions they make about what is discussed in the first instance.

Further research is needed on how teachers select CPIs for classroom discussion. In particular, a more in-depth exploration of how teachers define a CPI is necessary in order to deepen understanding of what teachers consider legitimate controversial content and how that content should be learned by students. Recall, that Joe conceptualized gay rights not as a CPI, but as a human rights issue. His personal views clearly influenced his definition of what constitutes a legitimate CPI. A larger and more diverse sample of teachers is needed to further investigate whether and how the distinction between human rights issues and CPIs is informing the content selection and pedagogical decisions made by teachers.

Additional Research

In addition to the research that stems directly from this study's findings with respect to teachers' conceptions, assessment, and issue definition and selection, there are other studies for which this dissertation has established a need. Here I briefly describe three additional studies.

This study sheds no light on the question of what, in fact, the students of these teachers learn from participating in CPI discussions, or whether the discussions do, in fact, improve their ability to participate, relative to their skills before discussion instruction. A study of what the students of skilled teachers learn from discussion, how they learn to participate more effectively in discussions, and how they experience their learning would contribute to the literature on classroom discussion.

If other researchers think, as I do, that these teachers are skillful, then the question is raised: How did they learn to teach in this manner? There has been virtually no research on the effect of teacher education or professional development programs on teachers' ability to teach for and with discussion. A study that investigated the effect of such programs on teachers' ability to teach discussion would be a particularly helpful contribution to the literature. A research question for this study is: In what ways do various professional development programs on classroom discussion influence teachers' discussion teaching practice?

Finally, another way to develop the initial theory presented in this study would be to have the three teachers whose conceptions and practices the theory was based on work together to examine, challenge, and extend the discourse among them about their conceptions and practices. The initial theory might thereby result in a revised theory that is richer and more satisfying than the one that I, as a single researcher, have constructed. I envision a research process that begins with sharing the entire dissertation with each of the three teachers. Then, each teacher would be asked to videotape a classroom discussion that is fairly representative of how CPI discussions work in his or her classroom. The videotapes would be viewed by all three teachers. Through the use of focus group methodology, the three teachers and I would work together to examine discussion practice and develop a revised theory.

Implications of the Theory for Practice

This study can contribute insights to the practice of teacher educators and secondary social studies teachers.

Implications for Teacher Educators

The problem this study was designed to address is that few secondary social studies students are given opportunities to learn to participate effectively in classroom discussions of CPIs. Teaching tomorrow's secondary social studies teachers how to teach students to discuss such issues is one way of addressing the problem. Teacher educators, then, have a responsibility to teach preservice students about classroom discussion. In what ways might this study help them fulfill that responsibility? Here I suggest three specific ways that teacher educators may want to use this study to inform their teaching of preservice students.

The portraits could be used to counteract the "apprenticeship of observation" (Lortie, 1975) that causes many teacher education students to lack a conception of what constitutes effective classroom discussion and/or to understand that discussion must be taught (Parker & Hess, forthcoming). One of the difficulties teacher educators face is that there are few models of skillful classroom discussion available to use when teaching preservice students. Although there are some videotapes showcasing the work of skilled discussion teachers, a finding from this study is that much of the work of discussion teaching occurs in the planning and preparation for discussion. Thus, viewing a videotape of a classroom discussion does not adequately explain the many steps that the teacher and students go through to prepare for discussion. Teacher educators may want to have their preservice students read and discuss the portraits as a way to bring to the fore the elaborate lesson plans that each teacher has developed.

The lesson plans embedded in the portraits could also be used by teacher educators who, themselves, lack confidence in how to teach their

teacher education students to participate effectively in classroom discussions. For example, if a teacher educator was interested in teaching one of the discussion models showcased in the portraits, the teachers' lesson plans could provide useful guidance in how to prepare for and structure the discussion. It may be that combining demonstration lessons using a particular model with reading the accompanying portrait could be helpful way of engaging preservice teachers in an analysis of both how they experienced the model and how it was used with secondary social studies students.

In addition to using the portraits, teacher educators may want to share the original and revised conceptual frameworks in this study with preservice students to help them think about possible interactions between teachers' conceptions, instructional plans and strategies, instructional practice, and contexts. By doing so, teacher educators would be working to complexify preservice students' ideas about the purposes and teaching of classroom discussion. My experience suggests this would be a helpful because many preservice teachers fail to grasp the reality that teaching students to participate effectively in classroom discussion is a challenging enterprise that requires intense preparation and focus.

Implications for Secondary Social Studies Teachers

The three portraits, and the propositions and initial theory that emerged from them, can help teachers reflect on, challenge, and strengthen their practices. By reflect on, I mean they can be used as "teaching cases" to bring to the fore issues related to teaching for and with discussion. For example, teachers may want to compare and contrast the cases as an avenue toward deeper understanding of important instructional decisions. Considering, for example, the difference between how Joe, Elizabeth, and Ann

approach the assessment of discussion may help teachers become aware of the tensions and tradeoffs embedded in this process.

By challenge, I mean the portraits and initial theory can be used to help teachers examine their own practices in relation to what these three teachers are doing. For example, one of the common problems in classroom discussion is that many students choose not to participate orally (Bickmore, 1991). Yet, in all three of these teachers' classes, few students are making that choice--in fact, quite the opposite. Teachers may want to analyze the three portraits as a way to challenge the oft-stated view that some students will just never participate orally in classroom discussion.

By strengthen, I mean that the portraits and the initial theory can be used to improve teachers' thinking about, and practice of, classroom discussion. For example, if a teacher is intrigued by seminar discussions, she may want to experiment with the detailed lesson plan that Joe has developed to teach his students to "do the work of" seminars. A teacher who is interested in foregrounding multiple perspectives in the social studies curriculum may want to adopt and/or adapt Ann's Town Meeting model.

Methodological Strengths, Limitations, and Issues

Focusing on and evaluating the credibility of this study illuminates its strengths and limitations. In making judgments about this study, I rely on how experts in qualitative research discuss the term "credible." Credibility is broadly synonymous with trustworthiness (Lincoln & Guba, 1985, p. 301) and is to the qualitative researcher what internal validity is to the quantitative researcher. As Miles & Huberman stated, credibility is concerned with truth value and can be assessed by asking such questions as: "Do the findings of the study make sense? Are they credible to the people we study and to our

readers? Do we have an authentic portrait of what we were examining?" (1994, p. 278)

Assessing the Credibility of This Study

In assessing the credibility of this study, I focus on two large categories – the quality of the data and how it was interpreted.

Data Quality. Beginning with the data on which the entire study is based, it is important to analyze whether the data was accurate and complete. As Maxwell states, "The main threat to valid description, in the sense of describing what you saw and heard, is the inaccuracy or incompleteness of the data" (Maxwell, 1996, p. 89). I assess the accuracy of the data in this study as high and the completeness of the data as relatively low.

I rate the accuracy of the data as a strength of this study because of the techniques I used to capture the data, the ways in which I triangulated the data, and the use of member checks to correct for mistakes in the data. I discuss each in turn. As explained in Chapter Three, various types of data were collected. For example, I interviewed the teachers, collected artifacts of their CPI discussion conceptions and practices, and observed many CPI discussions. All of the interviews were audiotaped and transcribed, which obviously allowed for a precise record of what was said during the interviews. All of the classroom discussions I observed were also either audiotaped or videotaped, and I took field notes as well. Thus, I had an accurate record of exactly what transpired during the discussions. Taping the discussions was an important check on the accuracy of my field notes. As a novice researcher, I recognize that taking good field notes is a craft that I must continue working to develop. By relying on the tapes along with the field notes, I could recreate

a more accurate and complete picture of what happened during the discussions.

Furthermore, the multiple types of data collected allowed for triangulation of the data. Triangulation, in a qualitative study of this type, means that different methods are used to collect data from the same source (Lincoln & Guba, 1985, p. 306). As a measure of credibility, triangulation is important because it decreases the possibility that the data collected is not representative of what one is trying to understand. For example, in this study understanding how the teachers defined effective classroom discussion was important. I first had the teachers define effective discussion through watching the videotapes of discussions in other teachers' classes. This strategy was designed to reduce the tendency for teachers to create a definition that fits their own practice. Additionally, in later interviews, I asked the teachers to assess the discussions I had observed, viewed, or listened to. Thus, I used different methods to solicit data that could help me understand what these teachers thought constituted effective discussion.

A third reason I rate the accuracy of the data in this study positively is because of the use of member checks. Member checks, "whereby data, analytic categories, interpretations, and conclusions are tested with members of those stakeholding groups from whom the data were originally collected, is the most crucial technique for establishing credibility" (Lincoln & Guba, 1985, p. 314). In this study, I used member checks by sharing with each teacher a draft of his/her portrait and having a conversation with each about the extent to which the portrait was accurate, or "rang true." As described in Chapter Three (p. x), the teachers' feedback was critical to the correction of some factual

errors which, if allowed to remain unchanged, would have decreased the credibility of the study.

One challenge to the accuracy of the data in this study was reactivity, defined by Maxwell as, "the influence of the research on the setting or the individual studied" (1996, p. 91). I felt this was particularly a problem for Elizabeth because I had been her teacher. Thus, I was concerned that she might parrot back what I had said while teaching the discussion class. If, in fact, she was simply describing her practice, that would not be a problem, but if she was changing the description of her practice to please me by incorporating what she had been taught in the discussion class, then the credibility of the study would be seriously impaired. I dealt with this in two ways. First, I told her I thought this might be a concern. We discussed why it was important for her to describe her conceptions and practice, not what she had heard from me when I was her teacher. Second, I carefully monitored my tone of voice and the phrasing of questions to guard against being perceived as judgmental.

Notwithstanding the overall accuracy of the data, its completeness is less impressive, thus raising what I consider to be a major limitation of the study. As discussed in Chapter Three, I learned about how the teachers enact CPI discussions through interviews and observations. The observations were critical to understanding what really happens in the discussions, as opposed to just relying on teachers' accounts. In a best-case scenario, I would have observed discussions in each of the teachers' classes. This was not possible, however, because Joe's class ended before the study began. As an alternative, I listened to audiotapes of all of the CPI discussions that occurred during Joe's nine-week class. While the audiotapes allowed me to learn what Joe and his

students said during discussions, I clearly missed other discussion-related behaviors that must be seen to be understood. For example, I could not make inferences about the students' body language or tell whether silent students appeared to be engaged or bored. Due to this lack of observational data for Joe's classes, I was unable to generate any propositions that dealt broadly with the types of student behaviors that must be seen to be understood.

There was a second way in which the data was not as complete as I would have liked. Interviews with the teachers suggested that much of the discussion-related teaching in the three teachers' classes occurs at the beginning of the courses. Given that I wanted to understand how the teachers were teaching their students to participate more effectively in CPI discussions, it was important to observe this instruction. This was not possible for all three teachers, however, because two of the classes had been in session for more than half of the school year when the study began. Therefore, it is logical to presume that some of what was important about the discussion teaching practice of two of the teachers was not observed--obviously a major limitation of the study. I addressed this limitation by asking the teachers to describe how they introduced CPI discussions to their students and by analyzing artifacts from the beginning of the school year.

Data Reduction and Interpretation. In addition to assessing the accuracy and completeness of the data, another way to judge the credibility of a study is to focus on how the data was reduced and interpreted. Recall that the data in this study was reduced and interpreted in two ways. First, I wrote portraits of each teacher's conceptions and practice and, second, I induced an initial theory consisting of seven propositions and relationships that exist among them. Here I will focus on the first level of interpretation, the writing of the

portraits. Miles and Huberman (1994, p. 279) suggest many questions that can be used to assess the credibility of how data is described. Two of them are: How context-rich and meaningful ("thick") are the descriptions? Does the account "ring true," make sense, seem convincing or plausible, enable a "vicarious presence" for the reader?

As descriptions of the teachers' CPI discussion thinking and teaching, the portraits are fairly thick, and offer context-rich first-level interpretations of the data. In using the term "thick," I rely on Maxwell's (1996) explication of the term made famous by Geertz (1973): ". . . I mean data that are detailed and complete enough that they provide a full and revealing picture of what is going on" (p. 95). To assess the thickness of the descriptions of the data in the portraits, I relied primarily on the reactions of people who read them. In addition to members of my dissertation committee, the portraits were read by a high school teacher interested in classroom discussion, a social studies expert, two adults who are non-educators, and the teachers themselves. All responded positively to my question about whether the portraits were detailed enough to provide the reader with a good sense of the teachers' conceptions and practices. Criticism of the portraits, in fact, suggested that they were too thick. One reader, for example, said a portrait was "cruelly long" and therefore less interesting than would be the case with a more succinct description. The teachers, of course, were most qualified to assess whether the portraits "rang true." As previously mentioned, through the use of member checks, I purposely solicited their feedback on whether they thought the portraits accurately captured and described their thinking and teaching relative to CPI discussions. Without exception, the teachers stated that the

first-level interpretations in the portraits were accurate portrayals of their conceptions and practices.

In addition to the first-level interpretation that occurred throughout the portrait writing stage of this study, the initial theory consisting of seven propositions also involved interpretation or meaning-making. Recall that Chapter Three explained the tenets of the grounded theory methodology used in this study to generate an initial theory about the CPI discussion conceptions and practices of skilled discussion teachers.

By describing the steps I undertook to create the initial theory (see Chapter Three), readers can evaluate how I moved from the raw data to the seven propositions. One of the limitations of this study is that it is difficult to describe a non-linear process (the constant comparative technique) in a linear way. By that, I mean that the recursive nature of data collection and data interpretation that is the hallmark of grounded theory does not lend itself to clear description. While I created a flowchart (Figure 5) to illustrate the steps I used in the study, I found it difficult to capture accurately the inherently complicated nature of generating grounded theory.

Ethical Issue

Disciplined inquiry, by definition, involves ethical issues that should be paramount to the researcher. Throughout the study, I was conscious of and concerned about creating a reciprocal relationship with the teachers. Although all three teachers indicated interest in participating in the study because they felt it would help them better understand their teaching, they all voiced frustration that the interviews were not conversations. On a few occasions, a teacher asked, "What do you think?" and then immediately said, "Oh, I forgot, I know you are not supposed to say." By not participating in a

genuine two-way conversation, I was trying to avoid influencing the teachers' views. After all, I already knew what I thought about CPI discussions--I was interested in learning what they thought. Toward the end of the interview schedule, two of the three teachers directly asked my advice about a part of their CPI discussion teaching practice that was troubling to them. Although some qualitative researchers, such as Grossman (1990), recommend not giving any "advice" to informants until after the study is completed, refusing to respond felt parasitic to me. When the two teachers asked my opinion about certain elements of their practice, I responded. However, this did not occur until the final interview, and I was very careful about what I said. Still, it raises an issue about the tension between reactivity and responding to the needs of the teachers. To avoid exploiting the teachers as informants (Glesne & Peshkin, 1992), I might have enhanced reactivity.

Conclusion

To close this dissertation, I take the opportunity to reflect on what I have learned about how secondary social studies teachers conceptualize and practice CPI discussion teaching and how I plan to use what I have learned in the future. At the beginning of this study, I expected to find that teaching students to participate effectively in CPI discussions is an extremely difficult and time-consuming enterprise. I expected this finding, based on my own experiences teaching students and teachers to engage in CPI discussions and based on a review of the literature showing that few teachers teach with, let alone for, discussion. Perhaps so few teachers use CPI discussions, I thought, because they are just so difficult.

While I still believe that it is not easy to teach with and for discussion, I have moderated my assessment of the difficulty of CPI discussions as a result

of what I have learned from the three teachers in this study. I do believe these are teachers who bring to the classroom a host of skills that some teachers may not possess. Notwithstanding this assessment of their abilities, it is apparent to me that one of the factors that matters most in their teaching of CPI discussions is the elaborate and sophisticated lesson plans they use when teaching discussion to their students. One of the factors that makes these teachers skillful is not any kind of classroom wizardry, but well thought-out and thorough lesson plans. While teaching people to be classroom wizards may be extremely difficult, learning how to develop sound lesson plans for CPI discussions is not. For those educators who believe that participating in CPI discussions is an important goal for their students, this is clearly good news. Thus, in my own teaching of preservice and inservice teachers, I plan to focus more specifically on lesson planning as it relates to CPI discussions.

I have also come to believe that it is important for educators to constantly reflect on what they value, for their values influence what they are teaching and what students are learning. The practice of the teachers in this study are clearly in line with what is important to them. I did not find this surprising. In fact, I expected it, given my personal experiences with other skilled teachers. What became more apparent to me, however, is that many tensions and tradeoffs are embedded in the practice of CPI discussion teaching. For example, I observed tensions between authenticity and accountability, and between equal participation and the focus on important ideas. These tensions are important reminders that teaching involves complicated choices between competing goods. Thus, in my own teaching and research about CPI discussions, I will be careful to bring to the fore these tensions so teachers are aware of the choices they are making.

Finally, I was heartened by how much the students in the classes of these three teachers seemed to enjoy the discussions. Often I heard students complain at the end of a discussion that they wished it did not have to end. They were having fun because they were engaged in talking about topics that are difficult and important. As a citizen in a democracy that relies on citizens' discussions of public problems, I found in the students' enjoyment of CPI discussions a bright ray of personal hope.

References

Alvermann, D. C., Weaver, D., Hinchman, K. A., Moore, D. W., Phelps, S. F., Thrash, E. C., & Zalewski, P. (1995). Middle- and high-school students' perceptions of how they experience text-based discussions: a multicase study (National Reading Research Council, Rep. No. 36). University of Georgia and University of Maryland.

Avery, P. G., Bird, K., Johnstone, S., Sullivan, J. L., & Thalhammer, K. (1992). Exploring political tolerance with adolescents. Theory and Research in Social Education 20 (4), 386-420.

Barber, B. R. (1984). Strong democracy: Participatory politics for a new age. Berkeley: University of California Press.

Barber, B. R. (1989). Public talk and civic action: Education for participation in a strong democracy. Social Education, 53 (6), 355-356, 370.

Barr, R. D. Barth, J. L., & Shermis, S. S. (1977). Defining the social studies (Bulletin 51). Washington, DC: National Council for the Social Studies.

Bickmore, K. (1991). Practicing conflict: Citizenship education in high school social studies (Doctoral dissertation, Standard University, 1991). Dissertation Abstracts International 52 (09), 3239A.

Baughman, J. E. (1975). An investigation of the impact of civics on political attitudes of adolescents (Doctoral dissertation, University of Maryland, 1975). Dissertation Abstracts International 36 (06), 3974A.

Blankenship, G. (1990). Classroom climate, global knowledge, global attitudes, political attitudes. Theory and Research in Social Education, 18, (4), 363-84.

Bogdan, R. C., & Biklen, S. K. (1992). Qualitative research in education. Boston: Allyn and Bacon.

Brody, R. A. (1994, March). Secondary education and political attitudes: Examining the effects on political tolerance of the We the People curriculum. Paper presented at the meeting of the Comparative and International Education Society, San Diego, CA.

Burbules, N. C., & Rice, S. (1991). Dialogue across differences: Continuing the conversation. Harvard Educational Review, 61, (4), 393-416.

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 266-298). New York: MacMillan.

Common, D. L. (1985). Teacher authority in the social studies classroom: Erosion of barren ground. Theory and Research in Social Education, 12, (4), 25-38.

Cousins, J. E. (1963). The development of reflective thinking in an eighth-grade social studies class. Bulletin of the School of Education, Indiana University, 39, (3), 74-112.

Dillon, J. T. (1994). Using discussion in classrooms. Philadelphia: Open University Press.

Ehman, L. H. (1969). An analysis of the relationships of selected educational variables with the political socialization of high school students. American Education Research Journal, 6, 559-80.

Ehman, L. H. (1970). Normative discourse and attitude change in the social studies classroom. The High School Journal 54, 76-83.

Engel, S. L. (1993). Attitudes of secondary social studies and English teachers toward the classroom examination and treatment of controversial

issues (Doctoral dissertation, University of Illinois, 1993). Dissertation Abstracts International, 54 (05), 1652A.

Evans, R. W., & Saxe, D. W. (Eds.) (1996). Handbook on teaching social issues, Washington, DC: National Council for the Social Studies.

Feiman-Nemser, S., & Floden, R. E. (1986). The cultures of teaching. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 505-526). New York: MacMillan.

Geertz, C. (1973). The interpretation of cultures. New York: Basic Books.

Gagnon, P. (1989/1996). History's role in civic education: The precondition for political intelligence. In W. C. Parker (Ed.), Educating the democratic mind, (pp. 241-262). Albany: State University of New York Press.

Gall, J. P., & Gall, M. D. (1990). Outcomes of the discussion method. In W. Wilen (Ed.), Teaching and learning through discussion (pp. 25-44). Springfield, IL: Charles S. Thomas.

Gehrke, N. J., & Parker, W. C. (1982, March). A 'grounded theory' approach to the generation of curriculum theory. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.

Glaser, B. G., & Strauss, A. I. (1967). The discovery of grounded theory: Strategies for qualitative research. New York: Aldine De Gruyter.

Glesne, C., & Peshkin, A. (1992). Becoming qualitative researchers. New York: Longman.

Goldenson, A. R. (1978). An alternative view about the role of the secondary school political socialization: a field experimental study of the development of civil liberties attitudes. Theory and Research in Social Education 6 (1), 44-72.

Goodlad, J. I. (1984). A place called school. New York: McGraw-Hill.

Goodman, J., & Adler, S. (1985). Becoming an elementary social studies teacher: A study of perspectives. Theory and Research in Social Education, 13 (2), 1-20.

Grant, R. W. (1996). The ethics of talk: Classroom Conversation and Democratic Politics. Teachers College Record 97, (3), 470-481.

Gray, D. (1989). Putting minds to work. American Educator, 13, (3), 16-23.

Gross, R. E. (1964). How to handle controversial issues. Bulletin of the National Council for the Social Studies. Washington, DC: National Council for the Social Studies, 3-4.

Grossman, D. L. (1975/76). Educational climates and attitudes toward dissent: a study of political socialization of conflict norms in adolescents (Doctoral dissertation, Stanford University, 1975/76). Dissertation Abstracts International 36 (12) 7980.

Grossman, P. L. (1990). The making of a teacher: Teacher knowledge and teacher education. New York: Teachers College Press.

Hahn, C. L. (1991). Controversial issues in social studies. In Shaver, J. (Ed.) Handbook of research on social studies teaching and learning. (pp. 470-480). New York: MacMillan

Hahn, C. L. (1996). Research on issues-centered social studies. In R. W. Evans and D. W. Saxe (Eds.), Handbook on teaching social issues (pp. 26-39). Washington, DC: National Council for the Social Studies Social Studies.

Hahn, C. L. (1998). Becoming political: Comparative perspectives on citizenship education. New York: State University of New York Press.

Harris, D. (1996). Assessing discussion of public issues: A scoring guide. In Ronald and David Warren Saxe (Eds.), Handbook on teaching social issues, (pp. 289-297).Washington, DC: National Council for the Social Studies.

Hess, D. (1997). Teachers' conceptions of effective discussions of controversial issues. Unpublished manuscript.

Hunt, M. P., & Metcalf, L. E. (1955/96). Rational inquiry on society's closed areas. In W. Parker (Ed.), Educating the democratic mind (pp. 97-116). Albany, NY: State University of New York Press.

Issues-Centered Education. (1989). The Social Studies, 80, (5).

Johnson, D. W., & Johnson, R. T. (1979). Conflict in the classroom: Controversy and learning. Review of Educational Research, 49, (1), 51-70.

Johnson, D. W., & Johnson, R. T. (1988). Critical thinking through structured controversy. Educational Leadership, 45, (8), 58-64.

Johnston, J., Anderman, L., Milne, L., Klenk, L., & Harris, D. (1994). Improving civic discourse in the classroom: Taking the measure of Channel One. Research Report 4. Ann Arbor, MI: Institute for Social Research, University of Michigan.

Kelly, T. E. (1986). Discussing controversial issues: Four perspectives on the teacher's role. Theory and Research in Social Education, 14, (2), 113-138.

Kettering Foundation. (1993). Meaningful chaos: How people form relationships with public concerns. Dayton, Ohio: no author.

Kohlberg, L. (1981). The philosophy of moral development: Moral stages and the ideas of justice. San Francisco: Harper & Row.

Ladson-Billings, G. (1994). The dreamkeepers. San Francisco: Jossey-Bass Publishers.

Langton, K. P., & Jennings, M. K. (1968). Political socialization and the high school civics curriculum in the United States. American Political Science Review, 62, 852-69.

Larson, B. E. (1995/1996). Teachers' conceptions of classroom discussion: A grounded theory study. (Doctoral dissertation, University of Washington, Seattle, 1995). Dissertation Abstracts International, 57 (02), 567A.

Larson, B. E. (1997, March). Teachers' conceptions of discussion as method and outcome. Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

Larson, B. E., & Parker, W. C. (1996). What is classroom discussion? A look at teachers' conceptions. Journal of Curriculum and Supervision, 11 (2), 110-126.

Lawrence-Lightfoot, S., & Davis, J. H. (1997). The art and science of portraiture. San Francisco: Jossey-Bass, Inc.

Leach, M. S. (1992). Can we talk? A response to Burbules and Rice. Harvard Educational Review, 62, (2), 257-263.

Levin, M., Newmann, F. M., & Oliver, D. W. (1969). A law and social science curriculum based on the analysis of public issues. Final Report, Project No. HS 058. Grant No. OE 310142. Washington, D.C.: U.S. Department of Health, Education, and Welfare, Office of Education.

Lincoln, Y., & Guba, E. (1985). Naturalistic inquiry. Thousand Oaks, CA: Sage Publications, Inc.

Litt, E. (1963). Civic education, community norms, and political indoctrination. American Sociological Review, 28, 69-75.

Lockwood, A. L. (1995). Controversial issues: The teacher's crucial role. Social Education, 60, (1), 28-31.

Lockwood, A. L., & Harris, D. E. (1985). Reasoning with democratic values: Ethical problems in United States history. New York: Teachers College Press.

Long, S., & Long, R. (1975). Controversy in the classroom: student viewpoint and educational outcome. Teaching Political Science 2: 69-75.

Lortie, D. (1975). Schoolteacher: A sociological study. Chicago, Illinois: University of Chicago Press.

Martin-Kniep, G. (1998). What am I doing this?: Purpose teaching through portfolio assessment. New Hampshire: Heinemann.

Massialas, B. G., & Cox, B. C. (1966). Inquiry in social studies. New York: McGraw-Hill Books, Inc.

Maxwell, J. (1996). Qualitative research design: An interactive approach. Thousand Oaks, CA: Sage Publications, Inc.

McNeil, L. M. (1986). Contradictions of control: School structure and school knowledge. Philadelphia: Metheun.

Meier, D. (1995). The power of their ideas: Lessons from a small school in Harlem. Boston: Beacon Press.

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: an expanded sourcebook (2nd ed). Thousand Oaks, CA: Sage Publications, Inc.

Miller, B., & Singleton, L. (1997). Preparing citizens: Linking authentic assessment and instruction in civic/law-related education. Boulder, CO: Social Science Education Consortium.

National Council for the Social Studies (1998). Minds on social studies (videotape). Washington, D.C.

National Education Association. (1916). The social studies in secondary education. A report of the Committee on Social Studies of the Committee on the Reorganization of Secondary Education. Bulletin #28. Washington, DC: Bureau of Education.

Newmann, F. M. (1988). Higher-order thinking in high school social studies: An analysis of classrooms. National Center on Effective Secondary Schools, Madison: University of Wisconsin.

Newmann, F. M. (1989). Reflective civic participation. Social Education, 53, (6), 357-361.

Newmann, F. M., & Wehlage, G. G. (1995). Successful school restructuring: A report to the public and educators. Madison, WI: University of Wisconsin, Center on Organization and Restructuring of Schools.

Oliver, D. W., & Shaver, J. P. (1974). Teaching public issues in the high school. Logan, UT: Utah State University Press. (original work published 1966).

Parker, W. C. (1996). Curriculum for democracy. In R. Soder (Ed.), Democracy, Education, and the Schools (pp. 182-210). San Francisco: Jossey-Bass Publishers.

Parker, W. C., & Gehrke, N. J. (1986). Learning activities and teachers' decision making: Some grounded hypotheses. American Educational Research Journal, 23, (2), 227-242.

Parker, W. C., & Hess, D. (1998). Teaching with and for discussion. Manuscript submitted for publication.

Parker, W. C., & Jarolimek, J. (1997). Social studies in elementary education (10th ed.). New Jersey: Prentice-Hall, Inc.

Patrick, J. L., & Hoge, J. D. (1991). Teaching government, civics, and law. In Shaver, J. (Ed.), Handbook of research on social studies teaching and learning (pp. 417-435). New York: MacMillan.

Roby, T. W. (1988). Models of discussion. In J. T. Dillon (Ed.), Questioning and discussion: A multidimensional study, (pp. 163-191). Norwood, NJ: Ablex.

Rossi, J. A. (1995). In-depth study in an issues-oriented social studies classroom. Theory and Research in Social Education 23 (2), 88-120.

Rossi, J. A. (Ed.). (1996). Teaching controversial issues. Social Education, 60 (1).

Schulman, L. S. (1983). Autonomy and obligation: The remote control of teaching. In L. S. Schulman & G. Sykes (Eds.), Handbook of teaching and policy, (pp. 484-504). New York: Longman.

Shavelson, R. J., & Stern, M. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. Review of Educational Research 51 (4), 455-498.

Shaver, J., Davis, O. L., Jr., & Helburn, S. W. (1980). An interpretative report on the status of precollege social studies education based on three NSF-funded studies. In What are the needs in precollege science, mathematics, and social science education? Views from the field. Washington, DC: National Science Foundation.

Singleton, L. R., & Giese, J. R. (1996). Preparing citizens to participate in public discourse: The public issues model. In Ronald W. Evans & David

Warren Saxe (Eds.), Handbook on teaching social issues, (pp. 59-65).Washington, DC: National Council for the Social Studies.

Social Science Education Consortium. (1997). Preparing citizens video series and user's guide. Boulder, CO.

Stiggins, R. J. (1997). Student-centered classroom assessment (2nd ed.) New Jersey: Prentice-Hall, Inc.

Strauss, A. L., & Corbin, J. (1990). Basics of qualitative research. Newberry Park, CA: Sage Publications.

Thornton, S. J. (1991). Teacher as curricular-instructional gatekeeper in social studies. In J. P. Shaver (Ed.), Handbook of research on social studies teaching and learning (pp. 237-248). New York: MacMillan.

Wade, R. (1994). Teacher education students' views on class discussion: Implications for fostering critical reflection. Teaching & Teacher Education, 10, (2), 231-243.

We the people. (1987). Calabasas, CA: Center for Civic Education.

White, C. S. (1982). A validation study of the Barth-Shermis Social Studies Preference Scale. Theory and Research in Social Education, 10, (2), 1-20.

Wineburg, S. S., & Wilson, S. M. (1988). Models of wisdom in teaching history. Phi Delta Kappan. 70, 50-58.

Wolcott, H. F. (1990). Writing up qualitative research. Newbury Park, CA: Sage Publications.

Zevin, J. (1983). Future citizens: children and politics. Teaching Political Science 10: 119-26.

Zinn, H. (1995). A people's history of the United States: 1942-present. New York: Harper.

Appendix A
Consent Form

Teaching Controversial Public Issues Discussions:
A Grounded Theory Study

Diana Hess, Ph.D. Candidate, Principal Investigator (206-523-6563)
College of Education, University of Washington

Purpose and Benefits

I seek your participation as well as your consent for participation in a research project aimed at understanding the instruction, and the thinking undergirding the instruction, of secondary social studies teachers who are skillfully teaching their students to participate more effectively in discussions of controversial public issues. I believe your participation will help you and other teachers and teacher educators better understand the characteristics of effective classroom discussion of controversial public issues. I have chosen to conduct this study in partial fulfillment of the requirements for a graduate degree in Education.

Procedures

I will interview you several times between in the winter and spring of 1998. The interviews will take place at your school, or another site you select. Each interview will last about an hour and a half, and will be tape recorded. In the interviews I will ask you questions about how your conceptions of social studies, democracy, and effective classroom discussions of controversial public issues; what issues you would and would not include in your curriculum and why; and for your reactions to videotaped excerpts of classroom discussions. I will observe (or listen to audio tapes of) three discussions your classes. After each observation, I will ask you questions about how you planned for the discussion, the instruction, and your assessment of the quality of the discussion. For these interviews, I will ask you to bring copies of any materials you assigned to your students to prepare for the discussion and any assessment instruments you used to evaluate the discussions.

Risks, Stress, and Discomfort

In any interview situation there is the possibility that some questions may cause confusion or stress. If you feel that any question is inappropriate or you do not wish to respond, you are free to refuse to answer.

Other Information

You will be assigned a pseudonym to protect your privacy. You have the right to review and delete any portion of the audio tapes made of the interviews. I will destroy the audio tapes on July 1, 1999. No one other than the investigator will have access to the audio tapes or to any data from the study that identifies you by name.

Signature of Investigator_____ Date_____

The study described above has been explained to me. I voluntarily consent to participate in this activity. I have had an opportunity to ask questions. I understand that future questions I may have about the research or about my rights as a subject will be answered by the investigator listed above.

Signature of Subject_____ Date_____

cc:
Participant
Diana Hess

Appendix B

Interview Protocols for Semi-Structured Interviews

Interview One: Context, Purpose of Social Studies, Democratic Citizenship. I will explain the purpose of the study, how confidentiality will be maintained, and the consent form. I will ask the participant if she/he has any questions about the study or the form. I will ask the participant to sign the form. Then, I will set up the tape recorders, explain what this interview will focus on, and that she/he should feel free to not answer any question that makes her/him uncomfortable, that she/he has the right to erase any part of the tape, and then begin the interview.

(Note to readers: The majority of these questions were followed up by others designed to elicit an understanding of the initial response. As such, the content of the follow-up questions varied).

1. Please describe what caused you to become a social studies teacher.
2. What kind of teacher education program did you go through? How would you characterize your teacher education experience?
3. How long have you been teaching? In what schools have you taught? If you changed schools, why did you do so?
4. Please describe and explain the school in which you currently teach.
5. Tell me about the students in your school.
6. Now, let's shift the focus from your school to the community in which it is located. Please describe the community.
7. Moving to the social studies curriculum in your school, please describe its design with a specific focus on the course(s) you teach.

8. Please draw a concept map picture (or some kind of visual representation) that illustrates what you want your students to know, be able to do, and want to do as a result of the social studies class you teach. Probe for explanations of what is on the concept map.
9. Now, draw a concept map (or some kind of visual representation) that shows what you think effective citizens in the United States should know, be able to do, and want to do. Probe for explanations of what is on the concept map.

Interview Two: Characteristics of Good Discussion. In an adaptation of the think-aloud method, the teachers view two short teaching tapes that showcase class discussion and explain their reactions to the discussions. The first videotape is of a high school United States History class discussing a moral dilemma drawn from an excerpt of Richard Wright's autobiography in Reasoning with Democratic Values (Lockwood & Harris, 1985). This class discussion is on the film Minds on Social Studies (National Council for the Social Studies, 1990). The second discussion is of six students in a high school American Government class discussing whether the electoral college should be abolished or reformed. This small group discussion is included in the film series Preparing Citizens (Social Science Education Consortium, 1997).
1. This interview will focus on how you conceptualize good discussion and the purposes of discussion in social studies. Together we will watch two short excerpts of video-taped discussions in middle and high school classes. During the viewing, I want you to think aloud about your reactions to what you are viewing. This is a kind of interactive viewing, where you can talk back about or comment on what you are viewing. At this point, I provide the teacher

with an opportunity to practice the think-aloud technique by showing a short excerpt of a kindergarten classroom. Practice continues until the teacher feels comfortable with the think-aloud process.

2. Introduce the first discussion the teacher will view by explaining what the discussion is about, where it takes place, the age of the students, the type of class, and the length of the discussion. As the teacher views the discussion, he/she will hold a remote control devise so the tape can be stopped to make comments. I will probe with follow-up questions designed to elicit a better understanding of what the teacher is saying throughout the think-aloud process. Repeat the entire process with the second discussion.

Interview Three: Rationales for CPI Discussions. Using labels drawn from the rationales for CPI discussions in the literature, the teachers rank those rationales to reflect the reasons they use CPI discussions.

1. We're going to focus on the rationales for including classroom discussions in social studies. Specifically, we're going to look at discussions of controversial public issues. I'm going to show you a number of index cards. On each card there is a different rationale for classroom discussion of controversial public issues. As you look at each one, I want you to first read what's on the card so we get a good transcript. Then, explain what this rationale means, in your own words, regardless of whether you agree with it or not. Probe for more fully elaborated explanations of each rationale.

2. Now I would like you to rank order the rationales from the one that most reflects why you use CPI discussions to the one that least reflects why you use CPI discussions. Probe for more elaborated explanations of the rankings.

3. On each index card, is one of the following rationales:

Improve interpersonal skills
Understand important democratic values
Participate in political life of democracy (e.g., voting, serve on jury)
Learn important social studies content
Learn how to think critically
Improve democracy

Interview Four: Issues Selection. This interview involves another selection and ranking exercise. The teacher organizes various CPI topics in terms of the likelihood they would be included in his/her curriculum.

1. I am going to give you another set of index cards to work with -- written on each one is a CPI topic. First, I would like you to shuffle through these and place them in two piles: those that you would include in your curriculum and those you would not. As you are doing so, please explain your reasons to me.

2. Now, I would like you to rank order the cards to reflect the issues that you are most and least likely to include in your curriculum. Again, as you are sorting them, please explain your reasons to me.

3. The CPI topics written on the index cards are:

Abortion
Affirmative Action
Balanced Budget
Gay Rights
Immigration
Legalizing Drugs
Physician-Assisted Suicide
Trade Policy
Vouchers for Private Schools
Welfare Reform

CURRICULUM VITAE

Diana Hess

PRESENT POSITION

Assistant Professor, Social Studies Education (appointment begins January, 1999)
University of Wisconsin-Madison
Department of Curriculum and Instruction
Teacher Education Building
225 North Mills Street
Madison, Wisconsin, 53706-1795

ACADEMIC PREPARATION

1998 Ph.D. Education, University of Washington, Seattle, WA.

Fields of concentration:
Curriculum & Instruction in Social Studies, Educational Policy, Culture and Citizenship, and Law.

Dissertation:
Teaching Controversial Public Issues Discussions: Learning from Skilled Teachers investigates how three secondary school social studies teachers teach their students to participate effectively in classroom discussions.

1995 M.A., Educational Policy, University of Illinois.

1983 N.E.H. program on political philosophy, Harvard University.

1979 B.A., Political Science, Western Illinois University.

PROFESSIONAL EXPERIENCE

1998 Lecturer during the fall quarter, Social Studies in the Secondary School, University of Washington.

1995-1998	Graduate student, teaching assistant for Social Studies in the Secondary School, developer and instructor of technology laboratory component of Social Studies in the Secondary School, lead teaching assistant (1997), research assistant on Historical Sense-Making Project at the University of Washington.
1997-1998	Instructor at Supreme Court Institute, a professional development program for high school teachers, co-sponsored by Street Law, Inc. and the Supreme Court Historical Society at Georgetown Law Center and the United States Supreme Court.
1996	Developed and taught graduate course: Teaching with Discussion, University of Washington.
1996-1997	Developed and taught graduate course: Discussion for Democracy, University of Colorado at Denver.
1987-1995	Associate Director of the Constitutional Rights Foundation Chicago. Responsibilities included developing political science, law-related and service learning curriculum and program materials, writing over four million dollars of funded grants, developing and implementing a national training of trainers program, designing and leading courses, institutes, and workshops for teachers, administrators, lawyers, judges, and police officers, providing technical assistance to national, state, and local civic education organizations, and hiring/supervising staff.
1990-1991	Adjunct Instructor: Developed and taught graduate level social studies methods courses for DePaul University and Roosevelt University, Chicago, IL
1979-1987	Social Studies teacher: Taught courses in political science, law, US. History, global education, coordinator of the Student Mock Trial team and developer/coordinator of the Government Internship Program. Community High School District #99 South, Downers Grove, IL.
1985-1987	President of Downers Grove Education Association, IEA/NEA. Responsibilities included organizing, bargaining, lobbying, enforcement of contract, and developing/supervising various association programs.

1976 Congressional intern for US Representative T. Railsback, Washington, D.C. Researched legislative issues, and responded to constituents' requests for information and service.

PUBLICATIONS

Parker, W. C., & Hess, D. (1998). "Teaching with and for discussion." Manuscript submitted for publication.

Hess, D., and students at East High School, Denver, CO. (1997) This video publication, Preparing Citizens, features students in a high school law class participating in a discussion using the public issues approach on the question, "Should physician-assisted suicide be legal in Colorado?" Hess facilitates the discussion. Available from the Social Science Education Consortium, Boulder, CO.

Hess, D. (1997) "Violence Prevention and Service Learning." Social Education, 61, (5) 279-281.

Hess, D., Miller, B., with Harris, D., & Arbetman, L. (1997) "The First Amendment in the Public Schools: Should Schools Address Intolerance Through Human Dignity Policies?" Lessons on the First Amendment for Secondary School Students (5-12), Denver, CO.: First Amendment Congress, 115-123.

Brandsberg, J., & Hess, D. (1997) "Is the V-Chip Legislation a Good Law?" Lessons on the First Amendment for Secondary School Students (5-12), Denver, CO.: First Amendment Congress, 151-165.

Hess, D., & Riddle, M. (1996) "Teaching Kids to Talk the Talk." Instructor 106, (2) 64.

Hess, D., & Senn, H. (1996) "Teaching Resources for Influencing Public Policy." I'm the People, Influencing Public Policy. Chicago, IL.: American Bar Association, 37-40.

Hess, D. (1994) Service Learning and the Social Studies. Chicago, IL.: Constitutional Rights Foundation Chicago.

Hess, D. (1993) "Law-Related Education and Cooperative Learning: A Rationale for their Merger." In Teaching for Tomorrow: Integrating LRE and the Social Studies. Chicago, IL.: American Bar Association, 2-7.

Pereira, C., Brady, S., & Hess, D. It's Yours: The Bill of Rights. Austin, TX: Steck-Vaughn Company, 1993.

Hess, D. (1989) "Drug Test Case Simulation." Social Education, 53 (4) 229-230.

Hess, D. (1989) "Teaching Strategies for the Fourth Amendment." Social Education, 53 (4) 232, 250.

Croddy, M., Pereira, C., & Hess, D. (1989) The Drug Question. Los Angeles, CA.: Constitutional Rights Foundation.

Rosenblum, W., Hess, D., & Pereira, C. (1989) From the School Newsroom to the Courtroom. Chicago, IL.: Constitutional Rights Foundation Chicago.

HONORS

Civic Education Exchange Program, 1998. One of six civic educators from the US chosen by the French-American Foundation to participate in a French-American exchange program.

Lead Teaching Assistant, College of Education, University of WA, 1997-1998.

Liberty Bell Award (1995), awarded annually by the Chicago Bar Association to a non-lawyer who has rendered service which strengthens the effectiveness of the American system of freedom under the law.

Nominated for the Isidore Starr Award (1995), awarded annually by the American Bar Association for national leadership in the field of law-related education.

Arkansas Traveler Award from Bill Clinton (1989), Governor of Arkansas, for working to create a law-related education center in Arkansas.

NATIONAL AND STATE CONFERENCE PRESENTATIONS (selected)

Hess, D., & McKenna, H. (1998). "Progress in Educating Tomorrow's Teacher Educators." Presentation at the annual conference of the National Network for Education Renewal, Seattle, WA.

Hess, D. (1998). "Preparing to Lead Professional Development Programs for Teachers." Presentation at the Japan Leadership Institute, Seattle, WA.

Hess, D. (1997) "Teaching Teachers to Lead Discussions: Notes from a Reflective Practitioner." Paper presented at the annual conference of the American Educational Research Association, Chicago, IL.

Hess, D. (1997) "Should Diversity be a Legal Goal?" Presentation at the state conference of the Colorado Law Project, Denver, CO.

Hess, D., and Pereira, C. (1997) "Controversial Discussions: Developing Habits for a Healthy Democracy." Presentation at the annual conference of the National Council for the Social Studies, Cincinnati, OH.

Hess, D. (1997) "So Many Cases, So Little Time: Selecting Which Supreme Court Cases to Teach." Presentation at the annual law-related education leadership seminar of the American Bar Association, San Francisco, CA.

Hess, D. (1996) "Previewing the Supreme Court's 1996-97 Term." Presentation at the state conference of the Colorado Law Project, Denver, CO.

Hess, D., and Miller, B. (1996) "Discussion for Democracy." Presentation at annual conference of the National Council for the Social Studies, Washington, D.C.

Hess, D., & Brandsberg, J. (1996) "Do Publicly-Funded Single-Sex Schools Violate the Constitution?" Presentation at annual conference of the National Council for the Social Studies, Washington, D.C.

Hess, D. (1996) "Teaching About Public Issues." Series of presentations at the Classroom Law Project Summer Institute, Portland, OR.

Hess, D. (1996) "Publicly-Funded Single-Sex Schools and the Equal Protection Clause." Presentation at the Michigan Social Studies Education Project Summer Institute, Lansing, MI.

Hess, D. (1996) "Using Legal Cases to Enhance Critical Thinking Skills." Presentation at the Michigan Social Studies Education Project Summer Institute, Lansing, MI.

Hess, D. (1996) "Discussing Public Issues." Presentation at the Michigan Social Studies Education Project Summer Institute, Lansing, MI.

Hess, D., & Miller, B. (1994) "Alternative Assessment in Law-Related Education." Presentation at the annual conference of the National Council for the Social Studies, Phoenix, AZ.

Hess, D. (1994) "Developing Assessments for Moot Court Simulations: The Experiences of Three High School Teachers." Paper presented at the meeting of the College and University Faculty Assembly of the National Council for the Social Studies, Phoenix, AZ.

Hess, D., & Croddy, M. (1993) "Rodney King Revisited." Presentation at the annual conference of the National Council for the Social Studies, Nashville, TN.

Hess, D. (1992) "Promoting Respect for Human Dignity." Presentation at the annual National Leadership Seminar of the American Bar Association's Special Committee on Youth Education for Citizenship, Los Angeles, CA.

Hess, D. (1991) "Teaching About Sex Discrimination." Presentation at the annual conference of the National Council for the Social Studies, Washington, D.C.

CONSULTATIONS (selected)

1997	Developed and facilitated professional development program on authentic assessment in K-12 social studies for the Edmonds School District, Lynnwood, WA. The program helps teachers form collaborative partnerships to improve curriculum, instruction, and assessment in social studies. A videotape showcasing the project and teacher-produced model assessments developed as a result of the program will be disseminated by the Edmonds School District in the summer of 1998.
1997	Developed and facilitated a professional development program for the education faculty at the University of Puget Sound, Tacoma, WA. The program focuses on how to improve class discussions in teacher education courses.
1992-1997	Provided input on program planning, presented at workshops, and provided feedback to teachers through the Authentic Assessment in Civic/Law-Related Education Project, Social Science Education Consortium, Boulder, CO.
1997	Worked with a team of educators to develop and implement Project REAL, an assessment program for social studies and science teachers, University of Washington, Seattle, WA.

1994-1995	Designed and implemented an evaluation of the Iowa Courts Curriculum Project for the Iowa Center for Law and Civic Education, Drake University, Des Moines, IA.
1994-1995	Worked with teachers and administrators to re-design the required US History course and create a Service Learning course for Mundelein High School, Mundelein, IL.
1992-1993	Script review for a film series on the Bill of Rights produced by Mississippi Educational Television, Jackson, MS.
1991-1992	Curriculum design for an after-school enrichment program on law-related education developed by the National Council of La Raza, Los Angeles, CA.
1989	Reviewer for revision of the Harvard Social Studies Project curriculum materials on the American Revolution and the New Deal for the Social Science Education Consortium, Boulder, CO.

PROFESSIONAL MEMBERSHIPS

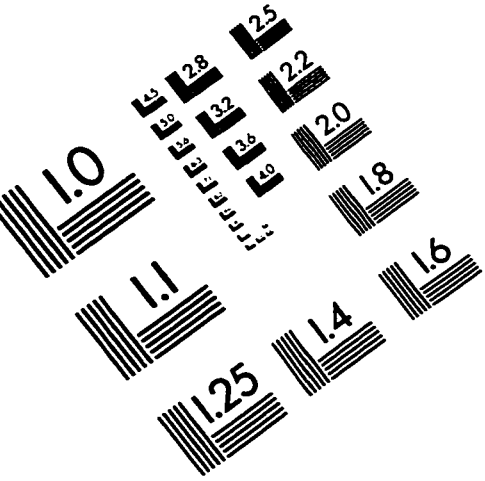
National Council for the Social Studies
Washington Council for the Social Studies
Association for Supervision and Curriculum Development
American Educational Research Association

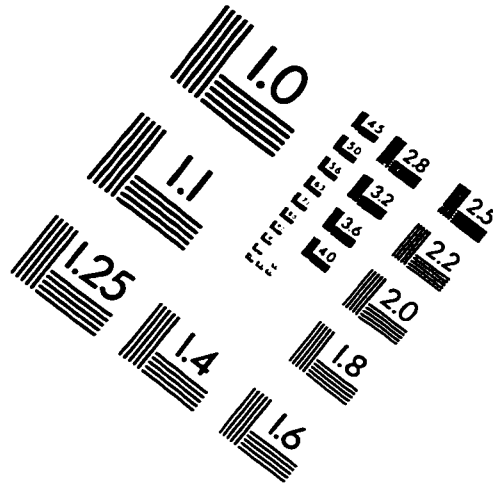
COMMITTEE SERVICE

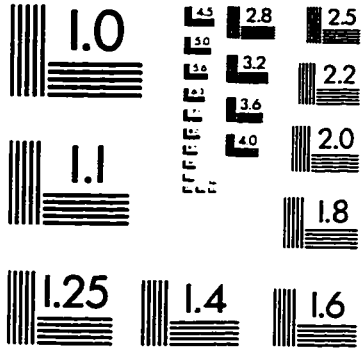
American Bar Association - National Law-Related Education Resource Advisory Committee (1992-1996)
National Council for the Social Studies Citizenship Committee (1988-1991)
Representative Fawell's (IL) Education Advisory Committee (1984-1995)
National Education Association Congressional Contact Team (1985-1987)
Illinois Education Association Women's Leadership Cadre (1985-1987)

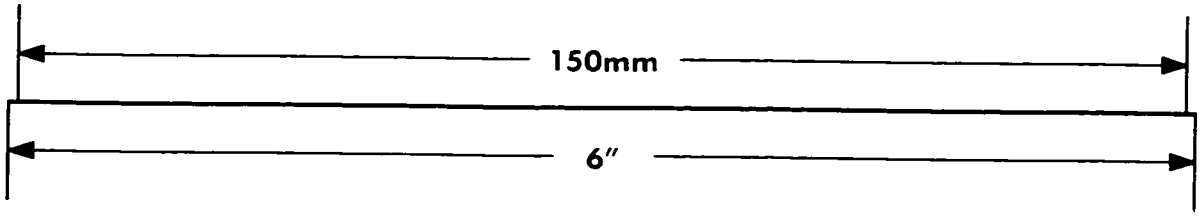
References Available Upon Request

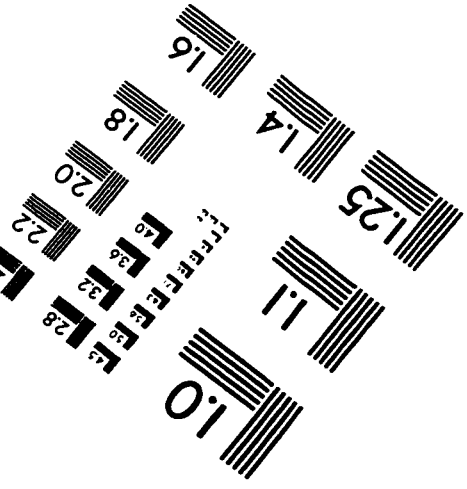
IMAGE EVALUATION TEST TARGET (QA-3)

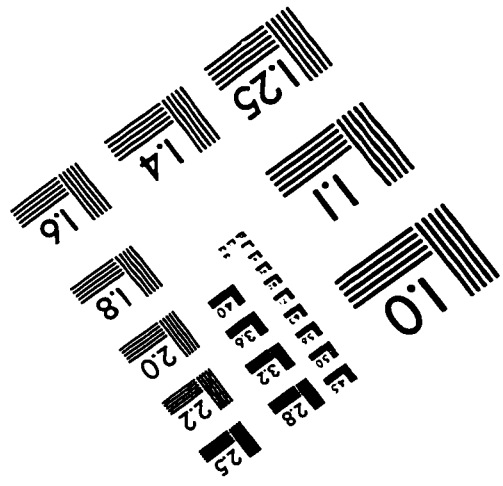












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved