8412376

Buttenfield, Barbara Pfeil

LINE STRUCTURE IN GRAPHIC AND GEOGRAPHIC SPACE

University of Washington PH.D. 1984

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark _√_.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _√_
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages _√_
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) __________ lacking when material received, and not available from school or author.
12. Page(s) __________ seem to be missing in numbering only as text follows.
13. Two pages numbered __________. Text follows.
14. Curling and wrinkled pages _____
15. Other __________________________________

Line Structure in Graphic and Geographic Space

by

Barbara Pfeil Buttenfield

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

Doctor of Philosophy

University of Washington

1984

Approved by John C. Sherman
(Chairperson of Supervisory Committee)

Program Authorized
to Offer Degree Department of Geography

Date March 8, 1984

Doctoral Dissertation

Signature Barbara Pfeil Butterfield

Date 2 March 84

University of Washington

Abstract

LINE STRUCTURE IN GRAPHIC AND GEOGRAPHIC SPACE

By Barbara Pfeil Buttenfield

Chairperson of the Supervisory Committee: Professor John C. Sherman
Department of Geography

The research reported in this dissertation has been based on the idea that a cartographic line is a probabilistic representation of the geographic feature which it symbolizes. Numeric parameters have been measured for two orders of structural relationships, and these parameters have been shown to provide significant distinctions between categories of cartographic line structure. The categories which have been developed are not intended as an exhaustive typology of line structure, but rather to demonstrate that meaningful categories of graphic structure can be defined numerically, and statistically verified.

Categories for both orders of structure have been summarized graphically, as structure signatures, and digitally, by storing parameters for each category as a computer look-up table. Structure signatures can be applied to cartographic line generalization in several ways which utilize the digital look-up tables. One application involves generating lines of predictable graphic structure, by stochastic modelling technniques. The other application does not serve to generata line structures, but to identify them, to provide a means by which threshold criteria may be automatically set and modified during computer generalization.

Line identification proceeds by matching measured parameters against parameters stored in the look-up tables. A possible problem arises when a line is identified which does not match any of the existing structure categories. An algorithm is presented which has the flexibility to incorporate new structures into an existing knowledge base, in effect, to learn new structures, and to become more proficient in line identification over time. Intelligent algorithms have been developed for pattern recognition by other authors, but the contribution of this research is to provide an intelligent algorithm for a specifically cartographic task, the automated modification of tolerance criteria during line generalization.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I have been very fortunate to have completed my doctoral work in a department which stresses independant thinking, and which provides moral support above and beyond the call of duty. I would like to thank the Washington Geography faculty for this; it was no easy task to take the blinders off a hard-nosed cartographer, especially one such as myself.

Much of this writing has been completed in California, during my first year as a junior professor. This is not a course of action which I would reccommend, looking back on it. I was so naive; but at the time, it seemed like an interesting challenge. Actually, it was good for me to do this: I have learned much about myself, and about my limitations. I fear that my committee members were less fortunate, because they had to put up with me. However, everyone has been extremely considerate.

Special thanks are offered to Dick Morrill, who has provided advice on statistics as well as on survival politics; and to David Hodge, who has been most helpful in discussing research design and analysis strategies. David has managed to keep my feet on the ground throughout this research, in spite of everything I have thrown at him. I would also thank Reg Golledge, who is not on my PhD committee, but is my department chairperson in California. He has backed me 100 per cent this past year, and I surely appreciate the encouragement and the moral support.

Two individuals have served as primary mentors for me in this dissertation. First is Tom Poiker (formerly Peucker), who has been more than hospitable throughout the research, offering guidance on all aspects of the dissertation, from data structures and computation to elements of

writing style... I can't even begin to enumerate. Tom must be credited with directing me to apply concepts of Artificial Intelligence to the identification of cartographic structures. I have learned so much in our discussions, about the nature of cartographic research, and about the specifics of cartographic structure; and I am grateful.

Next, I would like to express my deep respect for my chair, John Sherman, with whom I have spent many enjoyable hours discussing cartographic methods, and cartographic research. The most incredible thing about working with John is his capacity to listen to my ravings, with perfect equanimity, and to guide me each time back to the basic principles underlying my research questions. He has kept me on track, and has never stifled my ways of thinking: this is a difficult thing to do, and I hope I am as successful a chair for my own students, in the future.

Much of the computation involved in this research would not have been achieved without the patient advice of Fred Nick, at CSSCR, who kept telling me that I could teach myself Pascal (he was right) and who offered a great deal of help with analytic geometry. Liz Kohlenberg helped to polish the cluster analysis, and the discriminant design; she also has provided a respect and a cameraderie, a friendship. I would also like to thank Jois Child, a fellow cartographer, for many hours spent hashing and rehashing ideas presented throughout the dissertation. Last but not least, I must thank Thomas Jacobson, for late night debates on my research methods, and my algorithm designs, for proofreading whenever I asked, and for continual reassurance of my worth as an academic, and as a human being.

DEDICATION

So many people in my life have assured me I could accomplish anything I set my mind to do; but my role models come from my childhood days, and all are women in my family. I would dedicate this to Sarah, who taught me to be proud; to Fran, who taught me to be gentle; and to Gretchen, who taught me to be strong.

INTRODUCTION

The purpose of this dissertation is to consider cartographic generalization as a conceptual transformation of geographic detail into graphic representation. During the transformation, emphasis is placed on preserving the accuracy of the graphic information; also important is the preservation of the overall character, the visual structure of the geographic feature which is represented. The discussion will remain within the context of line generalization, and will focus in particular on computer-stored lines. Here, the limits of accuracy are not bounded by collection technologies, but rather by the costs of storage space and processing time. To say that cartographers collect more information than is needed for accurate representation is an oversimplification, as many generalization algorithms serve to reduce the volume of information which is collected -- efficiency in the cartographic process is only secondarily important to the discussion at hand.

The cartographer's intention in generalization is to preserve accuracy, of course. But it is also intended that the graphic representation which is produced still looks like its geographic counterpart. Thus accuracy may also be defined within a perceptual context, whereupon the preservation of recognizable characteristics is of major concern. This is more difficult to automate than might first be imagined: the inductive reasoning by which a cartographer generalizes the characteristic twists and turns of a fjord coastline, for example, are quite different than the decisions involved in representing the cuspate look of a sandy beach. A major stumbling block in expressing the kinds of decisions which are being made is the lack of a consistent and objective means to describe the character of a line.

It is proposed in this dissertation that the kind of information which defines the character of a line (be it a fjord or a sandy beach) is different from the information which preserves geographic accuracy of the line; and further, that it can be isolated as a set of statistical parameters. Previous efforts to derive measures of linear character have proved inadequate, in part because the measures as derived are dependant on the geographic and/or the perceptual accuracy of the line: a review of these efforts comprises Chapter 2 of the dissertation. In Chapter 3, the discussion focuses on the differences between recognition and representation in graphic space, on the relationship between linear character and linear structure, and the role of each in automating generalization.

The remainder of the dissertation is concerned with measuring certain aspects of cartographic structure, to demonstrate how linear character may be digitally identified for purposes of computer generalization. The approach involves an empirical procedure which is used to develop a parametric model of a line in graphic space. A sample of trial lines (from the World Data Base) is presented in Chapter 4, and parameters of structure are derived and measured. The measurements are analyzed in Chapter 5 to determine the extent to which these parameters distinguish the character of one line from the character of another. Finally, structural parameters are presented by which varying line characters may be identified within the computer, and the techniques used to measure these parameters are evaluated.

Several applications of this research are briefly discussed in the final chapter. The first is a direct application to automating line generalization tasks. When line features of differing character can be distinguished, tolerance thresholds for automated smoothing or for simplification can be modified without manual intervention. This may provide more consistent generalized representations of compound lines. Furthermore, tolerance criteria can be stored in-core with their associated structure parameters, providing a guideline for selection as well as modification of tolerance values. The notion of an algorithm which is designed to modify its generalization based on changing parameters of linear character falls within the realm of Artificial Intelligence. One possible algorithm by which to implement this kind of identification is presented in this chapter.

A second application marks the point at which generalization and line generation merge; the identification of line characters may be useful in stochastic modelling of cartographic lines. Work by Mandelbrot (1982), Carpenter (1981), Dutton (1981) and other researchers has shown that only a small number of coordinates are required to produce visually interesting geographies, many of which portray existing landscapes. The advantage of producing stochastic models is that only a fraction of the original coordinate set needs to be stored. Given the capability to isolate line character from line accuracy, the selection of critical points may be facilitated. Finally, there are obvious extensions from modelling lines to three-dimensional modelling and terrain representation, although this last application will not be considered in the current discussion. What follows in Chapter 1 is a brief discussion of traditional generalization concepts, and a more explicit statement of the problem at hand.

CHAPTER 1
ACCURACY IN GENERALIZATION

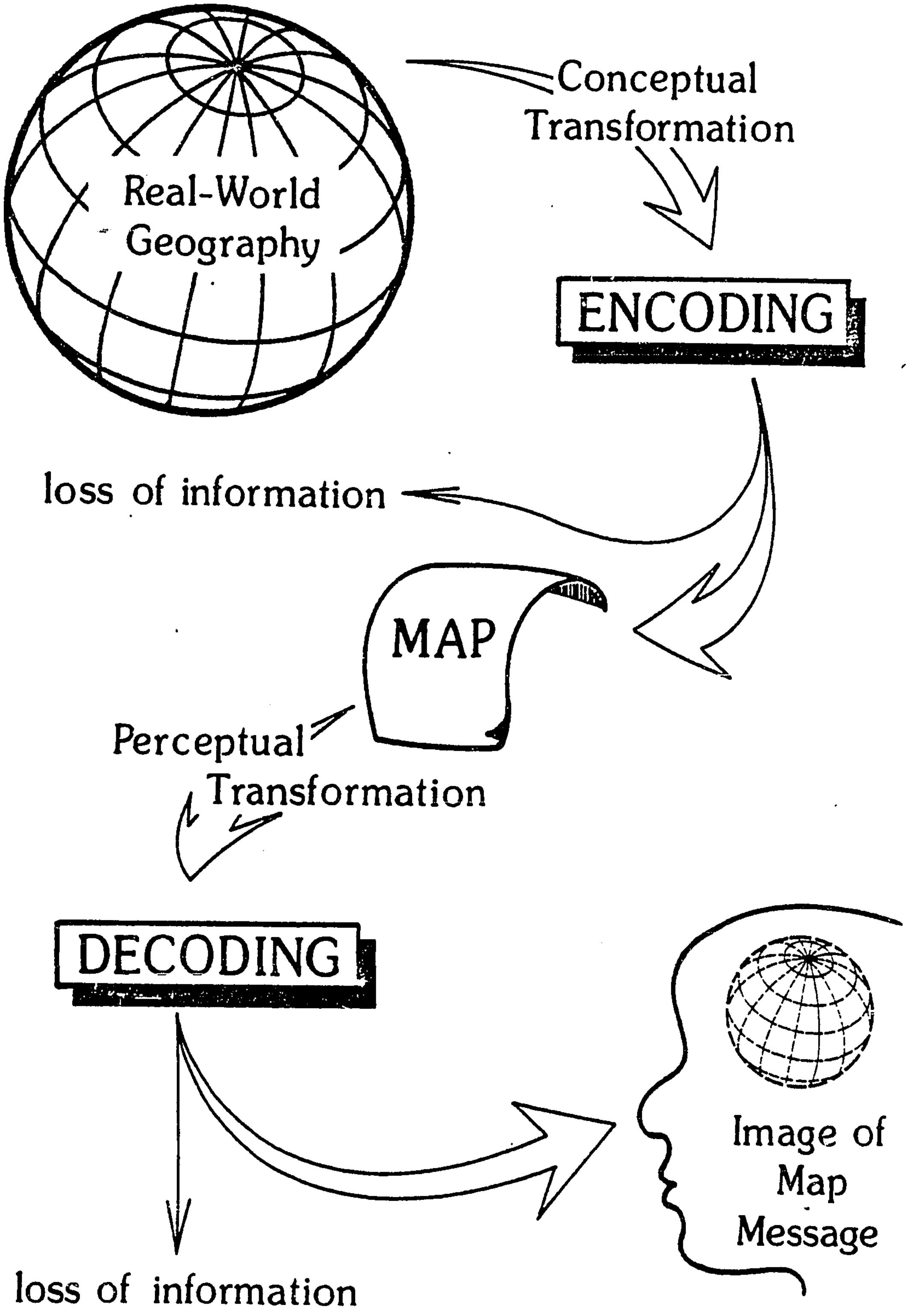
To the extent that geographic research is concerned with differentiations between information and error, the role of the cartographer focuses on maintaining a respectful balance between the two in graphic display. In the jargon, the intention is to 'maximize the clarity of the intended map message' by careful attention to what part is essential map information and what part is visual or conceptual noise.

The metaphor of information and noise stems from an information theoretic model of communication (Shannon and Weaver, 1949). In its strictest application to cartography, the model describes how information about some real world geographical phenomenon can be encoded into a cartographic format, and can then be decoded by a map reader to receive the intended message (Figure 1.1). The loss of information in the communication process will be reflected in discrepancies between the actual geographic pattern, and the realized graphic pattern of the map message.

Two aspects of the metaphor can be considered in the context of generalization and map accuracy. First, the reduction of information during the encoding process serves to extract details from the real world, with the purpose of communicating some particular aspect or spatial relation. This is the principle underlying the concept of the map as a graphic essay; and it is important to consider how the accuracy of the extracted detail may affect the communication process.

The second aspect of the metaphor is the reduction of information (or noise) during the decoding process. Perceptual limitations will prohibit a map reader from processing all of the visual information

Figure 1.1
CARTOGRAPHIC COMMUNICATION



present on the map: there are various reasons why this occurs, which are beyond the scope of the present discussion. What is important is that cartographic techniques exist by which information may be transformed to facilitate map communication (i.e., to construct a more 'accurate' cartographic representation).

The transformation is a kind of filtering process used to achieve a balance between information and noise; very often, the distinction between the two becomes somewhat amorphous. For example, if too much detail appears on a map, the map message may become confusing, as information becomes visual noise or clutter. Conversely, noise may be used to augment information. At some points it may become a valuable component in the map message: this point will come up again and again in the discussion. There are perceptual transforms, in which graphic compensations are made for those decoding problems which can be predicted: this is the realm of psychophysical research in cartography. There are also conceptual transforms, in which information is filtered to reduce unnecessary detail, while retaining or even emphasizing what is relevant to the message: and this is the realm of generalization.

Traditionally, generalization has been broken down into four components (Robinson and Sale, 1969). In the context of line generalization, the first component is simplification. This consists of two parts, selection of line characteristics to eliminate unwanted detail, and smoothing to modify or reshape a line to emphasize some aspect of its character. Another component is symbolization, or the graphic coding of the line character for map representation. Linear character is the information which identifies a line as a valid representation of a geographic feature. In a pragmatic sense, it is geographic location, or positional accuracy which must be symbolized. Perceptually, however, the information which should be retained has more relation to angular change (what is called sinuosity) than with precise coordinate locations (Attneave, 1954). One of the aims of this dissertation is to consider the balance between geographic and perceptual accuracy in cartographic lines.

A third component is classification, which is the aggregation or partitioning of information into categories. This is an important part of the generalization process, but will not be directly applicable to the discussion at hand. A fourth component is called induction; Robinson and Sale (1969) use this term to describe the logical assumptions made during generalization to improve the representation of the line, commenting that "...cartographic generalization is essentially a creative act." (Robinson and Sale, 1969, p.53) The inductive components provide real difficulties in automating the generalization process, for a variety of reasons which will be discussed throughout the dissertation.

A fifth component is proposed for this generalization scheme. It should be called enhancement, and defined as the purposeful and controlled introduction of detail which emphasizes a particular aspect of the map data. 'Introduction' is perhaps an inappropriate term here: enhancement is not the bringing in of unrelated information so much as the amplification of information which is already present. Enhancement may encapsulate those cartographic situations in which visual noise may add information to the map message, as follows.

Consider that automated data collection techniques (digitizing, for example) can provide more detail than is necessary to produce a recognizable representation of a geographic line at scale. For example, Jenks (1981) has shown that a line plotted with only a small subset of points in its original coordinate set will often be perceived as identical to the original line. What this implies is that more linear detail is being computer-stored than may be required for recognizable representations.

Consider also that the research of Attneave (1954), Kelley (1977), and of Marino (1979) has shown that perceptual information is not homogeneous along the length of a line, but is rather packed into the vicinity of points of maximum angular deviation. These perceptually critical points may or may not include coordinates required for geographic accuracy, for example the points which locate a town along a coastline, or the intersection of two roads. It would appear then that the recognizability of a cartographic representation is only partially determined by positional accuracy of coordinates; other components of information (such as sinuosity, or angular change) are important as well.

What all of this has to do with cartographic generalization relates to the concept of enhancement. More points are included in graphic display than may be required, in many situations; and the points which are mandatory for geographic and perceptual accuracy are easily identified and separable from other coordinates in the cartographic line. What kind of information comprises the remainder of the cartographic line? And also, if this remainder does not contribute to the accuracy of the line, then what kind of information does it provide?

When a representation includes only those points critical to the accuracy of the line, its visual appearance may be crude, blocky, even non-geographic. In other words, the line has a measurable accuracy, but it does not look like the geographic feature which it represents. One quickly concludes that the so-called critical points are necessary but not sufficient for representing cartographic lines.

It is proposed that this remaining information comprises the visual character of a line, that character is isolate from geographic and from perceptual accuracy requirements in a representation, and that line character can be considered as a statistical or probabilistic phenomenon, in contrast to the geometric or perceptual phenomenon which it has been considered in the past. To understand the worth of treating a line on a map as a probabilistic phenomenon, one should consider work being carried out in mathematics and geology.

Research using fractals and other kinds of stochastic models has demonstrated that easily recognizable representations of real geographic terrain can be produced with a very small set of points and a random number generator (although this is an oversimplification, it will serve for the purposes of the argument). These stochastic models are constrained to display lines and landscapes of varying degrees of complexity (Carpenter, 1981; Dutton, 1981). While perceptually similar to real world geography, the models are cartometrically useless, due to unpredictable repositioning of geographically critical points. Although it would be a fairly simple matter to anchor the position of certain coordinates, work of this nature has not been reported in the literature. Regardless, stochastic models have great value in many graphic situations, most notably, as illustrations for simulating physical processes (Mandelbrot, 1977; Mandelbrot, 1982; Fournier and Fussell, 1981).

The stochastic models would indicate that for cartograhic lines, the remaining information (i.e., what is left after isolating points required for accuracy) can be approximated probabilistically, as a kind of graphic filler. What makes these lines look like the geographies they simulate is the constraints on the probabilities, rather than on any particular coordinate positioning. In other words, lines and surfaces of predictable visual characters can be produced by quantifiable constraints, and the same set of coordinates can be used to generate lines of diverse visual characteristics.

Line generation is merely another component of generalization. It is presented as an example of enhancement, and of the amplification of information by visual noise. It stands to reason that if lines of differing visual character can be generated from a probability distribution, it should also be possible to generate probability distributions from measurements of differing characters of cartographic lines. The value of achieving this lies in the ability of a computer generalization algorithm to include digital guidelines by which to identify the type of line character which should be retained during the process of representing it on a map.

Guidelines to identify what information is relevant and what is unnecessary visual noise have not yet been precisely defined, although much research has been devoted to perceptual and geometric definitions of line character; these will be discussed in Chapter 2. Consequently, there are few consistent guidelines to be followed in generalizing a line: this explains in part why generalization has proved so difficult to automate. And while many algorithms for line generalization exist, it is well known that each algorithm deals with some kinds lines of lines more accurately than with others. McMaster (1983) has shown that the amount and kinds of accuracy retained by an algorithm may be determined quite specifically, and this knowledge will facilitate the computerization of generalization tasks.

However, it is unclear whether such accuracy measures can be related between lines of differing visual character: one cannot assume that all cartographic lines contain a standard amount of information, and a standard amount of error. Until the amount and kinds of information available in a cartographic line have been expanded beyond the necessary accountings of accuracy, the success of a generalization can become definable only <u>after</u> the generalized line is constructed. This is a very passive sort of analytic strategy, for two reasons.

First, in order to generalize different parts of a compound cartographic line and retain all of the various recognizable charcteristics, the line must be broken into pieces, where the line changes its visual character. Each piece is generalized as a different line, using different tolerance values. Decisions as to where the line character changes are not only arbitrary, and subjective, they are plainly inefficient. This is the first problem of the dissertation research, to demonstrate a means by which lines of differing characters may be distinguished numerically.

Secondly, the process of breaking up a cartographic line implies that differing visual characters may be manually identified; the second part of the dissertation analysis will consider how this aspect of generalization may be automated. The final goal of the research is not to delineate some exhaustive typology of cartographic line structures, but rather to demonstrate a methodology by which such a typology may be developed.

CHAPTER 2
TREATMENTS OF THE CARTOGRAPHIC LINE

In Chapter 1, a position was established for the development of consistent guidelines for determining the character of a cartographic line. It was proposed that linear character is probabilistic in nature, because its components may be used to constrain stochastic modelling of the lines in graphic representation. Measurements of line characteristics may also facilitate the automation of line identification tasks, so that tolerance criteria may be set and modified during cartographic generalization. In this chapter, a review of the literature will be presented to justify this position.

Certain trends have developed over time: most notable are the trend towards quantification, and measuring lines as opposed to merely describing them; the shift away from determinism, in considering not only the line itself but the graphic region immediately surrounding it; and the trend to view generalization as a process including perceptual as well as mathematical components. These three trends have developed to a great extent from the research of other disciplines, ranging from geomorphology to political science. It is important for the reader to keep in mind that a certain amount of background information is necessary to review these developments, if at times the discussion seems to range far afield of the topic at hand.

It is also important to realize that the cartographic line has been treated within a variety of geometric contexts, as a set of points, a line, and even as an areal feature. Other researchers work within a spectral model, in which a cartographic line is broken into components of frequency, amplitude, and phase. The reason for this variety seems closely tied to the cartographer's attempt to cope objectively with a basically inductive task, namely, retaining the character of a geographic feature as it is represented at various cartographic reductions. What should become readily apparent in the course of this review is the difficulty involved in accurately retaining something (namely 'linear character') which has not yet been comprehensively defined.

The Line as a Set of Equiprobable Points

One of the earliest attempts at objectivity in generalization proposed a formula by which to compute the amount of information to discard for a given scale reduction. In the Radical Law, Töpfer and Pillewizer (1966) provided a mathmatical rule formulating a geometric relationship between the amount of information contained in a feature (a river, for example, or a coastline) to the scale at which that feature is represented. For lines, information is defined by coordinate locations. The Radical Law states that when scale is reduced, the information (i.e., number of coordinates) should be reduced in proportion to the square root (hence 'radical') of the scale change. Three constraints on information reduction can be included to control the reduction according to the cartographer's choice of map purpose. Without going into the mathematics, one can say that the rule works, in practice. But an homogeneous information field is assumed along the length of the line, as each coordinate carries equal importance. The rule dictates how many points to remove, but not which ones.

Srnka (1970) derives a similar relationship, without the specification of scale. His reduction formula is thus suited to functional as well as scale generalization, although his discussions are limited to the latter type. The mathematics of this and the previous work indicate guidelines for the exponential decrease of information with linear decreases in scale. This exponential kind of relationship will turn up again and again, albeit in different forms.

Automated algorithms which treat the line as a set of equally important points include all of the techniques which fall within the realm of 'coordinate weeding', or removing certain coordinates on the basis of some consistent criteria. Some examples of this include removal of every n^{th} point along a line, or removal of points selected at random; sampling points along a line at some even interval is another possibility (Lang, 1970). In these algorithms, the probability that a coordinate will be removed or retained is dependant on its location in the sequence of points, rather than on its location in the context of a line feature. Thus, the coordinate which anchors the terminus of a sand spit has as much chance of removal as any point along the length of the spit. It is easy to understand how these algorithms could produce drastic (if unintentional) modifications of geographic shapes.

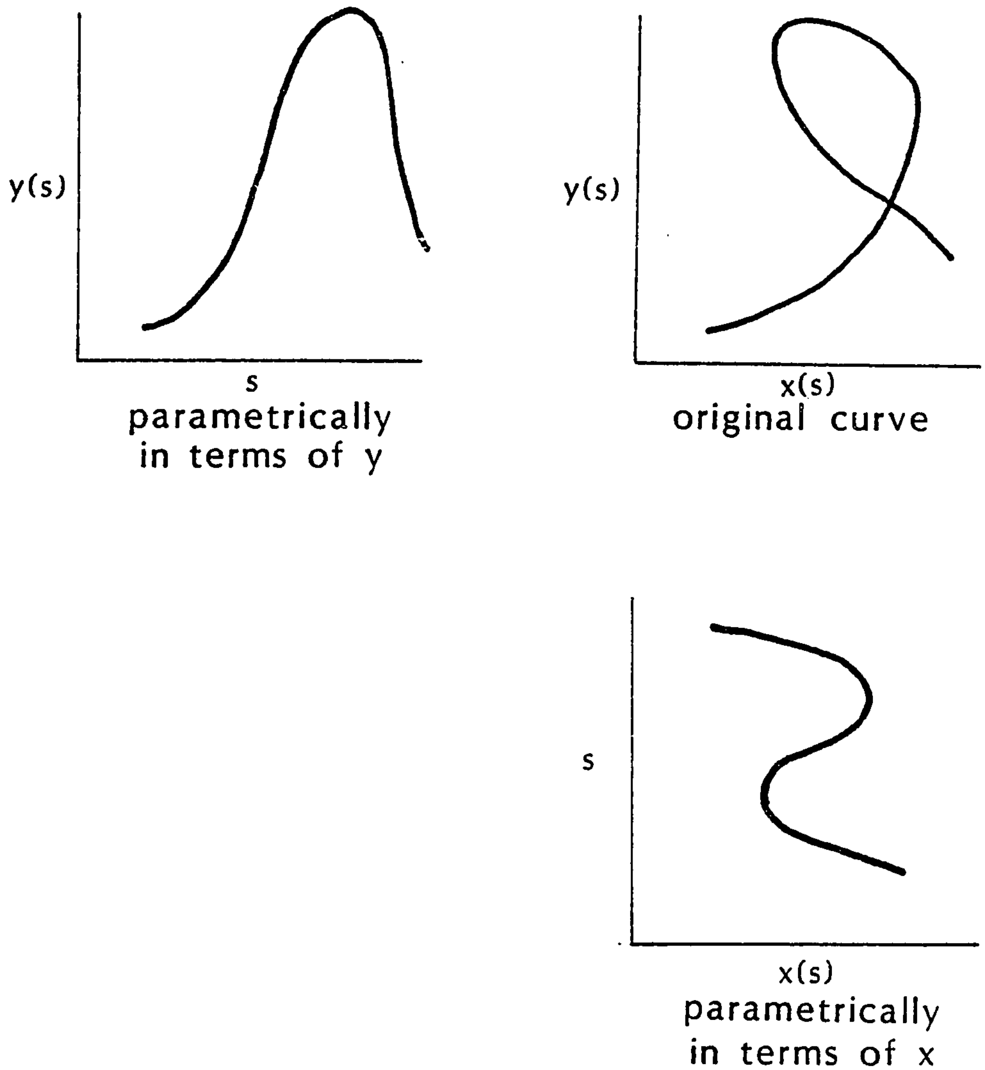
Smoothing a set of points can be accomplished as well, in treating a set of equiprobable points. Koeman and Van der Weiden (1970) suggest application of a simple moving average, in which some or all of the original points may be replaced. They compare the results achieved by exclusive runs and by overlapping runs of the moving average. In these techniques, as in coordinate weeding, no emphasis is placed on constraining the generalization according to specific line characteristics; a certain bias is implied by this, and also by the somewhat arbitrary choice of a starting point.

Boyle (1970) suggests an alternative solution, in preserving points on the line which are 'more important' to the accurate generalized representation. For examples, he cites submerged hazards on hydrologic charts, and the importance of accurately poisitioning their coordinates. These 'critical points' are not labelled as such, but are considered in a hierarchy by assigning weights (1-5) to each coordinate in the point set, to help a cartographer decide which point locations must be retained.

The Line as a Linear Feature

In another geometric context, the line has been treated as a linear feature. It has no width, only length: one line can be distinguished from another by comparing curvature, "wiggliness", or sinuosity (Maling, 1968). "The much more usual task of the geographer... is to measure the length of an intricately sinuous line, such as the longitudinal profile of a stream, a stretch of coastline or even the road distance between two places". (Maling, 1968, p.148). Maling suggests that sinuous character can be measured by running a smoothed mathematical function through the line. But consider that most empiric or free-form curves in space cannot be easily represented with a single-valued function -- the mathematics becomes too complicated to be of practical use. Parametric description of these curves simplifies matters, allowing the path of each component (in this case, x and y) to be expressed in turn.

For example, the graph A_s in Figure 2.1 appears simple, but it is



Re-Expression of a Complex Curve

FIGURE 2.1

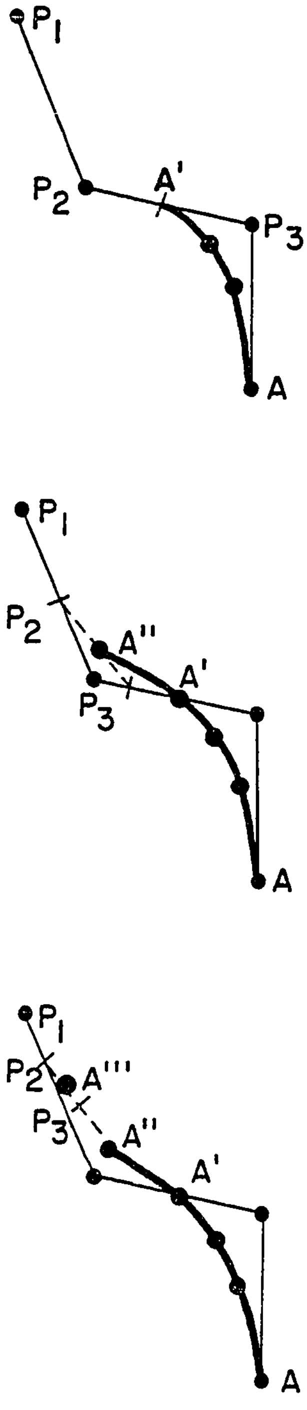
difficult to describe one variable in terms of the other (i.e., y=f(x) or x=f(y)). However, the curve can be expressed in parametric form (f(s)) by cubic equations, if each variables' path is described in isolation. Empiric lines such as rivers and coastlines are much more detailed than this simplistic example, of course. The value of parametric equations in describing approximations of geographic lines lies in their simplification of otherwise complex mathematical expression. It is for this reason that parametric equations are so often used in generalization by smoothed mathematical functions.

One method which automates parametric smoothing applies a spline function to the line. Classic splines are achieved by piecewise polynomial approximations of segments of the line, based on the derivative of each segment taken in turn. A large number of computations is involved, as a new polynomial equation must be derived for each segment in the spline. Chaikin's algorithm (1974) provides a somewhat faster computational alternative: he generates a curve by weighting the midpoints of a segment to generate a subsegment, then weights the midpoints of the subsegments, saving midpoints every time (Figure 2.2). Chaikin's algorithm is recursive, producing a sequential list of points to constitute the smoothed curve. It permits discontinuities as well as non-closure and self-intersection. Riesenfeld (1975) comments that Chaikin curves approximate a quadratic B-spline but are much more quickly constructed.

Another method which has gained recent popularity is the Bezier curve, developed originally for the automated milling machinery at Renault in France (Gordon and Riesenfeld, 1974). This is another polynomial

CHAIKIN'S ALGORITHM

(CHAIKIN, 1974)



$A' = (P_2 + P_3)/2$
STORE A'

$P_3 = (P_2 + A')/2$
$P_2 = (P_1 + P_2)/2$

$A'' = (P_2 + P_3)/2$
STORE A''

$P_3 = (P_2 + A'')/2$
$P_2 = (P_1 + P_2)/2$

$A''' = (P_2 + P_3)/2$
STORE A'''

FIGURE 2.2

approximation, a series expansion of a binomial weighting function:

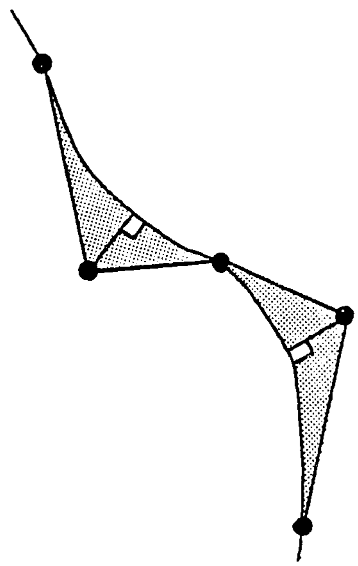
$$P(t) = \sum_{i=0}^{n} P_i \frac{n!}{i!(n-1)!} t^i (1-t)^{n-i}$$

$$= (1-t)^n P_0 + nt(1-t)^{n-1} P_1 + nt^2(1-t)^{n-2} P_2 + \ldots + t^n P_n$$

(taken from Chasen, 1978)

The series is expanded once for x and once for y coordinates, and the number of terms determines the number of points which will be generated. The resulting curve is similar in appearance to Chaikin's; except that Chaikin computes midpoints for successively small segments, working from the center towards each segment endpoint. Bezier computes a probability surface which changes weights from one end of the segment to the other, effectively 'pulling' the smoothed curve towards the original as it moves along. Either of these algorithms would be adequate in Maling's application of a smoothed function to a cartographic line. The smoothed function may be used as a cartographic representation, or to compare the sinuosity of two different lines by measuring discrepancies in length or area between the two (Figure 2.3).

Two fundamental problems arise in this kind of treatment, as Maling (1968) himself admits. First is the question of which smoothing function to use A moving average, a spline function, and a Bezier curve will produce quite different visual representatons of the same line, as an example. Furthermore, there is a problem of accuracy, as the smoothed function will contain only a small subset of the original coordinates. Second, and this follows directly from the first point, is the question of measuring the length of a geographic line. In order to determine distance or areal discrepancies between geographic features and their counterparts,



Measuring Discrepancy by Distance and by Area

FIGURE 2.3

one must first be able to define their lengths (or the lengths of their boundaries). For cartographic and geographic lines alike, this presents an interesting dilemma, which warrants a brief review.

Digression: Peculiarities of Geographic Length

Consider that the linear feature is continuous, and that the units used to measure its length are discrete. It turns out that the more precisely one measures an empirical (i.e., naturally occurring) line, the longer that line becomes. Several Europeans (Shokalsky, 1930; Volkov, 1949; Steinhaus, 1950, 1954) pointed this out, that the series of lengths obtained by repeated measures using smaller and smaller units does not converge.

If the length of the Puget Sound coastline is measured on a LANDSAT image, its length will tally at some (rough) multiple of 79 meters, assuming that image pixels are 79 meters on a side. (This is not absolutely true; but it will serve for the purpose of the argument.) Features of the coastline which are smaller than this will not be present in the image, and so will not be measured. A high-altitude (70,000 feet) photograph of this area will include some of these features, however; the length of the coastline in this image will equal the length as measured on the LANDSAT representation plus the length of all additional features resolved on the larger scale photograph. The coastline will therefore be longer, by definition. A low altitude (20,000 feet) photograph will include still more features; adding the length of these to the coastline measure will increase the length once again. And so on.

The Polish mathematician Hugo Steinhaus suggested that a practical solution to the paradox of length would employ measurement techniques which define only a lower limit for the length of a geographic feature (Steinhaus, 1950). He defined the length of an arc [L] as the limit of the summed lengths of straight line segments [$\sum l(i)$] making up the arc, as the segment size [l] becomes infinitely small:

$$L = \lim_{l \to 0} \sum_{i=1}^{n} l(i)$$

He maintains that this limit always exists; if infinite, however, the arc is said to be non-rectifiable. Steinhaus (1954) developed a "longimeter" to approximate measures for the length of a curve lying in a plane, as follows. He cuts the plane with a series of equidistant parallel lines, then counted the number of lines which intersect the curve. The measure achieved by this method is said to be "length of order n", where n equals the line spacing of the longimeter.

Other disciplines were interested in this problem as well. Some of the best known work in the area was done by Lewis Richardson, who investigated the paradox of geographic length when sidetracked from political science research relating the length of political frontiers to the occurrence of national boundary disputes. He used map dividers at various separations to measure coastlines and rivers in political boundaries on atlas maps. He found that apparent boundary length increased steadily with the decrease in divider separation, and the rate of increase differed somewhat for each measured line (Richardson, 1961). When repeating this measurement process for a regular polygon such as a circle, Richardson found that here, total length measurement stabilized quickly with decreasing units of measure.

Figure 2.4 shows a comparison of lengths for various coastlines, and for a circle, values which Richardson measured and plotted as logarithms. The negative trend of the plot led to the following derived relationship:

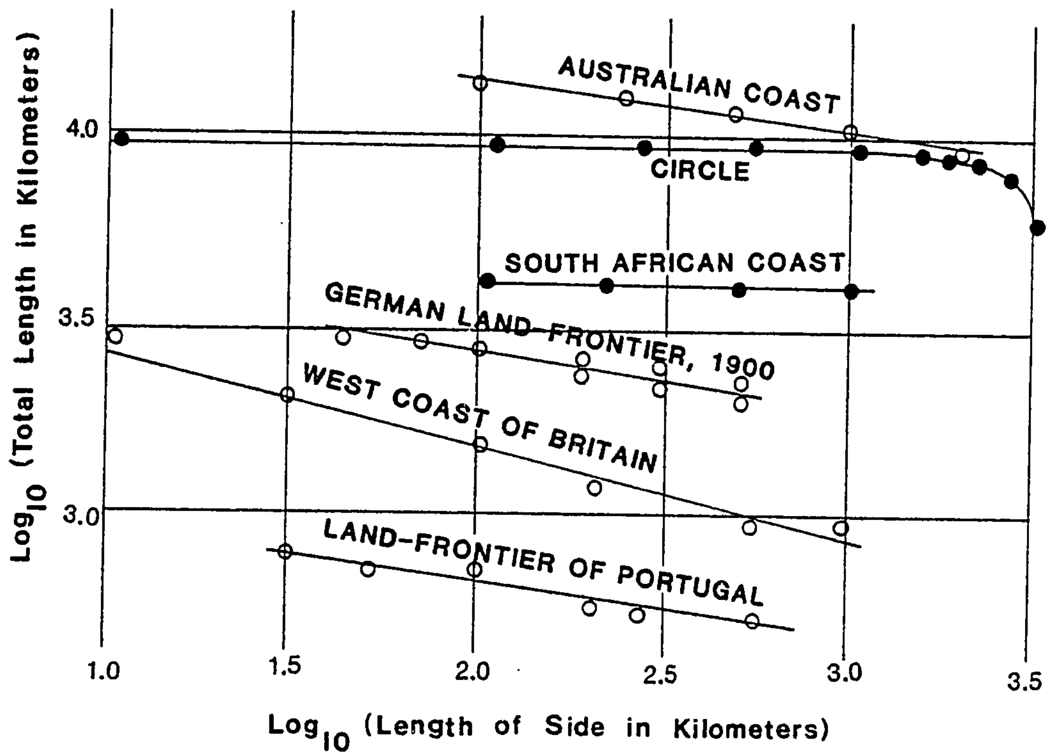
$$\sum l = l^{-\alpha}$$

$\sum l$ represents total measured length, and l represents the magnitude of the measurement unit, or divider separation. Richardson states " α is a positive constant, characteristic of the frontier" (Richardson, 1961, p.170). He goes on to say that α may be positively correlated with the perceived irregularities of the coastline, or its lack of smoothness, concluding that the differing slopes of his plots correspond to the characteristic irregularities of the coastlines they represent (Table 2.1).

TABLE 2.1
Values of α for Selected Coastlines
(after Richardson, 1961)

Britain	0.25
Germany	0.15
Spain-Portugal	0.14
Australia	0.13
South Africa	0.02

Two facets of this work are important in looking at the development of cartographic line generalization: first are Richardson's notions about accuracy. He comments, "It is doubtful whether the total polygonal length of a seacoast tends to any limit as the side of the polygon tends to zero." (Richardson, 1961, p.170) This is not to imply that the most accurate length is derived from the most precise measurement, but rather that accurate measurement becomes undefinable at the limits of precision (when unit length approaches zero). The problem of deciding which level of



RICHARDSON'S EMPIRICAL DATA
ON THE RATE OF INCREASE
OF COASTLINES' LENGTHS

FIGURE 2.4

precision will produce the greatest degree of accuracy is left entirely to Richardson's readers -- he is only pointing up the fallacy of over-reliance on precision in cartometric tasks. Chronologically, this work provides an early step in the shift away from determinism.

The second facet of the work which is important is the derivation of the exponent α , which is a preliminary quantification for the characteristic irregularities of a sinuous line. European research had also focused on the validation of map-measured distances (eg. Shokalsky, 1930). In these studies, too, regression-like equations were formulated by which 'actual' geographic length could be predicted, based on the particular units of measurement. All of the formulas include a term variously identified as a measure of line character, irregularity, or sinuosity; but this term was most often left as a by-product of the derivation. Richardson considered it a trivial finding; his interest remained focused on the accuracy of geographic measures. However, the topic (and the exponent) will be returned to further along in this discussion.

The Line as an Areal Feature of Finite Width

As discussed before, the Polish mathematician Steinhaus suggested that a practical solution to the paradox of length would employ measurement techniques which define only a lower limit to the length of a geographic feature. The technique as developed by Perkal (1956, 1966a, 1966b) is directly applicable to the study of geographic process, and provides a good example of how a cartographic line may be treated as an areal region, and not simply as a line.

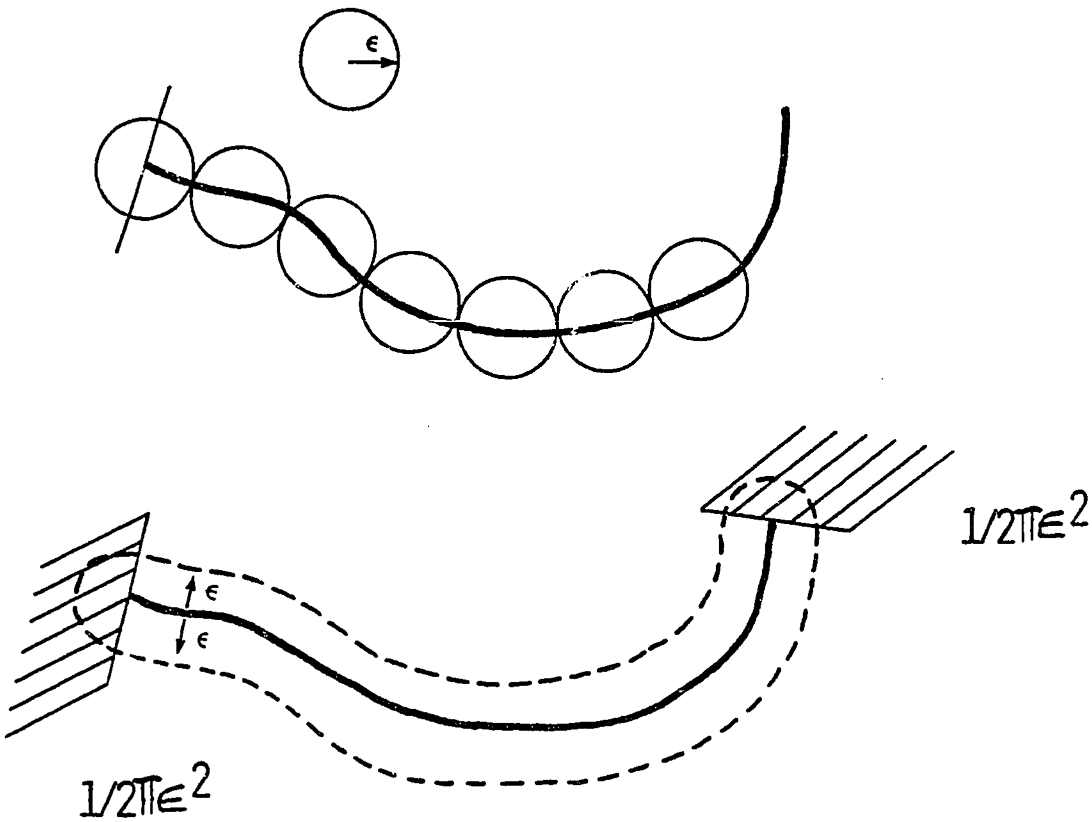
Perkal sidesteps the problems involved in measuring a (possibly infinite) geographic length, working instead to determine at what scale of measurement the length becomes finite, or rectible. He proceeds by constructing an "e-neighborhood" ('e' for empiric) around a curve. The -neighborhood is the set of all points on the plane for which the distance from the curve is less than or equal to some small but finite value epsilon. The effect is similar to that of rolling a circle along a curve, letting the path of the circle define an envelope or smooth closed polygon surrounding the curve. Perkal then computes the area [$A(e)$] of the e-neighborhood, subtracting the area of one half circle [$\frac{1}{2}\pi e^2$] at each end. (Figure 2.5) The length of the curve will be directly proportional to the area of the e-neighborhood, and inverse to the diameter $2e$ of the circle, thus:

$$L = \lim_{e \to 0} \frac{A(e) - \pi e^2}{2e}$$

The final step is to simplify the areal calculations, following from Steinhaus, replacing the area term with a simple count (n) of the number of (non-overlapping) circles of radius e which the curve intersects. He computes an 'epsilon length' for curve X as

$$L_e(X) = n - (\pi/_2)e$$

The units of e-radius determine the units of epsilon length. Thus the accuracy of the measured length is in fact determined by the scale of the measurement. Notice that an insignificant change in epsilon will not change the epsilon length substantially: the measure is in this sense more robust than the length measures of either Steinhaus or Richardson. It is



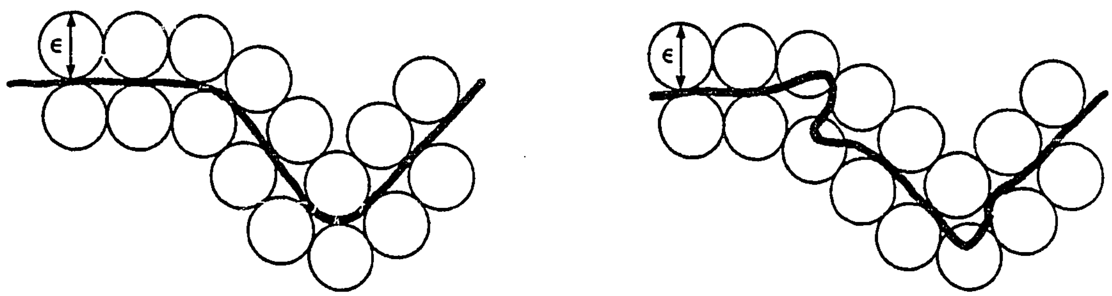
PERKAL'S ϵ-NEIGHBORHOOD

FIGURE 2.5

also interesting for the moment to consider e as a diameter instead of as a radius (Figure 2.6). A curve is said to be e-convex if the string of tangent circles (of diameter e) are tangent to the curve only at a single point. The curve on the left is called e-convex, for this particular value of epsilon; the curve on the right is not. However, the curve on the right <u>will</u> be e-convex for some lesser value of epsilon, which allows discernment of the smaller arcs in the curve. Perkal suggests using the envelope of the e-neighborhood as a generalized version of the empiric line. In a mechanical context, this is exactly what happens during some stages of the cartographic process.

Consider, for example, the minimum turning radius of a polar planimeter, a plotter head on a Calcomp flatbed, the magnetic grid size of a graphics tablet, or even the pixel size on a satellite image. These are effective e-values for the respective instruments: all lines plotted, all lengths measured will be a function of this value. It may be functional in a particular research design to specify a value of epsilon instead of relying on the tolerance thresholds of the machine. Maling (1968) considers the size of the white dot on the stereoplotter platen in this context. Boyle (1970) makes similar comments on the generalizing effects of hardware limitations in computer-drawn maps. Brophy (1973) designed a computer algorithm which effectively rolls a circle along a line, generalizing by removing details which fall within the boundaries inscribed by the circle.

Another interesting point here is that the e-neighborhood is not always symmetrical on both sides of the empiric line: this is in part due to the magnitude of epsilon. It is also a function of the sinuous twists



ε -CONVEXITY

(after Perkal, 1966)

FIGURE 2.6

and characteristic irregularities of empiric lines. And herein lies a direct application to the study of geographic process.

> "... The boundary of the ε-neighborhood may be thought of as the trace of points nearest the coastline left by the path of a circle of ε-radius which is rolled along the coastline on both sides. Because of the shape of the capes and bays, the length of this trace will not be the same on each side of the seacoast. Ships swing wide to avoid the capes but trains swing inland to avoid bays and estuaries. The coastline, then, appears to have an inside length and an outside length for purposes of movement along its perimeter, depending upon the technical requirements of the vehicle used. This statement may be made more general. The boundaries of the ε-neighborhood of a line are ordinarily not equal; the one-side length equals the other-side length only if the line is ε-convex."
>
> Nystuen (1966, no page)

Perkal's techniques provide a broad methodology by which to investigate the geographic notion of a linear boundary as region. Nystuen considers the effects which regional shape may have upon a spatial process: non-symmetry of the ε-neighborhood may facilitate studies of a process which affects one side of the boundary but not the other, studies of flows across a boundary, and studies of processes which exist only at the boundary (i.e., within the ε-neighborhood). Nystuen calls this application the concept of local convexity, or "convexity-in-the-small".

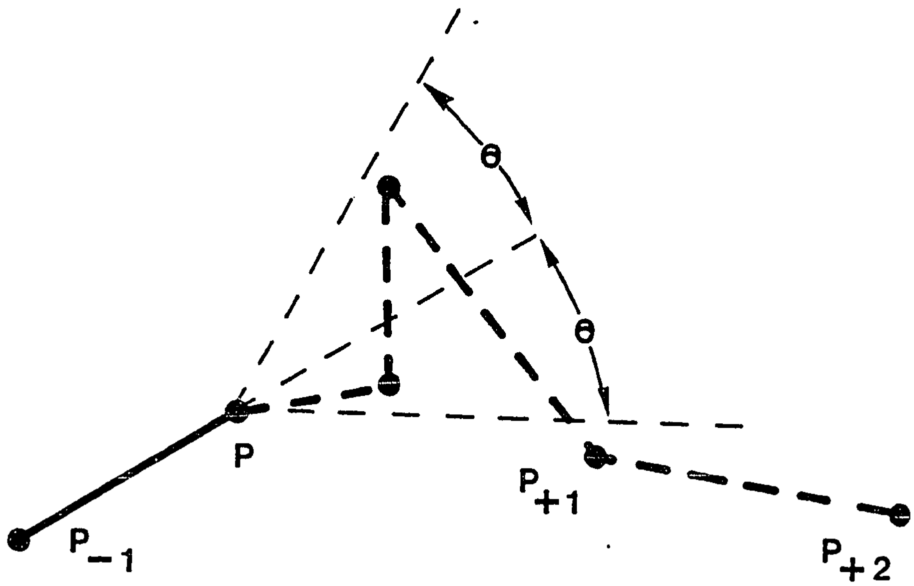
There are direct implications for line generalization as well. To return to the coastline example, a cartographer might generalize one side of the ε-neighborhood of the California coast in mapping seasonal traffic flows along Highway 101, and/or generalize the other side of the coast to map fishing rights along the three-mile international boundary.

<u>The Line as a Spectral Feature: Filters and Frequencies</u>

Applications of Perkal's method to line generalization are for the most part indirect; Nystuen's reference to ϵ-neighborhoods was generally neglected by cartographers. The generalization algorithms which were developed during the 1970's tended to focus on simplifying the existing line, rather than replacing coordinates by means of smoothing functions and splines. With the increasing prevalence of automation in map-making, cartographers could rely on improving the accuracy of their displays. Accordingly, the objective in generalization shifted somewhat: the purpose was not so much to reduce information as it was to filter it. For the most part the filtering process was computerized by setting arbitrary tolerance limits within which points along the line could be eliminated. Two kinds of tolerancing were pursued. In each one, an envelope was defined surrounding the line, in a manner similar to Perkal's ϵ-neighborhood.

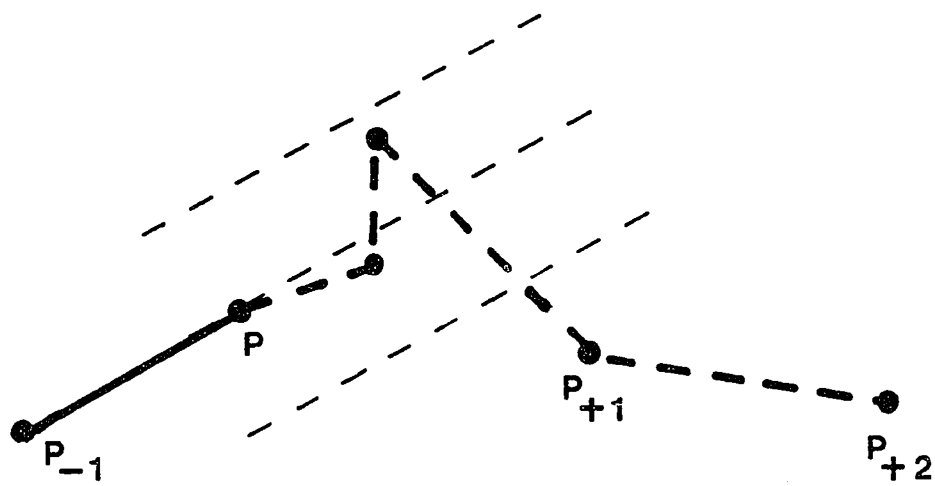
In angular tolerancing, a cone of angle 2θ is projected forward from each point along the line, in succession (Figure 2.7). Points ahead of the vertex are discarded, until a point is found which falls outside the threshold of the cone. The cone is then moved to this point, and the process repeated. The disadvantage of the algorithm is its tendency to give long, gradual curves a blocky appearance; because of this, the cone is usually constrained to a fairly short forward search.

Distance tolerancing is quite similar in nature, eliminating points on the basis of a threshold width for a corridor extended along either side of the line (Figure 2.8). This method also requires a limiting distance for forward searches, and is best suited to the elimination of small errors



ANGULAR TOLERANCING ALGORITHM

FIGURE 2.7



DISTANCE TOLERANCING ALGORITHM

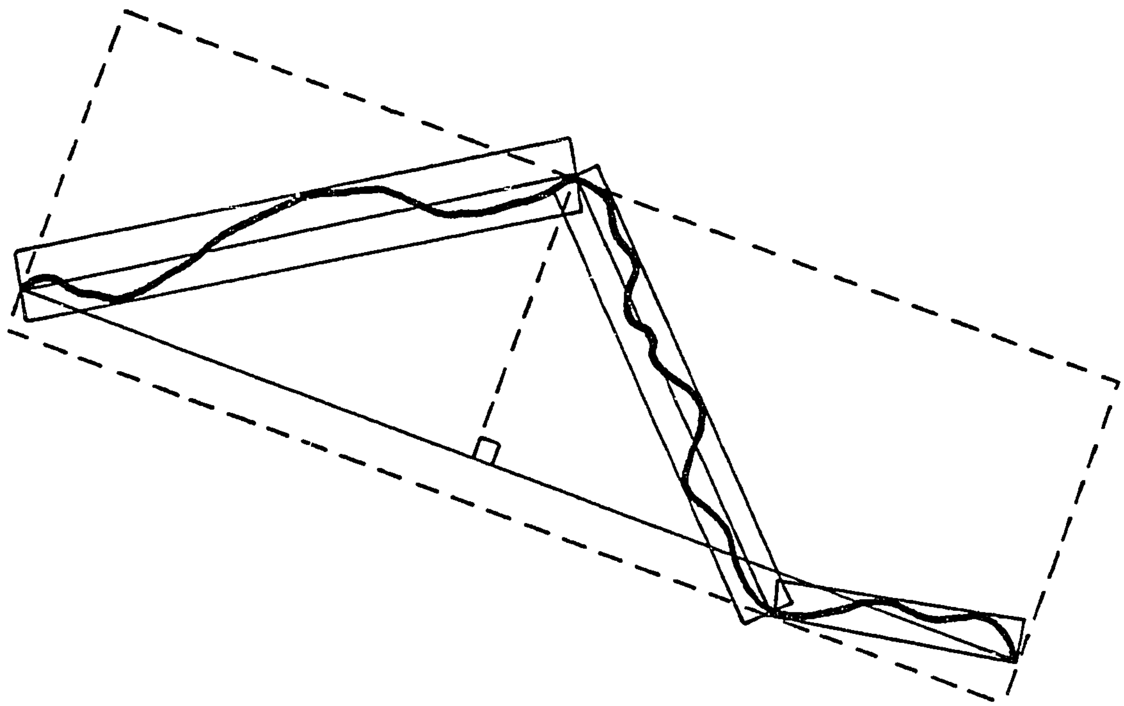
FIGURE 2.8

from sets of stream-digitized coordinates.

Tolerancing was formalized and refined by Peucker (1975) (now Poiker), who combined the notion of an areal band or corridor with a consideration of linear details which occur at a specific frequency. Each frequency has an associated bandwidth by which it may be represented; in the graphic adaptation of this electronic metaphor, decreasing bandwidths can be used to selectively isolate portions of fine detail (so-called high frequency information) from the line. The bandwidth model is operationalized as a data encoding scheme. Endpoints of the line form the length parameter for the band, and its width is defined by points of maximum deviation on either side of the straight line segment connecting the endpoints (Figure 2.9). These four points are used to define bandwidths of the next higher order frequency, as shown in the figure. The process of encoding continues until the desired minimum bandwidth has been achieved.

Ballard (1981) has automated the bandwidth encoding scheme as a data structure which he calls Strip Trees. A modified version of this data structure will be discussed in Chapter 4. Poiker cites several cartometric applications for bandwidth encoding, including computation of line intersections, matching of digitized lines, and searching tasks within polygons. The importance of his work lies in the ability to tie the accuracy of the application to some finite bandwidth which encompasses a particular scale of representation.

The technique is also useful in line generalization, which in this context becomes a simple process of selective bandwidth filtering for high, medium, or low freqency details. Douglas and Peucker (1973) developed one



BANDWIDTH ENCODING
(after Peucker, 1975)

FIGURE 2.9

of the few tolerancing algorithms which selects coordinates, as opposed to eliminating them, based on ideas which were subsequently refined into this bandwidth model. (It is interesting to note that both Ramer (1972) and Duda and Hart (1973) devised similar techniques at about the same time.) Although it seems a trivial distinction, there is a significant perceptual difference in a line generalized by selection of points: this distinction, and the algorithm which produces it, will be discussed in a later section of this chapter.

An earlier discussion of generalization as a filtering process is provided by Tobler (1966). In considering cartographic information as occurring at various spatial frequencies, he views generalization as a filtering of components such as amplitude, wavelength, and phase. Tobler suggests using Perkal's epsilon to identify threshold frequencies at which a geographic feature is sensitive to the filter -- it can be smoothed, and then designing a spatial filter to accomplish the task. The filters he describes are matrices of probabilities, which are used as smoothing weights. For example, a binomial weighting function provides a high-frequency filter which has a strong generalizing effect on small details. However, a Fourier series should provide another valid filter, and might be designed to smooth information at a variety of different frequencies.

As a matrix manipulation, the frequency response of a spatial filter is computable. Tobler has derived a concept of generalization as one more kind of transformation which is objective, predictable, and in some cases, reversible: as a matrix transformation can be computed, so can its inverse, if it exists.

> "As a priori criteria one would expect that generalization... should somehow eliminate 'small scale' features. If generalization is considered as a transformation, it is natural to inquire whether this process can be reversed. If topographical maps are generalized the result should conform to appropriately modified accuracy standards. Finally, certain statistical parameters...should be preserved by the transformation."
>
> (Tobler, 1966, p.1)

The importance of this work in the development of generalization becomes clear if all the threads of this research discussion are tied together. According to Steinhaus, and to Richardson, an empiric line can only be approximated in graphic representation, because its geography can never be precisely measured. Thus any graphic representation provides at best only a generalized version of a linear feature. Perkal's approximations of lines by e-neighborhoods indicates that the set of information which is included in the generalized representation will be a function of the scale (measurement units), and the purpose of the measurement (which side of the line is measured). It will also be dependant on the places of e-convexity, on the sinuous curves characteristic of the line. Sinuosity has still not been defined, although Richardson's work points to a possible index.

What Tobler has added to this thread of inquiry is that these generalized representations can be arrived at through a filtering process, which in theory can be designed to remove or to restore information, to smooth or to enhance the graphic representation, in effect. He also establishes the importance of preserving statistical characteristics of the line as it is filtered, in order to preserve the character of the representation. It seems appropriate at this point to digress once again for a brief review of one methodology which uses statistical parameters as a relative measure of line sinuosity.

<u>Digression: Lines of Fractional Dimension</u>

Let us return for a moment to the work of Lewis Richardson (1961), who found that small decreases in unit size result in often substantial increases in measured length. The rate of increase was found to vary for specific coastlines which were measured. Richardson commented that these rates were possibly related to the sinuous nature of each coastline, but saw no great importance in his findings. It remained for Mandelbrot (1967) to derive its theoretical basis. His work also provides a comprehensive example of generating figures which retain certain statistical properties.

Mandelbrot began by reexpressing Richardson's relationship $(\sum l = l^{-\alpha})$ as an equality between the increase (r) in total line length and the size (1/N) of N equal steps required to travel the length of the line.

$$r = (1/N)^{1/D} \qquad D >= 1.0$$

What this formula describes is really quite basic — if a straight line is broken into N equal pieces, the ratio of the length of any piece to the length of the whole line will be 1/N. This is called the ratio of self-similarity: because it is linear, the figure is said to be empirically one-dimensional. In two dimensions, a plane can be subdivided into N equal squares (as in tiling). Each square will have an area of ratio $(1/N)^{1/2}$. In three dimensions, the similarity ratio will be in cube roots, or $(1/N)^{1/3}$. In general form, the value which D takes on is mathematically equivalent to the dimension of the feature.

Solving for D:

$$\ln r = \ln [(1/N)^{1/D}]$$
$$\ln r = 1/D \ln (1/N)$$
$$D \ln r = -\ln N$$
$$D = -\ln N / \ln r$$

Note: substitutions between negative logarithms and logarithms of fractions will simplify arithmetic below; the computational formula is as follows:

$$D = \ln N / \ln (1/r)$$

As a computed ratio, D can take on fractional values: Mandelbrot's derivation expands the concept of 'dimension' into the realm of real numbers. Thus a figure will have a topologic dimension (as a point, line, area, and so forth) which describes to what class of figure it belongs; and it will have an empirically computed dimension which describes its "degree of complication" (Mandelbrot, 1967, p.636), or the characteristic way in which the figure fills up the space which it occupies. D is conceptually equivalent to Richardson's (1961) exponent α in describing this property, in that it varies according to the irregularities of the line being measured. In contrast to Richardson, however, Mandelbrot does not concern himself with problems of absolute length, focusing instead on the rate of its increase. An example may serve to explain the concept of empiric dimensionality; and then attention can be returned to a more geographic application.

Take for example the mathematical curiosities such as Peano curves (Figure 2.10), or Koch Triadic Islands (Figure 2.11), as they are constructed. As linear features, the curves are topologically one-dimensional. For the Koch Island, replace each side of the original unit triangle by a four-sided open polygon (4 sides is arbitrary; 5, 6, or n

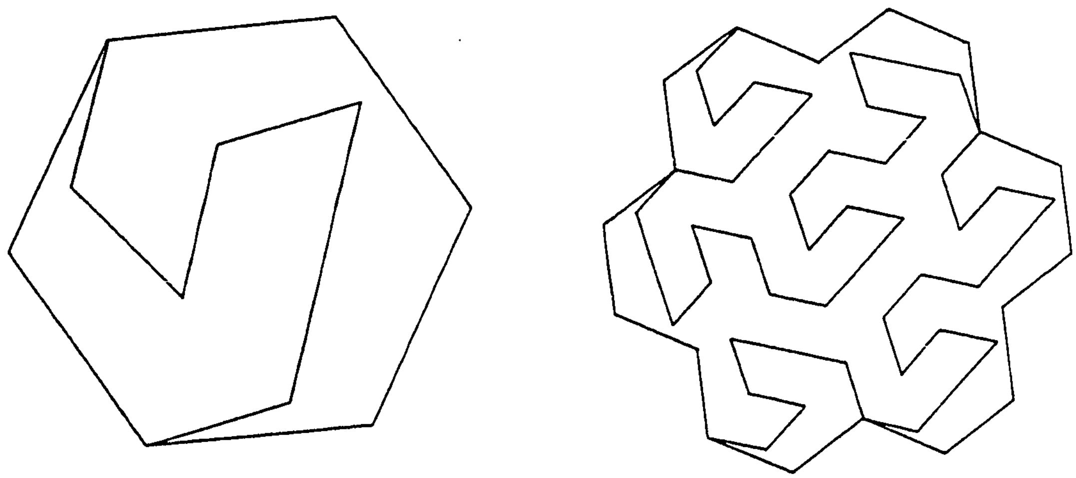


Figure 2.10
PEANO CURVE
(after Mandelbrot, 1982)

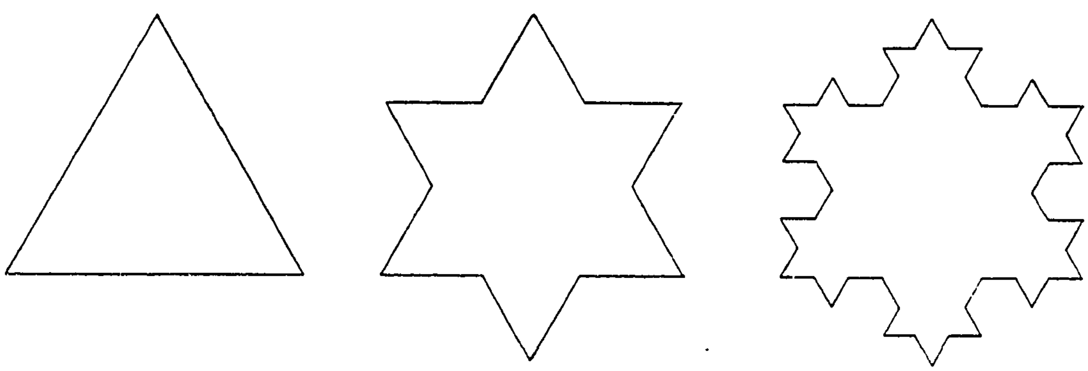


Figure 2.11
TRIADIC KOCH ISLAND
(after Mandelbrot, 1982)

sides are also possible); thus N = 4. Notice that the length of this side has increased by 1/3 in making this replacement; thus r = 1/3.

$$
\begin{aligned}
D &= \ln(4)/\ln(1/(1/3)) \\
&= \ln(4)/\ln(3) \\
&= 1.2618
\end{aligned}
$$

In effect, this value represents the rate at which the curve will fill the plane in which it lies. For the Peano curve, D = 1.1291, which means this curve will fill the plane less quickly than the Koch Island, for equal values of N. It should be obvious in looking at the two curves that this is so: D provides an index of the mathematical complexity of the figures, which is related to their visual sinuosity as well.

Mandelbrot calls these figures fractals, because their empiric dimension is fractional. Fractals differ from regular curves and polygons in that a regular geometric figure will have an empiric dimension which is an integer. As shown by the plot of the circle in Figure 2.4, the integer dimension reflects the fact that overall length of regular polygons stabilizes very quickly to some finite value, while the same cannot be said for fractals. Coastlines appear to have fractal characteristics, then, because their length does not stabilize: Mandelbrot (1967) computes the fractal dimensions for Richardson's coastlines to be 1.25 (Britain), 1.15 (Germany), and 1.02 (South Africa), concluding that Richardson's exponent is computable from his fractal dimension as

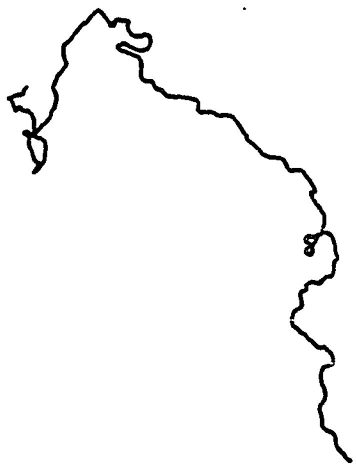
$$\alpha = 1 - D.$$

For the Koch Triadic Island, then, each unit side of the figure is replaced by a four-segment curve. Each of these four segments becomes a new unit side in the resulting figure, and the process is repeated. As each segment in the curve is replaced by small scale replicas of the entire curve, the figure is said to become self-similar; and the fractal dimension provides a ratio describing the rate at which the self-similar replication process increases the length of the curve.

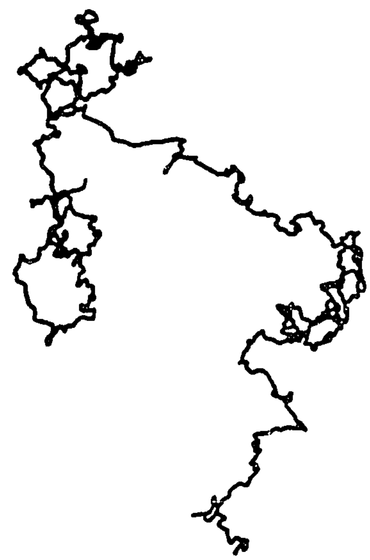
Mandelbrot maintains that the self-similarity concept has logical validity for coastlines as well as for mathematical constructions such as space-filling curves. The self-similarity for Peano curves and Koch Islands is termed 'precisely equivalent', because any piece of the curve can produce an exact replica of the whole. (As an aside, a holographic image is precisely equivalent in an optical sense, in that if the holographic glass plate negative is shattered, any piece can be used to reconstruct the entire image (Fleischer, 1982).)

On the other hand, statistical equivalence is probably a better term to describe the self-similarity of coastlines and other naturally-occurring features. Mandelbrot demonstrates this by generating a curve of random Brownian motion, and then generating a Brownian curve constrained by two different values of D (Figure 2.12). The stochastic process produces lines which bear some visual resemblance to coastal sinuosity; and further constraints produce other kinds of apparently geographic lines (Figure 2.13).

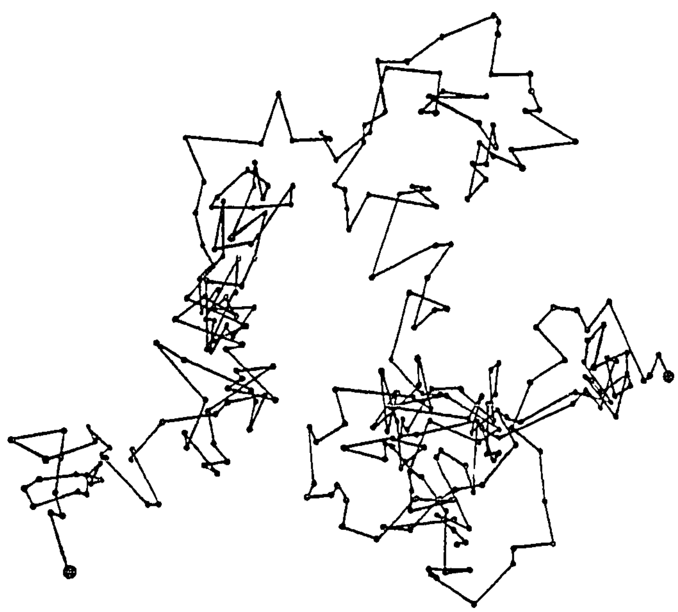
Dutton (1981) has developed an algorithm which produces fractal models of actual cartographic lines. The process involves standardizing angles according to four predetermined tolerance parameters. His intention is to introduce self-similarity to the representation, by standardizing its shape irregularities. "One wishes for a measure of geometric complexity and



D = 1.1111



D = 1.4285



D = 2.0000

GRAPHICAL MEANING OF D

FIGURE 2.12

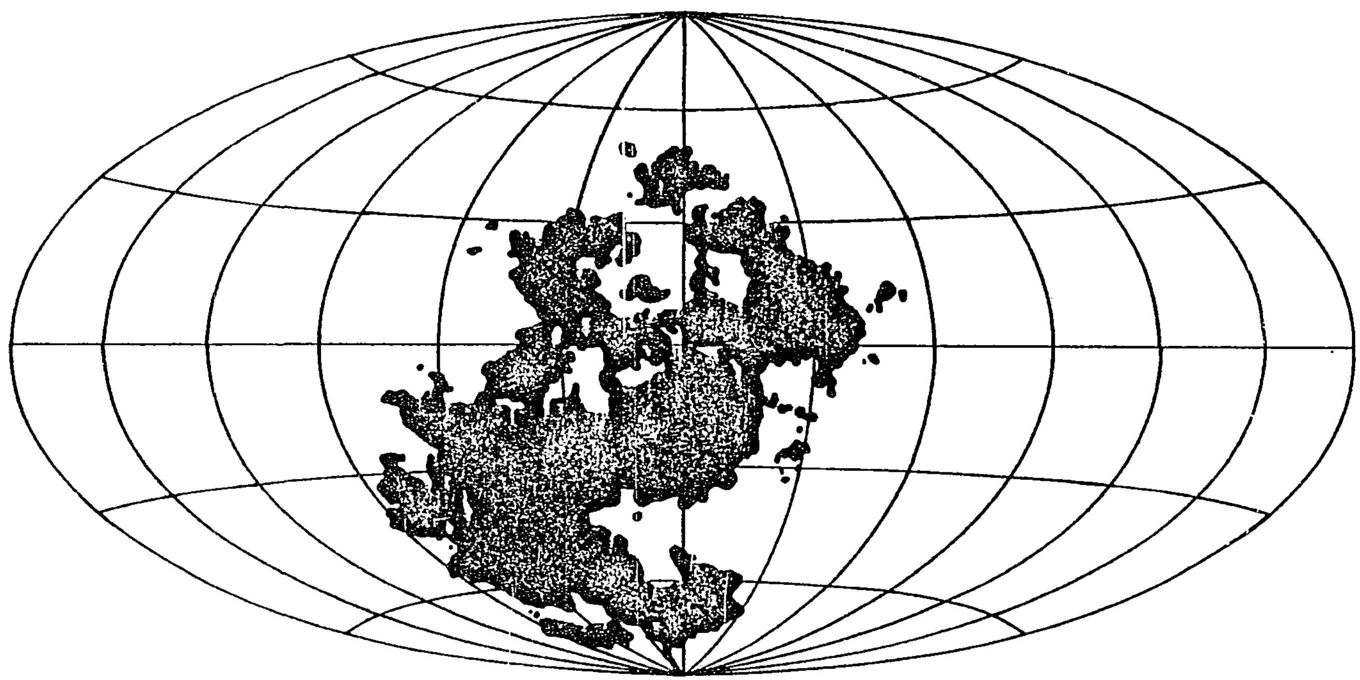
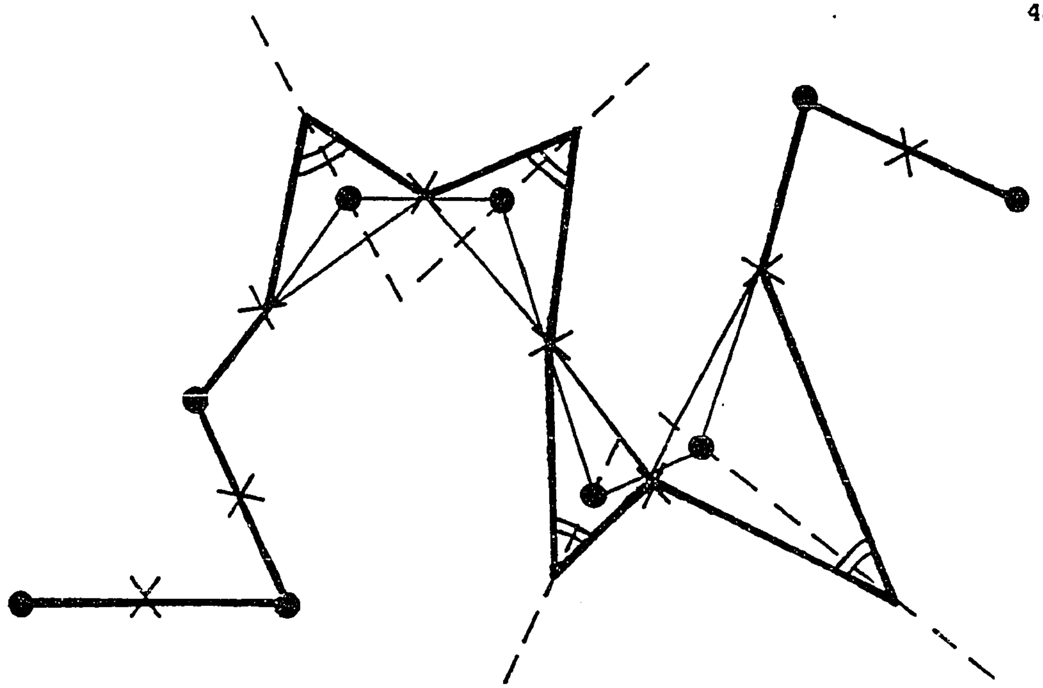


Figure 2.13
STOCHASTIC PANGAEA
(after Mandelbrot, 1982)

irregularity that is as general as that of entropy in thermodynamics. Fortunately, foundations for such a vocabulary and for such measures have been developed." (Dutton, 1981, p.24) Once the angles have been standardized, any portion of the line will exhibit the same complexity which characterizes the line as a whole: this is a pseudo-approximation of the self-similarity which Mandelbrot proposed.

The algorithm computes midpoints for two adjacent segments, connecting these to form a triangle with the common endpoint (Figure 2.14). At the apex of each triangle is the common endpoint -- this is the vertex to be moved. The vertex is slid along the apex bisector until its angle becomes (nearly) equivalent to some predetermined value. Then midpoints of the two segments become fixed vertices, and the process is repeated. The geographic accuracy of the line is obliterated by moving the original vertices, of course; but this problem could be easily remedied by holding constant the vertex positions and moving only the midpoints.

Four parameters constrain the fractalization. The first parameter, SD, determines the angle of standardization, which Dutton says is equivalent to Mandelbrot's fractal dimension (D). The second, UC, is a uniformity coefficient: ranging between -1.0 and +1.0, it determines the degree to which the angles are equivalent. At +1.0, verticies are moved the entire distance needed to produce equivalent angles; at +0.5, they are moved only half of the necessary distance, and so forth. For negative values, movement occurs in the opposite direction. Two other parameters respectively define the maximum and minimum applicable segment lengths to fractalize, controlling for straghtness (ST) and smoothness (SM) in the final reresentation. These are similar to the (distance and angular)



● = movable vertex (triangle apex)

X = fixed midpoint (triangle base)

DUTTON'S FRACTALIZING ALGORITHM

FIGURE 2.14

thresholds used in the tolerancing algorithms discussed previously . For example, they would insure retention of long gradual curves during fractalization. Even with these last two, some smoothing is necessary subsequent to the enhancement process, to 'fine tune' the look of the line.

> Unlike [splining] and other methods for coordinate reduction and chain smoothing, fractalizing permits features to be exaggerated and smaller scale features to be introduced into digitized curves, as well as allowing features to be eliminated. The exaggerations and additions are not arbitrary forms introduced to the chain, but are caricatures and recursions of forms already found there.
>
> Dutton (1981), p.25

Dutton's algorithm works to preserve linear character, then, by reexpressing shapes which are already present, and by introduction of standardized caricatures at smaller and smaller levels of detail. He offers no guidelines for choosing the values of his four parameters to construct lines of varying character, except to say that his parameter SD is operationally equivalent to the fractal dimension. And so while it has been established that linear character can be replicated, it has still not been defined. As Dutton admits, the algorithm is quite sensitive to localized deviations in sinuosity, which are not accounted for by the constant value of SD.

This is not to say that fractals can provide geographically accurate representations, but rather that fractal constraints on stochastic motion may be useful in modeling some of the visual components which are present in geographic representations. In other words, what you see in a fractal image is similar to what you expect to see in a geographic representation -- the visual character of the two images are quite similar. The computer graphics industry has capitalized on this effect, using fractal geometry to generate all sorts of generic landscape objects, including clouds, trees, and even mountain range backdrops. "The major problem in making a realistic computer animation is capturing the right balance between order... and disorder." (Fleischer, 1982, p.52). It is the randomness of Brownian motion which has been most commonly used to generate fractal landscapes (Plate I) which appear to be realistic geographic representations (Plate II). (Carpenter, 1980; Fournier & Fussell, 1981)

This approach has been criticized in the literature, most notably for the fallacy of considering geographic features to be self-similar. Goodchild (1980) comments that geographic features are generated by geomorphic processes which are scale-dependant in nature, and which supply visual clues to help determine the scale of the feature in representation. These clues are not present in a line or a surface which is its own replicate at every scale. Goodchild adds that this does not apply to, for example, lunar landscapes, whose form is related to very different morphogenic processes. This may be the reason why fractalized landscapes appear so "other-worldly" (Whitted, 1982).

Derivation of the fractal dimension has not resolved the paradox of empiric length, but has provided a numeric index for lines of relative 'complicatedness' (Mandelbrot, 1967), 'irregularity' (Richardson, 1961), and/or 'sinuosity' (Maling, 1963, 1968). Time and again, researchers have come up against the problem of distinguishing between lines and linear characteristics, in an objective manner. Until line character can be analytically identified, it will remain a major stumbling block in automating generalization tasks.

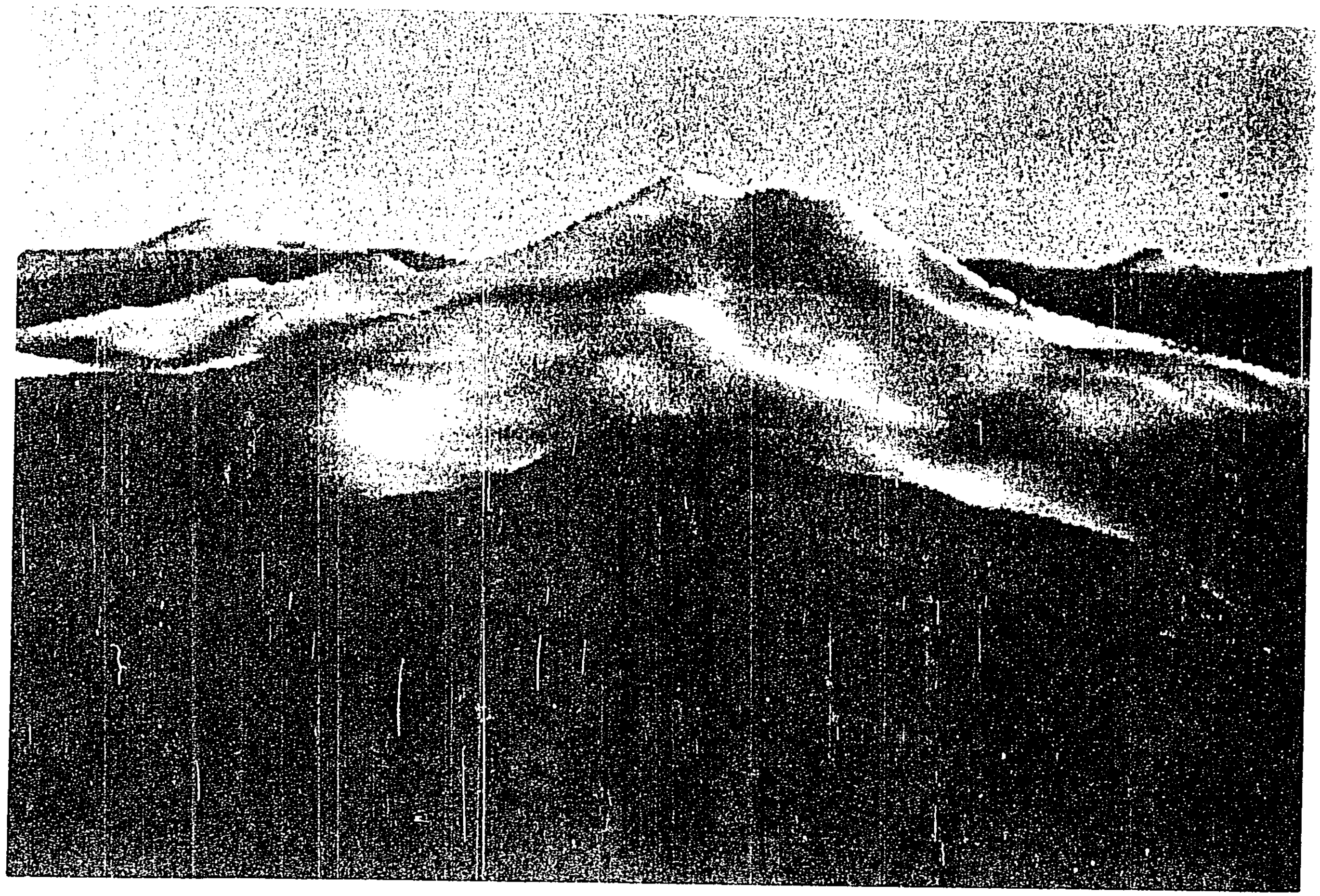


PLATE I
STOCHASTIC RENDERING OF MOUNT RAINIER
(Loren Carpenter, 1980)

PLATE II
OBLIQUE AERIAL VIEW OF MOUNT RAINIER (18 JUNE, 1932)
(Courtesy University of Washington Library, Special Collections)

One goal of this dissertation is to create and test a set of guidelines by which line characters may be distinguished. Fractal dimension will be computed for several trial lines, and used as one measure in the quantification process. It should be obvious, however, that attachment of a single metric to a set of geographic lines is not sufficient as an analytic distinction of line character; and this is especially true in a cartographic context, where the visual effect of a display can change quite drastically with changes in scale, projection, resolution, and so forth. A review of pertinent research on the perception of linear information should provide a perspective on the map reader's view of a generalized line, and should lead to a clearer definition of line character in succeeding chapters.

The Line as a Perceptual Phenomenon

From the map reader's point of view, the reason for line generalization is to allow a geographic feature to be recognized in cartographic representation, to use a map for the purpose for which it was intended. The importance of feature identification on thematic maps has long been a major focus in the study of shape generalization (Dent, 1972; Muehrcke, 1969; Dobson, 1980; Prince, 1977; Turner, 1977; Muller, 1979; Duncan, 1981) and in studies of map complexity (MacEachren, 1982; Brophy, 1980; Monmonier, 1974; Lavin, 1978; Muller, 1976; Olson, 1975). The general line of thought has been that the success of the graphic communication is based on the facility with which the generalized feature is recognized as a specific geographic feature. It would appear, then, that line character includes a definite perceptual component.

Accuracy, too, is an important part of the recognition process. Consider for example the importance of accurate coastline recognition for a sailor navigating in a triangle race around the islands of Puget Sound; or the measurements which a tactical pilot makes in planning his approach path for a high speed reconnaissance flight; or the accurate comparisons involved in cruise missile feature recognition. Accuracy can be retained during generalization by prohibiting coordinate movement for some or all of the original points. Preservation of perceived recognizability may be less directly achievable. For the map reader, generalization should be considered as a transformation which preserves a delicate balance between measurable accuracy and recognizable character.

> As a form of picture abstraction, the process of map generalization is probably a simple case of a more general problem of pattern analysis, which in turn may be considered to be the basis for the important inductive approach to any knowledge.
>
> Tobler (1966), p.1

The basis of pattern analysis is pattern recognition, or the identification of objects and features. It is based to a degree on information theory (Shannon and Weaver, 1949), which defines information probabilistically as the reduction of alternative outcomes. Availability of more outcomes in a situation implies more available information; for a continuous phenomenon, the maximum amounts of information will be available at those 'events' in which the phenomenon is changed.

The cartographic application is clear: for surfaces, 'events' are measured as slope or shading (color), and the inflection points or most abrupt change in surface shading will be the most informative points on the surface. For lines, which are topologically one-dimensional, informative events are defined by a change in direction of the line. Perceptually speaking, then, visual information in a graphic display is concentrated along contours or sudden breaks in elevation or shading, and along these breaks, information is concentrated at points of angular inflection (Attneave, 1954). It seems logical to assume that to preserve perceptual information on a line, one would insure that the points of angular inflection are preserved during the generalization process.

Two cartographers have tested this, in isolated studies: it is interesting how similar are their verifications, given their different research designs. Marino (1979) presented two groups of people with a series of cartographic lines, one at a time, accompanied by a box of dressmaker's pins. One group had previous experience in compilation and generalization tasks, while the other did not. People were asked to put pins in the lines at points which they felt should be retained in a simplified version of the line, in order to preserve its visual character. No time limit was set on the task.

The experiment was repeated three times; in each successive trial, the same line was presented, but with only half as many pins available. Marino found a significant consensus in the set of points chosen for each line, and further, found no significant differences in response between the two groups. She concluded that cartographic training has no apparent effect in determining which points are critical to perceived line character; and also that there is a hierarchy in the definition of line character, in that points chosen for the trial with fewest pins were often chosen in all three trials. The major shortcoming of her study is that Marino never summarized angular deviations to determine whether her trial lines had distinguishable angular characters, but rather made an apriori assumption that this was the case.

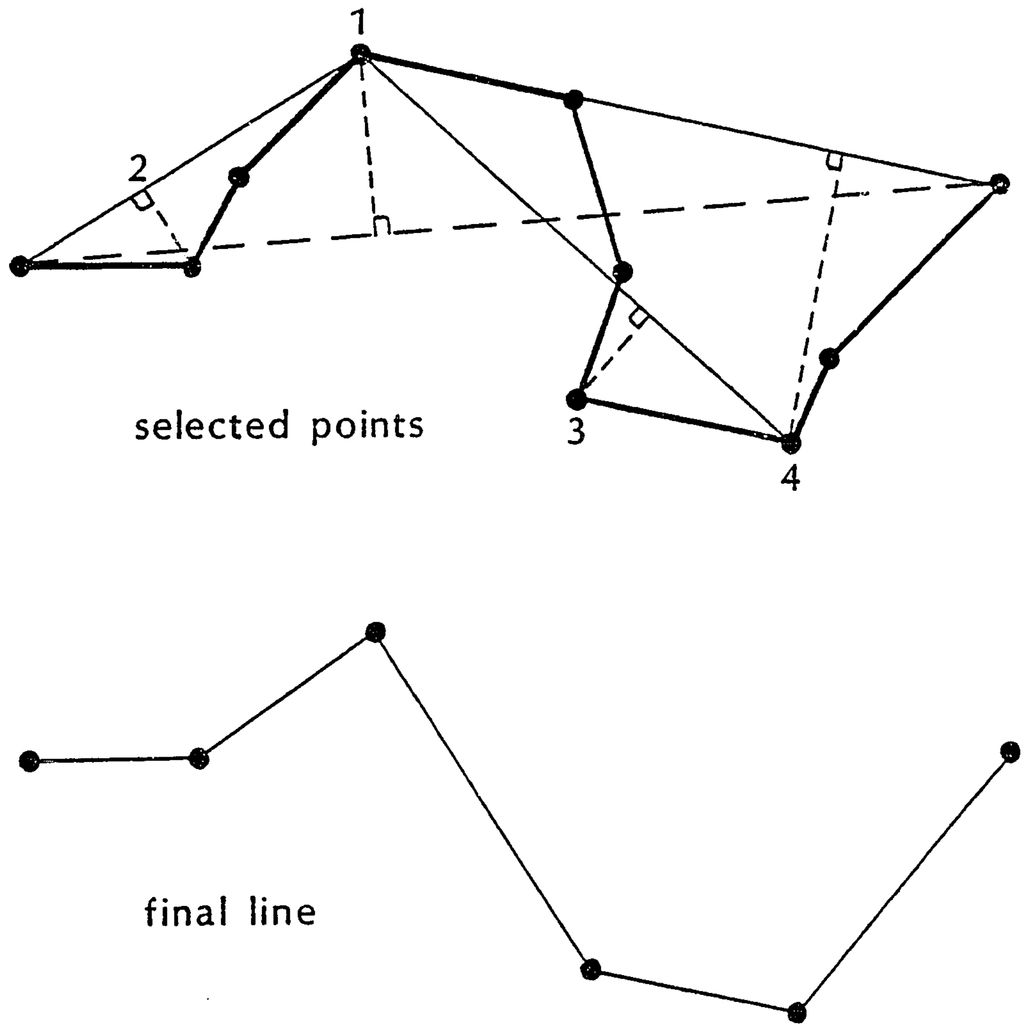
Kelley (1979) was interested in functional as well as scale generalization. Functional generalization does not require a reduction of information expressly for reduction of map scale, but rather involves the emphasis and/or deemphasis of features when generalizing between maps to serve different purposes (as in collecting information for a thematic map from a topographic data base). Kelley follows a fairly strict interpretation of the Shannon-Weaver (1949) information-theoretic model, assuming that line continuance has an associated probability function which is inverse to the severity (degree) of angular change. He tests the hypothesis that information which is important to the map reader will be found near points of maximum angular inflection.

Two points distinguish his experimental design from Marino's. First Kelley embeds his trial lines in a map context, and tests for the constraining effects of background noise, familiarity with the mapped feature, and open versus closed lines. Second, he measures the rate of changing direction along all lines tested, and thus determines where points of maximum angular inflection occur. Analysis is fairly direct, and involves comparing locations chosen by perceptual response to locations computed mathematically. Kelley finds a significant correlation between the two, and a strong consensus in point choice as well; neither of these is affected significantly by closure or by background noise. Kelley's findings on familiarity are weak, and he is not confident of his interpretations.

It would appear that these studies confirm the work of Attneave (1954), mentioned earlier, that information critical to the recognition of a line can be found in the vicinity of points of maximum angular deviation, and further, that most people will choose the same set of points along a line as being critical to its recognition. The conclusion is that a generalization algorithm which consistently selects these critical points will produce a more easily recognizable cartographic line.

The algorithm developed by Douglas and Peucker (1973) provides a good example of this, as it tends to select points of maximum angular deviation, indirectly. Further, it has been tested by generalizing Marino's lines (Jenks, 1982, personal communication), and the set of points which resulted bears a great similarity to the set of points chosen consistently by Marino's test subjects. The mechanics follow from Peucker's ideas on bandwidth filtering, as described in a previous section of this chapter.

Basically, the endpoints of the line to be generalized are connected by a straightline segment (Figure 2.15), and perpendicular distance from this segment is measured to each point. The point which is farthest from the segment is stored. This becomes a new endpoint for a subsegment (actually, two subsegments -- one to each of the previous endpoints), and the process continues until a threshold distance between all points and the straightline segments is reached. Incidentally, each defined segment identifies a specific bandwidth; as discussed earlier, the concepts developed in this algorithm formed the basis for Peucker's bandwidth theory of the cartographic line. This algorithm will be used in Chapter 4 to break out critical points for the set of test lines, in order that the



DOUGLAS-PEUCKER ALGORITHM

FIGURE 2.15

stochastic processes can be anchored to some standards of accuracy.

Summary

One can see that there are widely divergent viewpoints on how to measure a cartographic line, and how to construct its valid cartographic representation. What this chapter has established is that common to all approaches to generalization is the stumbling block caused by lack of a consistent definition of line character. The problem may be similar in difficulty to the problems of measuring length and sinuosity: there is no unique solution. And in reviewing this body of research, an approach does seem to present itself, if only in rough form.

Linear information is composed of two types of components. One is tied to the accuracy of the line. Geographic accuracy involves careful positioning of coordinates: the limits of positional accuracy are determined by the scale at which is measured the line; and points selected for positional accuracy are related to the geographic context at hand. For example, city locations should be carefully positioned on the land-side of a coastline or river bank. Other points should also be carefully positioned along the line, in part to anchor the trend of line direction to as large a point set as possible; equally important, however, the generalized version of the line should retain its visual clues for easy recognition. Perceptual accuracy is related hierarchically to the angular deviations along the line, with maximum importance placed at points of maximum inflection. Thus both kinds of accuracy can be easily applied to anchor the trend of a line in cartographic representation.

The other components of linear information are more statistical in nature, and not directly tied to absolute coordinate position. One component is visually recognizable as the characteristic twists and turns which ascribe a generic type of feature to a cartographic line ("oh, yes, that's a fjord, and this, over here, is a meandering stream"). Attempts to describe this line character have proved largely inadequate, partly because (as with Maling's sinuosity) the defining measures as constructed have been tied to geographic accuracy; the associated difficulties have been discussed at length in this chapter.

A more promising approach comes in light of two notions, first that line character is statistical, and secondly in development of methods which introduce stochastic information to a line. Partial success of these can be seen in construction of fractal lines and surfaces, which bear strong visual resemblance to geographic features. Allometric models may also provide good representations, without the limitations imposed by assumptions of self-similarity (Gould, 1966).

Discussion in the next chapter will relate perceptual aspects of line recognition to mechanical aspects of producing a graphic representation. In Chapter 4, attention will be focused on defining certain parameters which might allow different kinds of line features to be distinguished in a consistent manner. These parameters will be applied to a set of cartographic lines, to determine the extent to which such distinctions can be accomplished. The purpose of the parametric descriptions is twofold: first, to isolate components of line character from components of positional accuracy; and also, to establish procedural guidelines by which line character distinctions may be automated.

CHAPTER 3
COMPONENTS OF LINE INFORMATION

As discussed in Chapter 2, attempts to measure and generalize the information contained in a cartographic line have developed along a range of approaches, from purely descriptive to nearly inferential. Automation of these approaches has so far enjoyed only limited success, for a variety of reasons. At a conceptual level, the difficulties are related to a need to define line character more precisely. There are operational problems as well, with each existing approach, which may be useful at this point to discuss.

The descriptive approach is geomorphic in nature. It was commonly used in generalization prior to the advent of the computer, and has proven more easily accomplished in a manual than in an automated context. A generalized feature is depicted in this approach to capture the so-called 'look' of the landscape feature. Guidelines for producing the early pictographic representations have been provided by Pannakoek (1962), by Steward (1974), and to a lesser extent by Raisz (1948). The visual form of the graphic line was considered by these writers as a reflection of the geoglogic processes which produced the feature. Thus effective depiction required artistic talent and some intuitive interpretation, as well as a firm background knowledge of geomorphology and physiography.

The geomorphic approach has proven difficult to teach, because of problems involved in expressing consistent guidelines for an essentially intuitive process. For this reason, early textbooks in cartography (and in geomorphology, as well) contain page after page of landscape renderings: the purpose of these was most likely to teach by example, by illustration.

The disadvantage of such an approach to generalization becomes obvious when one considers that no text can illustrate an exhaustive typology of landscapes; a generalization task should always be anticipated which cannot be solved by mimicry.

The reason the geomorphic approach has proved difficult to automate has to do with the visual cues which allow identification of the particular process at work. These cues are scale-related, and because different processes can be seen to operate at different scales, the proper inclusion of these cues allows a trained observer to identify the type of line and the scale of the representation as well (Goodchild, 1980). Automation of this descriptive approach would involve comprehensive descriptions of the visual cues required to identify each process, and these for each desired scale. The sheer volume of required information is formidable, not to mention the task of providing consistent definitions for a set of essentially intuitive guidelines which may vary from one cartographer to the next.

With the rise of automation, more mathematical approaches have developed, as discussed in the literature review. Many of these are still commonly applied in line generalizaton, but the same intuitive limitations are in some ways still present. Mathematical approximations by spline functions provide adequate representations, but many splines require large memory and processing capacity to compute derivatives for each line segment generated. Fourier series approximations provide a less costly generalization, provided that coordinates are equidistant along the length of the original line. However, regular sampling is an inappropriate digitizing strategy for naturally-occurring features: the points along a line whose location must be preserved for geographic accuracy (for example, city locators, or points marking stream intersections or the mouth of a harbor or bay) do not often occur at equally spaced intervals along a line. Unless the chosen interval is very small, some of these critical points will not be sampled.

The alternative mathematical approach assumes that geographically critical locations are given. Then tolerance values of distance and direction are used to eliminate coordinates along the line. The effect is to simplify rather than to approximate; and many of these algorithms provide easily recognized generalizations, with varying degrees of accuracy (McMaster, 1983). For a very complex linear feature, such as the outline of the United States, different parts of the line require different tolerance levels for consistent accuracy and recognizability from place to place. To account for the differences in simplifying, for example, the coastline along the Chesapeake Bay as opposed to the adjacent coastline along the east coast of the Carolinas, not only do the specific tolerance values for each type of coastline have to be defined for the computer, but a decision rule must also be provided as to where to break up the line during generalizaton. These decisions are largely intuitive, and may also vary from one cartographer to the next.

Consistent guidelines by which to determine at what point such a complex (perhaps compound is a better term, here) line changes in visual character have for the most part focused on empiric testing for line recognition, or for psychophysical ranking of many lines to produce a continuum of visual complexity (Jenks, 1981). This is a distinctly perceptual approach to generalization, in contrast to the descriptive and the mathematical approaches which led to its development. The common thread which runs through all of these approaches is the focus on preservation of recognizable characteristics. And because this focus has not changed, the same problem is encountered for each approach to automating the generalization process.

In focusing on recognition, researchers have tried time and again to formalize a relationship between visual character and relative coordinate positioning along the line as it is generalized. Formalization by geomorphic function, by mathematical function, and by psychophysical function have all tended to confound rather than to isolate the types of information available in a cartographic line. The accuracy of geographic position and the characteristics by which the geographic identity of the feature is determined form two separable components of cartographic information. Until these components are dealt with one at a time, we will encounter difficulties in automating the process of generalization.

The rest of this chapter will suggest an alternative conceptualization of line generalization which attempts this kind of separation. It will identify a procedural approach to distinguish between types of cartographic lines, for automated construction of graphic representations. The purpose here is to circumvent the logical fallacies of the previous approaches, by distinguishing between line characteristics in terms of graphical rather than perceptual aspects. It should be emphasized that the intention is not to ignore the perceptual aspects of the map-reading process, but rather to focus on computer generalization as a process of construction as well as a process of recognition.

Recognition and Representation

The logical disadvantage of a strictly perceptual approach results from the strong perceptual connotations which the term 'visual character' carries in defining recognizability of a line. Consider that a line which is accurate in terms of its relative coordinate positioning may be represented with such distortion as to become geographically unrecognizable. For example, the longitudinal expansion in Figure 3.1 renders a much more sinusoidal view of the Hudson Bay coastline than is normally encountered (Figure 3.2), yet both maps display a mathematically consistent projection of spherical coordinates onto a plane. Conversely, infinitely many representations of a geographic line may be constructed which are visually recognizable, but cartometrically useless, due to repositioning of coordinates. Sketch maps hand-drawn from memory provide one example of this kind of generalization; the enhancement techniques proposed by Brophy (1973) and by Dutton (1981) provide another.

Further complicating the relationship between recognition and coordinate positioning is the fact that no two people can be expected to have the same impressions by which they characterize what they perceive. In other words, they may rely on differing visual clues to recognize the same graphic representation. In focusing on recognition for generalization, then, one needs to develop a set of finite and consistent definitions of recognizable characteristics for an audience using inconsistent and possibly non-finite sets of perceptual criteria. In order to automate the generalization process, in order to treat different 'kinds' of lines in different kinds of ways, one is confronted with the dilemma of having to define the visual character, or the set of clues by which each

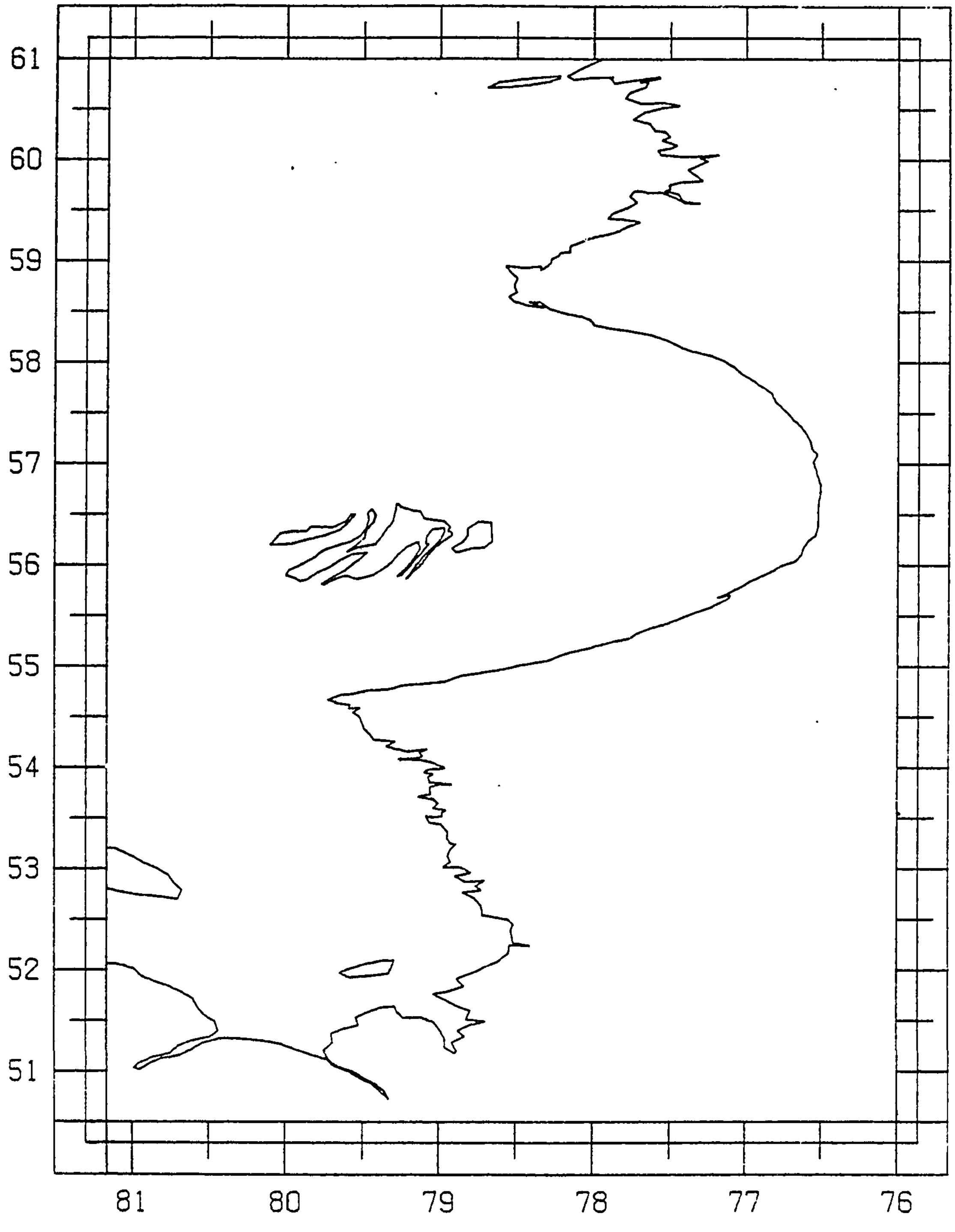


Figure 3.1
A VIEW OF HUDSON BAY

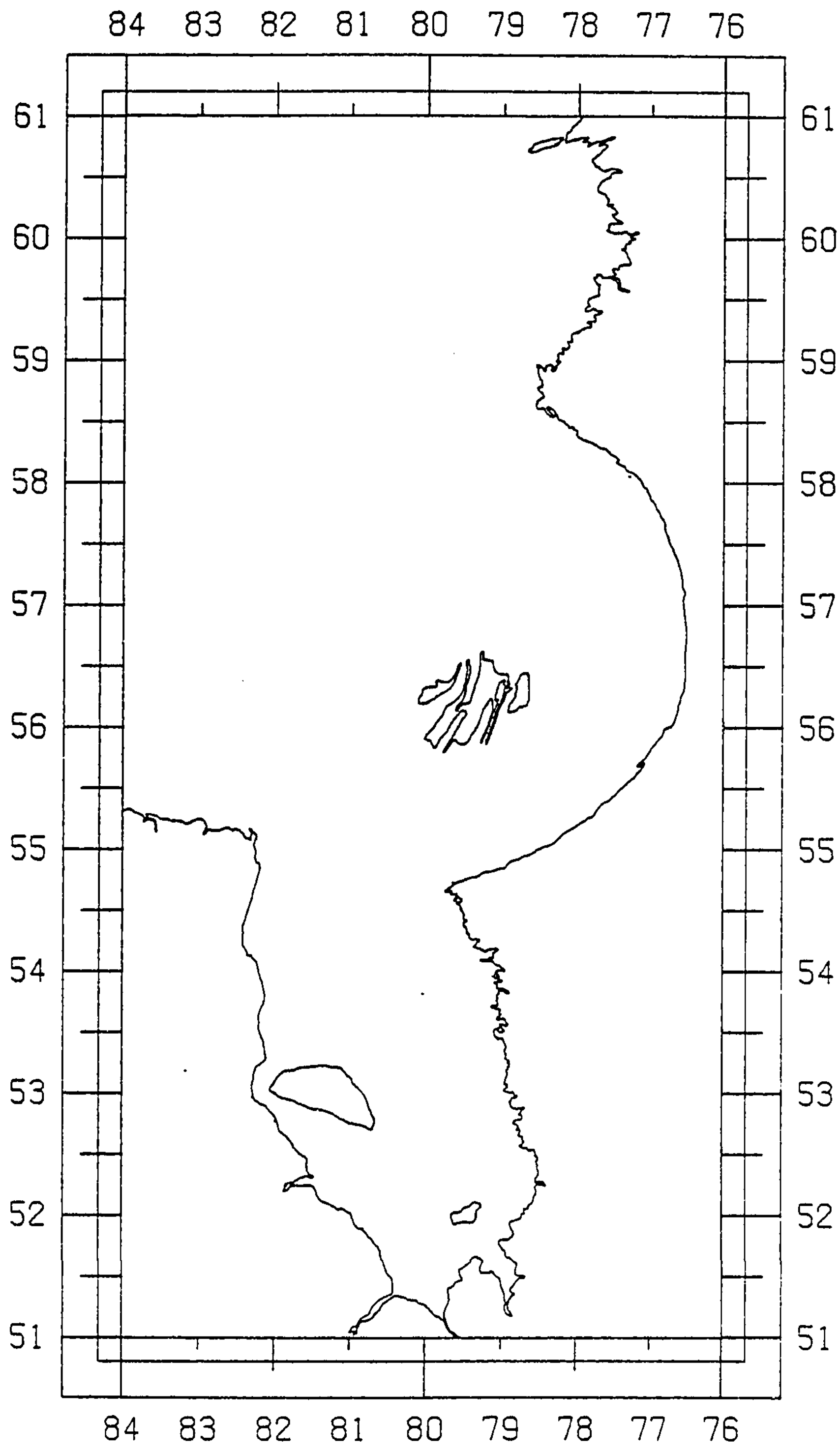


Figure 3.2
ANOTHER VIEW OF HUDSON BAY

kind of line feature is recognized. Plainly, consistent guidelines for the definition of these perceptual clues have not been developed for geographic features, and have not been successfully automated for line generalization.

The resolution of the dilemma is to focus on the representation of the line, rather than on its recognition. For computer representation, this can be parametric, and a sequential process. First, a feature is generalized in graphic representation. Only when the representation is complete does the question of recognizability arise; and this can be empirically tested, using a range of parameter values as a validity check on the procedures used to construct the representation. The two aspects of line representation and line recognition are not independent of each other, but are merely operationalized in separate steps; the computer generalization is still intended to maximize the recognizability of the linear feature in graphic space.

In the most simple terms, the computer doesn't need to 'know' anything about visual character, or recognizability of the geographic feature; it only needs information on how the line should be represented in graphic space, and how representations at various scales can be constructed from a common coordinate set. This is a procedural approach to generalization, in counterpoint to the strictly perceptual approach. It works to build up a representation in the structuralist sense, in effect putting pieces of a line together, as opposed to eliminating information while preserving recognizability.

The procedural approach can be accomplished by defining numeric parameters which approximate the geographic line, in much the same way as a series of terms can be used to closely approximate a mathematical function.

The advantage here is that parametric descriptors can take on a range of values, to accomodate inconsistencies of the perceptual and cognitive attributes of recognizability. Once constructed, the representation can be empirically tested, in order to fine-tune the range of values which each parameter may take.

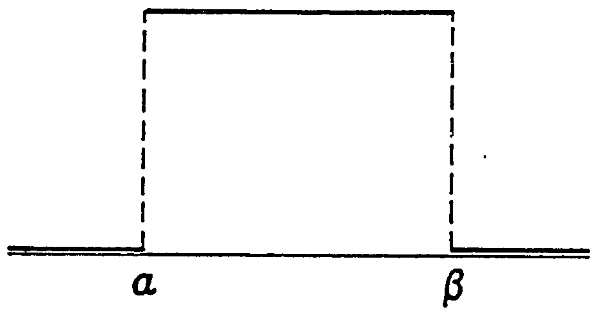
Generalizaton is thus conceptually transformed into a modelling process. Because the parametric value ranges must be constructed inferentially, one can say that the procedural definition of the linear information is probabilistic in nature, or at least that parametric construction can proceed by probabilistic approximation. Therefore the procedural approach provides effective isolation of the two types of cartographic line information: one type is related to the accuracy of the line, and tied in with coordinate selection and position. The probabilistically defined parameters compose the other, and these parameters can be used to approximate details about the kind of geographic feature which the graphic line represents. This second kind of information has been called line character by some authors, or line structure, by others. It seems useful at this point to mark a clear distinction between the two terms.

Character and Structure in a Probabilistic Context

If you detect a difference in listening to two different musical pieces, say, the opening phrases of the Moonlight Sonata and the opening phrases of Beethoven's Ninth Symphony, you may ascribe these differences to certain charcteristics of the music. In describing how the pieces differ, mention might be made of the strong percussive tones of one, and the softer, arpeggiated phrasing of the other. Words such as 'forceful' or 'soothing' might distinguish another friend's impressions. In communicating your ideas, however, you might find that still another friend would describe one as 'empassioned' and the other as more 'modulated' in character. Most interesting of all is the fact that even though you and your friends describe the character of either piece by using different terms and concepts, chances are that each of you can recognize the presence of the other concepts in listening to the music, and understand your friend's use of these terms in describing the character of the music.

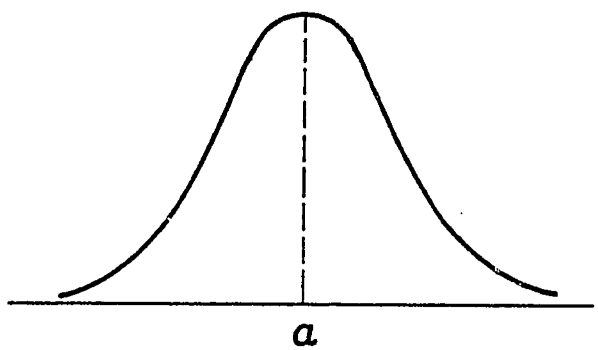
A skeptic might argue that this point is interesting, but trivial. Because of familiarity with Beethoven's music, people will tend to hold common cognitive associations which have nothing whatsoever to do with their perceptual impressions of its character. This criticism would hold true, for any example of musical impressions requested, except one. If you are presented with musical sequences produced from various random distributions, then your impressions of those sounds will not have prior associations, because the music will differ somewhat each time it is generated.

The most common types of random noise are generated from uniform distributions, as white noise; from normal distributions, as Brownian noise; and from harmonic distributions, as '1/f' noise. (Figure 3.3) Each of these distributions produces a distinct accoustic pattern, according to the parameters of its probability density function. No opportunity to provide such acoustic patterns presents itself in the context of a dissertation; but the visual analogy of random noise is random motion, which will provide graphic rather than accoustic impressions of differing



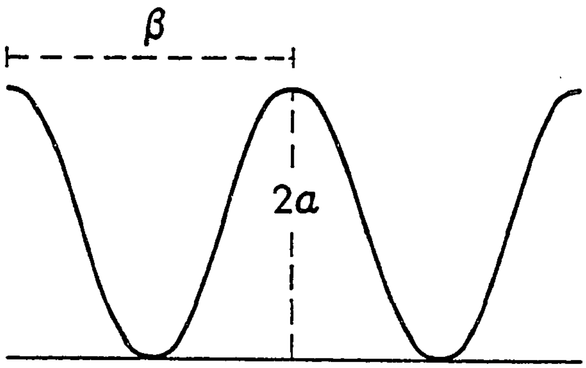
Uniform Distribution

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \alpha < x < \beta \\ 0 & \text{elsewhere} \end{cases}$$



Normal Distribution

$$f(x) = \frac{1}{\beta - \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\alpha}{\beta}\right)^2}$$



Harmonic Distribution

$$f(x) = \alpha + \alpha \sin\left(\frac{2\pi x}{\beta}\right) + \alpha \cos\left(\frac{2\pi x}{\beta}\right)$$

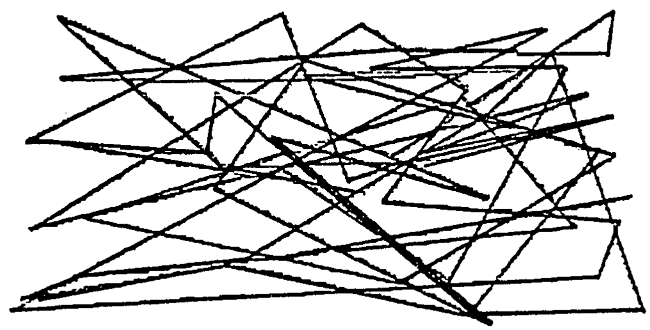
Figure 3.3
EXAMPLES OF PROBABILITY DENSITY FUNCTIONS

character. Plots of random motion will additionally provide a ready comparison between impressions of visual character and parameters comprising its structure.

Figure 3.4 displays plots for the three kinds of random distributions mentioned above. Once again, the visual characteristics of each pattern are obviously distinct, and could be described by expressive terms, and by concepts as varied as were used for Beethoven's music. However (and this is fundamentally important) the computer which generated these plots did not need any informaton concerning impressions of visual character. Its requirements were rather a random number generator and an algorithm, a procedure, by which the random numbers should be constrained.

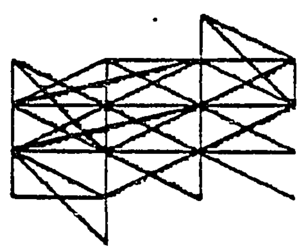
In white motion, each coordinate in the plot has an equal probability of being picked -- the only constraint is a limit for the size of the plot. Brownian motion constrains not only plot size, but individual segment size as well: the probability of a coordinate choice is dependant on the location of the previous chosen point. '1/f' or harmonic motion extends the location constraint one step farther, by restricting coordinate choices to partial or complete multiples of a predefined stepsize in x and in y. The resulting visual regularity can be compared to the harmonic progression of notes included in a musical chord.

These parametric descriptions of the various plots are structural, and refer to the way the segments of the representation fit together. By changing the parameters, one can modify the size, extent, or even shape of the plot: but the overall structure, the way that the segments fit together, will remain the same. The information contained in these parametric descriptions is quite different from the descriptions of visual



White Motion. . . . Uniform Distribution

Brownian Motion . . . . Random Walk



1/F Motion. . . . Harmonic Distribution

Figure 3.4
PLOTS OF RANDOM MOTION

impression which one might use in recognizing a plot, or in describing its visual character to a friend. It is important to realize that while the description of character can be derived from the structural description, this relationship is not a bi-directional mapping: one cannot so easily derive the information required to construct one of these plots given only descriptions of how the plot is recognized.

Given all of this, it may be an oversimplification to conclude that 'structure' as a term relates to representation in the way that 'character' relates to recognition. Think of it rather that structural definitions operationalize procedures by which (in this case) graphic images may be constructed to produce a variety of recognizable visual impressions. And consider character as the set of perceptual impressions received which facilitates recognition of the image as representing a specific object, landscape, or scene.

'Character' provides generic categories of features, for example road or river, fjord or sandy beach. No demands are made on specific geographic location, as in "Route 66", or "The Mississippi River", but rather an assurance that the feature can be recognized within some generic category. In common sense terms, one might consider line character as a set of visual clues which allow distinction between those generic categories on a map (eg., to identify a line as belonging to a coastal region or to an inland river valley). This kind of visual clue helps to orient a map reader to the geographic region encompassed by the map, and it is a primary responsibility of the cartographer to retain this kind of information during generalization.

It may also serve as partial indication of map scale: for instance, there are certain features which one would expect to see on a map of the Norway coast, at varying scales. The long narrow valleys should be apparent as pairs of roughly parallel lines at one scale. At a smaller scale, however, one would expect that some of these valleys would be displayed only by irregular indentations, or not at all. In another example, one would expect that with changing scale that a graphic depiction of the Mississippi River would widen from a single line to a double line, and finally into an areal feature whose smooth but uneven sides depict increasing details of alternating erosion and deposition of the alluvium. The information of 'character' is thus necessary for recognition, but not necessarily sufficient for construction of the generalized feature.

The term 'structure', on the other hand, refers more directly to the rules and procedures required to construct the representation, rather than to any apriori knowledge about erosion and deposition, or about the look of the linear feature which results from these processes. A structural definition of the irregular indentations mentioned above would more likely include parameters for length and width of the indentations at scale, and the frequency with which these features occur.

A part of the structural definition must include guidelines for scale-related modifications of a representation. For example, with decreasing scale, the large river narrows from an areal strip to a single line; however, the parallel indentations of the fjord behave somewhat differently. At first, the two sides become more closely spaced, and at a certain scale both sides disappear altogether from the representation. Structural definitions for fjords and rivers differ in this respect, which a cartographer takes into account in deciding at what scale the graphic metamorphosis takes place. Intuitively speaking, two lines can be said to have different structures if they behave differently at different scales. "Behavior" in this case refers to the values which parameters in the structural definition take on in generalizing the representation at various scales. In differentiating fjord from river in this example, the parameter in question has been the width of spacing between roughly parallel lines bounding the areal strip of river bed or fjorded valley.

In the next section, this parameter will be explained in detail. For now, it is enough to see that structural definitions are concerned with fitting together smaller pieces to form a whole image, in this case on how far apart should two segments be spaced at a given scale. No perceptual connotations are explicitly involved here; there are simply questions of how to build the graphic representation. This is the fundamental difference between line character and line structure: separating one concept from the other provides a procedural approach to automated generalization, by first constructing the representation parametrically and <u>then</u> checking the accuracy of parameter values on the basis of recognizability of the generated line.

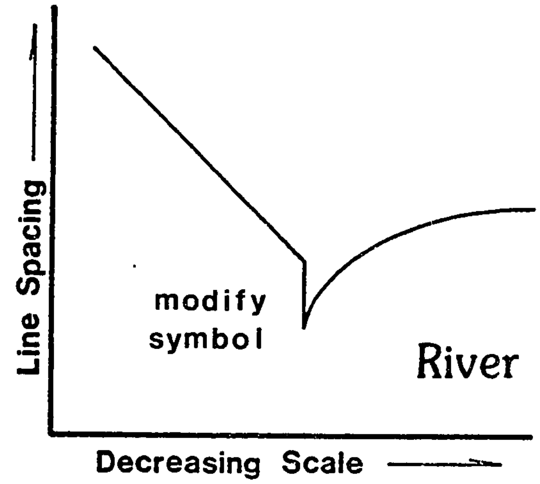
<u>A Parametric Definition of Line Structure</u>

The previous discussion of line character and line structure included a proposal that structural parameters can be selected to distinguish between types or categories of linear features. This can be accomplished by measuring parameters for each line type as it is represented at progressively smaller scales. Lines whose progressions differ could be said to have different structures. The example which was discussed included a paramter for width of spacing for certain double line symbols.

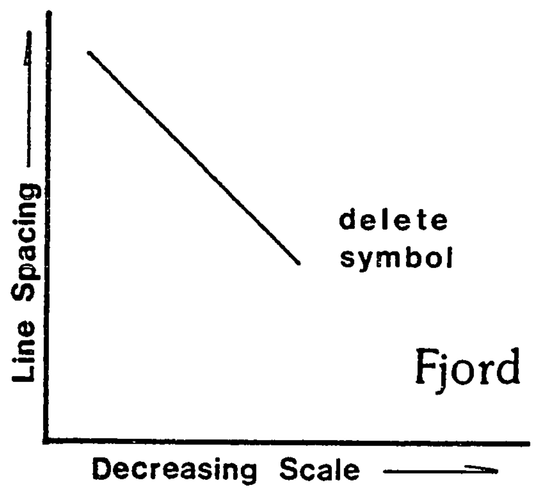
It is possible to plot hypothetical spacings for categories of line features, at various scales, to provide a graphic comparison of the relationship between size in graphic and in geographic space. Three categories of lines will be discussed; each category is commonly represented by means of a double line symbol. Figure 3.5 displays the line spacing plots for these three categories. Each plot will be discussed in turn, to show how the spacing of the graphic symbol reflects geographic size of the feature at various scales, and also to illustrate how parametric descriptors can be used to distinguish between categories of features.

(Figure 3.5A) In representation at large scales, the width of the river in graphic space is proportional to the actual width of the geographic feature which it represents. With scale reduction, the river width on the map should decrease, and this in linear proportion to the scale change. This kind of generalization is similar to the situation of measuring river width on air photographs taken at increasing altitudes. Thus the double line feature can represent scaled width of the river, as long as the size of the river permits resolution of each side of the river bank at scale. (Realize that the limits of resolution will occur at a very different scale for the mouth of the Columbia River, for example, than for its headwaters. Thus it is the shape of the line spacing plot which is important, rather than an absolute delineation of map scales across the axis.)

A.



B.



C.

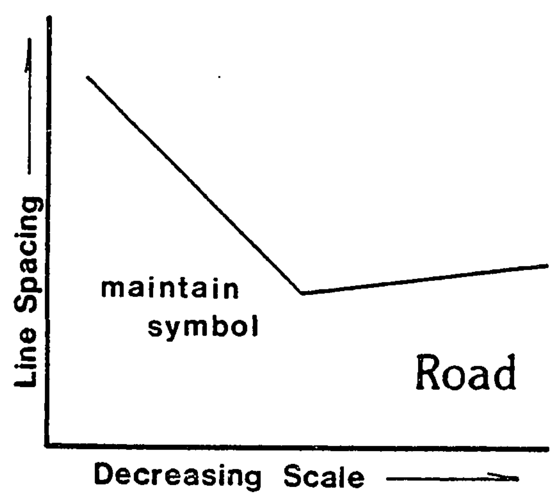


Figure 3.5
PARAMETRIC DEFINITIONS OF LINE STRUCTURE

Below this limit of resolution, the river should most appropriately be symbolized by a single line. The relationship between plotted width and actual width will change suddenly with the single line representation, and this is shown by the vertical jog in the plot. The single line may be narrower than the feature which it represents, at this scale. With further reduction, however, and especially if line width is kept constant, the relationship between graphic and geographic width may rebound somewhat, as shown in the plot, and then tail off towards some equilibrium value at which the cartographer decides to remove the feature altogether from the map. This equilibrium may or may not be an accurate depiction of river width at scale. More likely, it will be a function of the importance of the river to the particular map purpose.

(Figure 3.5B) It can be seen that the shape of the plot for spacing of line symbols of a fjord differs somewhat from the plot of a river. They appear similar at large scales, and there is the same linear relationship between graphic and geographic size for the representation of the fjord. At the limits of resolution, however, both lines of the fjord symbol will be removed: by cartographic convention, fjord structure is not defined by a single line. This convention is reflected in the abrupt termination of the structure plot.

(Figure 3.5C) The third plot stands in juxtaposition to the first two, because roads are man-made rather than naturaly-occurring features. For fiords and rivers, the double symbol is irregular at larger scales: the pair of lines are only roughly parallel, because they represent irregular terrain. For parallel lines symbolizing a road or highway, of course, this is not true. The sides of the line symbol in this case will be quite regular, to reflect this structure on the map. At certain scales, a road might be indicated as a pair of lines whose spacing drops linearly with scale reduction (as with the river and the fjord symbols). At the limits of resolution, the representation may be retained as a double line, as is the representation of interstate highways on a road map. With further decrease in scale, the symbol is not deleted, nor is its width reduced. Line spacing will eventually become much larger than the geographic width of the road, and the road on the map will become "larger than life".

An important aspect of all these plots is that each one contains an identifiable jog, or elbow, at which the direction, shape, and/or definability of the relationship changes quite suddenly. In each case, the change occurs at the limits of resolvable detail for the particular feature in question, and the graphic depiction for each kind of feature differs accordingly. For rivers, the two-dimensional areal feature is compressed into a one-dimensional representation. The fjord, on the other hand, disappears altogether. The road is not changed in representation, but the graphic symbol loses all proportionality with the geographic feature which it depicts.

For a topographic mapping situation, one can identify quite specifically the scale at which the elbow in the plot occurs. The limits of resolution are quite strictly defined as half the Map Accuracy Standard, or 0.01" (1/100 inch) at scale. This means that the elbow should occur at the scale for which river width, road width, or fjord width reached .01" on the map. For a 1:62,500 series quadrangle, fjorded valleys smaller than about

50 feet across will not be represented. At 1:125,000, the limit of resolution should be twice this. The resulting effect on rivers will be the symbolizaton of all channels less than 100 feet wide by a single line. And in looking at topographic maps at these various scales, it should become quickly apparent that these standards are consistently followed.

The plots shown in Figure 3.5 describe relationships between three hypothetical features and their graphic representations at progressive scales, in terms of a single parameter called line spacing. The cartographic application demonstrates how such a parametric definition may be used to predict the structure of graphic representations for the category called "river" on topographic maps. An analogous situation to this type of prediction may be found in satellite image interpretation.

One can predict the composition of land cover in an area as being composed of grass or asphalt, by measuring spectral reflectances for each pixel in an image of that area. The reflectance values for a cover material are not measured at a single intensity level, but for a range of intensities, or bandwidths. The pattern of reflectance values across all bandwidths is graphically depicted as a spectral signature (Figure 3.6). As reflectance values serve as parameters for land cover categories, so line spacing values serve in the present discussion for categories of linear features. The plots could most appropriately be called structure signatures, each for the type of structural relationship (for example, line spacing) which it represents.

Of course, a single structure signature may not provide sufficient distinction between categories of cartographic representations, in practice; other relevant parameters should also be included. Some

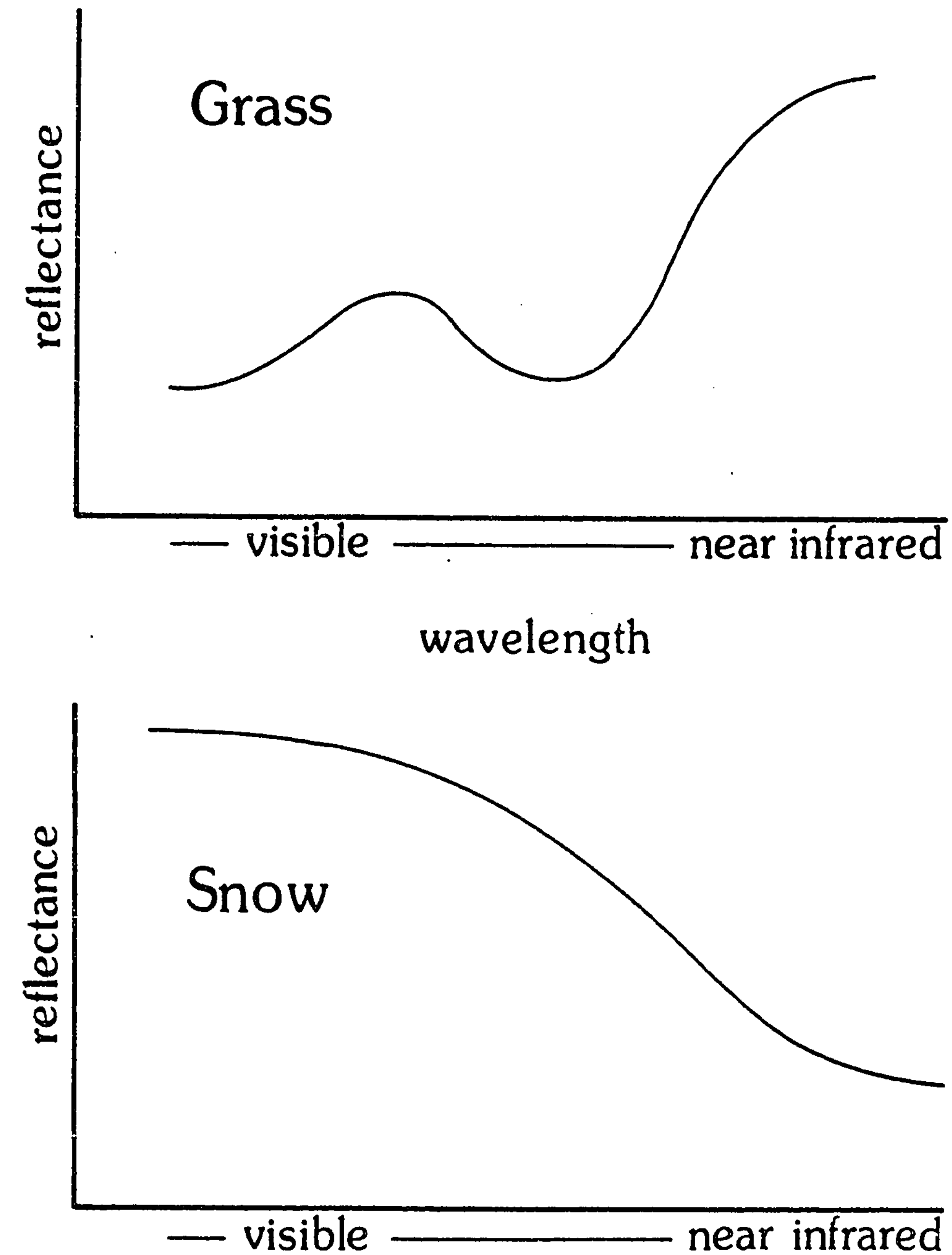


Figure 3.6
SPECTRAL SIGNATURE EXAMPLES

consideration should be given to measures of error variance, for example, or fractal dimension, to identify various sinuosities. Other structure parameters might include a bifurcation probability, to compare the branching networks of linear features. This probability could also be used to parameterize the frequency with which indentations or irregularities deviate from the trend of a cartogaphic line: other measures of monotonicity or periodicity might also be derived.

The important point is that some parameters will best distinguish certain kinds of features, but not others. In other words, no single structure signature, or even a single set of signatures, can provide sufficient distinctions between representations of all geographic lines, at all scales of representation. Realize also that parallel line spacing, bifurcation frequency, and fractal dimension are examples of structural parameters which are of a different order than components such as error variance and monotonicity. Topologically speaking, the first three parameters measure relationships between line segments, while the second pair of parameters measures relations between specific coordinates. As a result, the structural definitions at these two levels should be constructed somewhat differently. Signatures can be derived for both kinds of parameters, however, and it will be shown in this dissertation that the parameters in a signature take on a distinctive range of values for different categories of features.

It makes sense to begin with the lower order of dimensionality (points, as opposed to lines), with parameters such as monotonicity and error variance, and then use these parameters to build up a higher order structural definition. The next chapter describes the procedural definition of lower order structures for four different line features, sampled from the World Data Bank II, in three steps. First, the coordinates required for geographic and for perceptual accuracy will be isolated; this subset of points will be organized in a data structure. A series of parameters will be measured, to distinguish structural components for each line at varying levels of resolution. Finally, the parameters will be used to build a structure signature of the cartographic line feature from the components in the data structure.

It is proposed that separating one component of cartographic information from another, namely separating geographic accuracy from graphic structure, will facilitate a definition of cartographic structure. The development of structure signatures provides a systematic technique for graphic display of these definitions. The procedural approach is intended to provide an empiric means of testing the success of the parametric definitions in distinguishing between different types of line representations. The parametric values may also be useful in generalization applications, as a guide by which to set and to modify tolerance criteria.

CHAPTER 4
A PROCEDURAL DEFINITION OF LINE STRUCTURE

It has been argued throughout this thesis that overreliance on the perceptual aspects of cartographic line generalization has inhibited its automation. Existing algorithms provide parametric techniques for approximating a linear feature, by simplification, selection, and smoothing, and more recently by enhancement techniques. But the fundamental problems remain: namely, how to set parameter values, and how to develop guidelines for treating different types of cartographic lines in different ways. These are hurdles which a perceptual approach to automation cannot yet surmount, because there is little consensus in perceptual definitions of line type.

The purpose of this chapter is to discuss an empiric procedure by which types of cartographic lines may be distinguished. As discussed in Chapter 3, the term "type of line" implies a generic category of linear feature, for example fjord or sandy beach. Also in Chapter 3, the concept of line character was differentiated from that of line structure, to insure a focus on the structural characteristics of a graphic representation rather than on perceptual aspects of line recognition.

This is not to say that the perceptual aspects of generalization can be ignored. To the contrary, a sound algorithm will retain recognizability as well as line accuracy. However, information about recognizability is not sufficient, in and of itself, as a starting point for creating a sound algorithm: it is merely one step in the process. This is an important distinction, which may affect not only the values chosen for the algorithm's tolerance parameters, but also the kinds of parameters which are used.

For example, although a particular coastline may be recognizable as a fjord as it is generalized for different map scales, the requirements for representing its detail, its 'fjord-ness', may change from one scale to the next. Thus to consider the structural characteristics of a line is to consider the way that details fit together in graphic representation, and the way in which these graphic relationships change with changes in scale. This is the basis of the procedural description to be applied in this chapter, to study graphic relationships between coordinates, and then to consider the higher order relationships between line segments.

It should be understood that the power of any given description technique should lie in the facility with which it may distinguish between different categories of line features. One hypothesis of this thesis is that the pattern of these changing relationships will be distinct from one category of cartographic line to another. The purpose of this chapter is to research this hypothesis, using four different cartographic lines. The procedural description as developed will incorporate the previously discussed components of geographic and perceptual accuracy, as well as the statistical information which has been proposed to measure cartographic line structure. Accordingly, the parameters developed in this chapter will be analyzed for their discriminating ability, and developed as structure signatures for each of the four cartographic lines.

The Sample Cartographic Lines

Four line features were sampled from the World Data Bank (WDB II), a CIA-produced cartographic data base containing roughly six million coordinates in five separate files. All points used for this study were sampled from the North America file, at an initial resolution of approximately one point per mile. Figure 4.1 displays the line features at a scale of 1:3,000,000.

The scope of each linear feature is limited by the display window provided by the high resolution graphics memory of an Apple II microcomputer (screen size is 280 by 190 pixels). All lines were transferred to the Apple's UCSD Pascal operating system, and all line measurements were computed from these microcomputer files. Table 4.1 lists the geographic locations for each line feature, along with some nominal indices for each line file. Figure 4.2 displays inset locations for each line, to better illustrate its geographic location.

TABLE 4.1
GEOGRAPHIC INFORMATION FOR THE SAMPLE LINES

File Name	Lat/Long	Range (stat.mi)	Areal Extent (sq. mi)	Number of Points	Mean Point Density
FJORD	50°00' - 51°11'	81.56	12,333	500	4.50
(British Columbia)	124°10' - 127°36'	151.22			
HUDBY*	59°30' - 61°39'	148.84	12,386	306	2.75
(Hudson Bay)	76°16' - 78°40'	83.22			
SIOUX	42°40' - 43°58'	89.74	13,154	331	2.89
(Sioux Falls, S.D.)	97°02' - 99°54'	146.59			
TEXAS	27°32' - 28°50'	89.51	11,541	531	4.94
(Galveston, Texas)	95°35' - 97°41'	128.94			

NOTE: distance computations according to Appendix D of Robinson, Sale, and Morrison (1973)
NOTE: Mean Point Density = number of points divided by the square root of the areal extent

* File HUDBY has been rotated approximately 31° counterclockwise, to fit within the window provided for by the Apple II high resolution graphics memory.

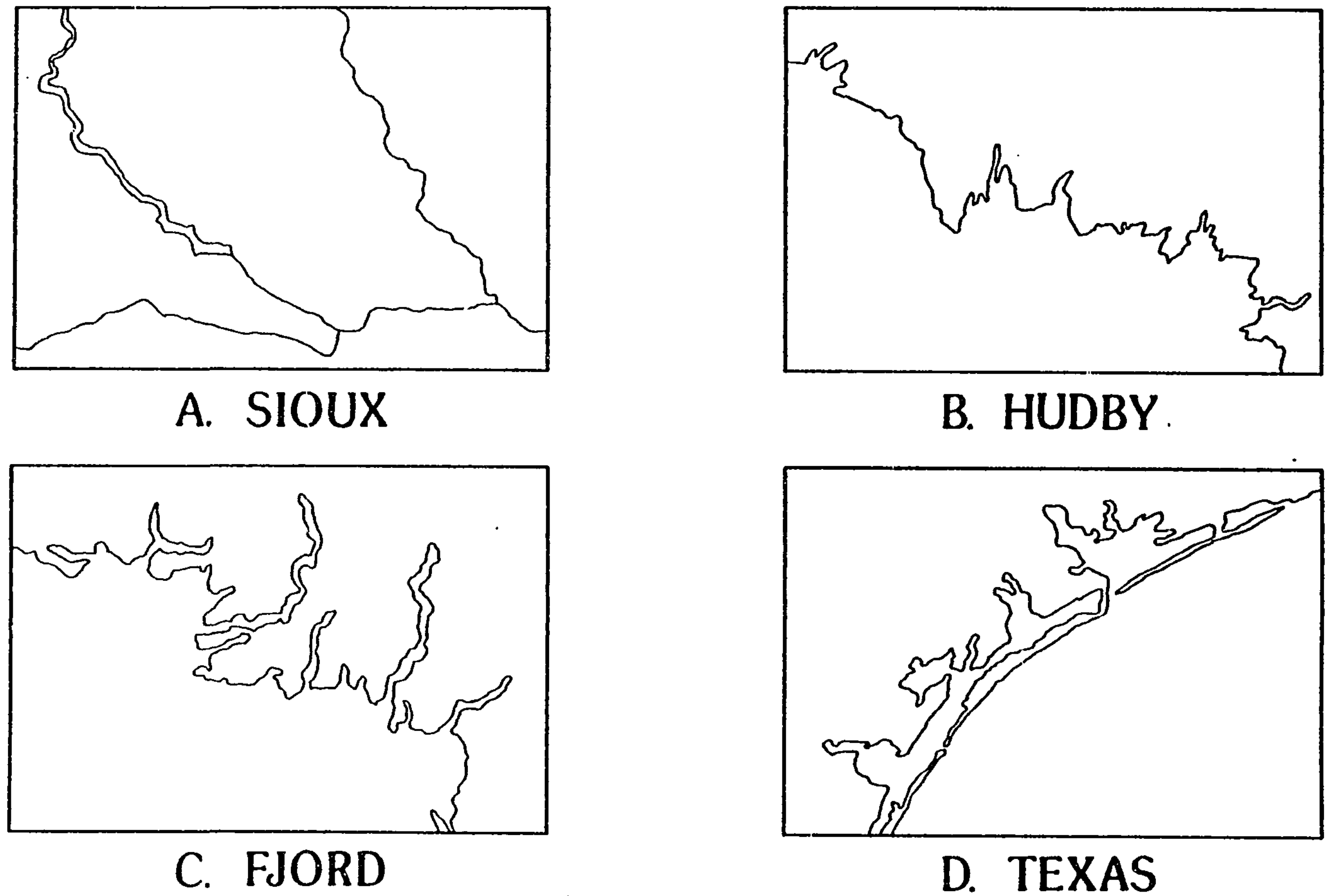


Figure 4.1
SAMPLE CARTOGRAPHIC LINES

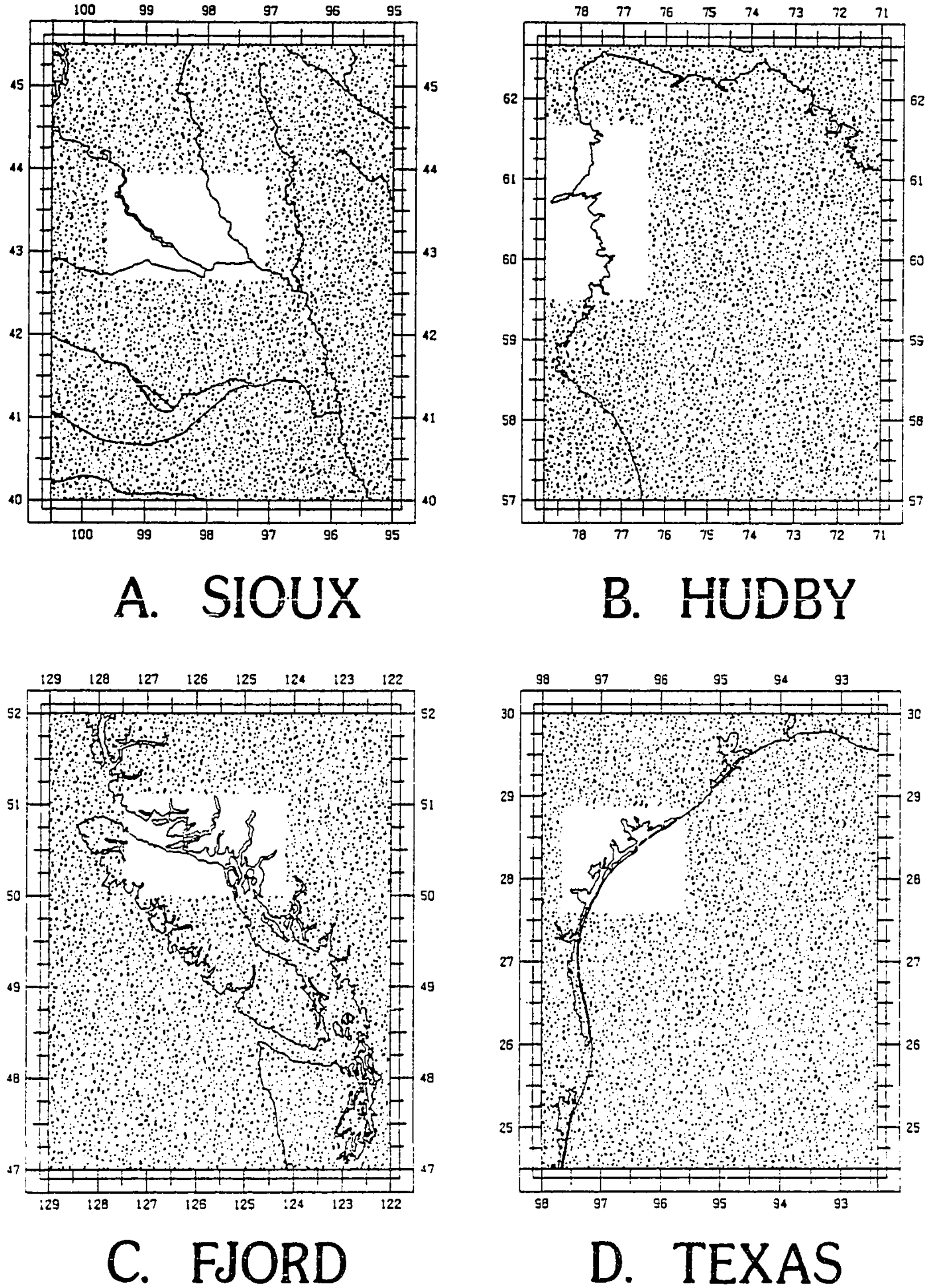


Figure 4.2
INSET LOCATIONS FOR SAMPLE LINES

Before deveoloping the procedural descriptions for these four sample line files, it may be useful to demonstrate several other techniques for line structure description which are commonly applied in a cartographic context.

Three Alternatives for Structural Description

Each of these four lines can be described in a variety of ways. They were chosen because they differ in terms of at least two kinds of commonly applied descriptive techniques, namely geomorphic and mathematical descriptions; and these lines should offer a range of procedural descriptions as well. To explain, one could consider each coastline as a plot of some (very complex) mathematical function, most likely a series of some sort. The function which describes the Hudson Bay (hereafter, the lines will be referred to by their file designations -- the Hudson Bay line will be referred to as HUDBY) coast would need to be monotonic overall, and to include terms which introduce periodic oscillations which occur with increasing frequency at increasing resolutions of display. A Fourier harmonic function might provide a very good approximation for this particular coastline.

The dendritic river included in SIOUX is also very monotonic, or at least each of its segments is monotonic; and its mathematical description should express this characteristic structure. But the most notable aspect of this line feature is the bifurcation of one river channel into several. Notice that none of the periodic oscillations of HUDBY is apparent in this line feature. Accordingly, the kinds of terms, perhaps even the number of terms, and thus the series function as a whole, should prove different from one line feature to the next. These two examples should give the reader a feeling for the application of a purely mathematical description of structure. Here, the line is an isolate, an abstract object in numeric space. Other descriptive techniques can be applied to the sample lines, however, which attend to the surrounding environment as well as to the line.

The cartographic lines might also be considered in terms of the geomorphic process by which they were formed. For example, the drowned river valleys apparent in FJORD reflect the submergence of a coastal area subsequent to the glacial erosion of its river beds. A description of this type would explain quite well the processes at work in forming this particular scene. HUDBY similarly reflects a glacial process, specifically the emergence of a coastal area by isostatic rebound, a tectonic compensation for recession of glacial icepack from the region.

The TEXAS scene is not glacial, but is rather a depositional feature: sand and silt deposited at the mouths of rivers have been moved by coastal wave action, and transformed into longitudinal spits and barrier bars. The SIOUX river basin displays still another geomorphic process. This scene displays the branching of a drainage system over a consistent bedrock material, accompanied by deposition of alluvial materials along the path of the meandering river bed. The branching pattern should become dendritic when the river flows over harder bedrock materials, as here in SIOUX; it should form alternatively a trellis pattern when passing over softer bedrock, as is the case for example along the floodplain of the Connecticut River (Strahler and Strahler, 1978).

Several comments about geomorphic line descriptions are appropriate at this point. First of all, these descriptions reflect a system at work, rather than (as in the case of mathematical descriptions) an isolated set of linear objects. The terrain which lies adjacent to the coastline or river will determine in part the local shape of the graphic line, and a geomorphic description will take this into account, which is an obvious advantage of this sort of description.

On the other hand, it should be apparent that the various geomorphic processes may work to produce very similar looking features, for example, the reservoir in SIOUX, the barrier spits in TEXAS, and the river valley shapes in FJORD. Because the processes which formed these features are so different (respectively, erosion, deposition, and submergence), the features' geomorphic descriptions are different. As described, then, none of the visual similarities are made apparent. This is plainly an aberration, in the context of automated generalization: the computer-rendering for all of these features might proceed more efficiently from similar or equivalent sets of construction rules. Effectively, the geomorphic description alludes to the feature, and not necessarily to its cartographic representation.

When generalizing map features, a cartographer works not in geomorphic space, nor in purely mathematical space, but in a kind of graphic space, which partly overlaps the two. If the measurements and descriptions of the features to be generalized are made within this graphic framework, then it may be possible to avoid such conceptual aberrations, in computer generalization and computer generation alike. This is the purpose of the procedural description, to provide rules by which to reconstruct line characteristics by mimicking those visual cues we expect to find in a graphic representation at a given scale. A procedural description is also aimed at generating rules which a computer can use to distinguish between types of graphic structures. Once again, the four sample cartographic lines may serve as examples of this third type of description.

The visual appearance of the FJORD line is characterized by pairs of long irregular lines, which are roughly parallel to the trend line from which they deviate. Spacing of these pairs of lines is relative to the scale of the representation. It is interesting to note that this description of line pairs would also describe the visual character of a representation of river beds along a drainage network at a fairly large scale, or of a man-made resevoir above a river dam at a smaller scale. The dammed lake in the SIOUX provides an example of this; but note that the similarities are somewhat scale dependent.

The geologic processes which have formed the terrain around the river beds in SIOUX may be different from the geologic formation of the coasts in FJORD; the features are quite different geomorphically. However, the cartographic representation of the features is quite similar; at certain scales, then, their graphic structures may converge. To define these structures in strict terms of geomorphic process may by useful for some purposes, but misleading or even unnecessary in the cartographic context. Notice, however, that some degree of overlap does exist: the procedural description includes several references to purely geometric relations (parallel and perpendicular, eg.) -- and also to the geomorphic features which are being described. The TEXAS file will provide further evidence of this overlap.

The barrier beach in TEXAS is identifiable by the short river mouth bays which occur at regular intervals along a slightly scalloped trend line, and by the smooth and narrow longitudinal bars which repeat the scalloped pattern of this trend. The indented bays are wider at a given scale than are the indentations of the FJORD coast, and tend to bifurcate into multiple channels more often than do the fjords. Nevertheless, the barrier beaches are visually similar to fjords, at certain scales; but perhaps this is due to the parallel aspects of both indentations, and to the nearly perpendicular angles at which the indentations occur.

The dendritic river network is most prominently characterized by linear bifurcations, and by the occurrence of dammed lakes at infrequent intervals. The coast of the HUDBY line displays no such bifurcation, at any of the sample scales; but it does have a somewhat sinusoidal appearance. At some display resolutions, HUDBY looks very much like the lines produced on an oscilloscope, with the oscillations increasing in frequency at finer levels of detail (refer back to Figure 3.1).

Of course, in the context of a verbal description, from one person to another, a procedural approach may appear trivial, or even ridiculous. But the purpose of the approach is not fulfilled by these simple verbal descriptions; they serve as introduction for the reader to the quantifications and analysis which will follow. The purpose of the procedural description is to provide a means for computer-generated descriptions of line features, and for computer-generated distinctions between them. Ambiguities will be counterproductive in an automated procedure, and thus the various components of accuracy and recognizability must be separated from the structural components during the description.

The terms used will not be verbal, but will be numeric parameters, which are as a rule more specific and more easily automated.

Finally, the procedural definition must remain flexible to conditions in which cartographic structures are described and/or represented: a given category of feature (say, a fjord) can be graphically described with varying degrees of accuracy, at varying scales, and levels of detail, depending on the purpose of the graphic display. For example, is the fjord to be generated for an arcade game, displayed in a flight training simulator, or stored in memory for a navigation check in the path routing algorithm for a cruise missile launch? Accuracy, recognizability, and structural details play very different roles in each of these types of computer generalization; and while the same computer line file may not be equally useful in all three situations, the procedures by which the three line files are generated and generalized should be quite similar.

The rest of this chapter will discuss parametric descriptions of graphic structure for the four cartographic lines, by procedural means. The procedures will be developed in two stages: first to be derived will be parameters based on relationships between coordinates in the line, that is, parameters of low order structure. One example of this kind of parameter is the error variance of a line. Low order parameters may be used to build structure signatures, and to analyze the extent to which the four sample lines can be distinguished from each other.

The second stage of the procedural description builds on the first, by augmenting the low order measurements with parameters which are based on relationships between segments of the line. One example of this kind of parameter is the spacing between roughly parallel lines, as described in the previous chapter; another is the fractal dimension, or the rate of increasing line length as more and more segments are resolved. Whereas the low order parameters can be used to distinguish between cartographic lines, the higher order structure parameters can be applied in line identification tasks.

Taken together, the parameters may be used procedurally to describe and differentiate between categories of cartographic lines; in this way, the procedural approach may be used to automate line discrimination tasks. Once able to distinguish categories of features, a computer algorithm may be designed which applies procedural descriptions to identification and representation of cartographic features similar to those already described.

Justification for the Procedural Description

Measuring those characteristics of a cartographic line which help to distinguish between generic categories of lines is the first stage in this study of structure. Here, the focus will be on the coordinate relationships, as opposed to focusing on the higher order line segment geometry. Attention will be paid to how the measurements differ between the lines at any single display scale, and also how the measures for each line change for finer units of measurement.

This is an approach similar to the work of Richardson (1961), who sought to determine accurate coastline lengths by repeating his measurements many times with smaller and smaller units of measure. Mandelbrot (1967, 1977, 1982) has also used this kind of piecewise concept to study the rates at which coastline lengths increase, and to develop his concepts related to fractal dimension. Line lengths will be measured in this study as well, in an attempt to replicate Richardson's results. But other measurements will also be made, and repeated, for smaller and smaller pieces of the line. These measurements will be based on Peucker's (1975) theory of the cartographic line, and also based on the concept of epsilon (ϵ) length developed by Perkal (1966) (see also Chrisman, 1981).

Peucker (now Poiker) considers line segments as having an areal extent which tends to decrease with increasing scale of display. It is probabilistic, in the sense that the 'true' line is more closely approximated in representation at larger scales. Thus the representation of the line becomes 'thinner' as one is able to resolve more and more of its detail. The theory is well applied to line generalization: the thinnest possible line will be displayed at the maximum possible resolution, i.e., when every original coordinate is included in the graphic representation. The importance of the theory is that only rarely does one need to display all of the original coordinates. In fact, many cartographic tasks are facilitated by operating on a subset of these points. Most notable as examples here are the common sorting and searching tasks, such as line intersection procedures, and point-in-polygon searches.

The subset of points is chosen hierarchically, to define for each line segment an overall direction, a length, and a width; Poiker defines these geometricaly as a minimum bounding rectangle surrounding the line segment. As the line is broken into smaller subsegments, the bounding rectangles (which Poiker calls 'bands') will decrease in size. The width of the rectangle is considered akin to a bandwidth. Within a particular bandwidth, then, certain frequencies of line detail can be resolved.

Poiker's theory is similar to, but not based on, the research of Julian Perkal (1966), which was discussed in Chapter 2. As a brief review, Perkal suggested that any graphic representation of a line is at best a probabilistic approximation based on the resolution of measurement. In other words, no matter how precise is the measurement, certain details of the original line feature will be lost (unresolved). As a solution, he devised epsilon (ϵ) envelopes around each line -- the envelopes are of finite areal extent, and are intended to subsume the line and its (unresolved) detail. Perkal approximates the line length by measuring instead the length of the envelope, arguing that this value will approach the true length as the width of the envelope approaches zero (at maximum resolution).

Both Poiker and Perkal have encompassed an irregular line feature with a more regular polygon feature, and as a result have made their measurement tasks much simpler. Perkal was interested in measuring the length of a line, and Poiker concerned with searching, sorting and topologic relations. The verity of their measurements is predictable, as a function of the scale at which the measurement takes place. It is a logical extension from these two studies to the problem at hand, to approximate the structual characteristics of cartographic lines by measuring the structure of a simpler bounding figure; and further, to study changes in structure by changing the size of the bounding figure.

Choosing the Measurement Parameters

The bands developed by Poiker offer a simpler geometry than the more irregular sausage shapes of Perkal's epsilon envelopes, and these bands will be applied to the four cartographic lines. Five parametric measures will be developed: three of these refer to the structure of the band itself, and two others to the details of the line segment contained within. These measures are for low order structure, and the line will be treated as a sequence of related coordinates. Before describing the measures, it may be helpful to consider how these measurements will be organized in order to study and compare line structures.

Dana Ballard (1981) has developed a data structure by which to summarize line segment information. The data structure is used to facilitate graphic representations of irregular curves, and is called a Strip Tree. The structure is a binary tree: at each node is contained a record which includes fields for two (X,Y) coordinate pairs and two width measurements, as well as pointers to other record nodes in the binary tree. The coordinate pairs define the endpoints of a line segment, which Ballard terms a 'strip' (hence Strip Tree); this strip is conceptually equivalent to one of Poiker's bands, as the original curve can be closely represented by straight line strip approximation, at varying levels of resolution.

Ballard's contribution is to provide the mechanics for the bandwidth encoding theory. Figure 4.3 shows a Ballard strip, with labels to its associated data structure. In the figure, the last two fields of the record contain pointers to subsequent records in the data structure. Two pointers are included, to point at each 'half' of the subdivided strip. In Figure 4.4, an imaginary curve has been broken into subsegments and organized graphically as a full strip tree, to show how successive levels in the tree contain finer resolution representations of the curve. Because more of the

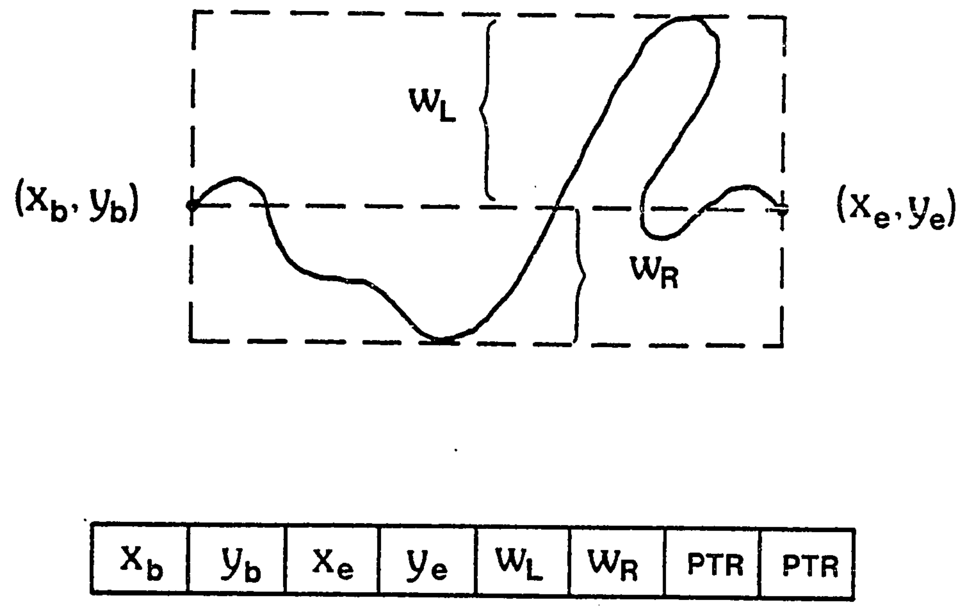


Figure 4.3
A STRIP TREE DATA STRUCTURE
(after Ballard, 1981)

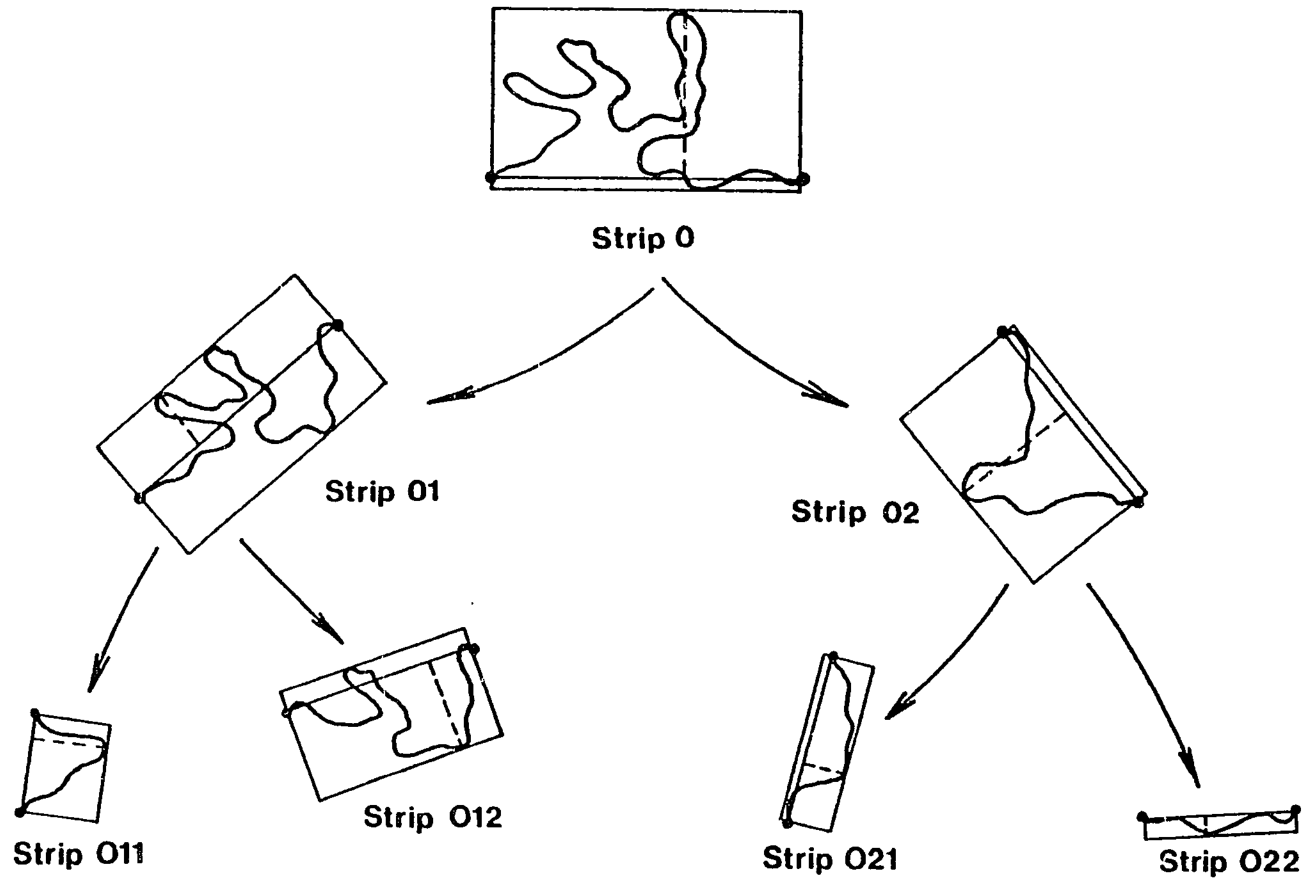
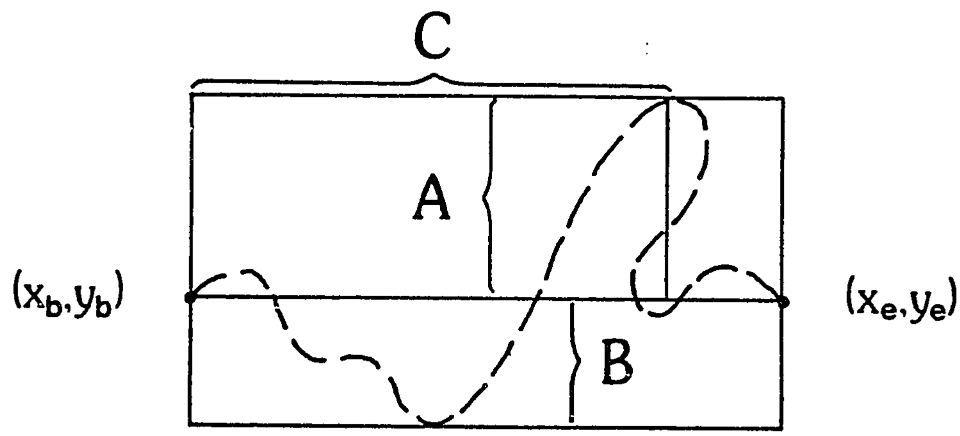


Figure 4.4
GRAPHIC REPRESENTATION
OF A BALLARD STRIP TREE

original points in the curve are included as strip endpoints at each successive level, representation will become more and more accurate at these lower levels.

A modified version of the strip tree will be used to organize the line measures in this study. The modification will enlarge each record, to include a larger number of fields in which to store the five structural parameters measured for each of the line segment bands. The parameters to be stored are illustrated in Figure 4.5, and have been designed to include not only the geometry of the band, but its relationship to the details of the line segment contained within.

Band geometry is defined by four parameters: first is the length, or the Euclidean distance between endpoints. This straight line distance is an anchor for other parameters, which will be expressed as proportions of this value; and the straight line connecting the endpoints will be referred to as the anchor line throughout the rest of the analysis. The second two parameters are closely related: the width of the band overall, and the signed value for the larger half of this width. This second parameter is computed as the maximum deviation of any point in the original segment from the anchor line. The sign of this value (negative or positive) determines on which side of the line the maximum deviation falls. Notice that if the maximum deviation is exactly half the magnitude of the overall width, the band can be seen to be symmetric about the anchor line, and further, consider that a symmetric band can be subdivided into non-symmetric bands. This is an example of the way in which band geometry can be used to study cartographic line structure, and of the manner in which these parameters should be interpreted.



Band Length	$L = \sqrt{(x_b - x_e)^2 + (y_b - y_e)^2}$
Band Width	$W = (A + B) / L$
Maximum Deviation	$M = A / L$
Segmentation	$S = C / L$

Figure 4.5
PARAMETERS OF BAND GEOMETRY

A third parameter included in the strip tree data structure measures the length from the beginning of the anchor line to that point along the anchor line from which the maximum deviation is measured. This measure will be referred to as the segmentation parameter; it is also standardized to the length of the band. In a way, this parameter measures longitudinal symmetry of the band, in similar fashion to the cross-sectional symmetry implied by values of band width and maximum deviation from the anchor line.

Two other parameters refer more directly to the line segment contained within the band. The first of these measures an error variance, and computed as the sum of the distances of all coordinates in the segment from the anchor line. The measure is constructed in general format for variance:

$$\text{error variance} = \frac{\sum(d^2) - \frac{(\sum d)^2}{n}}{n-1}$$

The term error variance is used, because this measure represents a probable displacement error which should accrue by maximum generalization. In other words, by representing the original set of coordinates by only two coordinates, the anchor line endpoints in the strip tree, what is the maximum error which can be measured in the generalized line? Error variance values are reported in coordinate units of the Apple screen.

The second parameter is called monotonicity, and is a measure of the probability that the $(n+1)^{th}$ point in the original line will fall on the same side of the anchor line as did the n^{th} point. It is computed as the number of times the original line crosses the anchor line, divided by the number of points in the original line contained within that particular band. Both monotonicity and error variance are considered as probability measures, a point which will be referred to again during the second stage of the procedural description.

Five parameters (band length, width, segmentation, error variance, and monotonicity) describe the geometry of the line. All of the measures (except length, which can be computed) are to be stored at each node in the modified strip tree. Also stored in the strip tree are the coordinate endpoints of each strip, and the number of coordinates in the original line encompassed by each strip. Taken together, these parameters provide representative summaries of the original cartographic line, at increasing levels of resolution.

It will be shown that the information in one strip tree may be quite distinguishable from the information in another, and this kind of distinction may be useful in identifying generic categories of line features. In terms of line generalization, this implies that the measures in the strip tree could provide preliminary computer-generated guidelines as to where along a complex line feature the smoothing or simplification tolerance values might best be changed. Before discussing this further, however, the extent to which the parameters make such distinctions possible must be determined.

Building the Structure Signatures

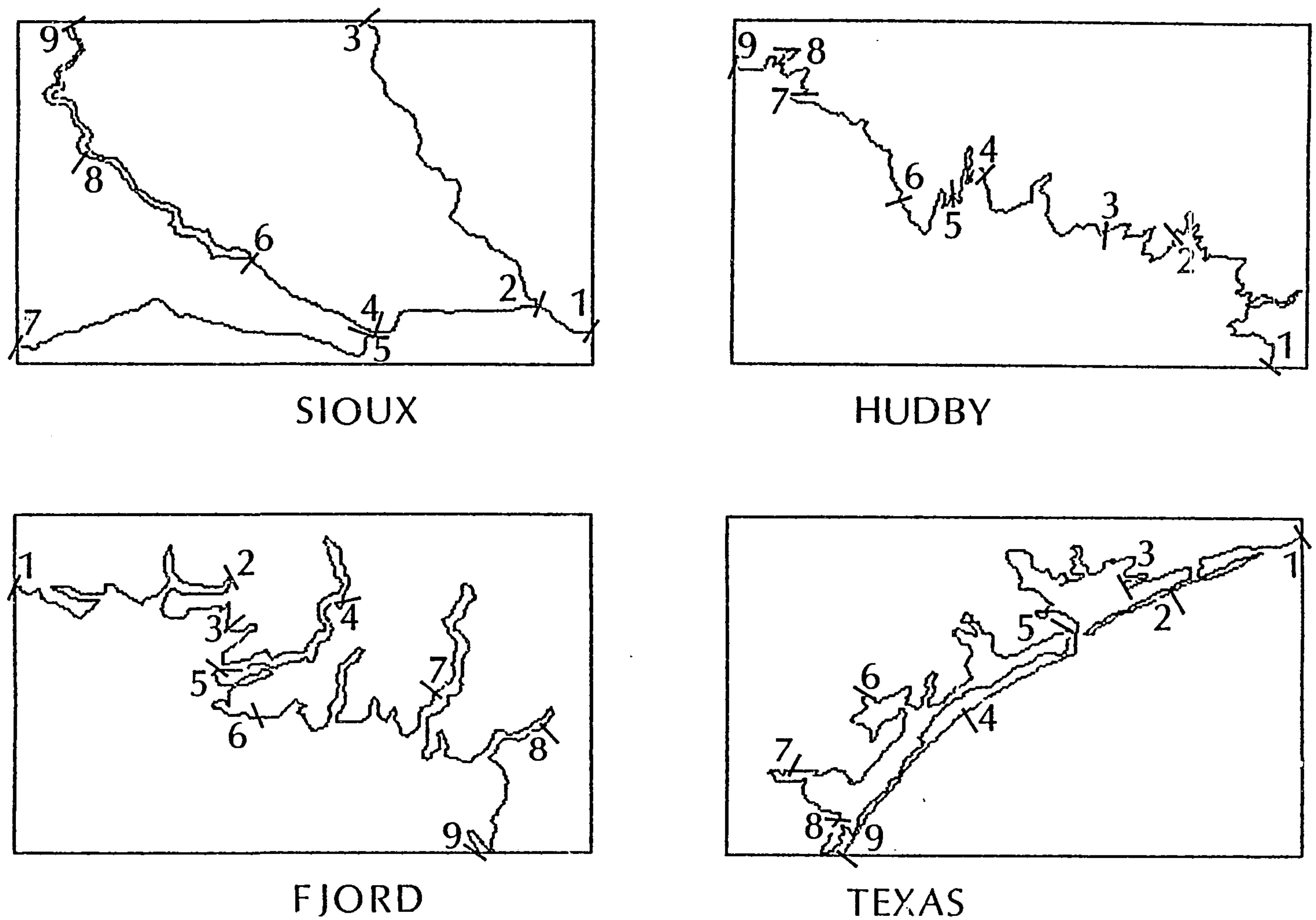
In generating a procedure by which to subdivide the four cartographic lines, attention must be paid to the purpose of this subdivision, and of the parameters derived from it. In the most general terms, the subdivision is aimed at distinguishing between categories of cartographic lines. Line distinctions are important for generalization in order to retain the balance between accuracy and recognizability of the cartographic line.

Accuracy is retained by specific coordinate position: some of the coordinates may define geographically important locations, for example, the intersection of two river channels, or the location of a city. Inclusion of such locations in generalization is closely tied to the purpose of the map, and to the geographic scope of the mapped information.

To simulate this aspect of generalization, a set of nine coordinates was marked in each of the four line files. These points included the first and last point in the file. In the SIOUX file, the seven points included intersections of river channels and the two extreme ends of the reservoir. Points were chosen randomly in the other sample line files, because no clear indication of geographically critical points was visually apparent. Figure 4.6 displays the locations of the nine coordinates chosen for each cartographic line.

These coordinates are intended to simulate critical positional information for each cartographic line; this means they should be retained regardless of the degree to which the line is generalized. Therefore, these points will be included at all levels in the strip tree data structure: the first level of the tree for each cartographic line is thus composed of not one but eight nodes. Each node represents a line segment band which is bounded on either side by a point required for geographic accuracy.

As these endpoints will be repeatedly included within each lower level of the data structure during the subdivision process, it will be possible to plot a representation of the complete line using only endpoints and anchor lines from any single level of the strip tree, and to rely upon the fact that the geographic accuracy will be constant for any level chosen.



POINTS CRITICAL TO GEOGRAPHIC ACCURACY

FIGURE 4.6

Appendix A includes illustration of the initial eight bands for all four sample lines, and their associated measures of low order structure. These bands form a starting point, from which all further subdivisions will proceed.

A cartographic line contains another kind of positional information as well, as discussed in Chapter 3. These coordinate locations are not absolutely important, but serve to define position in a more relative sense. The coordinates lie in the vicinity of maximum angular change; and their inclusion in a generalized feature has been shown to maintain recognizable shapes and perceptual accuracy in linear representations Attneave, 1954; Kelley, 1977; Marino, 1979). As opposed to the information required for geographic accuracy, in which all coordinates are required at all levels, this information appears to be hierarchic. That is, not all of the coordinates are critical to line character recognition at all levels of resolution, but those points which are determined to be critical at coarse levels tend to be chosen for finer resolutions as well -- the isolated studies of Marino and Kelley demonstrate this overlap quite clearly. In the present study, a hierarchy of points critical to perceptual accuracy will be selected as band endpoints for the rest of the line segment subdivisions. In conjunction with the geographic points, these should prove useful anchors on which to base measurements of the lines, as each band so constructed can be said to comprise a visual and/or geographic unit of the cartographic line.

To select these perceptually critical points, the eight initial bands of each line have been further subdivided using the Douglas-Peucker (1973) generalization algorithm, which selects points based on maximum perpendicular deviation from the band anchor line. This algorithm has two advantages: it can be implemented in a small amount of core memory, which is useful for working in microcomputer environments; and also, the algorithm has been shown to select points similar to the set which map readers would consider critical to retaining line recognizability during generalization, that is, to retaining perceptual accuracy (Jenks, 1982, personal comunication).

As a point is selected, it is entered into the strip tree as an endpoint for both child nodes of the given band. The other endpoint for each new band is replicated from one endpoint of the original parent band (Figure 4.7). As it is broken out, each line segment band has been measured according to the five parameters (band length, width, segmentation, error variance, and monotonicity) described previously.

The subdivision process terminates for any band when one of two conditions has been met. First, termination will occur if the current band encompasses fewer than five points in the original line; or second, it will occur if the bandwidth narrows to the resolution of a straight line. In either situation, the lines' irregularities are not resolvably different than the band which encompasses it, and further subdivision and measurement would not be meaningful. Figure 4.7 illustrates the second criterion for termination, for a hypothetical band. On the Apple screen, the minimum resolvable bandwidth is five pixels. Appendix B includes the measures collected for each band in all four cartographic lines, with an explanation of the algorithm designed for this task and a more specific description of the data structure.

Only five levels of subdivision were possible, or actually, four levels beyond the initial subdivision into eight geographic bands. Any band

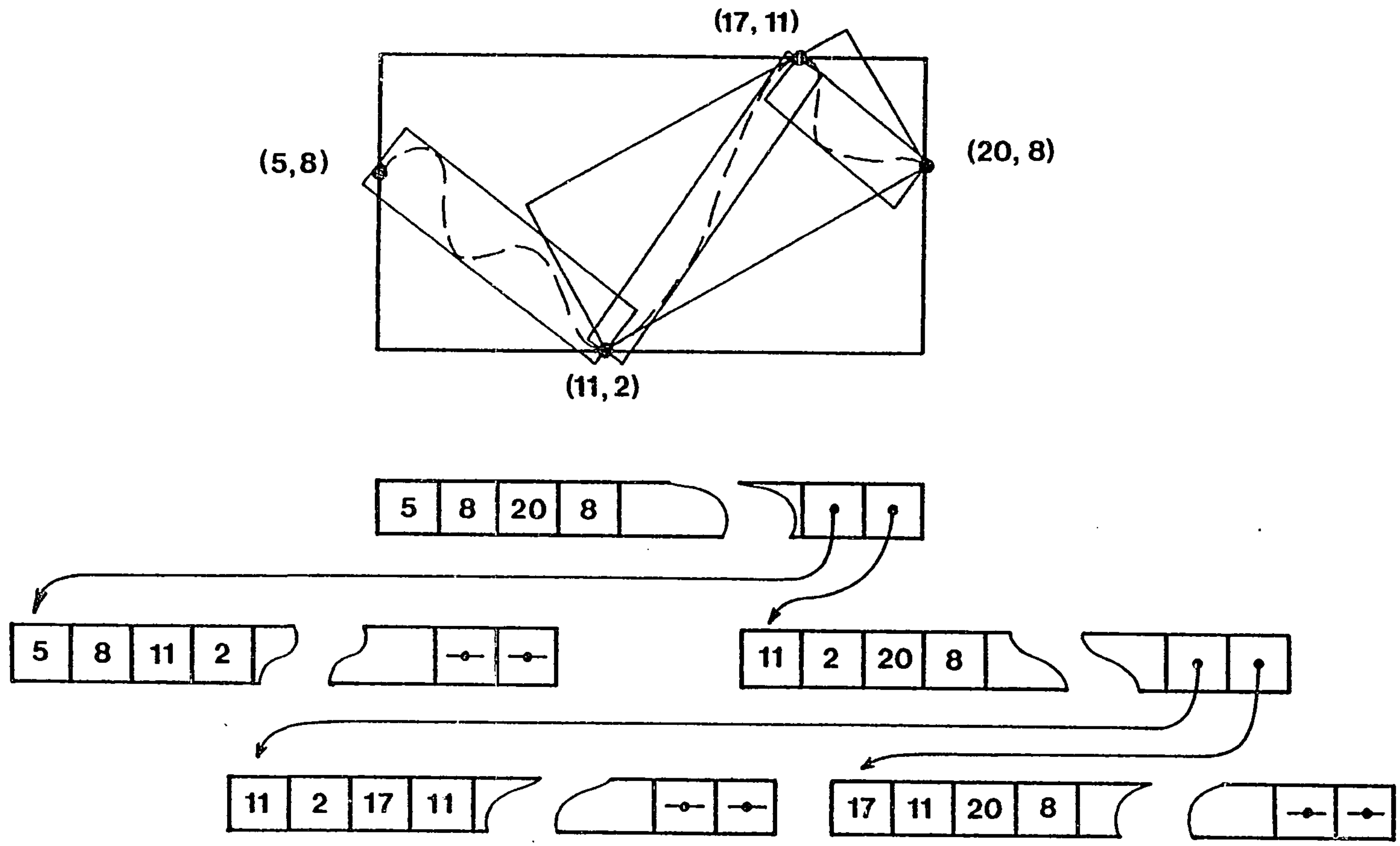


Figure 4.7
SUBDIVIDING THE CARTOGRAPHIC LINE

which reached the threshold criteria was not subdivided further, but was instead replicated as a single node at the next lower level in the tree, and at all levels below that one. Building the tree in this way preserves the integrity of the cartographic lines by resolution level: any single level of the strip tree can provide a strip reconstruction of the complete line.

One disadvantage of building the data structure in this way is the certain introduction of empty, or 'non-significant' nodes: when a single band is replicated in place of two nodes in the level below, it must stand in place of four nodes in the next level beyond; this is the nature of a binary tree. Conceptually, this replication implies that all of the significant details in the given segment have been resolved by the current band, and that requirements for recognition and representation of the given line segment will not change when the line is displayed at finer resolutions. Steinhous (1950), who first published the mechanics for enveloping line details within regular polygons, would say that at this point, the line segment has been "rectified".

As more and more segments are rectified, the strip trees will become increasingly sparse. In a full and balanced tree, with all nodes significant, the progression of the number of nodes in each level would be $2^{(n+2)}$, where n equals the tree level. This means that there should be $8=2^{(1+2)}$ nodes in level 1, $16=2^{(2+2)}$ nodes in level 2, and then 32, 64, etc., nodes in subsequent levels. In all four strip trees, though, only the first two levels were filled with the total number of significant nodes possible; beyond this level, some bands in every line became rectified. This is in part due to the original choice of (eight) bands, which was purposely randomized to simulate the cartographer's lack of control.

As each level in the tree is created, it is possible to predict how many bands can be meaningfully subdivided to create the next level of the tree (Table 4.2). By completion of the fifth iteration for the sample data, it was found that over 90% of the line segments had been rectified. Out of a possible 256 nodes, then, the next level in the tree (which would have been Level 6) for every line would contain less than 10% of the new bands possible; therefore, the subdivision process was terminated at this point (Figure 4.8).

The final task in building the structure signatures involves consolidating the information stored in each strip tree. For each line, the parameters have been summarized by computing a mean and a variance; a coefficient of variation has also been calculated, as a more meaningful expression of the very large variance values. Overall band width and the parameter for maximum deviation are quite similar: as a result, these have been converted into two related measures, negative width and positive width. These are equivalent to Ballard's parameters called 'width left' and 'width right': by either name, they refer to the partial band width falling on either side of the anchor line.

Appendix C displays all summary values for all parameters for the four cartographic lines. These summaries are also stored as a look-up table whose primary index is keyed to the name of the line -- FJORD, HUDBY, etc. The keyword FJORD will therefore access all summary parameters for any desired level of resolution. The purpose of this look-up table will be explained in detail in the second part of the analysis, and during reconstruction of the line features. But first, the parameters themselves

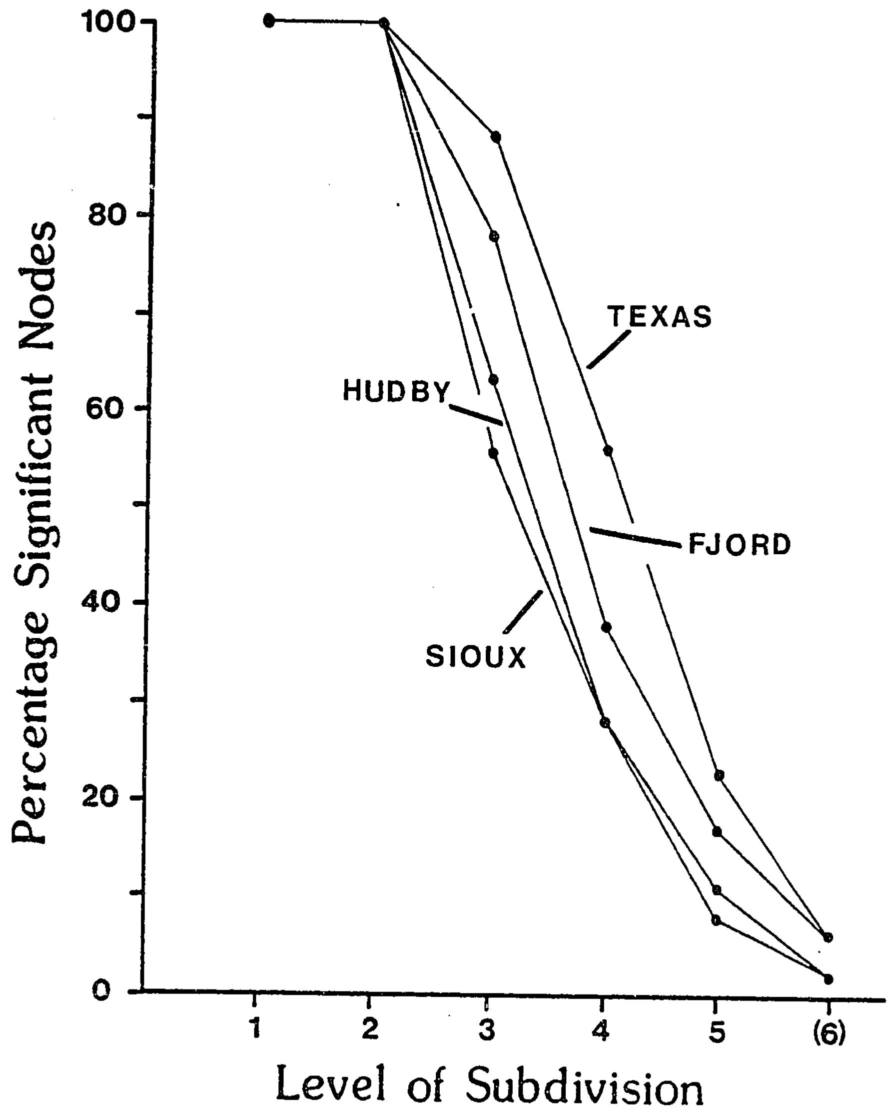


Figure 4.8
NEW NODES ADDED AT EACH LEVEL
(as percentage of total nodes possible)

will be studied, to analyze their power to distinguish between the four types of cartographic lines.

TABLE 4.2
SIGNIFICANT NODES AT EACH LEVEL IN THE STRIP TREE

Number of Nodes Added During Each Subdivision
(new nodes at each level)

File	Level 1 (8)	2 (16)	3 (32)	4 (64)	5 (128)	(6) (256)
SIOUX	8 (100)	16 (100)	18 (56)	18 (28)	10 (8)	6 (2)
HUDBY	8	16	20 (63)	18 (28)	14 (11)	6 (2)
TEXAS	8	16	25 (78)	24 (38)	22 (17)	16 (6)
FJORD	8	16	28 (88)	36 (56)	30 (23)	16 (6)

Total Number of Nodes After Each Subdivision
(new nodes plus rectified nodes)

File	Level 1 (8)	2 (16)	3 (32)	4 (64)	5 (128)
SIOUX	8 (100)	16 (100)	25 (78)	34 (53)	39 (30)
HUDBY	8	16	26 (81)	35 (55)	42 (33)
TEXAS	8	16	28 (88)	40 (63)	51 (40)
FJORD	8	16	30 (94)	48 (75)	63 (49)

NOTE: Numbers in parentheses refer to percentage totals, which are subscripted to each column heading.

Signatures for Low Order Structure

Summaries by Mean Value

The signatures displayed in Figures 4.9-4.12 illustrate mean parameter values for all five measures, at each level of subdivision. All measures except error variance range from 0.00 to 1.00; they are standardized either to band length or (in the case of monotonicity) to the number of coordinates in the original line. Only significant bands' measurements have been included in the summaries; thus the fact that strip width parameters plotted in Figure 4.9 approach zero is due to decreasing strip width, and not biased by the inclusion of increasing numbers of rectified bands. In Chapter 3, it was mentioned that the strip width measures indicate the degree of cross-sectional symmetry evident in the line at increasing levels of resolution.

In Figure 4.9, symmetry can be interpreted by comparing the magnitudes of the negative and positive widths for each line. The SIOUX signature shows cross-sectional symmetry at all levels of resolution, while the TEXAS line displays the least symmetry of the four lines. The strip width parameter shows an interesting progression across these lines: the rate at which the bandwidth narrows varies distinctly from one line to the next. Coupled with the interpretation of (non-) symmetry, one might imply from these plots that the SIOUX line is less convoluted than are any of the other lines, and that FJORD and TEXAS display the highest number of irregularities, at every level of resolution.

Plots for the error variance parameter (Figure 4.10) tend to confirm this interpretation. Mean error variance decreases for all four lines, from one level to the next; but the rate of decrease is highest for TEXAS.

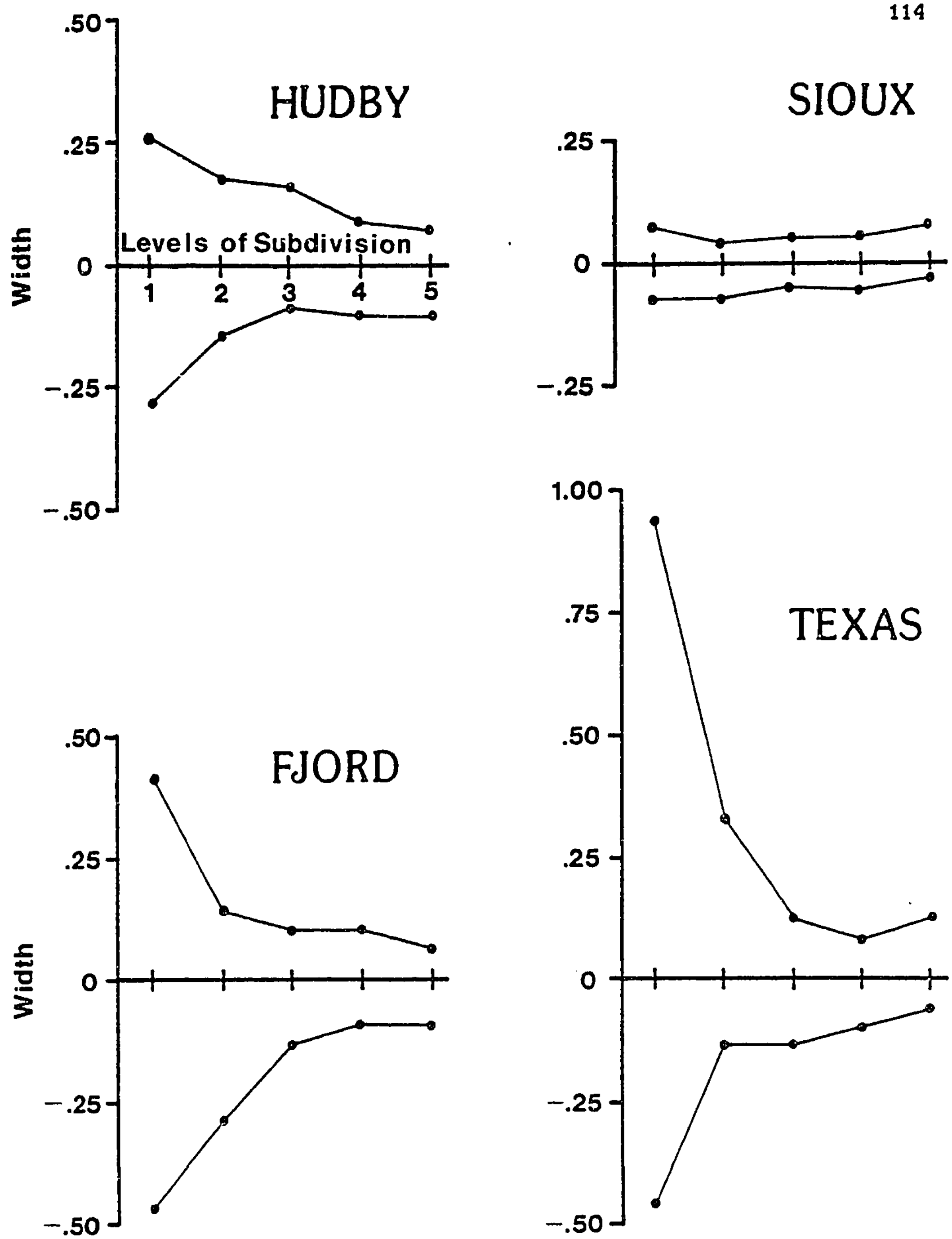


Figure 4.9
CROSS-SECTIONAL SYMMETRY
(band width)
Mean Value Structure Signatures

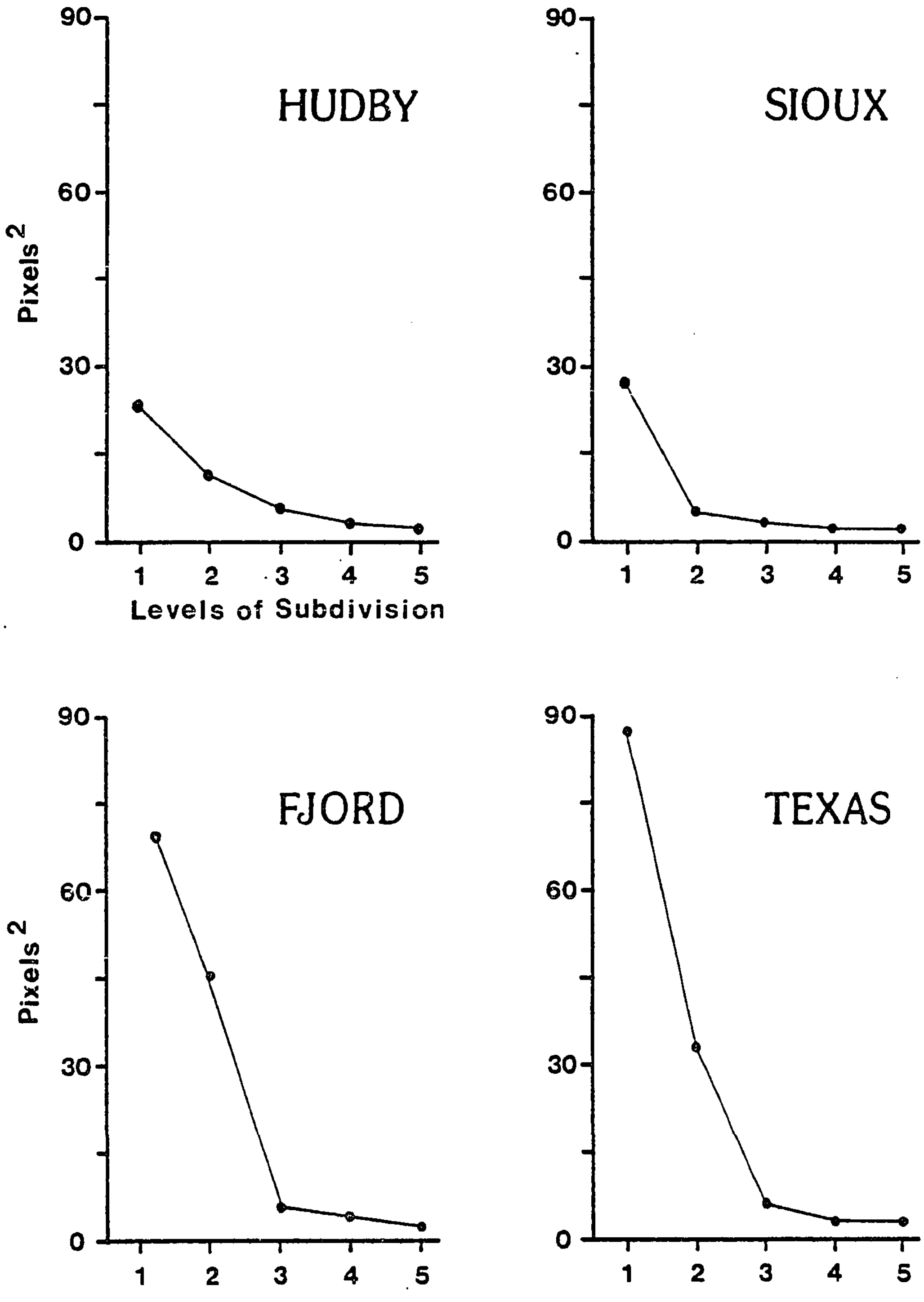


Figure 4.10
ERROR VARIANCE
Mean Value Structure Signatures

Furthermore, the progression for this measure is identical to the progression for strip width signatures (first SIOUX, then HUDBY, FJORD, and TEXAS). As a tentative interpretation, one might combine these two parameters to infer a progression of linear complexity in the four lines. Further evidence for this conclusion will be provided later in the analysis.

It is also interesting to note the distinct elbow, or jog, in three of the four error variance signatures. In Chapter 3, a jog was indicated for the line spacing plot example (Figure 3.5). The jog was said to be located at the limits of accuracy for the line spacing symbol. Below this scale, a change in the structure of the graphic line was indicated, to preserve an accurate visual relationship between the cartographic line and the feature it represents.

Here, the error variance parameter measures the sum of the squared deviations of original coordinates from the anchor lines, which is in effect a mean squared error accrued by producing straight line representations of the lines at coarser levels of resolution. The jog in the structure signature may be interpreted as indicating the level of subdivision below which simple anchor line plots can no longer provide a visually accurate representation of the original line at scale.

In this context, the error variance parameter can be considered as a measure of generalization accuracy for the Douglas-Peucker algorithm which directed the subdivision process. While it is no surprise that the amount of generalization error tends to decrease at finer resolutions, it is interesting that the pattern of generalization error differs distinctly from one type of cartographic line to the next, and that the jog occurs at different levels. Prior to this study, it has been commonly presumed that a sound algorithm such as this one produces generalizations of comparable accuracy for all types of line features.

Since parameters of low order structure have been measured by subdivisions based on perceptually critical points, one could go a step farther, perhaps, to hypothesize that the jog in the error variance plot marks a level of resolution beyond which line recognition occurs. Anchor line plots for all four lines are illustrated in Appendix D. In looking at these plots, notice the level at which the anchor line plot begins to look like the original sample line. For many readers, the SIOUX plot may become recognizable at a much earlier level than, for example, the FJORD plot. However, this sort of prediction cannot be certified without empirical tests, which are beyond the scope of the present dissertation. For the time being, then, this must remain as speculation, and as a suggestion for future research.

Structure signatures for the segmentation parameter are displayed in Figure 4.11. This measure has been standardized to the length of the anchor line; thus a mean value of 0.50 implies that bands are being split in half at a given level of subdivision. The structural interpretation is one by which longitudinal symmetry may be compared: line segments containing consistent magnitudes of deviations from one end to the other will tend to be broken at their midpoint. On the other hand, subdivisions will occur off-center for segments whose deviations are inconsistent.

In this context, one can see that the SIOUX file is quite symmetric, within this range of resolution. The subdivisions in the TEXAS file also oscillate about the anchor line midpoint, but here the pattern is more

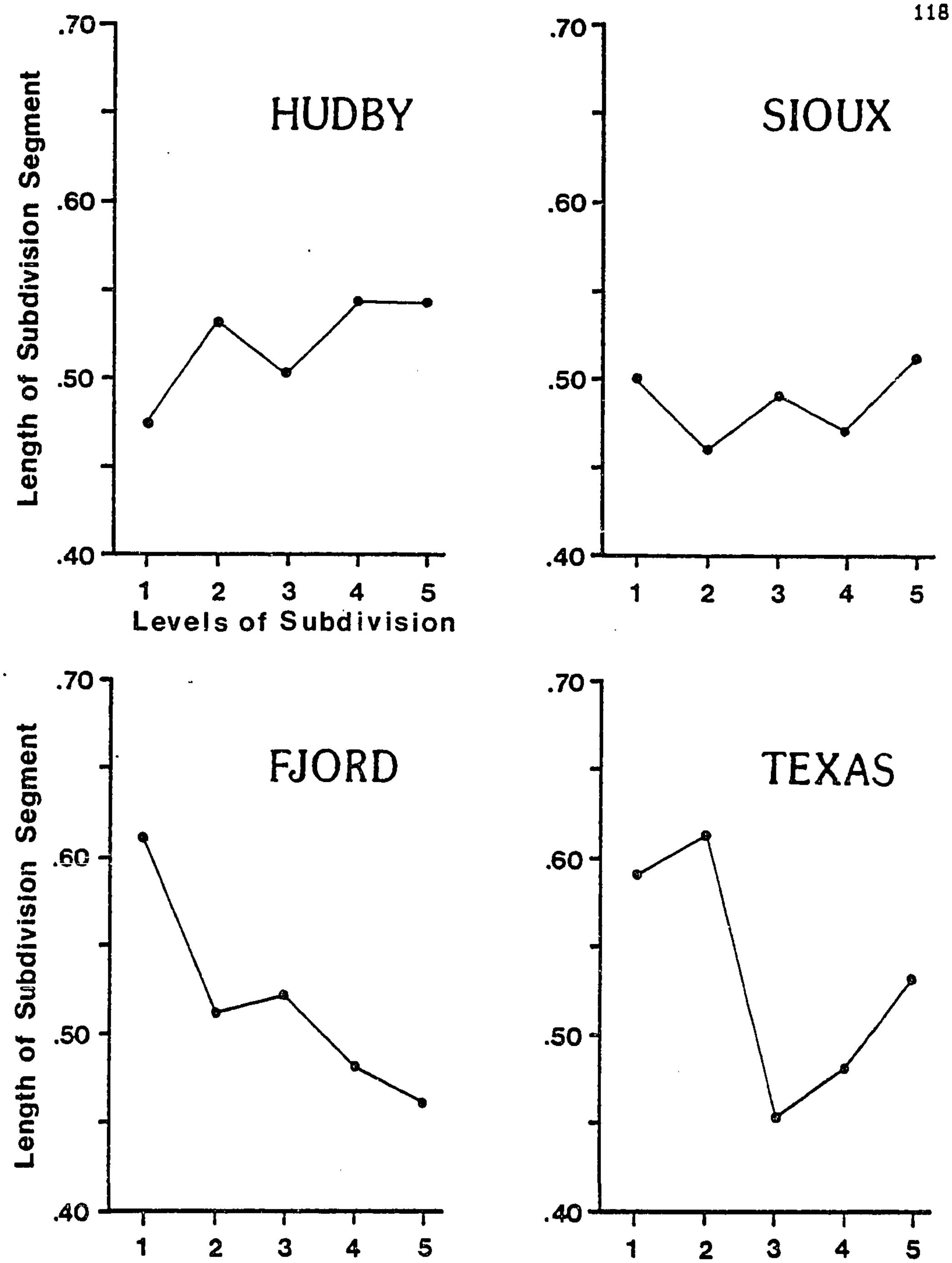


Figure 4.11
LONGITUDINAL SYMMETRY
(segmentation)
Mean Value Structure Signatures

pronounced. There is a definite trend in the segmentation of the FJORD file to occur towards the ends of the band at finer resolutions. HUDBY displays a similar, though weaker trend; however, this last is a cautious interpretation, as the trend is hard to verify given only five subdivision levels.

Because the segmentation parameter measures the place at which a maximum deviation occurs, it will of course be affected by the largest magnitude irregularity which lies within the band. If this magnitude remains in constant proportion at all levels of resolution, then segmentations will also tend to occur in constant proportion. The structure signature should reflect this consistency, as happens in the case of SIOUX. In the FJORD and TEXAS files, however, the segmentation process resolves several magnitudes of irregularity. The fact that the TEXAS signature rebounds to 0.50 at finer resolutions inplies that the smaller scale deviations are homogeneous, and consistent along the length of the line, even though the larger scale irregularities are not. The trend of the FJORD segmentation signature implies a different structure again: increasing irregularity is resolved with finer and finer subdivisions of the original line. As with Mandelbrot's fractal curves, this line seems to become more complex at finer levels of resolution.

The fourth structure signature (Figure 4.12) measures monotonicity, which is the number of times the original line subsumed by a band crosses that band's anchor line. This measure is standardized by the number of points in the line segment. One can think of monotonicity in a probabilistic sense, the probability that two adjacent points will fall on the same side of the anchor line. This parameter is not affected by the

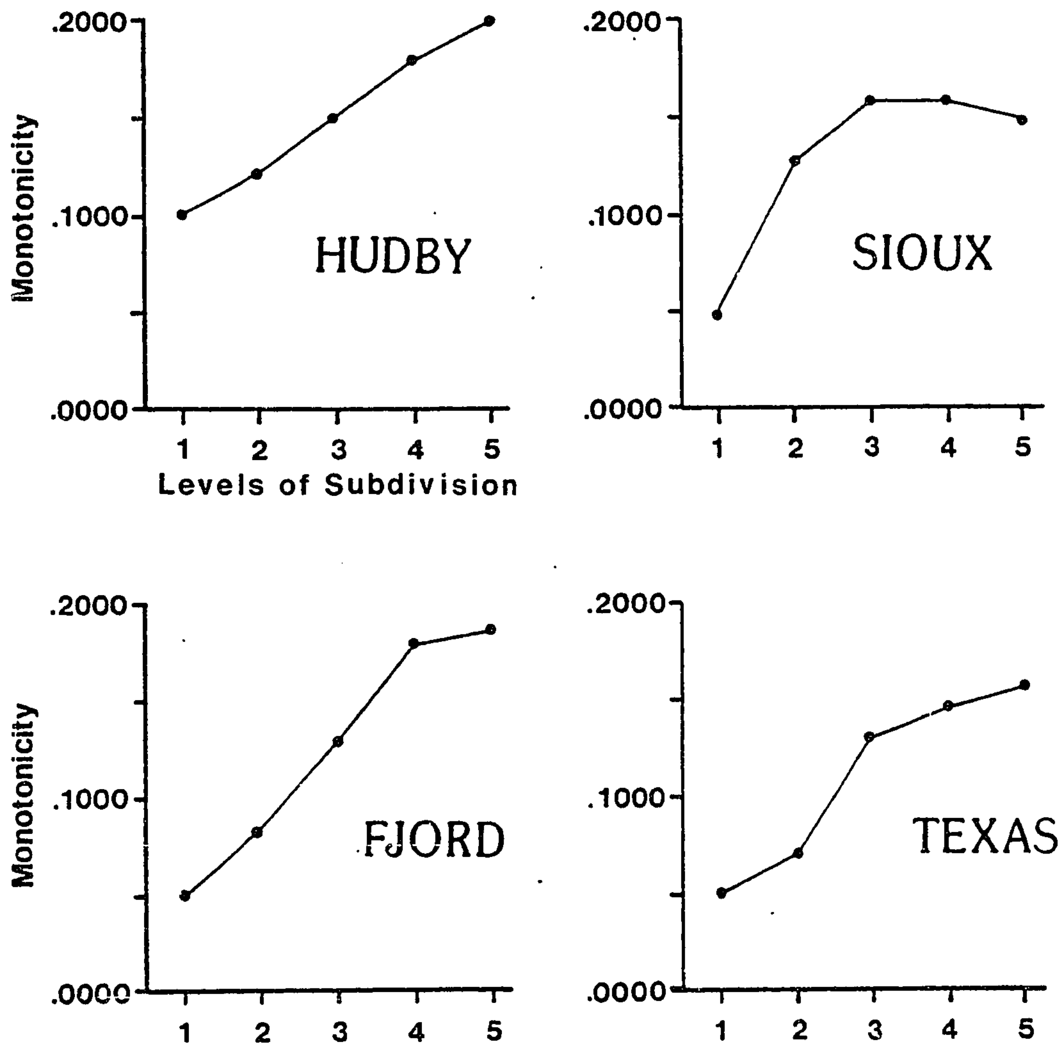


Figure 4.12
MONOTONICITY
Mean Value Structure Signatures

magnitude of the deviations along the original line, but rather by the freguency with whcih they occur. Keeping this in mind, then, the mean monotonicity signatures can be interpreted as follows.

For SIOUX, the monotonicity levels off early in the subdivision process. This does not mean that the frequency of deviations is constant at all levels of resolution, but rather that it begins to drop, as fewer coordinates are being subsumed in each band. There is a sharp rise in the signatures for both FJORD and TEXAS, where the coastlines display a distinctly smoother side and a more sinuous side. The rise in the structure signatures reflects the fact that with further subdivisions, the anchor lines more closely approximate the original coordinate set, by passing through the middle of this set, and not to one side of it. At the finest level of resolution, the signatures reach equilibrium, as did the structure of SIOUX.

The only signature for monotonicity which continues to increase through all levels is HUDBY, reflecting those oscillating deviations which characterize this line's graphic structure. The rising parameter value implies that the frequency of oscilations increases with finer resolutions of the line, even though the magnitude of those oscillations (as measured by previous parameters) may not be changing very much at all.

One may summarize the mean parameters for low order structure by saying that SIOUX displays the lowest amount of irregular convolutions, being symmetric both longitudinally and in cross-section. Of all four sample lines, SIOUX is probably best approximated at all levels of resolution by the simple anchor line plots. HUDBY displays more irregularity, which is symmetric in cross-section, and oscillates about the anchor lines. The oscillations appear to increase at finer resolutions. TEXAS and FJORD appear to be most similar of the four lines, in terms of the complexity of their low order structure. But while the cross-sectional symmetry of TEXAS becomes more apparent at finer resolutions, new irregularities in FJORD continue to appear with each new subdivision. In this respect, the low oreder structures of FJORD and HUDBY are most similar.

These four parameters (strip width, error variance, segmentation, and monotonicity) are summaries, average values computed for rows of the strip tree data structure. They focus on coordinate relationships of the original line, to measure low order graphic structure. However, two points must be kept in mind when interpreting the mean parameters. First, structure is being measured indirectly, and approximated by measurements of a regular polygon, the band, which bounds the original line. The fact that this approximation is repeated for smaller pieces of the line gives better credence to the summaries at finer levels of resolution. Secondly, the parameters have been consciously derived as probabilistic measures. For example, in the case of monotonicity, <u>where</u> the original line crosses the anchor line is not so important as the fact that it crosses it at all.

These are important points in analyzing low order structures, where summaries are made for all bands in a given level of the strip tree, without regard to their sequence in the level. In this way, if the characteristic irregularities of two lines are similar in magnitude, but different in frequency, the mean values summarized for the four parameters will still be similar -- the lines will be seen to have similar graphic structure. This idea is illustrated in the top half of Figure 4.13, for

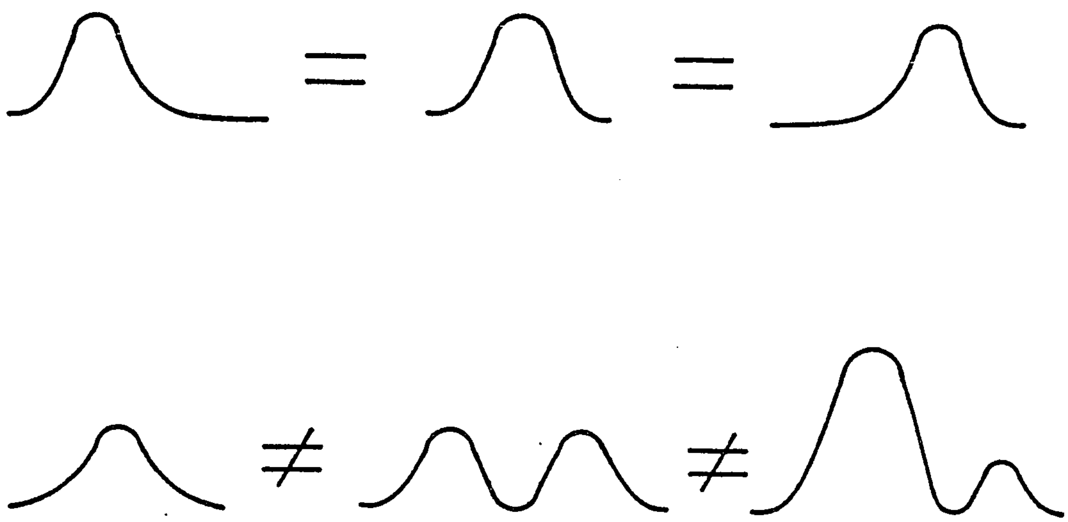


Figure 4.13
EQUIVALENCE
OF
LOW ORDER STRUCTURES

three simple lines having a single irregularity. The mean structure signatures would reflect equivalent values for these three lines, because their structures are the same.

However, mean values will also be equivalent for the three lines in the bottom half of the figure, which displays three different graphic structures. The solution is to summarize parameters by variance as well as by mean; it should be clear that in making structural distinctions, the frequency of occurrence of the structural irregularities is as important as magnitude.

At a more conceptual level, of course, one can say that as two cartographic lines have equal valued mean and variance for all parameters at a given level of resolution, the structures of those lines can be said to be similar. Similar, but not necessarily equivalent: this is a probabilistic equivalence, based on approximations of structure by regular polygons. Attention will turn to the parameter summaries by variance, as the second part of this study of low order structure signatures. After this, the parameters will be analyzed for statistical distinctions between the four sample lines.

Summaries by Variance

The variance values are relatively high for all structure parameters. This is due in part to tha small sample sizes involved in this study: each of the line files is composed of 300 - 500 coordinates. These are subdivided into a larger number of bands at each level of subdivision. By the fifth iteration, the median number of coordinates in each band has dropped below 15, in all four line samples, which could contribute to the high variance values. Therefore, the values have been converted to coefficients of variation, which is computed as the square root of the variance divided by the mean. This standardization to the mean should also improve comparability between variance parameters, which are displayed as structure signatures in Figures 4.14 - 4.17.

The interpretation of these illustrations focuses on the consistency of graphic structure across the bands in a given level, and on the degree to which the mean values offer appropriate summarizations. While the signatures display inflated values overall, one can discern clear patterns by which the four lines may be compared and distinguished. Keep in mind that as a ratio of the mean values, the illustrations should be interpreted cautiously: a rising or dropping coefficient of variation may reflect a changing mean as much as a changing variance.

In Figure 4.14, the negative and positive width measures have been recombined to analyze the variance in band width as a whole. One can see from the horizontal trend of its signature that the mean bandwidth for SIOUX is an equally representative summary at all levels of the strip tree. The FJORD line's structure is less consistent overall than is SIOUX, but here too the coefficient of variation is fairly stable. The peak in the HUDBY signature is difficult to interpret, as the mean band width is also dropping in the first levels. Inconsistencies in overall band width may be caused in part by the resolution of one or more of the graphic oscillations which seem to characterize this line.

In the case of TEXAS, the sharp decrease in the coefficient of variation is most likely caused by several outlying values, measured for at

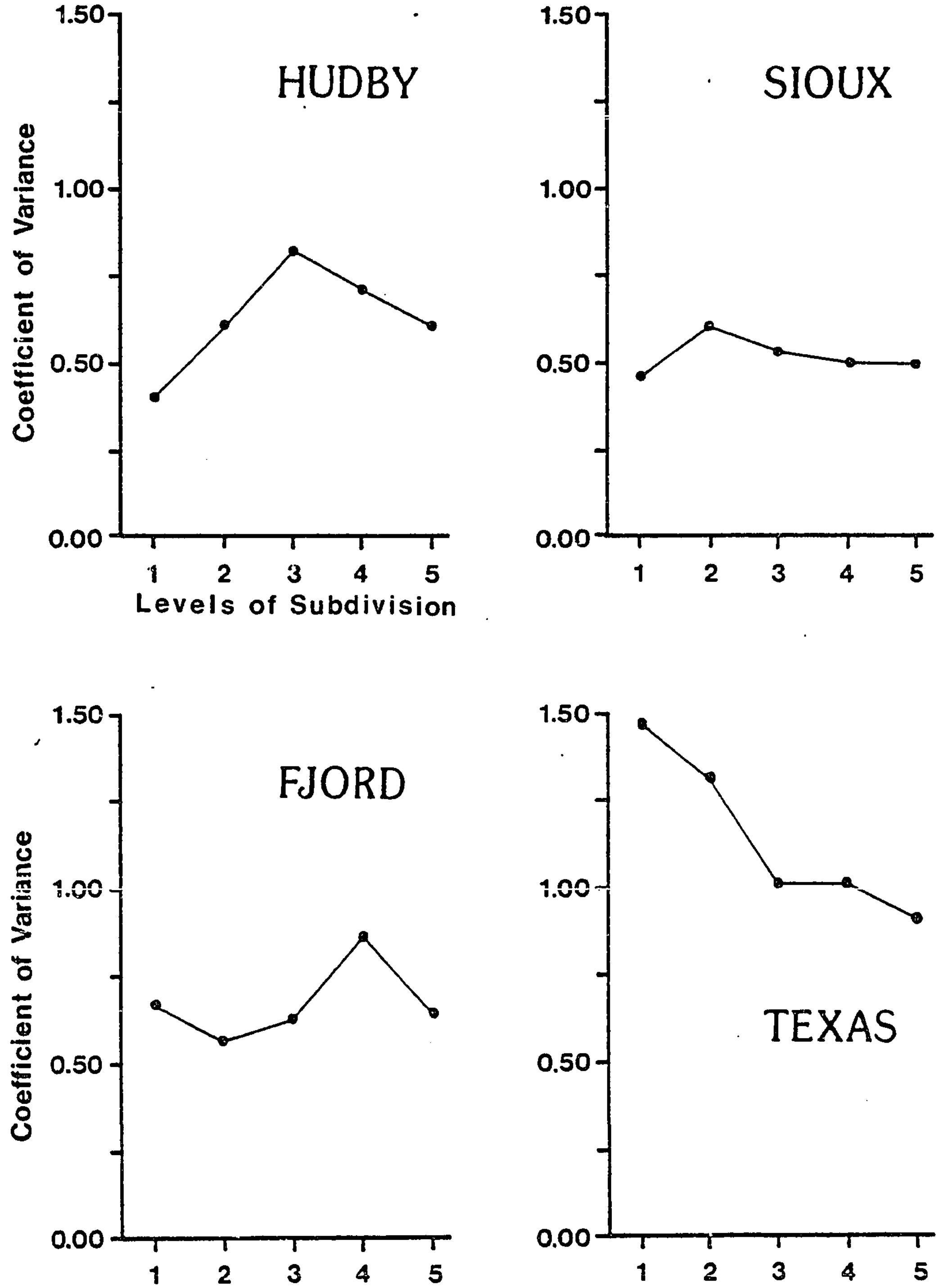


Figure 4.14
BAND WIDTH
Variance Value Structure Signatures

least one of the initial bands. As the line has been subdivided, a few extreme irregularities have been rectified, and the variation in strip width begins to drop. These outliers are evident in Table B.4 of the Appendix: several values in the width column can be seen to be magnitudes larger than the length of their respective anchor lines.

A plot for the first outlier value encountered in this table is displayed in Appendix A (see strip 8--9, for the TEXAS file). This band is notable in that it does not completely encompass its piece of the origianl line. Of course, these exterior convolutions become encorporated into a band as the subdivisions continue. In part, this anomaly is due to the arbitrary (although random) choice of initial bands; but it is also a reflection of the extreme sinuosity of the original line, and of a graphic structure in which band orientations tend to double back on themselves.

Most notable about the coefficient of variation for the error variance parameter (Figure 4.15) are the rising trends for both FJORD and TEXAS. These trends imply that inconsistency in these lines is increasing at finer resolutions. This was pointed out once before for the FJORD file, in reference to Mandelbrot's fractal curves. The variance is generally stable for the SIOUX file; once again, no ready interpretation can be offered for the HUDBY pattern.

The segmentation parameter's variance signatures (Figure 4.16) are more easily interpreted; the plots are similar for the sharp rise in all four lines, after the first subdivision. Recall that the initial assignment of bands was random, but that the subdivision algorithm worked on the distinctly non-random principle of maximum perpendicular distance. The implication here is that the variance parameter is reflecting the

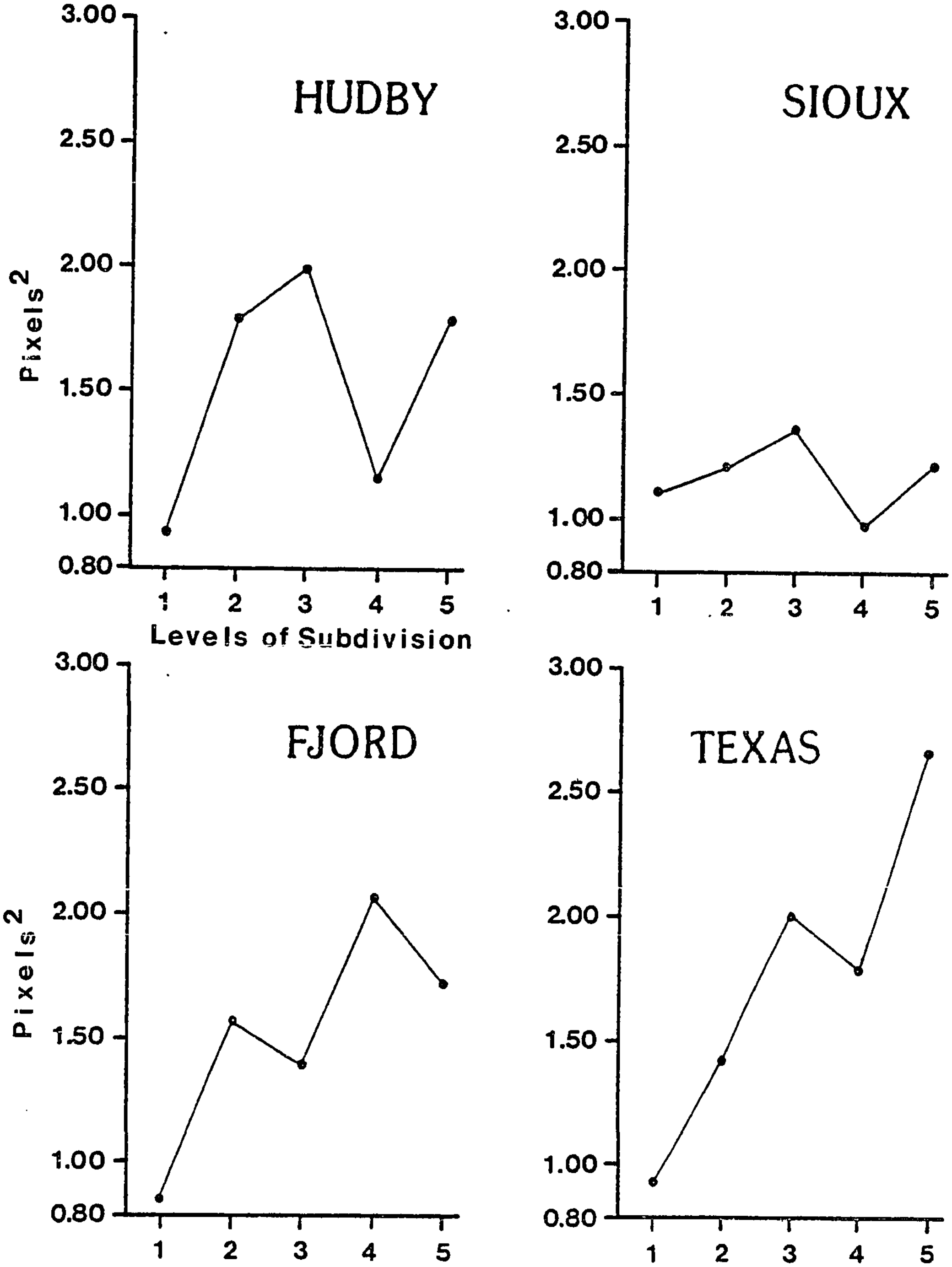


Figure 4.15
ERROR VARIANCE
Variance Value Structure Signatures

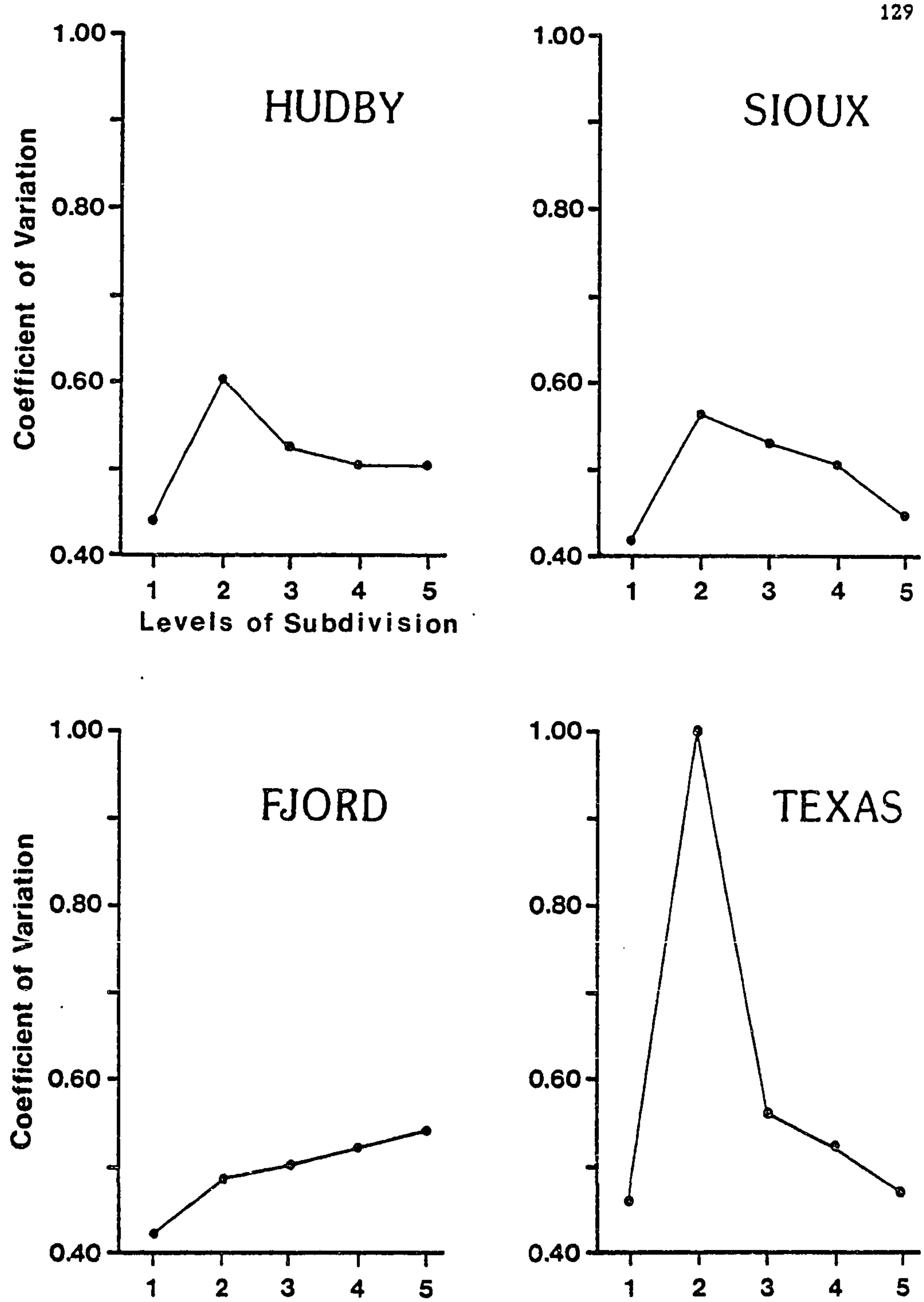


Figure 4.16
SEGMENTATION
Variance Value Structure Signatures

change in criteria for segmentation which occurred at this point.

The small peak in the FJORD plot is perhaps coincidental, as is the sharp peak in the TEXAS file. Much more interesting is the continued rise in the variance for FJORD. This may be due to the fact that segmentation is occurring at either end of the band, at these levels, rather than in the middle. The drop in the segmentation plot for TEXAS would similarly reflect its mean value oscillating more closely about the anchor line midpoint at finer levels of resolution.

The fourth signature for coefficients of variation displays the variance parameters for monotonicity (Figure 4.17). The signature for the SIOUX file appears to be rising; however, because this parameter is standardized by a mean which is decreasing, one can interpret a variance which is stabilizing in correspondance with its mean. Variance in monotonicity is actually decreasing for the HUDBY line, indicating that the mean parameter is becoming more reliable summary at finer levels of resolution.

Monotonicity variance also decreases for the FJORD line file with finer resolution, once again verifying the consistent increase in complex details to be found with increasing resolution. As with TEXAS, the peak at level 2 may indicate nothing more than the rebounding effects previously described for the segmentation variance. The rise in the TEXAS variance signature at level 5 is not clearly understood.

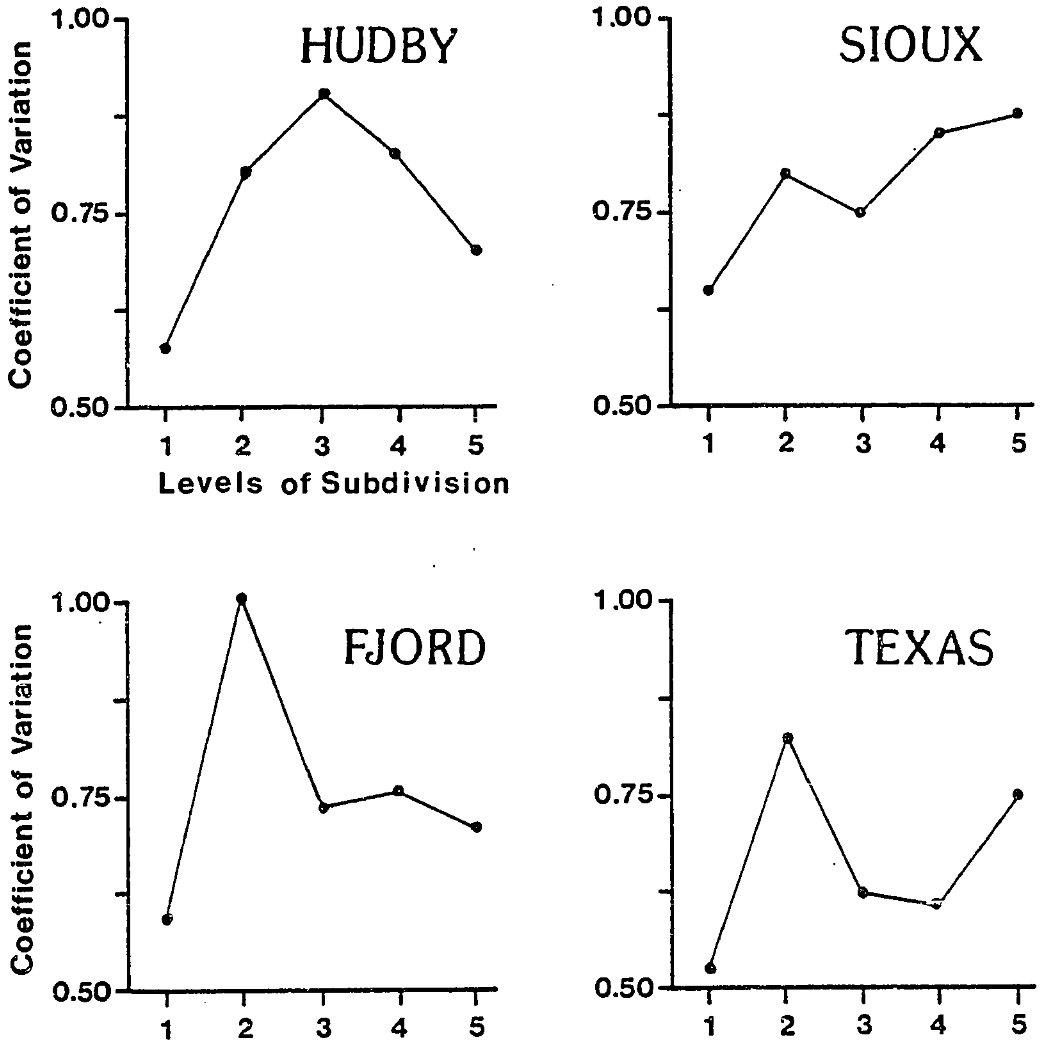


Figure 4.17
MONOTONICITY
Variance Value Structure Signatures

Summary

From these eight structure signatures, one is provided visual comparison between the sample cartographic lines of means and variance for parameters of low order structure. The graphic details which have been the focus here are based on relationships between specific coordinates, including endpoints of the anchor line and points at which maximum deviations occur. The specific endpoint coordinates have been selected by an algorithm which is acknowledged to retain points critical for recognizability; and the purpose of this chapter has been to describe and justify the subdivision process, and the graphic structures which were measured as a result.

One may conclude from the interpretations here aspects of graphic structure which one might expect to find in cartographic representations of these four categories of line features. For example, one should expect in the case of a feature such as HUDBY a line of strongly periodic deviations, which are fairly small in size, but which increase in frequency with increasing resolutions of representation. For a line such as SIOUX, one might also expect small magnitude deviations; one would anticipate that these deviations would be symmetric to their trend line, and that the visual complexity (as measured by error variance) in the representation should not increase markedly at finer resolutions.

The FJORD line type is characterized in its low order structure by a much more irregular line than either SIOUX or HUDBY, with deviations of many different sizes. In looking at this line at finer resolutions, more and more details become apparent. In comparison to the TEXAS line, the FJORD line is markedly nonsymmetric in cross section. In contrast, TEXAS' visual complexity increases for several levels of subdivision, and then begins to decrease once again.

Effectively, the low order structure parameters are designed to quantify those graphic relationships which are visually expected within line segments, and which may be useful for verification of the category of feature of which each line segment is a part. Earlier in this thesis, it was postulated that if two lines 'behave' differently at different scales, they can be said to have different structures. 'Behavior' was loosely defined at the time as a characteristic visual representation, akin to the so-called look of the line. This point can be articulated more succinctly now, by saying that if two lines have significant differences in parameter values, one may conclude that their low order structures are distinct within a range of resolutions. And this suggests a direct application of low order parameters to automating graphic representation tasks.

In a computer generalization, original line files might be checked by a similar subdivision process, performed piecewise along the length of the line, to determine at which bands these parameters display significant change in mean and variance. These would mark places along the line at which tolerance values for the smoothing or simplification algorithm might best be changed, to reflect the changing structure in the line file. Of course, the parameters by themselves might not be useful in choosing the actual tolerance value. In other words, the low order parameters cannot by themselves provide sufficient information to generalize these cartographic lines, but they may provide numeric information by which one kind of line may be distinguished from another. Focus in the next chapter will determine the extent to which such distinctions are possible, for the four sample cartographic lines. In the final chapter of the dissertation, the algorithm for line distinction tasks will be discussed in more detail.

CHAPTER 5
LOW ORDER AND HIGH ORDER STRUCTURE IN CARTOGRAPHIC LINES

In the previous chapter, a procedure was developed by which cartographic lines may be broken down into components of geographic information, including components critical to line recognizability, and other parametric components, designed to measure low order structure. Low order parameters measure relationships between points along the cartographic lines; and it would appear that these parameters can produce reasonable visual approximations of the specific lines which they have measured, as shown by the series of anchor line plots presented for each level of subdivision in the previous chapter. It has also been proposed that low order structure may provide a means of automating distinction tasks between types of line features, including modification of tolerance values during map generalization.

High order structures can also be considered parametrically; these parameters contrast with the low order parameters in measuring relationships between entire segments of a cartographic line. These parameters will be described in detail in this chapter; and it will be shown that the high order structure parameters for any given line feature can be used to identify that line by automated techniques, and to identify many other lines which represent the same type of feature. Before introducing the high orders of structure, however, attention will be turned to the low order structures, and to the utilization of low order parameters in discriminating between categories of cartographic lines.

<u>Relationships Between Parameters of Low Order Structure</u>

A preliminary step in applying the low order structure parameters to line distinctions is to analyze the distribution of the parameters as a whole. Two questions arise here: first, to what extent is a common pattern discernable in the parameter values for all lines? And second, to what extent are the parameters used in the study redundant expressions of low order structure? The first question can be addressed by comparing the distribution of parameter values between lines, at each level of resolution, and by variance computations. The second question can be addressed by analyzing correlations between parameters.

Analysis of the distribution of parameter values is an exploratory exercise, accomplished by computing a variance for the mean parameter values of each line. This is not the same as computing a variance for the original measurements, but is a variance computed directly from the means. The plots in Figures 5.1 and 5.2 display variances for all five parameters, and also for the breakdown of strip width into its positive and negative components.

Two groups of parameters present themselves. The first group is comprised of parameters which are directly related to strip width (including maximum deviation and the strip width measures). The patterns in all of these plots can be seen to be quite similar; one would expect that correlations between these measures should be significant. The pattern is easily interpreted, for the band widths should logically begin to drop with increasing subdivision of the line: as the mean parameter values approach zero, then, their variance values will also stabilize.

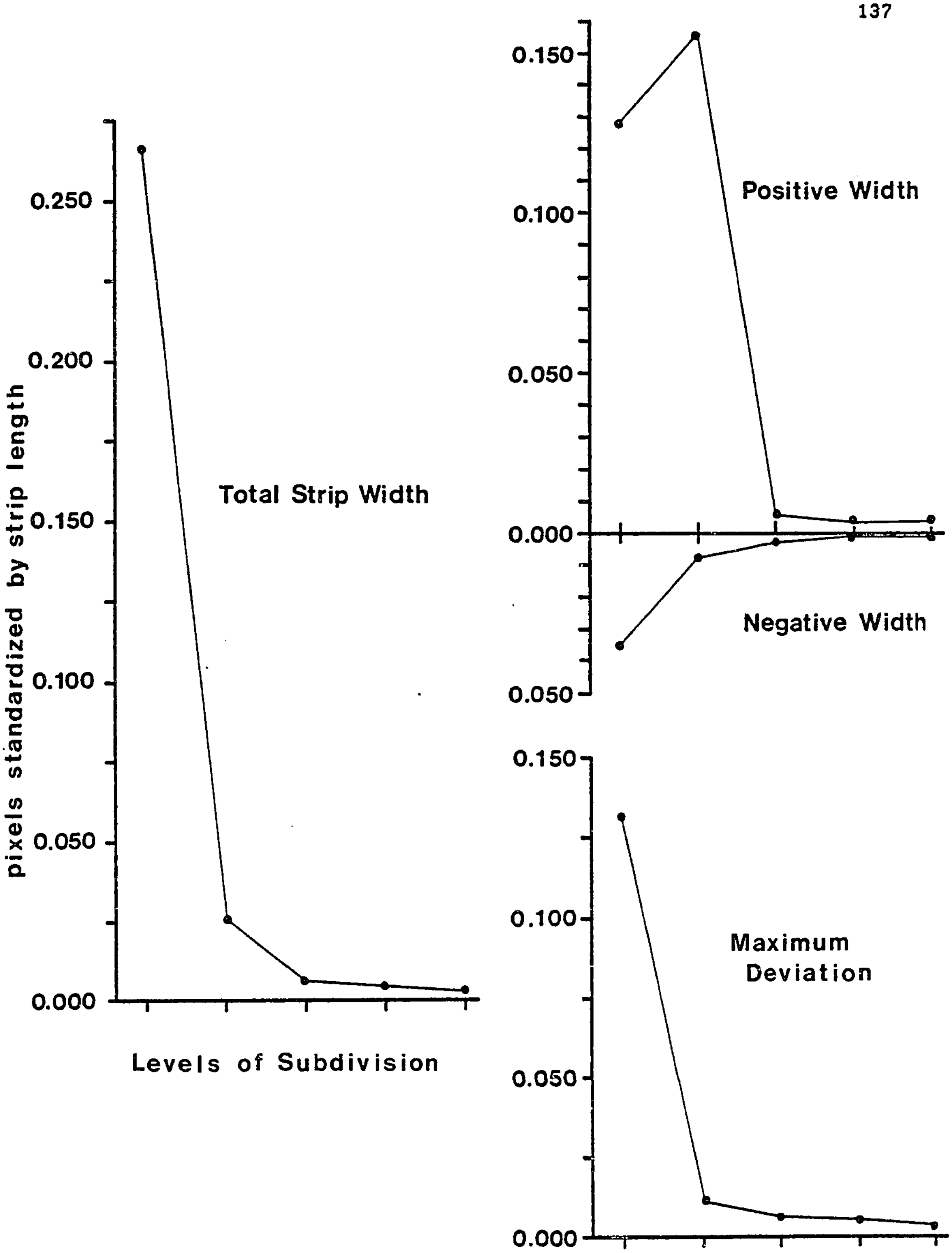


Figure 5.1
VARIANCE BETWEEN LINES
FOR MEASURES OF STRIP WIDTH

The other grouping of variance comparisons (Figure 5.2) are similar in that all are measuring displacements from the anchor line. This group includes monotonicity, error variance, and the segmentation parameter, none of which display similar patterns. Interesting to note are the apparent similarities between the error variance plot and the previous plots for strip width. Also notable is the low variance for the segmentation parameter at finer resolutions; this is interesting given the diversity in mean values and coefficients of variation for this parameter. Most likely, the segmentation parameter will not prove powerful as a discriminating agent, at finer resolutions.

Monotonicity presents another interesting variance pattern here. The first peak in this plot is probably due to the high value for SIOUX's mean monotonicity; and the final rise may be caused by the high value for HUDBY. It may be that monotonicity provides strong distinctions between these two lines, at certain levels of resolution.

In answer to the first question posed in this analysis, one can hypothesize that patterns between lines do appear to exist, most specifically for measures related to strip width and to segmentation. It is possible that the graphic structures of the four lines are most comparable in terms of cross-sectional and longitudinal symmetry, as these are measured by parameters whose variance drops consistently with further subdivisions of the lines.

The second question to be addressed concerns the redundancy of the parameters, and the extent to which the parameters overlap in describing graphic structure. This is an important question which must be considered prior to the discriminant analysis, to determine which parameters (if any)

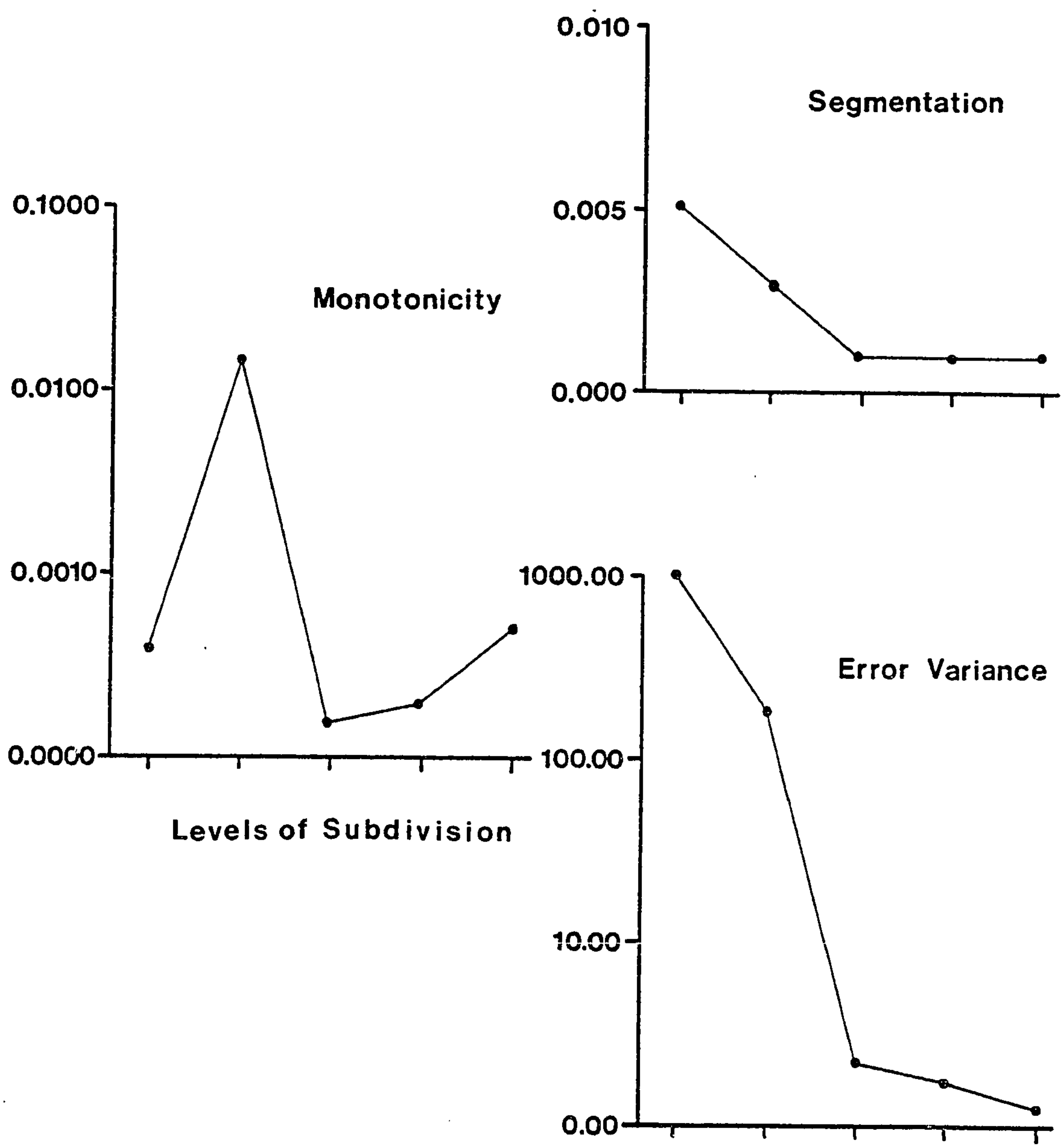


Figure 5.2
VARIANCE BETWEEN LINES FOR MEASURES OF DISPLACEMENT

may be discarded, and which should be included for maximum discriminating power. Correlations have been computed, to compare raw parameter values measured for each band. Computations for each level of subdivision display three distinct categories of parameter relationships. First are those parameters which correlate significantly at all five levels of subdivision; next are the parameters whose correlations are significant at some of the levels; and finally those parameters which do not correlate significantly at any level of subdivision. Results for significant correlations are displayed in Table 5.1; the threshold for inclusion is a significance of .01.

As anticipated from similarities in the variance plots, the error variance parameter is correlated with width and maximum deviation at all levels of resolution, with consistently high significance levels. Error variance is also correlated with negative and positive strip widths, but these relationships are not as strong, and they are not always significant. The two partial strip width measures are somewhat correlated, but this relationship is more likely caused by the fact that both parameters approach zero at finer resolutions. None of the other parameters display correlations which are significant.

From these analyses, one can conclude that parameters related to strip width directly (as in the case of maximum deviation and partial strip widths) and indirectly (as in the case of error variance) are redundant, to varying degrees. Because of the consistent overlap between error variance, overall strip width, and maximum deviation, the discriminant analysis will include only error variance. Positive and negative strip width will each be included, as their respective correlations are not consistently high from one level to the next.

TABLE 5.1
CORRELATIONS BETWEEN LOW ORDER PARAMETERS

Group 1 -- All Levels Significant

Parameters	Level 1 N=32	Level 2 N=64	Level 3 N=109	Level 4 N=157	Level 5 N=195
width by max.	r= .979	r= .979	r= .958	r= .970	r= .968
deviation	=.00001	=.00001	=.00001	=.00001	=.00001
error variance	r= .637	r= .613	r= .588	r= .710	r= .690
by width	=.00009	=.00001	=.00001	=.00001	=.00001
error variance	r= .562	r= .625	r= .581	r= .676	r= .722
by deviation	=.00101	=.00001	=.00001	=.00001	=.00001

Group 2 -- Some Levels Significant

Parameters	Level 1 N=32	Level 2 N=64	Level 3 N=109	Level 4 N=157	Level 5 N=195
error variance		r= .395	r= .305	r= .580	r= .671
by positive width		=.001	=.001	=.00001	=.00001
error variance	r=-.603	r=-.446	r=-.445	r=-.319	
by negative width	=.0002	=.0002	=.0001	=.00005	
positive by	r=-.694			r= .229	r= .351
negative width	=.00002			=.003	=.00001

NOTE: Numbers subscripted to each column heading refer to the number of cases included in each correlation. Non-significant correlations are not included in the table.

Distinctions Between Categories of Lines

Discriminant analysis involves a linear combination of variables which are constructed in order to account for a maximum amount of variation between categories of a dependent variable. Here, the variables to be combined include the five parameters of low order structure; and the dependant variable is made of four categories, from the four sample cartographic lines. Five discriminant analyses have been performed, one for each level of resolution.

The purpose of the analysis is to ascertain the degree to which the low order parameters can be relied upon to discriminate between categories of line structure. Each sample cartographic line is assumed to represent a different structural category, ' and this is based on differences interpreted from the structure signatures presented in the previous chapter. All parameters will be included in the analysis, except those which were discarded during the correlation analysis. Only significant bands have been included in the analysis.

One can presume that the discriminating ability of all the remaining parameters will drop with increasing levels of subdivision: although more bands are included at finer resolutions, the parameter values drop towards zero. Indeed, at the finest possible resolution, each band would theoretically surround only two points; monotonicity, error variance, and strip width would equal zero, and the low order structure for all lines would be identical. As a result, the accuracy of classifications of bands based on the discriminant functions will be meaningful only within a finite range of resolutions.

Stepwise entry of variables has been included, to better understand which parameters are the most powerful discriminators at each level of resolution. Table 5.2 displays the order in which parameters entered the analysis. The threshold entry value used in this analysis was an F-to-Enter of 1.000: all parameters which have entered show significant entry values ($\alpha = 0.05$), with the exception of error variance in Level 4. This would indicate that parameters related to bandwidth are overall most important, after the first subdivision. It is surprising that these parameters are consistently the first to enter, as the variance comparisons between lines had indicated that monotonicity should provide stronger discriminating power than band width, at several levels of resolution.

TABLE 5.2
STEPWISE ENTRY OF PARAMETERS INTO DISCRIMINANT ANALYSIS

LEVEL	PARAMETERS ENTERED	F-TO-ENTER	SIGNIFICANCE
1	monotonicity	3.129	.041
	error variance	2.598	.028
2	negative width	4.900	.004
	positive width	3.459	.004
	error variance	2.662	.007
3	positive width	2.988	.034
	negative width	2.549	.021
4	error variance	2.430	067
	negative width	2.189	044
	positive width	2.477	009
	monotonicity	2.168	013
5	negative width	3 477	.017
	positive width	2.686	.015
	error variance	2.665	.005

Table 5.3 displays statistical summaries for each discriminant function constructed during the analysis. The eigenvalues displayed here express the relative ability of each function to discriminate between lines: summing the eigenvalues provides a relative measure of the total amount of variance in the set of discriminating variables. In the next column of the table, each eigenvalue is reexpressed as a percentage of this variance, accounted for by the respective discriminant function. Notice the sharp drop in the eigenvalue sum at Level 3, indicating a marked decrease in the discriminating ability of the parameters included at this point. From the previous table, one can see that of the five parameters available, only the positive and negative widths have entered. It is likely, given the drop in eigenvalues, that the accuracy of classifying bands into line categories will also drop at Level 3; but this aspect will be discussed further on.

The canonical correlation figures in Table 5.3 measure the association between each discriminating function and the set of variables (listed in the final column of the table) used to construct it. The squared correlation values indicate the proportion of variance in the function which is explained by the categories of lines.

The canonical values can be seen to drop in Level 3 as well. What this implies is that the linear combinations of parameters are accounting for less and less of the variation between categories. In part this has to do with the fact that the parameters used to construct the functions are somewhat correlated: because the parameters are not independant, the discriminating power of the functions is reduced. Nonetheless, the significance level of these functions remains better than 0.05, until the fifth level of resolution. Here, the extreme narrowness of bands for all categories of lines inhibits the distinguishing capability of all five parameters. As predicted at the beginning of this analysis, there is a limit to the meaningful distinction of lines by low order structure, and this limit is dependant on the resolution of the measured line.

TABLE 5.3
STATISTICAL SUMMARIES FOR THE DISCRIMINANT FUNCTIONS

Level	Discriminant Function	Eigen Values	% Variance Explained	Canonical Correlations	Discriminant Function Coefficients > 0.70
1	1(.027)	.3510	60.48	.510	monotonicity
	2(.050)	.2293	39.52	.432	error var, monotonicity
		.5803			
2	1(.007)	.2555	61.44	.451	negative width
	2(.050)	.1076	25.86	.312	positive width
	3(.080)	.0529	12.71	.224	*
		.4159			
3	1(.020)	.0953	64.59	.295	positive width
	2(.070)	.0522	35.41	.223	*
		.1475			
4	1(.013)	.0768	44.40	.267	error var, pos/neg width
	2(.027)	.0587	33.91	.235	error variance
	3(.061)	.0376	21.70	.190	*
		.1731			
5	1(.005)	.1019	79.15	.304	neg/pos width, error var
	2(.280)	.0181	14.05	.133	*
	3(.198)	.0083	6.80	.093	*
		.1287			

* No coefficients are listed for non-significant functions.

NOTE: Numbers in parentheses refer to significance levels associated with each function.

Classification tables constructed in the discriminant analysis are presented in Table 5.4. This table presents the accuracy with which bands are correctly assigned to a particular line category by the set of discriminant functions. As predicted, the total accuracy tends to drop with increasing resolution. This was previously explained as due in part to the convergence to zero of bandwidth parameters which act as primary discriminators; it may also be caused by correlations existing within the parameter set.

TABLE 5.4
PERCENTAGE OF BANDS CLASSIFIED CORRECTLY

LEVEL	SIOUX	HUDBY	FJORD	TEXAS	TOTAL
1	62.5	75.0	12.5	37.5	46.88
2	81.3	43.8	37.5	37.5	50.00
3	68.0	42.3	6.3	21.4	33.03
4	61.8	28.6	33.6	25.0	36.31
5	53.8	50.0	23.8	13.7	32.82

The Level 4 increase in classification accuracy is probably due to the inclusion of error variance as a discriminating variable (recall that it was not included at Level 3); in the statistical summary table it is shown that this parameter accounts for sufficient additional variance to provide a second significant discriminating function.

Perhaps more interesting than the total number of accurate assignments are the breakdowns of classsification accuracy by categories of line. The SIOUX file maintains a fairly accurate assignment rate, as does HUDBY, with the exception of assignments made in Level 4. It is presumed that these lines are more accurately identified overall because they contain irregularities of roughly one magnitude; this aspect was discussed at length in the section on structure signatures. The majority of misclassified lines have been assigned to the SIOUX file, at all levels. Misclassifications between FJORD and TEXAS also occur, but mostly at lower levels of resolution, while assignments to the SIOUX category increase with each level of resolution.

One may conclude from this discriminant analysis that low order parameters have provided clear identification of the SIOUX bands, and also provided good distinctions for HUDBY (with the exception of Level 4). The classification accuracy is lower oeverall for FJORD and for TEXAS categories, but even here, the assignment accuracy remains better than chance (25%) in half of the analyses. It is also possible to conclude that at finer resolutions, more bands in all categories of lines contain low order structures which are similar to the SIOUX category, because increasing numbers of bands are assigned to SIOUX at each level of resolution (these figures are not shown in the table). This would imply that the low order structure of SIOUX is the least complex of the four line categories; this is also borne out by the early rectification of many SIOUX bands during the subdivision procedure discussed in Chapter 4.

The low values for overall classification accuracy suggest that while some categories of line features are easily identified, low order parameters are not by themselves sufficient to distinguish between categories of line features at all resolutions. A larger sample of lines, subdivided across a larger range of resolutions, may serve to improve the results of this discrimination task; it may also be that other measures of low order structure exist. However, these results seem adequate to demonstrate that low order structures do differ between categories of cartographic lines, and that they may serve as guidelines by which to identify and distinguish those categories.

It is also reasonable to presume that there is a higher order of graphic detail evident in the four sample lines, which may be useful in distinguishing between categories. The high order structures include details of the segment-to-segment relations, and are visually recognizable as features such as fjorded valleys, resevoirs, bays, and sandspits. At early levels of subdivision, these features have not yet become completely rectified. They may be contained within a single band, which (as shown in the case of TEXAS) can be identified by a high error variance and/or monotonicity which reflect the internal convolutions. High order structures have been shown to inflate the variance values for several low order structure signatures; they provide one possible explanation for higher accuracy values in the initial levels of the discriminant analysis classification table as well. The next section will focus on several aspects of high order structure, and on how identification of various categories of cartographic lines may be improved.

Aspects of Measuring High Order Structure

As previously discussed, the high order structure of a line refers to the topology of line segments, meaning how these segments fit together to form features along the line. Geographically speaking, the four sample lines display a variety of high order structures: FJORD has its fjorded valleys, TEXAS has river mouth bays and barrier spit beaches; SIOUX, on the other hand, contains dendritic river channels and a dammed lake. Without knowing the precise geographic location of any of these scenes, one can easily recognize the features which characterize each line's high order structure.

Structurally speaking, as before, these feature descriptions take on a somewhat different tone. A line is composed of a trend line and certain features which bifurcate from the trend; bifurcations occur with varying frequency for each category; the features (or rather the bifurcations) will be characterized by having a certain range of widths, lengths, and orientations, relative to the trend line. Also, the line may be expected to have a high order error variance, which measures the rate at which irregularities in segments are rectified: this can be parameterized as a fractal dimension. Taken together, these five parameters (frequency, length, width, orientation, and fractal dimension) can be used to measure high order structures for the cartographic lines. The parameters may also be used to identify similar line features on different cartographic lines, and this will be demonstrated for one of the line categories.

Limitations exist in the measurement of high order structure. Because only one sample line has been collected for each category, its representativeness for its category cannot be justified. Thus, there is no certainty that the FJORD file provides a best representation of features found in all fjords, nor that the river channels in SIOUX are the best possible sampling of dendritic features for its category. The problem of small sample size was encountered for the measurement of low order structures as well; and as before, this means that the technique will be described and illustrated, with the reader's understanding that the structure signatures which result are at best only a first approximation.

Another limitation exists, which was not encountered with the measurement of low order structures. It poses a more difficult problem, namely, the problem of locating the specific places where features begin and end along the line. Structural measurement of features requires that they first be isolated from the trend line. While it is clear that most people will identify the same kind of indentations in the FJORD line as being river valley features, a request to locate or identify the point(s) along the trend line at which a bifurcation begins and ends will probably not result in the same strong consensus. People will tend to break out features from the trend line in a variety of places, as shown in the top of Figure 5.3. Of course, the distinction in feature identification illustrated here should not make much difference, if it is carried out consistently from one line to another, because whatever bias is introduced will accrue systematically to all sample lines.

But the trend-feature separation is hierarchic in concept: in closing off a feature from the trend, a new trend is created, the trend line of the feature, which may itself have smaller features and bifurcations. In this way, high order structure can be seen to be recursive. Realize further that for some categories of lines, the recursion may extend for more levels than for other line categories, depending on the complexity of the line's high order structure. The problem is illustrated in the bottom half of Figure 5.3, in which the same feature is broken out in two ways from its trend line. This example can be seen as one feature which has two levels of recursive bifurcation, or as a trend line from which two features bifurcate at the same level of recursion.

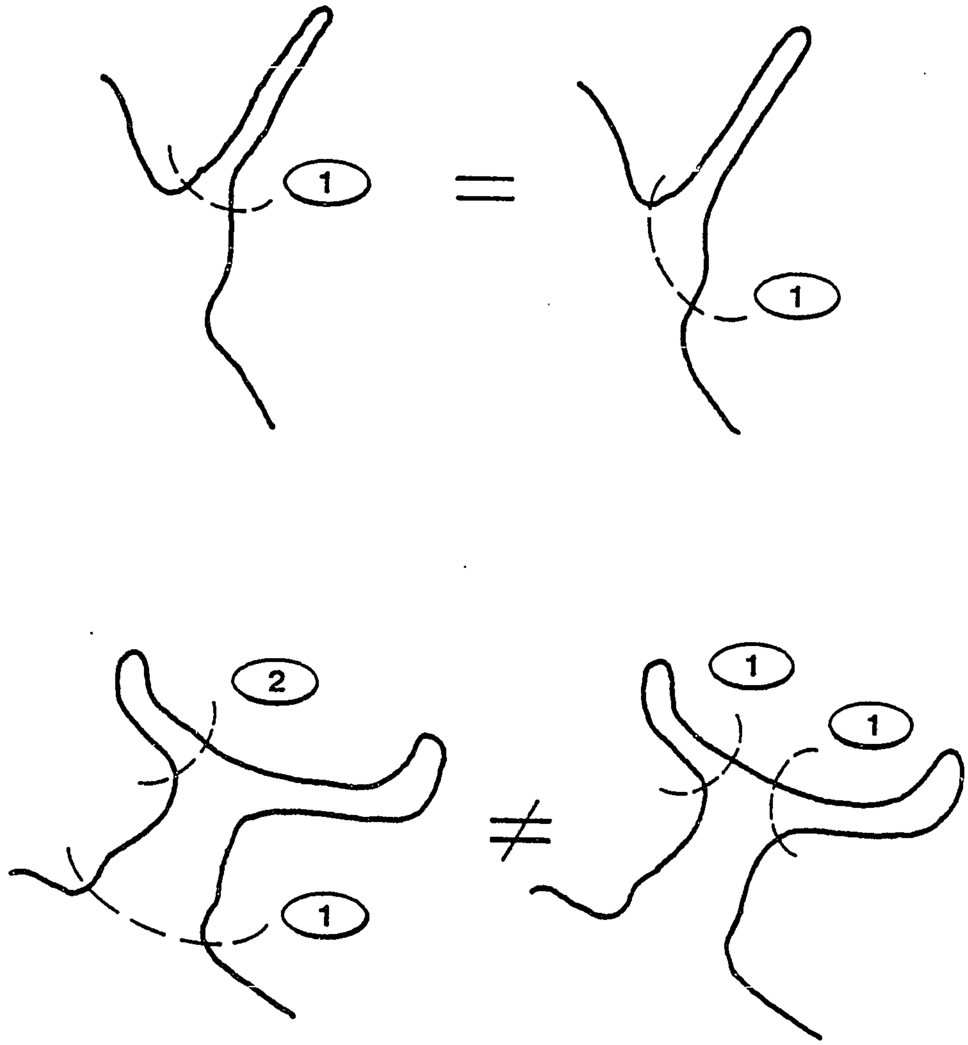


Figure 5.3
EQUIVALENCE OF
HIGH ORDER STRUCTURE

Plainly, the decision to bifurcate at a given level cannot be made arbitrarily in this case, as it may confound the structural measurements between categories of lines. This means that significant inconsistencies may be introduced to some trend lines and/or features, but not to others. And there is no way to verify the consistency after the fact, because the level of recursion will determine the measurement of high order structure. In the vernacular, this situation is called a 'Catch-22'.

Of course, a solution could be reached by means of a perceptual test, in which many people are asked to bifurcate the four sample lines, and some average location identified for each feature to be isolated. But this is the other half of the 'Catch-22', because one must be able to express a definition of trend line and of feature in order for people to locate a boundary between the two; and that expression is not yet defined. Enough questions are being raised here for a research project, a dissertation in its own right, well beyond the scope of the present study. It would be a digression at this point to focus on trend and feature distinctions for this particular small sample, especially when the current focus is on explanations of a structural measurement technique, rather than on explanations of specific structures.

For this reason, a trend line will be arbitrarally chosen, by the author's induction of the visual expectations associated with features for each category of line. The obvious disadvantage of this affects the confidence which can be attached to any absolute interpretations of structure, as mentioned before. The advantage is the provision of an anchor on which to base an exmination of the high order parameters: the parameters can be derived, and used to produce first approximations of structure signatures. And while numerical distinctions such as discriminant analysis cannot be relied upon for meaningful identifications of lines, the high order structure parameters can be evaluated by a different kind of identification task, more closely related to line generalization. Before discussing the task, however, the measurement technique will be explained.

Parameters of High Order Structure

The first measure to be applied to the four lines is akin to an error variance for segment-to-segment relationships. This is computed as a fractal dimension. As previously discussed, this parameter indicates a relative measure of how the total line length increases with decreasing average segment length. Of all the parameters of high order structure, the fractal dimension most clearly overlaps with the low order parameters, in that it can be measured directly from the strip tree data structure. It is also the only high order parameter which does not require distinctions between trend line and feature. Therefore, the parameter can be statistically verified after measurements are completed, to provide numeric distinctions between lines. For these reasons, the fractal dimension will be discussed separately from the other high order parameters.

Fractal Dimension of the Cartographic Lines

Mandelbrot's (1967,1977,1982) fractal dimension measures a rate of increase in length with increasing resolution. Recall from Chapter 2 that it is a ratio between the number of straight line segments contained in a figure, and the unit increase in total length of each segment; the ratio is expressed in logarithms. Using the strip tree, fractal dimension can be computed by relating the length of band anchor lines to the number of bands introduced at each level.

A problem arises in working through Mandelbrot's formulas, however, as the strip trees contain no constant increase in length to insert into the ratio. The length of each sample line does increase with increasing levels of resolution, and the rate of increase differs between lines; but the magnitude of increase differs within each anchor line, from one level to the next. The FJORD strip tree provides clear illustration of the problem (Table 5.5).

TABLE 5.5
RATES OF INCREASE IN ANCHOR LINE LENGTH
(FJORD FILE)

LEVEL	NUMBER OF NEW BANDS	TOTAL LENGTH	ABSOLUTE INCREASE	PERCENTAGE INCREASE
1	8	484.39		
2	8	734.00	249.61	51.53
3	14	958.71	224.71	30.61
4	18	1100.69	141.98	14.81
5	15	1196.40	95.71	8.70

Neither the number of bands nor the total length show a constant proportional increase; and consequently, the fractal dimension of the lines changes in ratio from one level of resolution to the next, with somewhat disquieting results. Here are the computations which Mandelbrot would require, for the FJORD file:

LEVEL	(N) NUMBER OF NEW BANDS	(R) PERCENTAGE INCREASE	LN(N)	LN(1/R)	FRACTAL DIMENSION
2	8	51.53	2.08	1.94	1.07
3	14	30.61	2.64	1.18	2.24
4	18	14.81	2.89	1.91	1.51
5	15	8.70	2.71	2.44	1.45

The fractal values which result are quite extraordinary. They imply that after Level 2, this line is becoming less complicated with increasing resolution, which is precisely opposite to all of the indications given by the signatures of low order structure. Secondly, the change in fractal dimension is much too drastic for such a small range of resolutions, as an exponential ratio. It might be interesting to consider these as a set of sampled fractal dimensions, and average the values computed at each level of resolution. One might speculate that this would provide an overall dimensionality for the line (here its value would equal 1.56); but no justification for this course of action appears in the literature, to this author's knowledge.

Mandelbrot implies that the fractal dimension should be constant across all sampled levels of resolution. He is ambiguous in describing his applications of the formulas to coastline features, saying only that coastlines display a 'statistical' rather than a 'precise' self-similarity; but he does suggest that fractal dimensions can be computed for coastlines having non-consistent increases in length (Mandelbrot, 1982). It would appear from this example that some modifications are required to manipulate the formulas, but Mandelbrot does not discuss the specific computation schemes he applies.

However, fractal dimension is computable from another, more indirect approach, which is distinctly statistical in nature. Mandelbrot's mathematics are based on Lewis Richardson's (1961) work relating coastline length to the magnitude of the units measuring length. As discussed in Chapter 2, Richardson found a logarithmic relationship between the two, and plotted these for various coastlines he had measured (refer back to Figure

2.5). Richardson made the comment that the slope of his plots indicates the degree of visual irregularity of the coast. By Mandelbrot's derivation, this slope (α) can be used to compute a fractal dimension (D) by:

$$\alpha = 1 - D.$$

Because of the difficulties encountered with Mandelbrot's formulas, Richardson's method has been used to compute fractal dimensions. First, an average band length is computed from the summed length of all anchor lines in each level of resolution, to approximate a unit length. Next, natural logarithms have been computed for total length and for average length; these computations are displayed in Appendix E. Finally, a least squares regression line has been fitted to each set of sample data, in an attempt to replicate Richardson's results.

TABLE 5.6
RESULTS OF LINEAR REGRESSION

LINE	R^2	INTERCEPT	SLOPE	SIGNIFICANCE	FRACTAL DIMENSION
SIOUX	.99	7.12	-.0826	.00002	1.0826
HUDBY	.99	8.35	-.6527	.0007	1.6527
FJORD	.93	9.36	-.7481	.0077	1.7481
TEXAS	.88	10.48	-1.0185	.0188	2.0185

Results of the regression are displayed in Table 5.6, which displays surprisingly high significance levels for such small data sets (there are five observations in each case -- one data point for each of five levels of resolution). The regression lines are also plotted as a structure signature in Figure 5.4. When this figure is compared with Richardson's

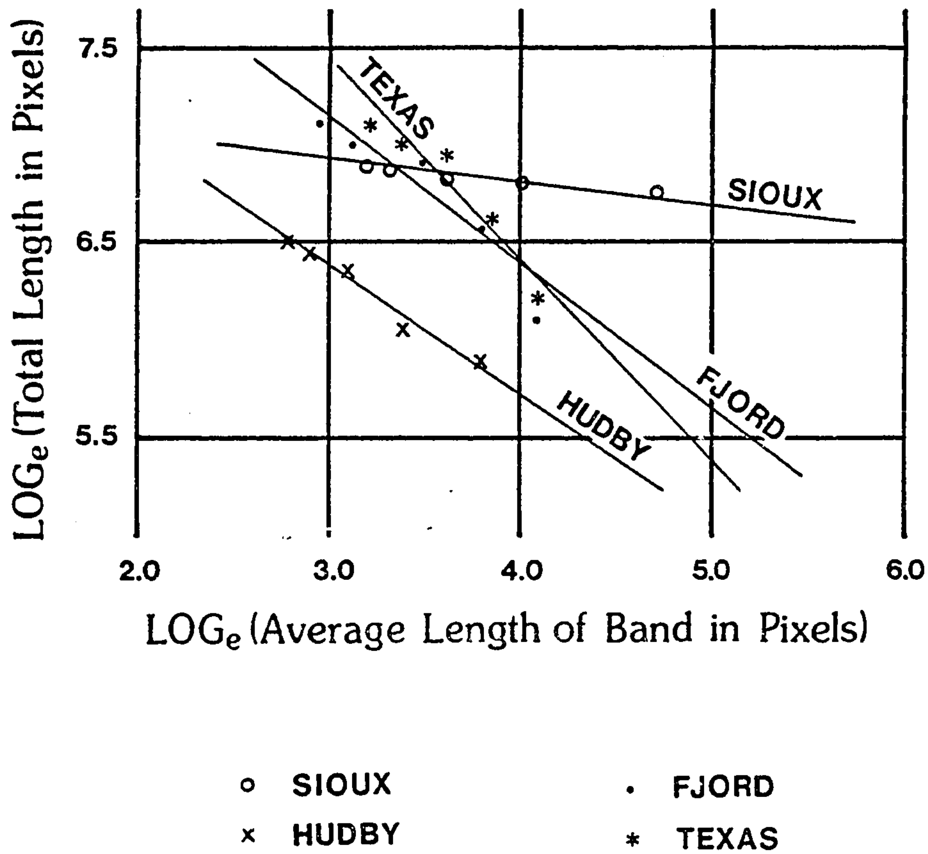


Figure 5.4
RATE OF INCREASE IN LINE LENGTH
(after Richardson, 1961)

results as plotted in Figure 2.5, it is easy to see that the magnitude of slopes for these lines are more closely attuned to the slopes found by Richardson in his study. Also of interest is the similarity of slopes for FJORD and HUDBY, the close proximity of FJORD to TEXAS, and the extreme singularity of SIOUX in position and in slope.

Interpretations of the fractal dimension are clearly related to concepts of high order structure. The nearly equivalent slopes of FJORD and HUDBY imply that even though the magnitude of irregularities (and thus of total line length) is much higher for FJORD, the relative rate at which new irregularities become resolved is similar for these two lines. This was also implied during study of the low order structure: recall the strictly increasing monotonicity plot for HUDBY, and the discussion of oscillations whose frequency increased at finer resolutions, in comparison to the discussions of increasing irregularity with increasing resolution for FJORD.

The plots for TEXAS and FJORD lie close together within this range of resolutions, implying that the visual irregularities of their features may be quite similar, at certain scales of display. However, the fractal dimension of TEXAS is greater than 2.0000. A topologic dimension of 2 is assigned to planar surfaces, and a fractal dimension of this magnitude implies the extreme sinuosities of a space-filling curve, a curve which would completly fill the plane on which it lies if it were represented at some maximum (theoretical) resolution. The high order structure of TEXAS appears to be much more complex than is the structure of any other line.

The slope of the SIOUX line is almost horizontal, reflecting only minimal changes in line length with increasing resolution. This is also in keeping with interpretations made for low order structure of this line. Not only did this line display a large number of rectified bands early in the subdivision process, but the mean parameters for error variance and for strip width both dropped quickly to zero, and coefficients of variation were relatively low at all levels. Structurally speaking, this line displays a certain homogeneity from one level of resolution to the next.

The fractal dimension parameter has been shown to meaningfully distinguish between the four sample lines, by means of linear regression, and by graphic means, as a structure signature. This is the only high order parameter for which statistical validation is possible, due to the nature of the sampling process for the four cartographic lines. Before elaborating on the other parameters, the cartographic lines must be broken into their separable components of trend line and features.

Thinning the Cartographic Lines

The inductive criteria used to isolate features from trend lines will be quite straightforward. First, a feature will be identified at locations where the original cartographic line either splits, or doubles back on itself. The intention is to provide the smoothest and most simple approximation of a trend line, and to isolate as many convolutions and irregularities as possible into features. One might think of the trend line as the smallest scale representation of the original line, a generalization to the 'n^{th}' degree. A feature is defined to have a width of at least one pixel on the Apple II screen: this provides for definition of single line features, such as river channels, as bifurcations which have a width of zero (the sides of the feature touch).

Features must also have a resolvable length, which can be defined by a series of one or more straightline segments, travelling along the longitudinal axis of the feature. The process used to derive these straightline segments will not be inductive; instead, a line-thinning algorithm will be applied to each cartographic feature in turn. This process will be discussed and illustrated.

The feature-trend definition is intended to be recursive, as mentioned above. The set of thinned segments along the axis of each feature may be considered as a sub-level trend line. At sub-levels, the trend lines may be seen to bifurcate in one of two ways. One type of bifurcation will be seen where the width of a feature shows a sudden significant change: this could be a narrowing, as in the case of a bay at a river mouth; or it could be a widening, as in the case of a dammed lake along a river channel. The second type will occur if a feature itself bifurcates, that is, if a single thinned line can no longer adequately represent the feature axis. At these locations a new level of recursion will be defined, with associated axis segments serving as trend lines. The process of bifurcation will be continued to the farthest extremes of all features resolved along a sample line.

Figures 5.5 - 5.8 display the first level of bifurcation for all sample lines. Dashed lines in the figure mark the next levels of bifurcation for various features. It should be made clear that the levels of recursion are not equivalent to the hierarchy of resolution produced for low order structures, as features of all sizes are included at every high order level. In the case of SIOUX, the largest order of river was used as

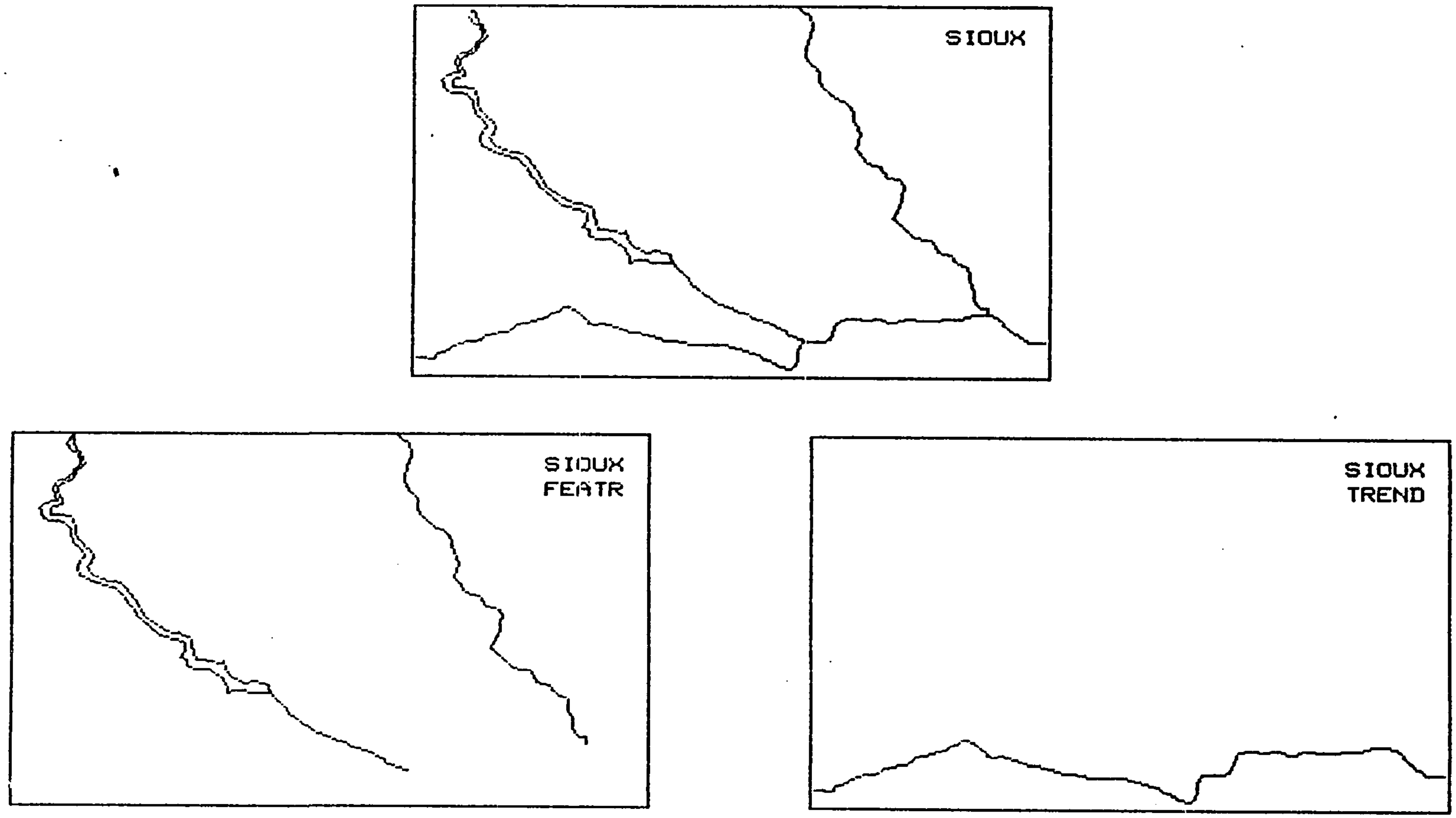


Figure 5.5
COMPONENTS OF SIOUX

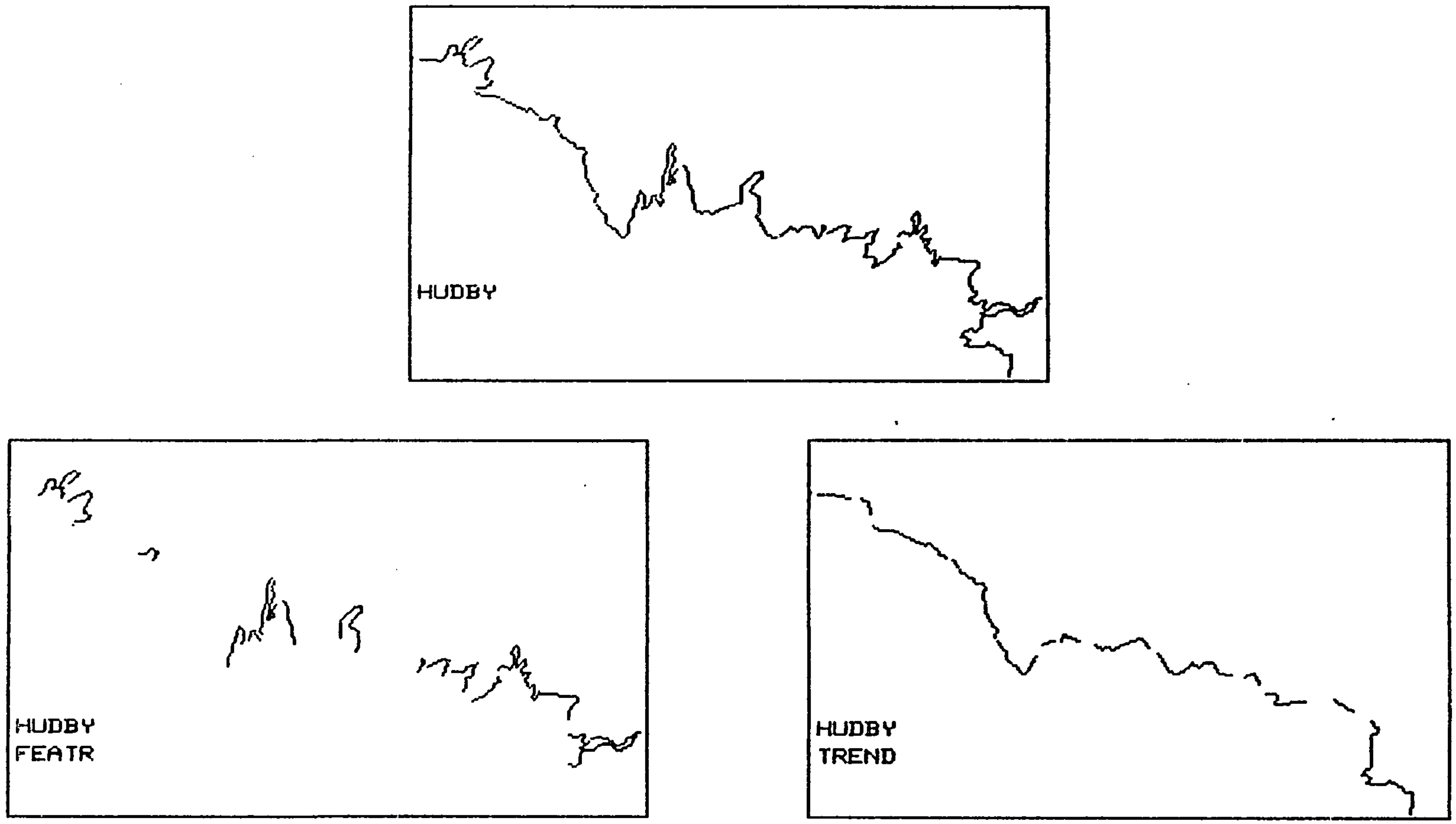


Figure 5.6
COMPONENTS OF HUDBY

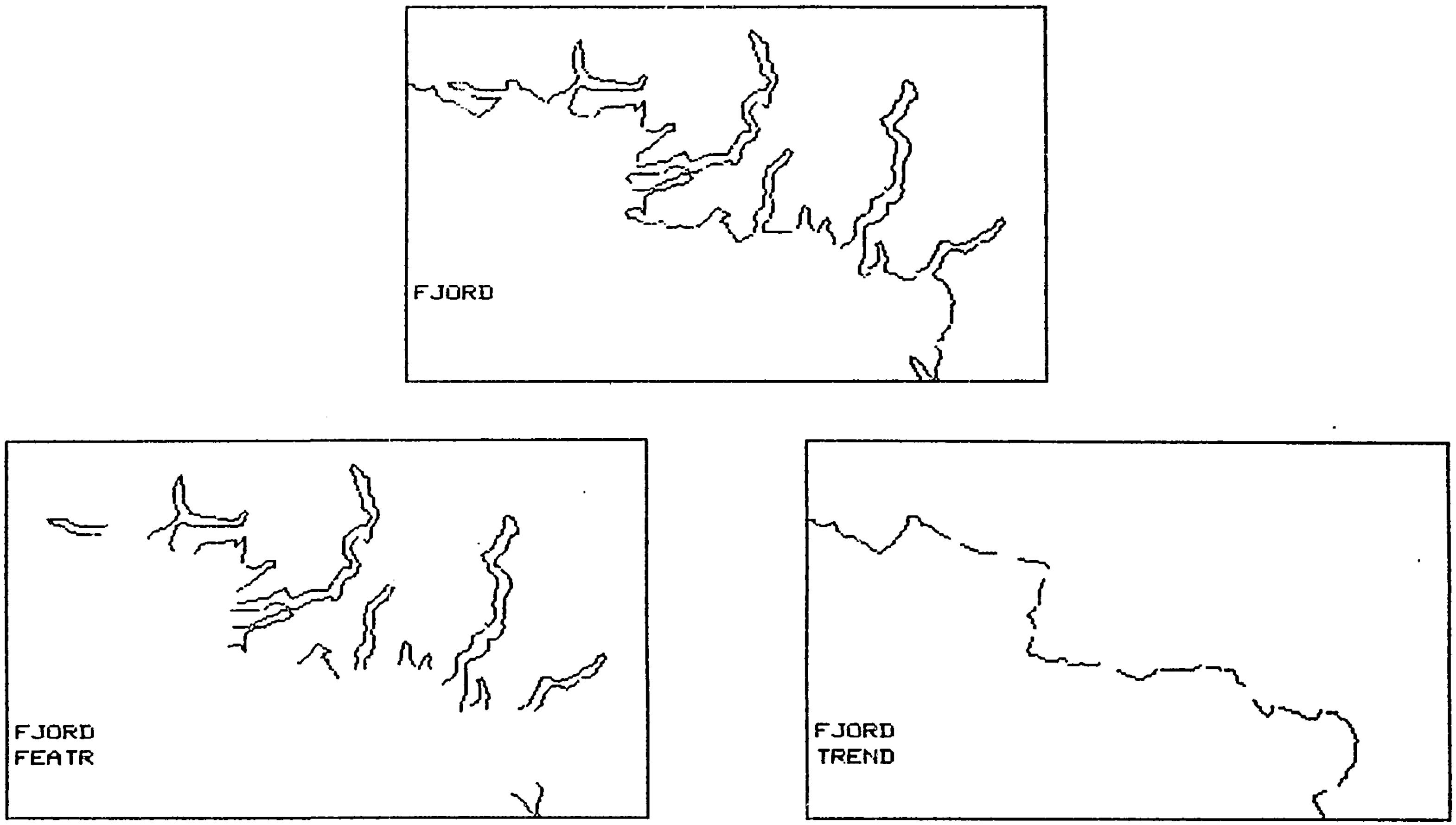


Figure 5.7
COMPONENTS OF FJORD

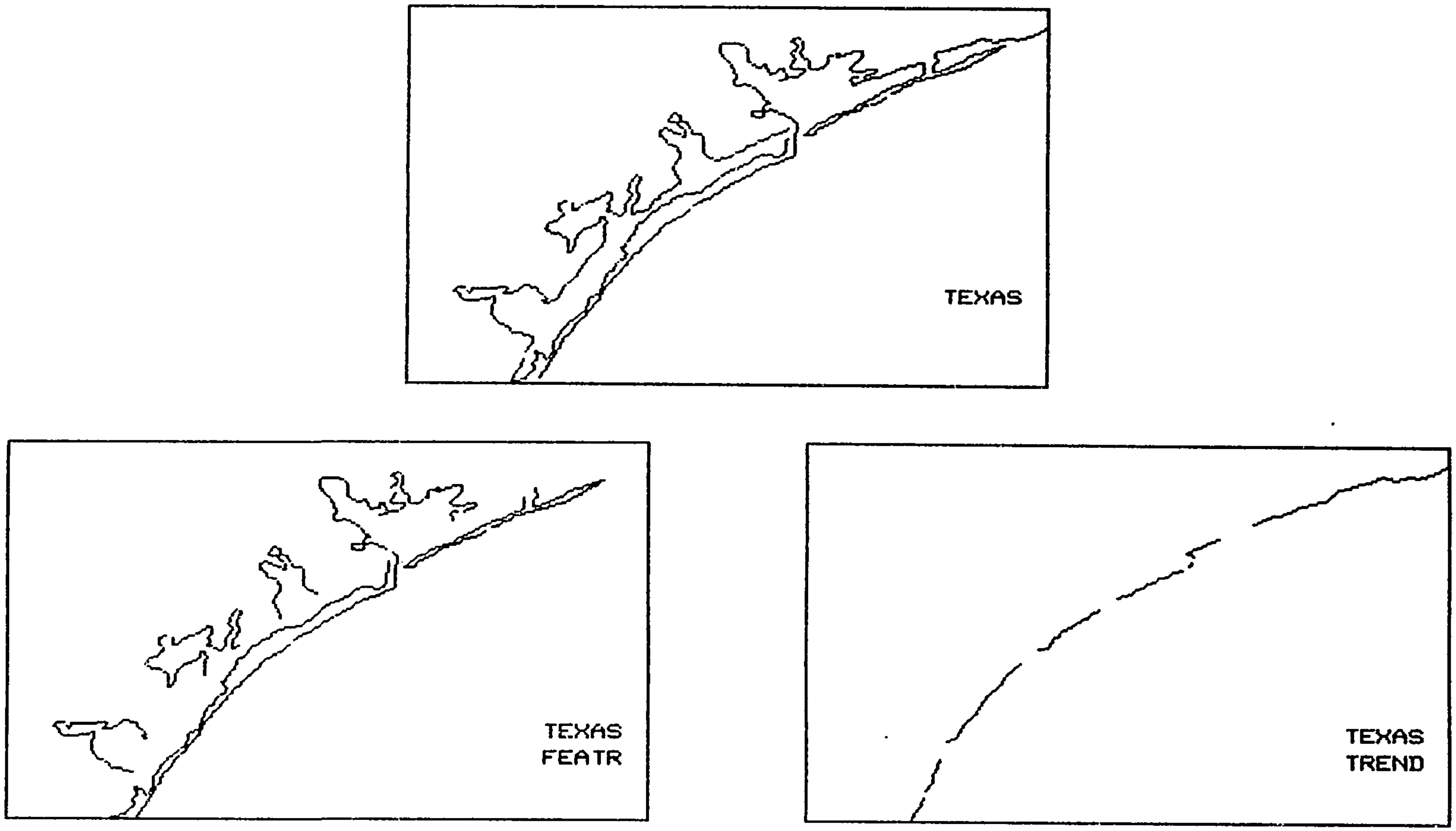


Figure 5.8
COMPONENTS OF TEXAS

trend line, with tributaries considered as first level features. In all other lines, the features were separated from the trend line according to the definition set forth above. Breaks in the trend line plots indicate the locations of feature bifurcations. Many of these features contain further bifurcations, which have been measured; but the first level of bifurcation serves to illustrate the process.

The reader may feel that the four trend lines may still contain irregularities definable as features (notable examples of this are the two large indentations in the FJORD trend). It could be argued in response that these two indentations are evidence not of sub-level features, but of super-level features, occurring at a level of resolution above this study's scope. Either argument is reasonable, but neither can be justified beyond pure inductive reasoning, which is the very basis for these original bifurcations. This is a limitation in the measurement process, as previously discussed. But the purpose of discussion here is not to describe the structure of a fjord, per se, so much as to distinguish it from the structure of a wide river mouth bay, or a barrier beach spit. In other words, the focus here is evaluation of the measurement technique about to be described, rather than on the specific content of the measurements which are made in the course of the evaluation. It stands to reason that if the high order measurement techniques are robust, then inconsistencies in feature definitions which do arise should improve the evaluation, by pointing out where a technique might be modified.

Once the bifurcation process is completed for all sublevels, two measurements can be made for features which have been identified. First is the frequency of bifurcations, the number of features which can be identified at each level. The second is which side of the trend line the bifurcations occur. As with the low order parameter called monotonicity, each trend line has a positive and a negative side. In certain lines, for example TEXAS, it may be useful to contrast the structure of features which occur on each side. Taken together, frequency and monotonicity constitute parameters of feature occurrence. It will be shown that these parameters are conceptually related to the fractal dimensions for these lines.

In order to complete the other high order measurements, namely length, width, and the angle of the feature bifurcation, the features must be reduced, or thinned, to their longitudinal axis. Line thinning algorithms are most commonly applied in image interpretation problems, as one method of edge enhancement, or in region measurement tasks, for boundary delineation. As a pattern recognition technique, thinning algorithms are most often oriented to raster geometries (Rosenfeld, 1975; Dyer and Rosenfeld, 1979). The features is this study are stored as vector representations, which is to date a more common data structure for cartographic lines. Furthermore, each feature has been defined as having a roughly constant width; when the width changes, a new level of feature is defined. Rosenfeld and Kak (1982) offer a simple algorithm for this special case of line thinning.

> Another method that can be used if S is an arc of constant thickness is to initiate two border-following processes at one end of S that traverse opposite sides of S. If the distance between the border followers gets significantly larger than the width, we stop one of them until the other catches up; thus they always remain approximately alongside one another. The midpoint of the line segment joining the border followers thus traces out a skeleton of S.
>
> Rosenfeld & Kak (1982), p.234

Figure 5.9 displays the thinned versions of all levels of features for the four lines. The twists and curves of the thinned lines occur at a short lag behind the twists and curves of the original feature; the length of the lag is a function of the distance threshold between the border followers. But a visual comparison between the features and the thinned feature plots should illustrate how strongly the algorithm reflects irregularities characteristic of the original arc.

Once thinned, the TEXAS features display a stronger resemblance to dendritic river patterns than did the original features plot. This provides a good example of how line thinning may be used to emphasize certain aspects of high order structure, which would not otherwise be graphically apparent. On the other hand, the barrier spits are visually similar to some river valleys, but do not appear to reflect the strong sinuosities of the long river valleys in FJORD, nor the resevoir in SIOUX, as was originally suggested; but numeric measurement can better evaluate the differences and similarities between features.

The numeric measurements which can be made for thinned features include the feature geometry, including the length of the thinned axis, and its orientation, the angle at which each sub-level bifurcates from its respective trend line. A third parameter measures the width of the original feature; this is computed during the line thinning task, by averaging a series of width measurements made while walking along the feature. Parameters for all measurements on each of the features are listed in Appendix F. Each type of structure parameter will be discussed in turn, beginning with frequency and monotonicity.

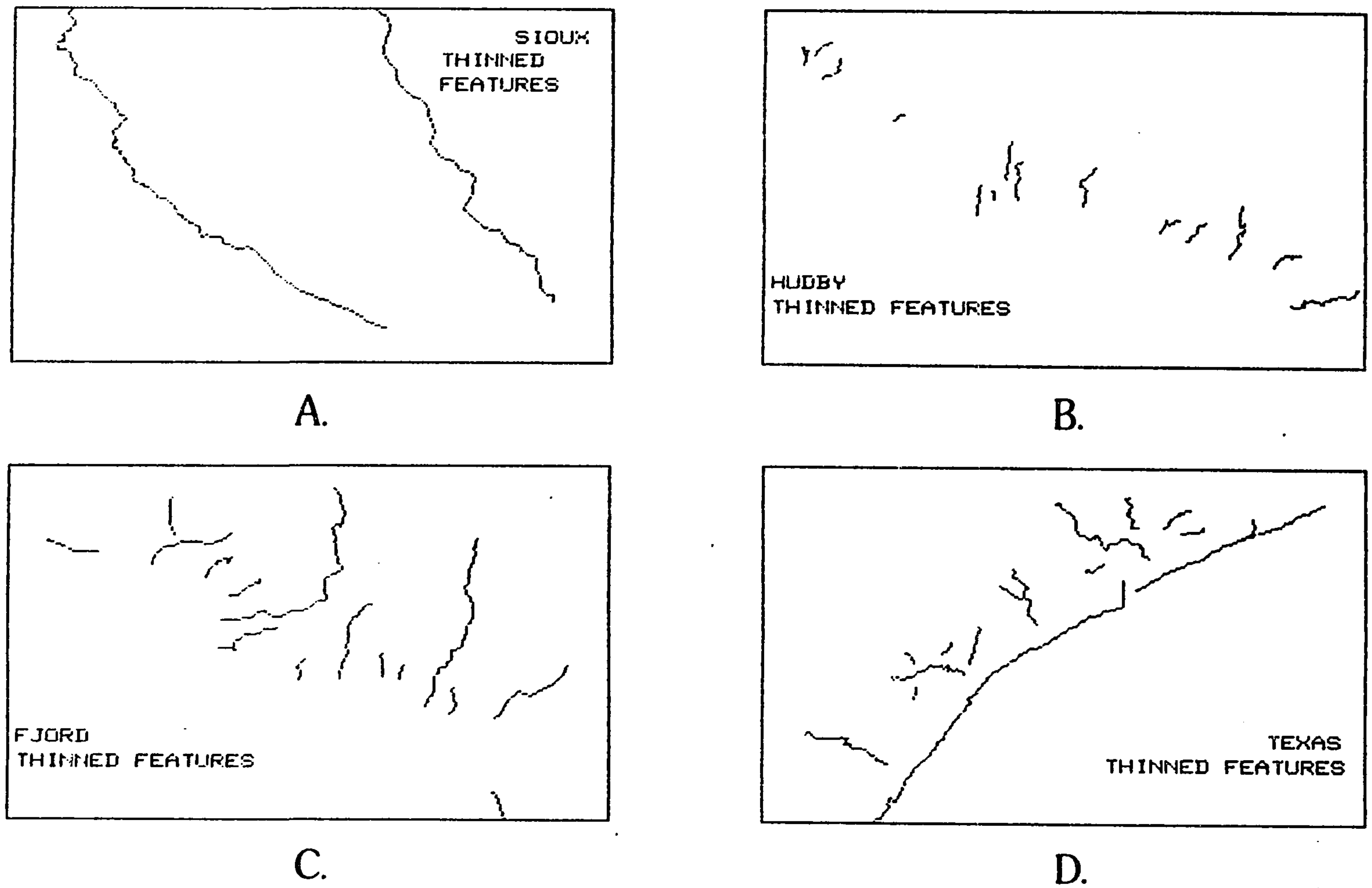


Figure 5.9
THINNED FEATURES

Parameters of Feature Occurrence

The frequency of feature bifurcations is illustrated in Figure 5.10, and includes a breakdown count of features identified at three levels of recursion. The structure signature for SIOUX displays the flattest curve, implying that the number of bifurcations does not change with increasing resolution. This implies that SIOUX is least complex of the four lines in terms of high order structure, mirroring the findings for fractal dimension previously discussed. HUDBY and FJORD display almost equivalent patterns of bifurcation frequency; it is TEXAS which displays the most anomolous pattern. Recall that the computations of fractal dimension displayed the same ordering of the lines. As an index of segment-to-segment relations, the fractal dimension also provides an indication of bifurcation complexity.

Other indications of this complexity of high order structures are also present. Earlier in this discussion, it was suggested that the high order structure might differ significantly on either side of this line, and such a discrepancy might cause erratic patterns in structure measurements. For this reason, the features for TEXAS will be measured in two parts: the 'landside', including the shallow and irregular river mouth bays, and the 'seaside', including the long and narrow barrier spits.

This treatment in two parts is presented as example of the second parameter of feature occurrence, namely, the monotonicity of high order structure. In Figure 5.11, high order parameters for all of the TEXAS measures are displayed, to illustrate how different are the two sides of this line in graphic space. On the 'landside', features tend to include more coordinates, and to be wider. The frequency plot reveals that the

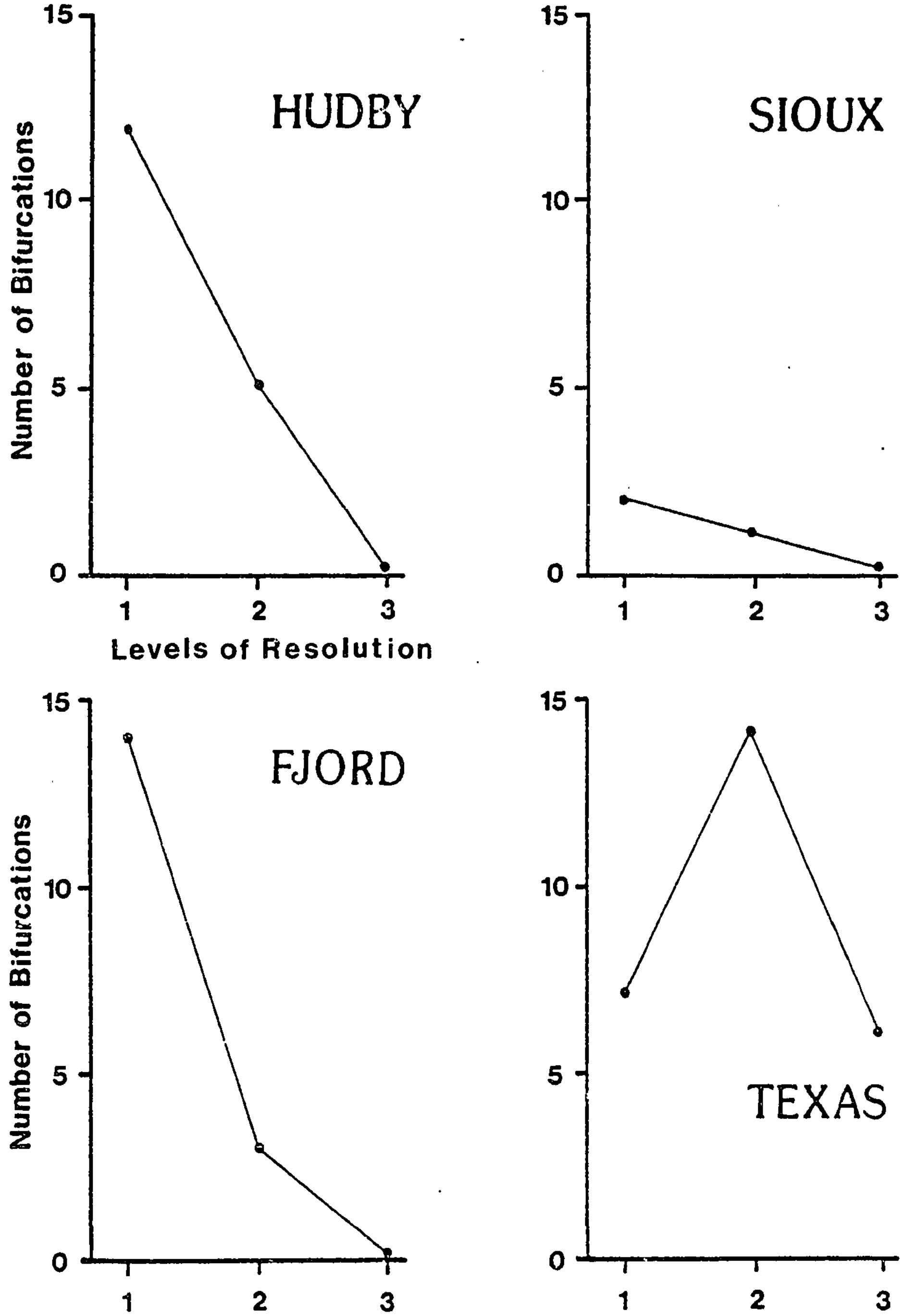


Figure 5.10
FREQUENCY OF BIFURCATION
High Order Structure Signatures

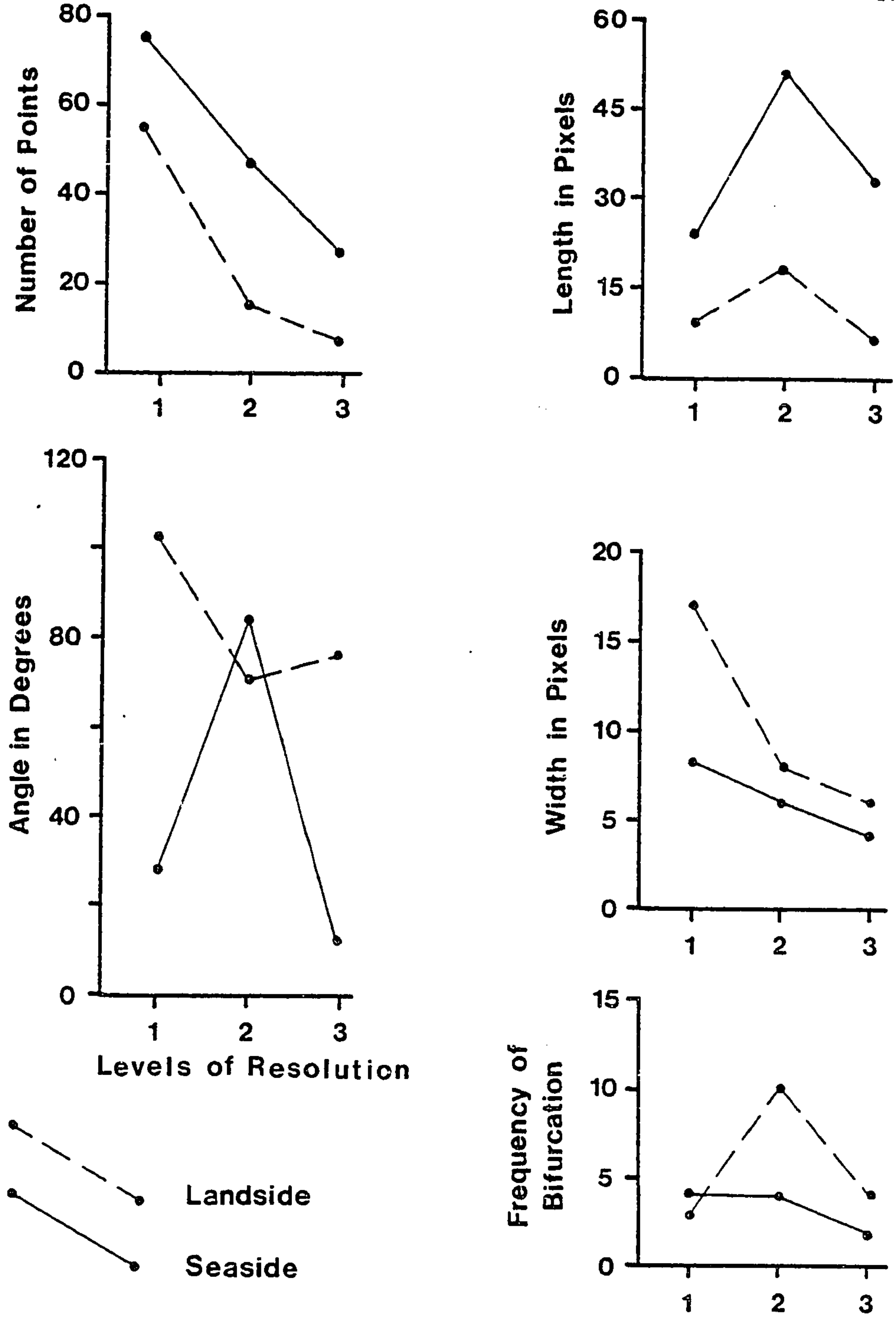


Figure 5.11
THE TWO SIDES OF TEXAS
High Order Structure Signatures

'seaside' has a simpler pattern of bifurcations, similar to that of SIOUX; the 'landside' frequency is a reiteration of the pattern of bifurcations displayed for TEXAS as a whole.

The two sides of the line tend to differ most in terms of length and width; overall, the 'seaside' will be shown to be more similar to HUDBY or to FJORD than to the 'landside' of TEXAS. The 'seaside' is composed of longer but more narrow features, and is dominated by the barrier beach spits. The two sides sides of this line appear most similar in that they both exhibit in increase in average feature length at the second level of bifurcation. It is important to realize that successive bifurcations do not necessarily imply smaller features, but rather features of decreasing complexity, although by the third level, values for all high order parameters are decreasing. (Orientation is the exception, but an angle of zero degrees does not imply the same concept as a length or width of zero.)

Nystuen (1966) discussed the concept of a two-sided geographic line, in the context of studying spatial boundaries, and processes which occur along them. He commented that many processes are modified by the character of the boundaries along which they flow, including traffic flows, migratory flows, and flows of commerce. The coastline of TEXAS is a good example of such a two-sided line, one whose high order structure changes quite drastically from one side to the other. In a cartographic context, these two sides should most likely be generalized according to different tolerance criteria, in any conventional smoothing or simplification task. The monotonicity parameter indicates the differences in the high order structure, and could be used in an automated generalization task, to identify such a two-sided line.

Parameters of Feature Geometry

Because none of the other sample lines displays a strong degree of monotonicity, their high order structures will be summarized for each line as a whole. It should be noted that none of these lines include any bifurcations beyond level two, which is a major distinction with both sides of TEXAS. Once again, this lack does not indicate a limit to resolution, but rather an upper limit to complexity. Figure 5.12 illustrates the size of features for these lines, that is, the number of coordinates. This figure may be useful in a computer identification task, for it provides information not available in measures of length and width. The number of coordinates provides a direct index of how much in-core storage is required to deal with a feature as a whole. It may also prove useful as a guideline in searching through a computer file from one end of a feature to the other, in effect, a constraint on how far to 'look ahead'.

The number of coordinates may or may not be correlated with the overall length of the line; only SIOUX displays an identical pattern for feature length, as shown in Figure 5.13. While the feature lengths in FJORD display a steady drop, the sharp increase in the HUDBY line is similar to both sides of TEXAS; this increase is most strongly indicated by the SIOUX plot. Here, the second level feature (there is only one) is a long, narrow resevoir: one should not conclude from this that the SIOUX lengths are similar to TEXAS features; as a matter of fact, it will be established that this resevoir is structurally equivalent to a fjord, later in this analysis.

Widths of features are displayed in Figure 5.14. Once again, the patterns of HUDBY and OF FJORD are quite similar, and the SIOUX signature

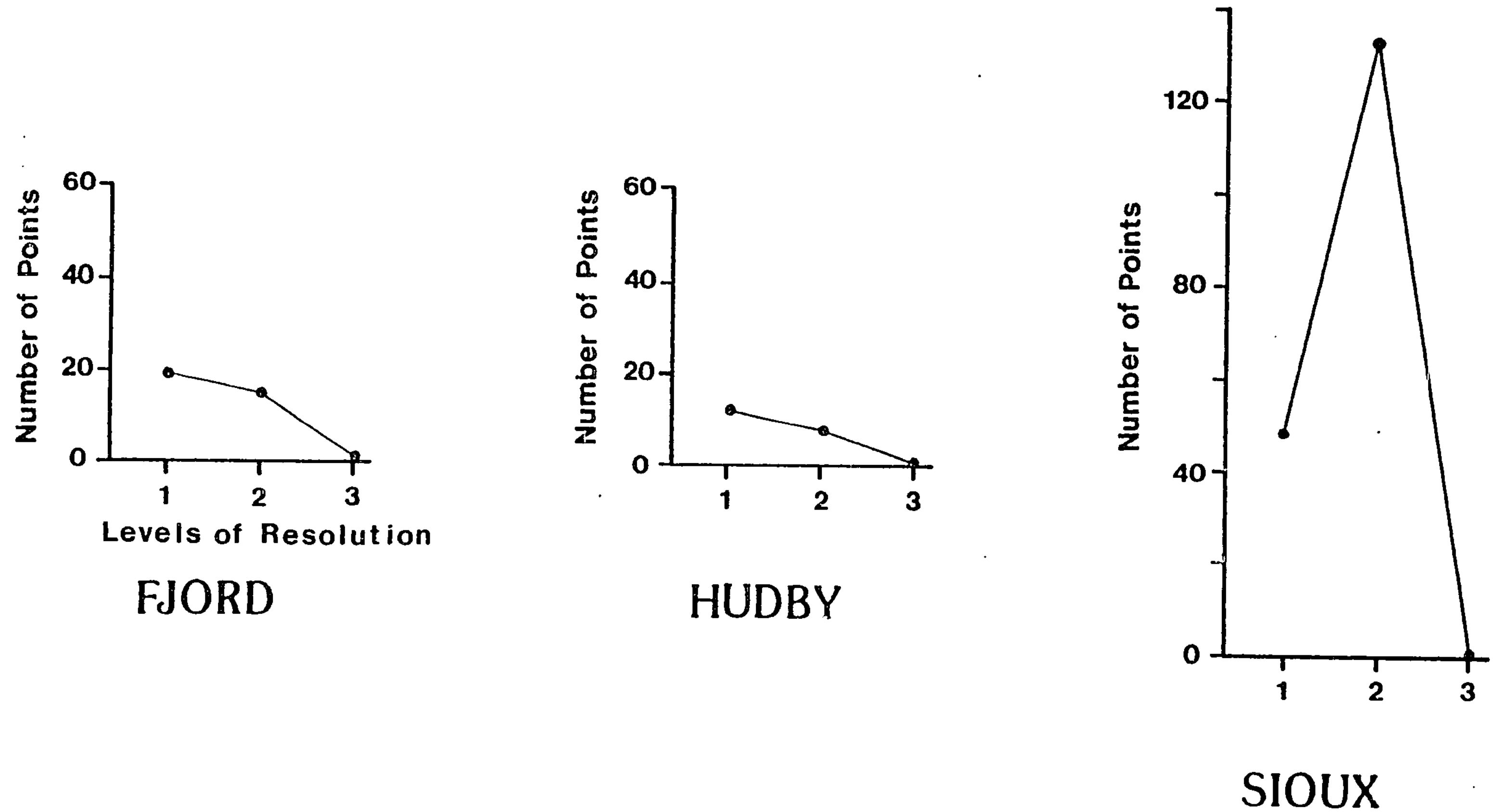


Figure 5.12
SIZE OF FEATURES
High Order Structure Signatures

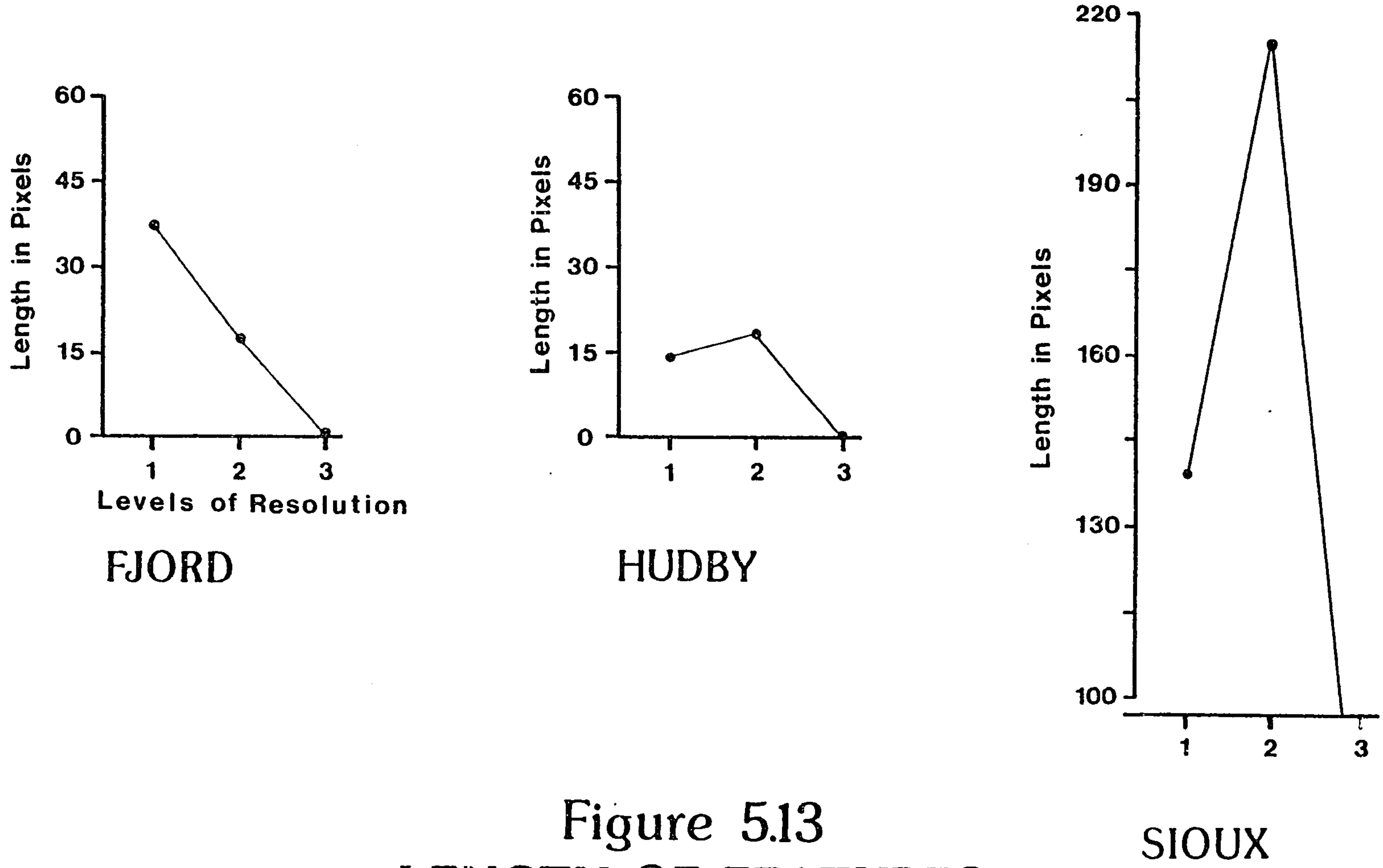


Figure 5.13
LENGTH OF FEATURES
High Order Structure Signatures

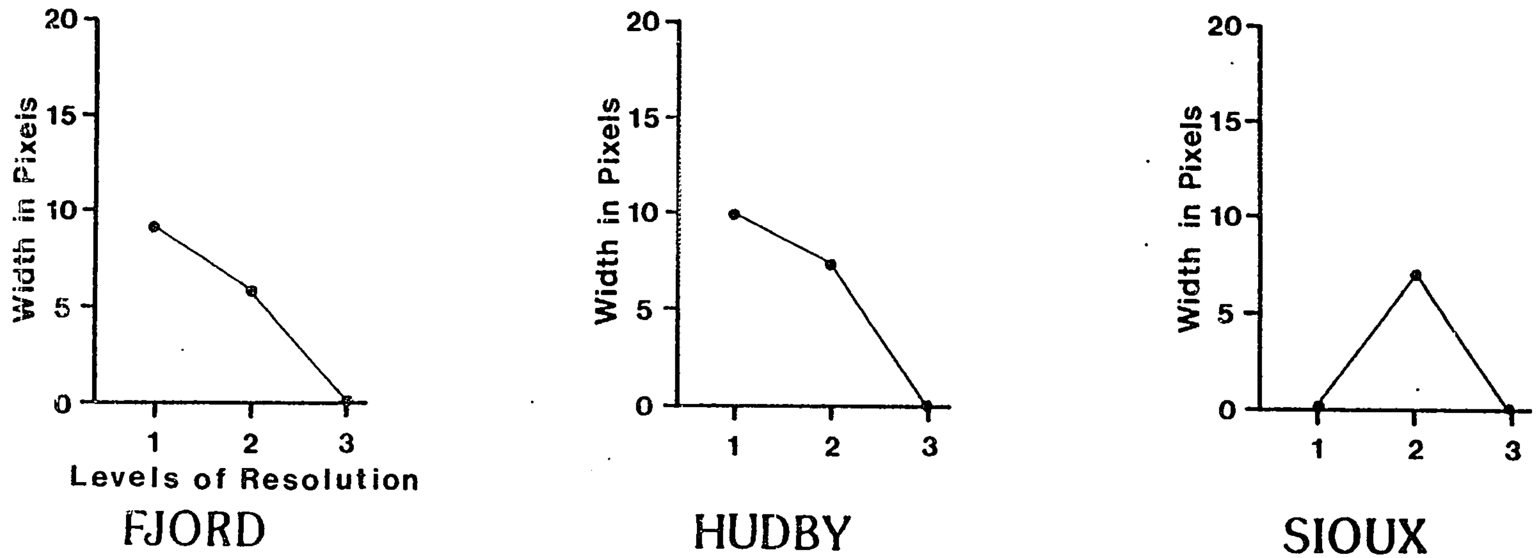


Figure 5.14
WIDTH OF FEATURES
High Order Structure Signatures

is heavily biased by the resevoir in level 2. In Figure 5 15, the orientation of the three features is a truncated graph: as before, the scaling of angular measurements differs from measurements length and width in terms of the concept of a zero value. Becuase no features were defined at the third level, no angular measurements were taken, and as before, angles of zero degrees do no indicate a lack of orientation. However, it can be seen that the average orientation of features does not change markedly from line to line.

As a matter of fact, the set of high order structure signatures as a whole does little to distinguish betwen lines, except in the non-monotonic case of TEXAS. High order structures for HUDBY and FJORD are almost identical, except for length; and the SIOUX signatures reflect only three parameters, being strongly biased by the single feature in level 2. Plainly, high order structures such as these will not prove powerful in distinguishing between these four lines. However, the problem may lie in the organization of features rather than in their measurement. The assumption all along has been that high order structures will differ between lines as a whole, as was the case for parameters of low order structure. However, it may be more meaningful at this level of structure to distinguish between classes of features, instead of between classes of lines.

Consider that the very long narrow resevoir in SIOUX looks like the very long and narrow river valleys in FJORD; and to a lesser extent, these both resemble the longitudinal spits of TEXAS. On the other hand, the HUDBY features may not necessarily appear fjord-like. It may be that the features in certain parts of each line are unique, but that some lines

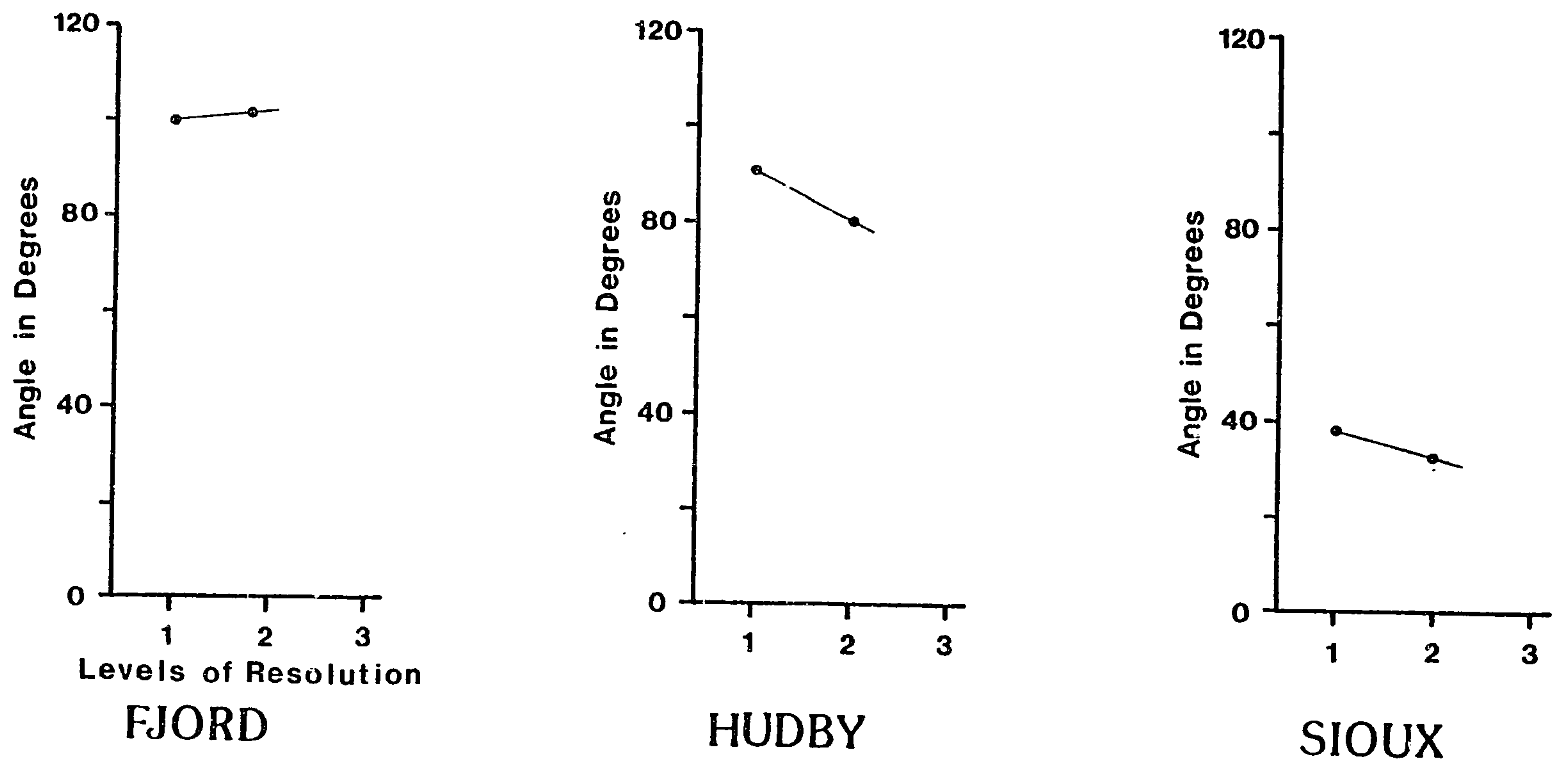


Figure 5.15
ORIENTATION OF FEATURES
High Order Structure Signatures

share features of similar high order structure, and further that the features which are similar from one line to the next in terms of high order structure are also similar visually. If this is true, then the high order parameters can be reorganized according to types of features, to provide a more powerful means of structure identification.

Classifying Categories of Features

The purpose of classification analysis, or cluster analysis, is to form groupings of a set of observations such that variation within groups is minimized, and between group variations are maximized. The variations are based on variance computations for a (set of) variable(s), which are most commonly standardized to allow for comparisons of the amount of variance between variables remaining after the classification is complete. Cluster analysis is similar to factor analysis in that composite sets of variables are created in both techniques; a major difference is that in factor analysis, the composite variables are ratio-scaled, while the scale of composites created by clustering is nominal.

The purpose of the clustering analysis in this study is to explore the hypothesis that high order structure parameters can be more meaningfully organized by types of feature than by the four categories of lines. The keyword here is exploration: it is not clear how many types of features are present, nor what specific size of feature groupings is most appropriate. But it is peceptually evident in looking at the line feature plots that some sort of similarities exist; and these similarities may be more meaningfully organized by the high order structures of specific features, than by trying to deal with the high order structure of an entire cartographic line.

The clustering is intended to show what cartographers call 'natural breaks' within the five variables. If they exist, these natural breaks should become apparent in looking at the dendogram. Pre-processing and data massage techniques will be kept to a minimum: the primary intent is not a search for statistical significance so much as an evaluation of content in the high order parameters.

The variables used in the analysis include length, width, and angle of feature orientation. Also included for each feature is the number of coordinates, and the number of bifurcations from each feature, as a kind of fragmentation index. For example, the bays on the 'landside' of TEXAS are distinctive in that many small bays bifurcate from larger ones; these features are much more fragmented than the bays in HUDBY, which are visually different but of a similar length and width.

A preliminary hierarchic clustering was performed, using Ward's criterion to minimize within-class distance for the system as a whole. This constraint tends to produce a larger number of smaller groups, as opposed to other minimum distance constraints which may cluster fewer, larger groups. The intention at all times was to facilitate interpretation of the feature clusters. Originally, all variables were weighted equally, but this was found to confound the interpretation of clusters: as a result, the length and width variables carried double weight in all classifications. All variables were converted to Z-scores, to facilitate cross-variable comparisons of within-class variation.

One major limitation of the hierarchic strategy is that once an observation has been aggregated into a group, it cannot be reassigned during a later iteration. As a result, some clusters may contain certain anomolous members, even though the system variance is minimized as a whole. The solution is to improve the hierarchic solution by allowing individual members to move between groups in an effort to further reduce the within-class variance. This iterative clustering technique is included as an option in many hierarchic clustering routines, and can be initiated by inputting desired class means or by initializing class membership for each observation.

The iterated solution includes nine clusters, all of which are clearly interpretable. For example, the first four categories can be seen to include short, medium, and two classes of long narrow features (respectively, Figures 5.16 – 5.19). 'River valleys' might serve as one possible generic label for these four feature categories, or 'fjords', except that each of these labels carries with it distinctly geomorphic connotations; in graphic space, such connotations have traditionally biased cartographic decision-making. The purpose of this entire research is to propose a move away from these biases, and therefore such generic labels for clusters will be used sparingly, if at all.

The visual similarities between the reservoir, the barrier spits, and the fjorded river valleys are clearly reflected in these clusters, at several scales of feature size; and other feature categories display visual similariites as well. It may be reasonable to aggregate clusters #2 and #3, which differ only in terms of length (Figures 5.18 – 5.19). Aggregation might also reduce clusters #5 and #7, or even #8 and #9, as the features in each of these pairs are visually and numerically quite similar.

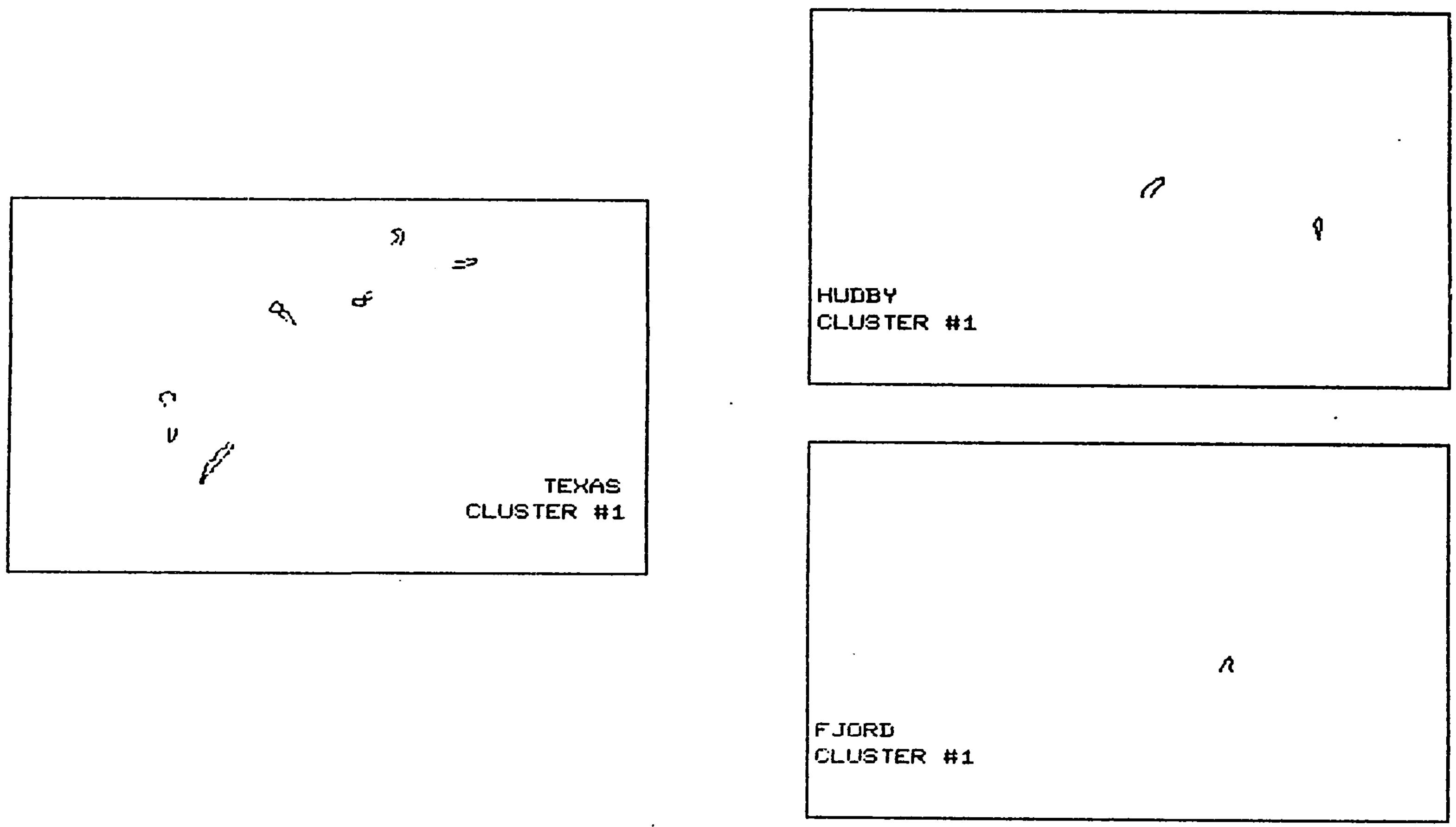


Figure 5.16
FEATURE CLUSTER #1

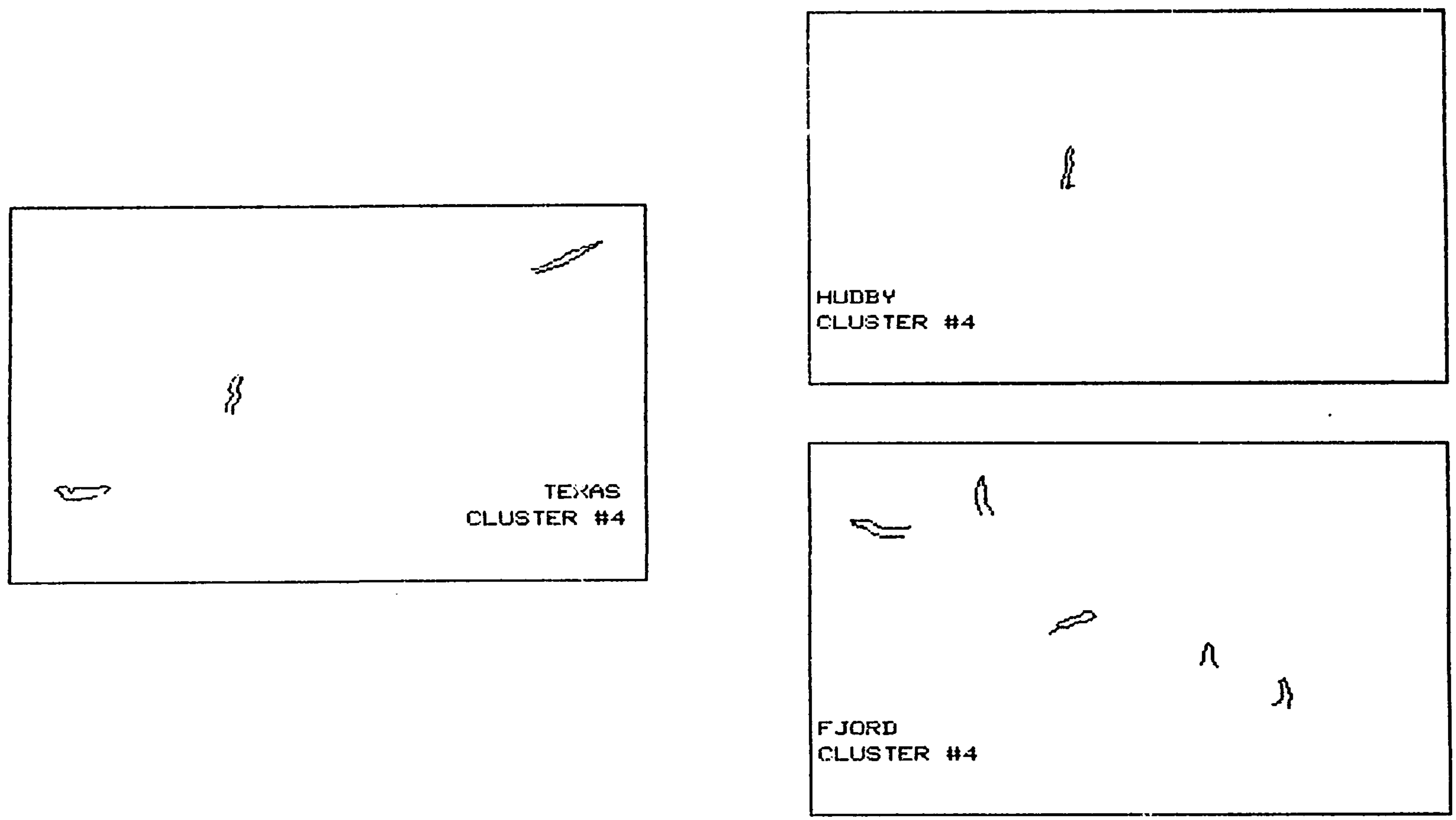


Figure 5.17
FEATURE CLUSTER #4

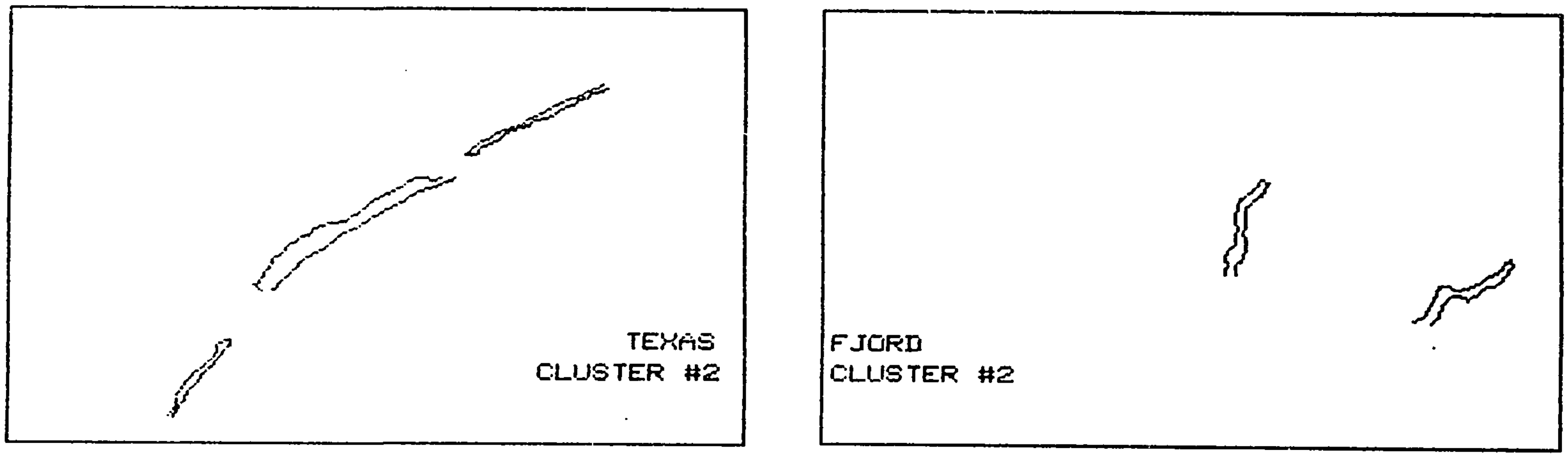


Figure 5.18
FEATURE CLUSTER #2

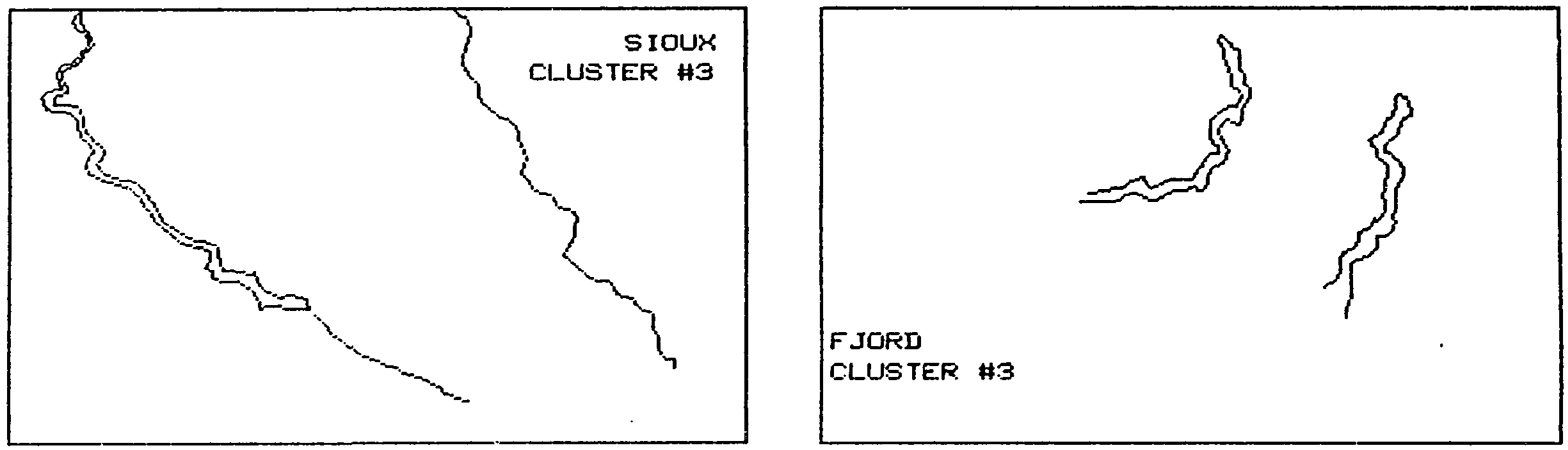


Figure 5.19
FEATURE CLUSTER #3

In demonstration of this, means and standard deviations for each cluster are presented in Table 5.7. A justification for leaving these groups isolated will be presented further along in the discussion.

TABLE 5.7
CLUSTER MEANS AND STANDARD DEVIATIONS

CLASS	NUMBER OF POINTS	LENGTH	WIDTH	BIFUR-CATIONS	ORIENTATION (IN RADIANS)
1 [10]	9.60	11.45	5.28	0.00	0.86
	(3.67)	(4.66)	(1.10)	(0)	(0.50)
2 [5]	34.40	58.32	5.26	0.20	1.09
	(6.38)	(18.99)	(1.92)	(0.40)	(0.48)
3 [5]	107.00	146.15	7.87	0.20	1.33
	(35.78)	(54.54)	(8.23)	(0.40)	(0.85)
4 [9]	17.00	21.39	6.06	0.00	1.81
	(4.32)	(5.81)	(1.38)	(0)	(0.32)
5 [9]	13.67	14.15	10.33	0.22	2.43
	(5.83)	(3.76)	(2.45)	(0.41)	(0.30)
6 [5]	21.60	27.18	13.23	0.20	0.58
	(7.42)	(12.61)	(1.24)	(0.40)	(0.27)
7 [7]	11.00	11.56	7.48	0.14	1.53
	(4.21)	(4.30)	(0.94)	(0.35)	(0.27)
8 [12]	46.50	21.96	10.65	1.50	1.12
	(20.5)	(11.18)	(3.10)	(0.65)	(0.51)
9 [2]	80.50	21.73	19.15	5.50	1.36
	(13.50)	(7.42)	(9.49)	(0.50)	(0.12)

Numbers in parentheses ( ) are standard deviations.
Numbers in brackets [ ] reflect number of features in each class.

Only half of the clusters are dominated by the long and narrow river- and fjord-like features. Three other clusters display groupings of small, medium, and larger bays (Figures 5.20 -5.22). Two final clusters also include bays, but these are differentiated by the large number of bifurcations emanating from each feature (Figures 5.23 - 5.24). In other words, these features are similar to each other in that each one has several smaller features which fragment its trend line. The features in cluster #9 are so fragmented as to be unrecognizable out of the context of the line; a copy of the feature file is included in the illustration (Figure 5.24), to identify the two bays in question.

However easy the categories are to interpret, specific features in several clusters appear visually unrelated. For example, the inclusion of the FJORD feature in cluster #8 (Figure 5.23) is anomolous, as it is not a wide feature in relation to the other bays in this class. It may be that the parameter is not width but bifurcation which has forced membership. This is the only feature in FJORD to bifurcate off to one side, and this singularity may have separated the feature sufficiently from other groupings of fjords. It may also indicate an error in the original feature definitions: a more appropriate trend/feature breakdown here might have included two small fjords bifurcating from a single small bay.

Another anomoly can be seen in the HUDBY cluster #6 (Figure 5.22), by the inclusion of a small and sinuous feature in the lower right hand corner. The other members of this class are not so narrow as this one, although they appear to be of a common length. On first glance, one might subjectively place this feature into cluster #4. However, a closer look indicates that features in cluster #6 are fairly sinuous; the number of

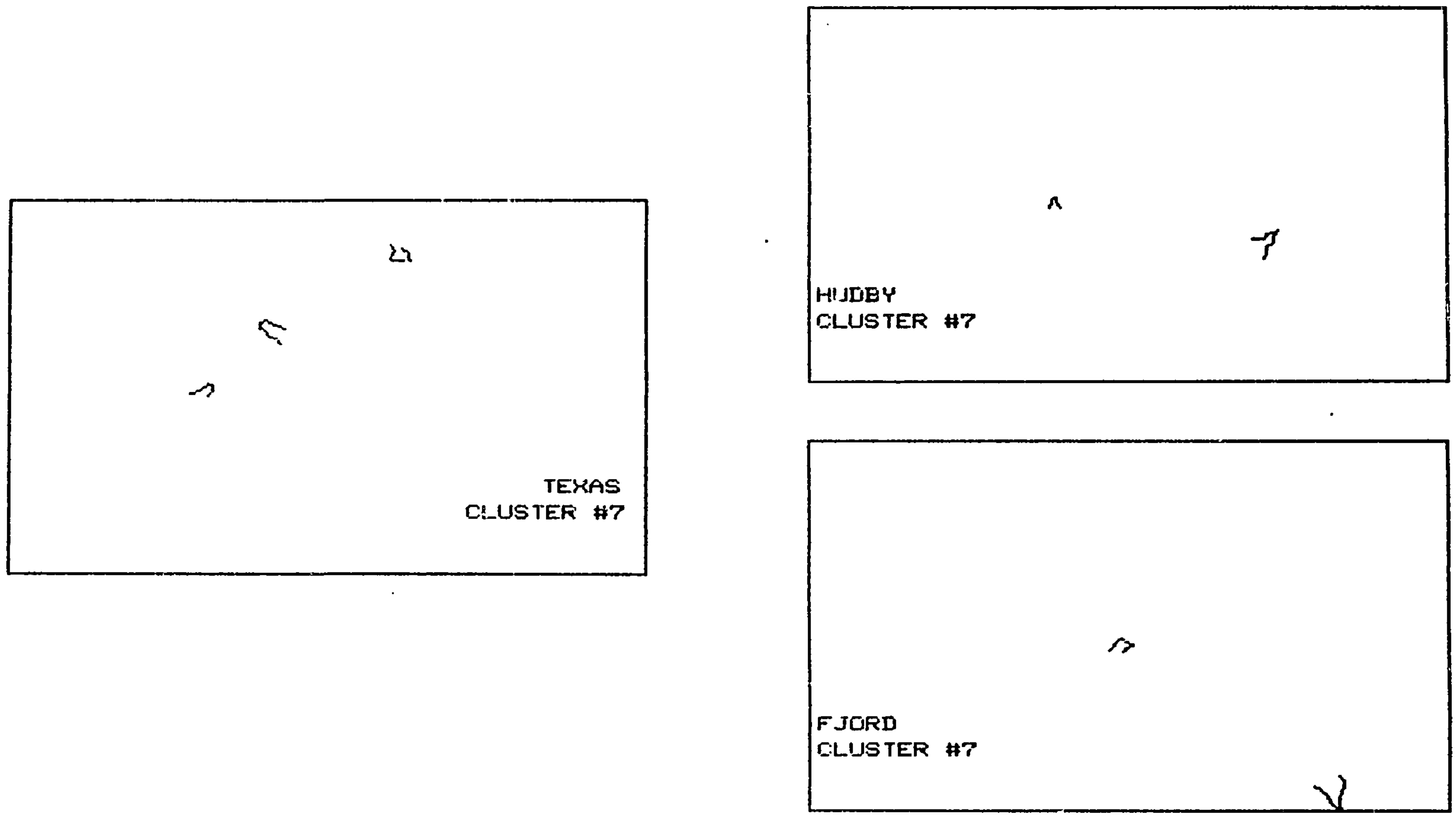


Figure 5.20
FEATURE CLUSTER #7

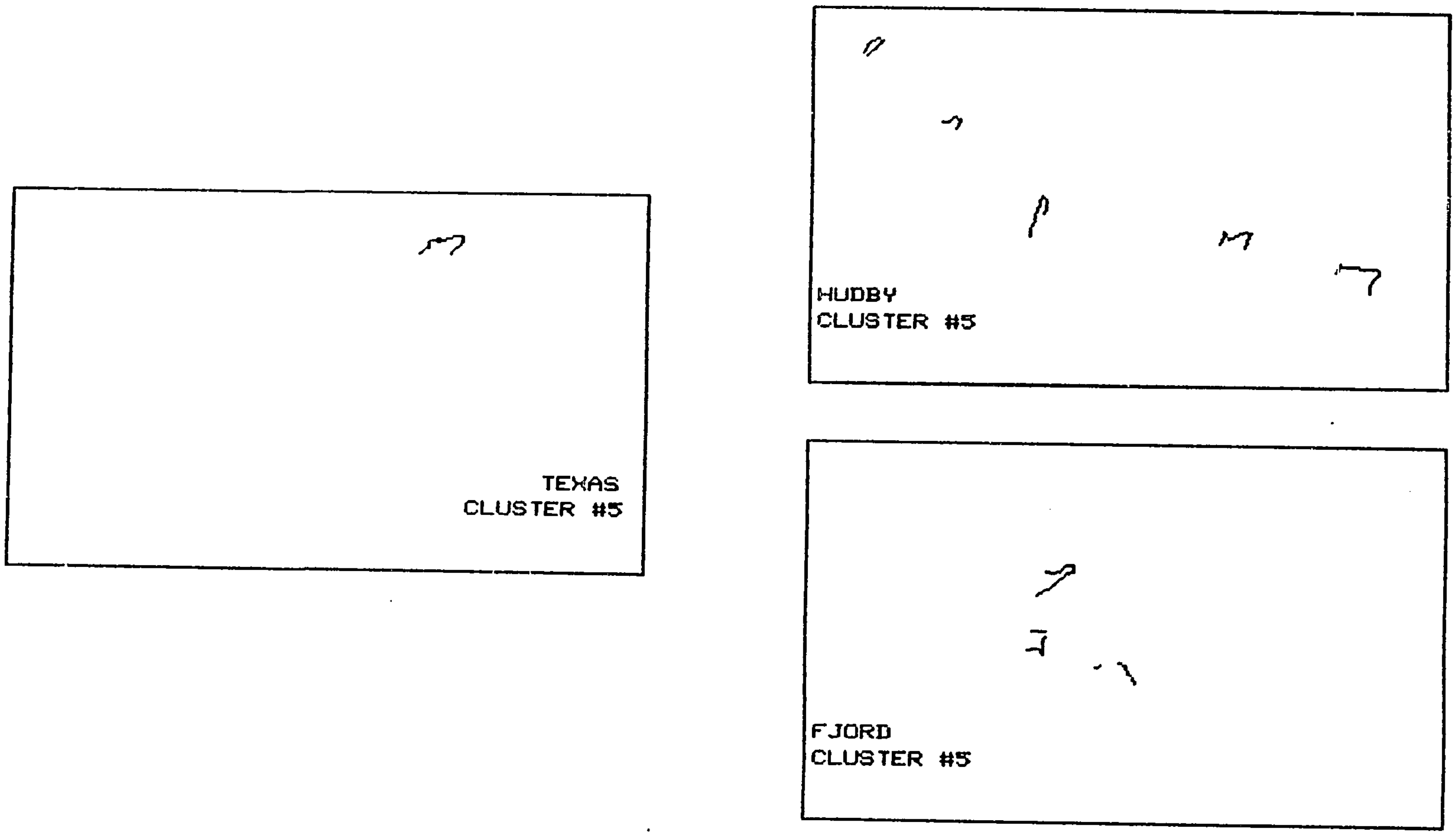


Figure 5.21
FEATURE CLUSTER #5

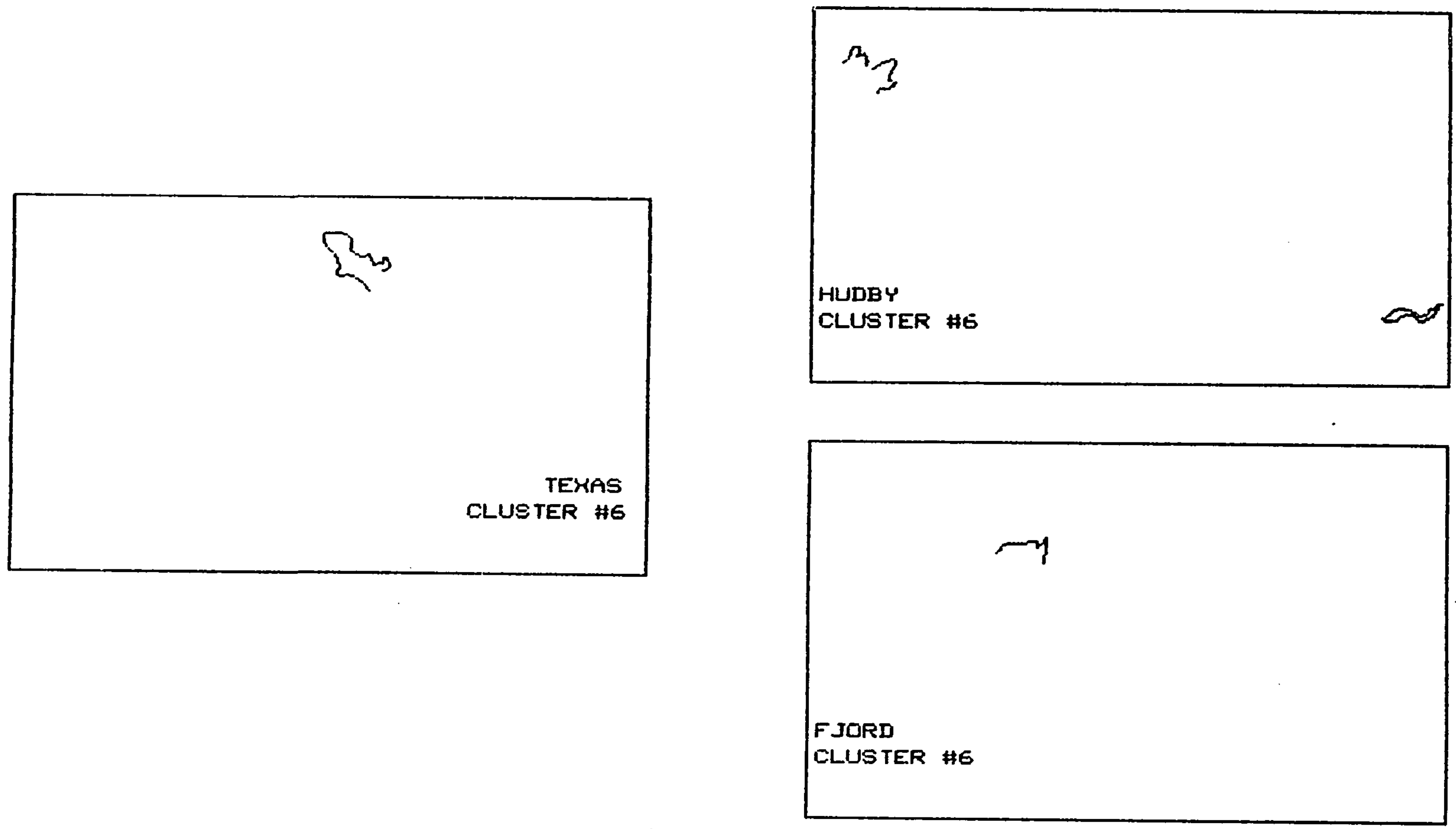


Figure 5.22
FEATURE CLUSTER #6

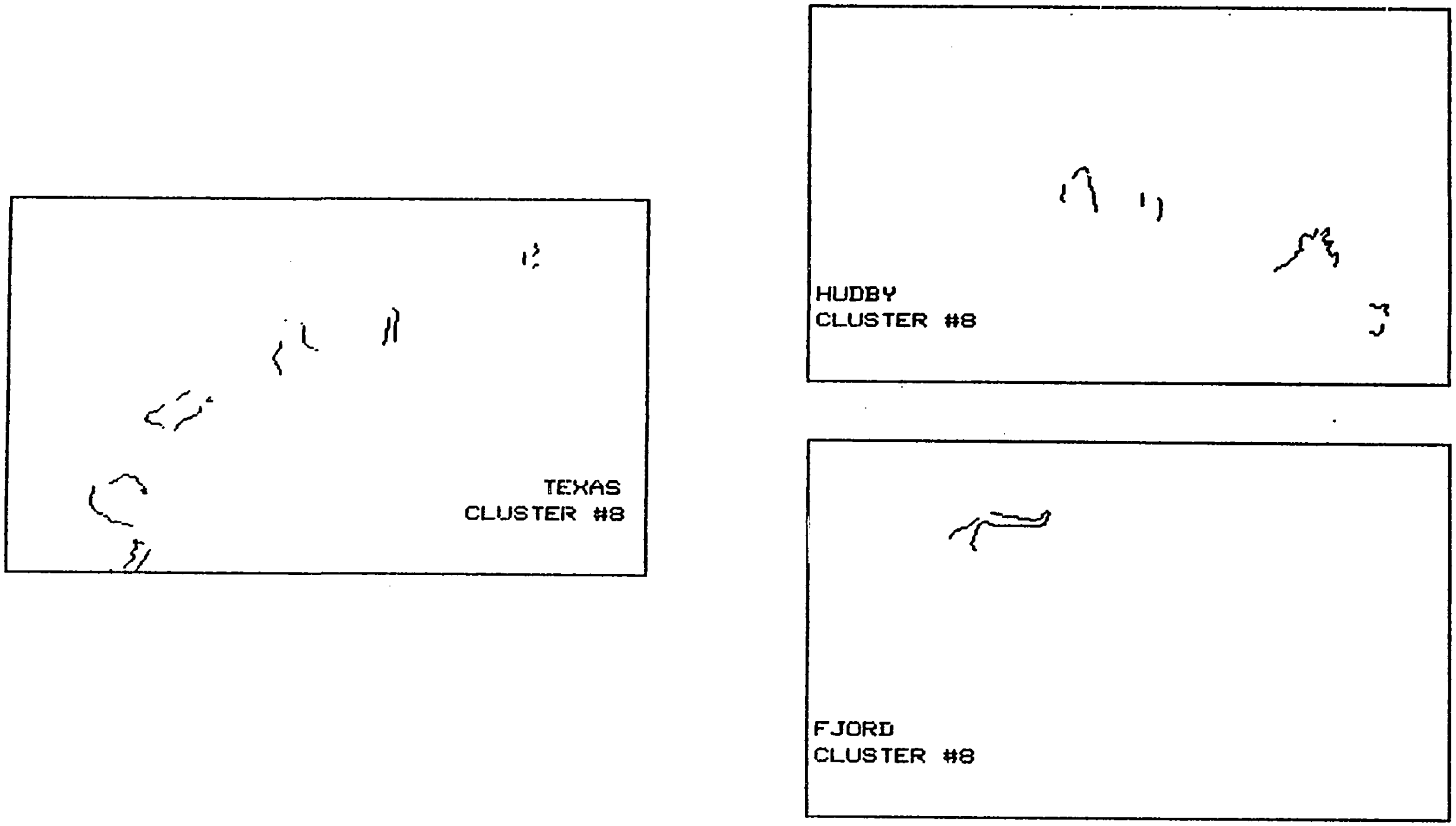


Figure 5.23
FEATURE CLUSTER #8

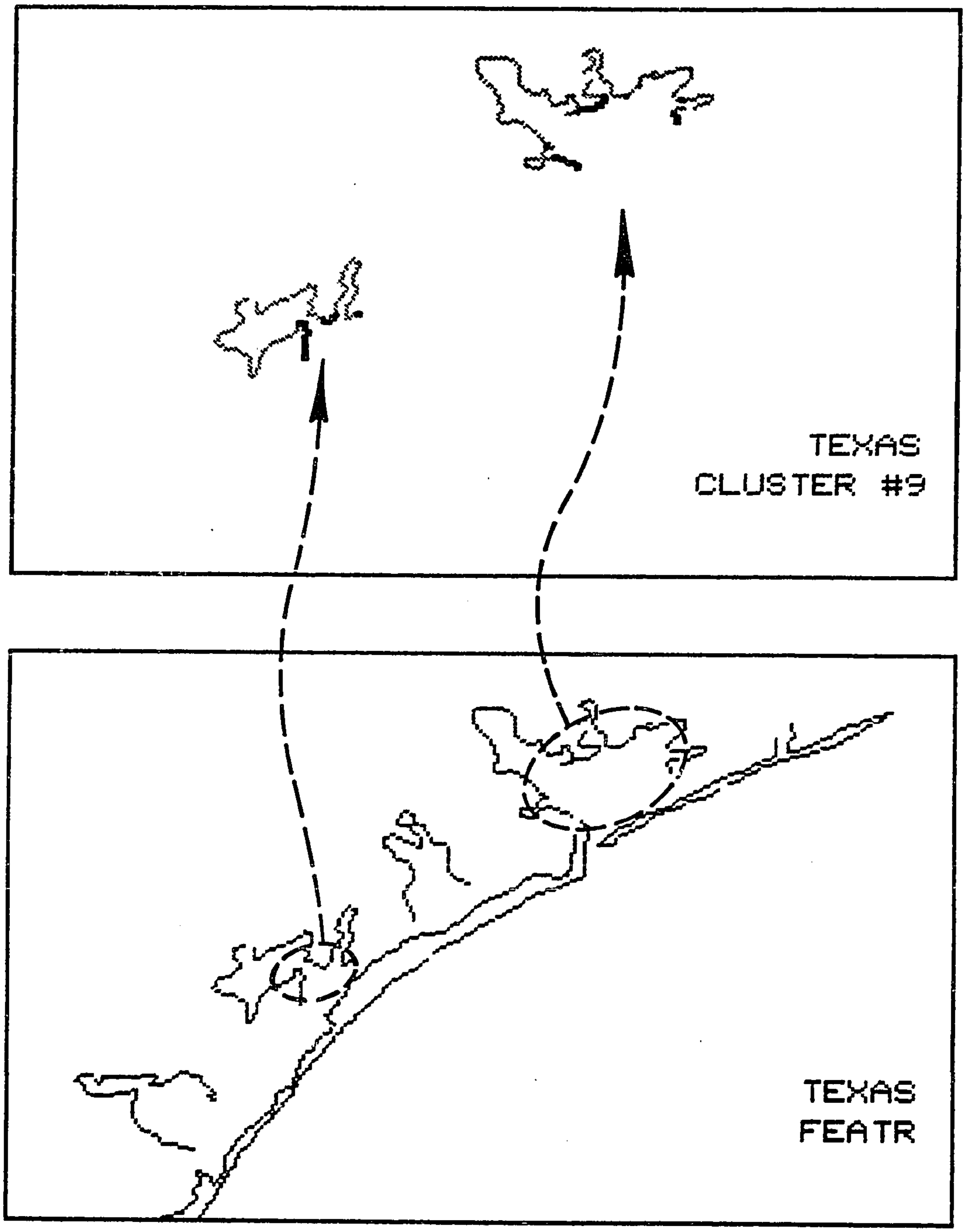


Figure 5.24
FEATURE CLUSTER #9

points in these features coupled with the factor of length have most likely caused this particular assignment.

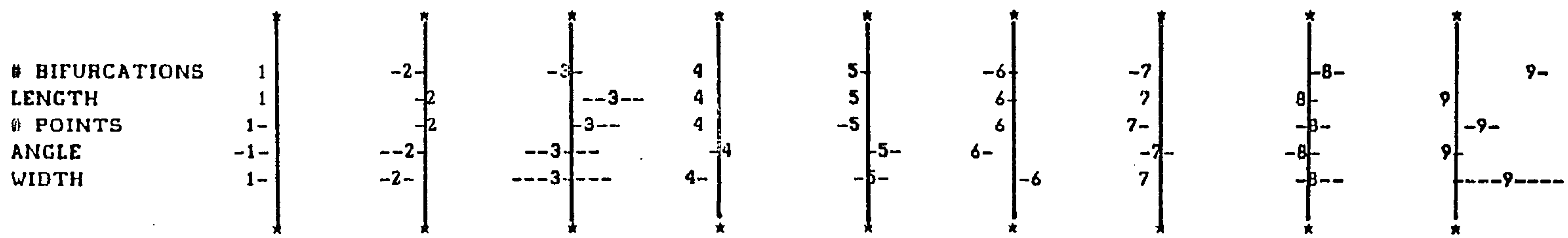
Other anomolies are also present: the inclusion in cluster #1 of a TEXAS barrier spit, and the single bifurcated feature included in cluster #5, provide two examples. As before, the point here is not to produce necessarily optimal clusters, but to explore the feasibility of organizing high order structures across categories of lines, by feature. For this reason, these clusters will not be improved upon, but will stand as unconstrained examples of how readily such an organization can be accomplished.

Table 5.8 displays a graphic depiction of each cluster, to better express the dispersion of each grouping in relation to each variable, and in relation to the means of the data set as a whole. One can readily see that relatively compact groupings of features have been achieved by straightforward clustering of high order parameters.

The cluster analysis has shown that high order parameters can be meaningfully combined and summarized by type of feature, rather than by categories of lines. One might construct a table of cluster frequencies, to indicate the percentage of features identified for each line falling into each category of high order structure. Each of the columns in Table 5.9 represents a particular cartographic line; the numeric values are the number of features sampled from each line which fall into a given category; and the decimal fraction beneath each integer value reexpresses the integer as a percentage of all features sampled from the given line.

TABLE 5.8
CLUSTER PROFILES AND DISPERSION
(taken from the output of BMDP K-MEANS)

variables are ordered
by F-ratio magnitude



Each column is a cluster.
The cluster number is printed at the mean of each variable.
Each dash indicates one standard deviation above or below the mean

All data values are reported as Z-scores.
The asterisks and vertical line indicate the grand mean for the data set as a whole; thus the location of each cluster within the data set can be related, and compared from one variable to the next.

TABLE 5.9
CROSS-TABULATION OF FEATURE CATEGORIES WITH LINES

CLUSTER	SIOUX	HUDBY	FJORD	TEXAS	TOTAL # FEATURES
1		2	1	7	
		(.12)	(.06)	(.26)	
2			2	3	
			(.12)	(.11)	
3	3		2		
	(1.00)		(.12)		
4		1	5	3	
		(.06)	(.29)	(.11)	
5		5	3	1	
		(.29)	(.17)	(.04)	
6		3	1	1	
		(.17)	(.06)	(.04)	
7		2	2	3	
		(.12)	(.12)	(.11)	
8		4	1	7	
		(.24)	(.06)	(.26)	
9				2	
				(.07)	
	3	17	17	27	64
	.04	.27	.27	.42	

The percentage tallies for the four lines are plotted in Figure 5.25, to illustrate how these clusters may provide clear visual distinctions between lines. These plots will be used in place of the original high order structure signatures. The category frequencies listed in Table 5.9 will also be stored on disk as a look-up table; they may serve as guidelines for predicting line identity from sampled features, as follows.

For a feature sampled at random from the group of 64, if that feature were measured for high order structure and assigned to cluster #4, for example, the chance of that feature belonging to the FJORD line would be much greater (29%) than its possible membership in any other line (29% of the features in Cluster #4 are FJORD features). One might work with a

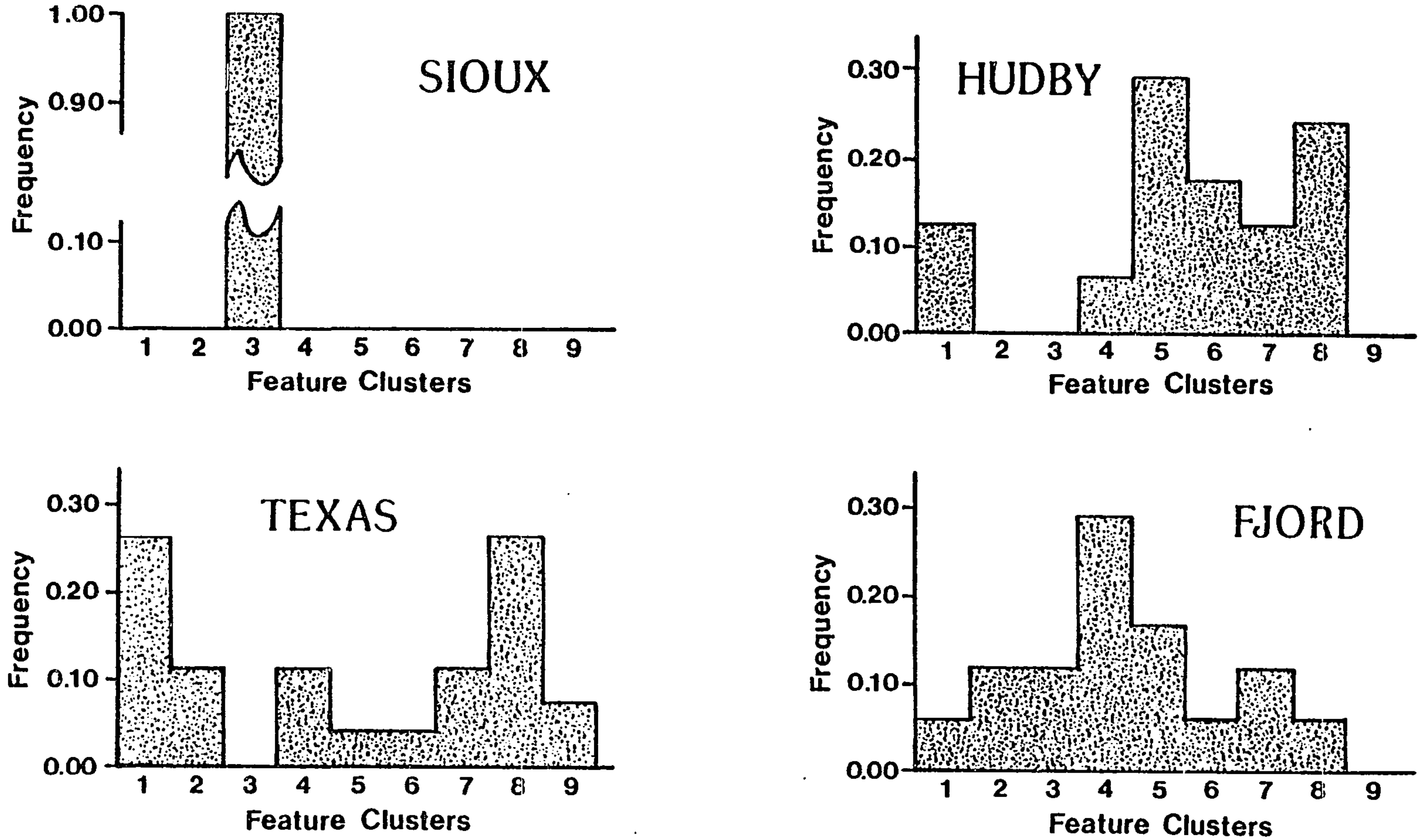


Figure 5.25
FREQUENCIES OF FEATURE CATEGORY
By Line

sequence of features known to belong to a single line, comparing each in turn with the table of frequencies, to make identifications with a higher degree of confidence. The real power of such a look-up table would lie in sampling many features, instead of only one; it is also reasonable to consider adding new features (and new feature categories) by sampling features from other previously identified cartographic lines. This predictive application of high order parameters will be discussed in more detail in the next chapter, when the high and low order parameters are directly related to the question of automating line generalization tasks.

Summary

This chapter and the previous one have discussed two kinds of structural relationships which can be measured in cartographic lines. The first reflects relationships between points along the line, and the second reflects segment-to segment details. The parameters of low order structure have been shown to provide good distinguishing capabilities between the four sample cartographic lines, within a finite range of graphic resolution; these measures have the additional advantage of being based in major part on certain perceptual criteria used commonly in manual cartographic generalization tasks.

Parameters of high order structure are somewhat different in nature than the low order measures. Two significant distinctions are that first, the high order measures are not tied to discrete levels of resolution, or to a hierarchy of feature size, but seem instead to be based on some hierarchy of line complexity, as shown by the relationship between trend line/feature distinctions and fractal dimension of the lines.

The second major difference is that the high order parameters are not homogeneous within a single line, but are more appropriately organized by categories of features which may be present in several very different kinds of lines. In other words, it is the structure of individual features which is distinguished by these parameters; the cartographic lines are in this context merely combinations of high order structures. Differing lines should be expected to have different feature combinations, just as the lines have different fractal dimensions as indices for varying degrees of high order complexity. This has been illustrated in the frequency plots for feature categories by line (Figuree 5.25).

The low order and high order parameters are similar in that both have apparent perceptual overtones: the points which bound each low order anchor line are selected by an algorithm which works on a principle of maximum angular deviation, similar to the inductive decisions that cartographers make in generalizing linear features. And those features which are most similar in terms of high order parameters tend to be quite similar visually; a very simple and non-constrained clustering technique produced groupings which are visually logical. One obvious question which arises here is the extent to which these perceptual overtones can be empirically derived, by testing map reader responses to lines along a continuum of low and high order structures.

One other question is related more directly to the second similarity between the two kinds of structure. Both orders may be useful in distinguishing between the four types of cartographic lines; in both cases, the distinctions are probabilistic, and based on sampled points or on sampled segments, depending on the order of structure to be considered. To what extent can these structure signatures be applied to computer discriminations between lines? This question has direct implications for the selection and modification of tolerance criteria in automated generalization tasks. The final chapter of this thesis will consider this question, and present an algorithm which may serve to implement such discriminating capabilities.

CHAPTER 6
WORKING WITH STRUCTURE IN CARTOGRAPHIC LINES

The research reported in this thesis has been based on the notion that a cartographic line is a probabilistic representation of the geographic feature which it represents. Numeric parameters have been measured for two orders of structural relationships, and these parameters have been shown to distinguish between categories of cartographic line structure. The categories which have been developed are not intended to form an exhaustive typology of linear structure, but rather to demonstrate that meaningful categories of structure can be defined numerically, and that the categories which are defined include structures which are visually logical.

Categories for both orders of structure have been summarized graphically, as structure signatures, and digitally, by storing means and variances for the parameters in each category as a look-up table in the computer. The look-up tables are listed in Appendix C (low order parameters) and in Table 5.7 (high order parameters). To reiterate Chapter 5, the original summaries for high order parameters did not indicate strong differences between the four sample lines. Instead, the high order parameters were reorganized by feature category; these structure signatures display distinct means and variance values, and will be more useful summaries as a result.

The value of the structure signatures for line generalization can be seen in several applications, all of which utilize the digital look-up table versions of the structure signatures. One application involves generating lines of known graphic structure, and is called stochastic modelling. The other application does not serve to generate line structures, but to identify them. In both cases, the algorithms will be discussed but will not be developed as code. However, the application of the structural parameters can still be demonstrated, if only by verbal example.

Stochastic Modelling of Cartographic Lines

Several examples of stochastic models have been presented in the course of this research, all of which have been shown to produce visual displays of predictable degrees of structural complexity. The models have ranged from the single lines produced by Dutton (1981) to the fantasy landscapes produced by Voss (Mandelbrot, 1982), to the geographic simulation of Mount Rainier illustrated in Chapter 2 (Carpenter, 1981). The major disadvantage in all of these stochastic models for cartography has been the lack of control for accuracy, due to repositioning of coordinates. The algorithm described here avoids this limitation.

In a stochastic modelling technique, the look-up table of measured structural parameters can be referenced according to the desired type of low order line structure, and the means and variance values for each parameter can be used to constrain upper and lower limits for random displacements of straight line segments. Alternatively, a desired degree of visual complexity might be specified, and this value compared to the parameters for fractal dimension. The choice of a particular line structure would then be directed by whatever look-up table entry contains the most similar fractal dimension value. In this way, the two orders of structure can be used in combination, although the final stochastic model should be generated more easily from the low order parameters.

The stochastic model should be initialized with a simple line segment, which is subjected to a series of random displacements. Each displacement is controlled by parameters from the digital look-up table. For example, the width parameter constrains the length of the displacement. The segmentation parameter constrains where along the straight line segment the displacement occurs: not all displacements should necessarily occur at a segment midpoint, as was demonstrated by the segmentations of the four sample lines presented in Chapter 4. Monotonicity may be used to constrain the consistent direction of displacements required for stochastic modelling of long, gentle curves. Other parameters constrain the number of displacements, and the number of new coordinates to be generated during each displacement. Carpenter (1981) suggests that stochastic models can be generated from a variety of graphics primitives, including straight line segments, and also arcs of circles, ellipses, or parabolas. The displacement process is similar regardless of the initializing segment.

Two immediate advantages are served by the parametric format of the look-up table. First, the same initializing segment can be used to generate lines for many different kinds of graphic structures, using a modular generating algorithm. Secondly, the generation of random numbers can be parameterized, to produce further stochastic variations. For example, in a stochastic model which relies on Brownian motion, the number generator is a random walk, produced by a normal (Gaussian) distribution.

Any of the look-up table mean and variance values can be plugged in to the random number generators directly. Different subroutines can provide formulas for different probability density functions, which will produce different displacement patterns, and thus different graphic structures.

Instead of a normal distribution, one might prefer to utilize a gamma distribution for the stochastic model, to see what differences in structure result. Or probability distributions could be interchanged, depending on the parameter in question: in the case of the segmentation parameter, for example, a normal distribution would tend to cluster displacements closely about the segment midpoint. A gamma distribution would distribute displacements towards the endpoints, and has the additional advantage of favoring the alternate endpoint when inverted; thus two displacements can be produced in a single iteration.

Regardless of the type of probability function, the basic algorithm is the same. The process inputs an initial segment, plots the first endpoint, and enters a recursion controlled by the number of desired displacements. First, the low order look-up table is accessed for a requested set of parameters, and then the process calls the probability density function. The values which are returned indicate the location, distance, and direction of the displacement, which can then be computed from the endpoints of the segment by simple algebra.

No trigonometry is required, as all displacements occur at right angles, making the algorithm very fast. As each coordinate is computed, it can be stored in a temporary array, or plotted immediately. At the end of the loop, the final endpoint is plotted; and the process reads in the next segment. The strip tree data structure provides a ready means of insuring that endpoints are critical to accurate representations, as long as displacements occur within strips, and not between.

One should not assume from this discussion that the parameters measured for any of the four sample lines should be used to provide a stochastic model for itself. For instance, regenerating the FJORD line successfully from its own low order parameters would not validate the algorithm. One might argure that it serves as a preliminary test, a kind of identity matrix operation. The degree of 'identity' is questionable, however, given that the modelled line is randomly generated, and will be somewhat different each time it is produced.

More importantly, regeneration of the line would not serve to validate the low order parameters: the recreation of a particular line such as FJORD does not imply that generic renditions of fjord-like structures are necessarily achievable, because other fjorded coastlines (eg., the coasts of Greenland, Norway, or Chile) may contain additional graphic structures which are not yet represented in the look-up table. In order to evaluate the algorithm, other line samples should be collected, and modelled according to the existing look-up parameters. A further evaluation might be provided by perceptual tests of the stochastic models, to evaluate map reader reactions to the generated line structures.

Computer Identification of Cartographic Lines

The second application of structural parameters to line generalization is also probabilistic in nature, but can be considered more closely with traditional generalization techniques, namely, the elimination of detail from a cartographic line. Here, the bottleneck to automation relates to the inductive decisions which a cartographer makes in choosing tolerance values by which to control the particular simplification or smoothing algorithm which is applied. These decisions are not easily automated; and because of this, the values are most often set manually.

A particular problem arises when generalizing a compound line, such as the coastline of the United States. For example, the tolerance criteria required for appropriate simplification of the Puget Sound may be significantly different than the values which should be applied to the Oregon coast. When the values are set by hand, the algorithm must be stopped at some point, and the tolerance values modified. Alternatively, the coastline must be broken into pieces, and generalized as separate lines. In either case, the decision of where to break the line or where to change the parameters is inductive, based on some visual approximation of where line characteristics change. This task is also difficult to automate.

The structural parameters developed in this research can be applied to the modification problem, because it involves distinguishing between lines of different structure. On the other hand, choice of a tolerance value is closely related to the identification of line structure; and each of these applications will be discussed in turn.

Line Discrimination Tasks

The choice of where to change a tolerance parameter may be easily automated by checking for significant changes in the low order structure of a line, as these parameters have been shown to provide fairly accurate discriminations between lines of differing visual structure. It would not be necessarily required to match some sample of low order structure with look-up table values, because the important aspect here is not specific content but specific change. What is required instead is an algorithm which compares low order parameters for overlapping segments of the line, and tests adjacent overlaps for significant differences. The tolerance criteria should be modified at places where these differences occur.

Overlapping segments can be determined for one of several levels of resolution, by storing line information (coordinates and low order parameters) in a strip tree data structure. A sequence of strips may be compared to an adjacent overlapping sequence by comparing variance values computed for all parameters in each sequence, in a ratio. If the values are similar, the ratio will be close to unity; the algorithm will move across the strip tree to evaluate the next sequence of strips. When the ratio deviates from 1.0, the variance values (and the low order structures) differ. Here, tolerance values should be modified.

The two important controls in such an algorithm are first, a control for possible drifts in the progression of low order parameter values, and secondly, a criterion by which to determine how many coordinates should be included in each overlap, and the degree of overlap required. The problem of parameter drift can be controlled by several means: a very simple solution is to compare current low order parameters with the first set of low order parameters computed for the line, and to use this first set as a standard until some significant change is encountered. At this point, the discrepant set becomes the standard, until another change is measured.

The size of an overlap, and the number of strips common between overlaps can most appropriately be controlled by the same variance computations. For line segments (groups of strips) whose low order parameters display large variances, the size of segments should be kept small; conversely, those structures which display homeoscedascity in several low order parameters can most likely be traversed in a few very large sequences with only a small degree of overlap. One could also develop a very complicated algorithm, in which the size and degree of the overlap change at each iteration, based on the first derivative of the variance progression for all parameters; but this may be unnecessary. Nonetheless, the problem of how big a piece of the cartographic line to analyze at a time remains one of the most interesting problems in any application of structural measurements, and the issue will be raised again during the discussion of structure identification.

Structural Identification Tasks

The problems involved in automating the choice of a particular tolerance criterion are very much more difficult than the decisions involved in its modification, although neither of the two is a trivial algorithm. The actual value which is chosen is only one aspect of the problem, and most appropriately determined by psychophysical research: as previously discussed, the measurement of the four sample lines does not provide the author confidence in discussing the particular structure of any one; and the actual tolerance value will most likely depend on a more exacting measurement, based on a larger sample of lines. However, this is one possible direction for further evaluation of the structural measurements derived in this research.

The more relevant aspect of the problem is how to identify a particular line structure, in order to assign a given tolerance criterion.

The high order parameters can be used for this application, to identify what kind of line structure is being dealt with; and the given tolerance value(s) can be stored in the look-up table for each identified category of high order structure. The problem is shifted somewhat by this approach, from choosing a particular tolerance value to the determination of a particular structure category. For this task, the digital look-up table for high order parameters will be used.

High order structure signatures are not stored in a strip tree, but are organized by the frequency of feature occurrence for each category of line. A number of features need to be identified, before the overall structure of the line can be determined. One might consider the feature categories stored in the look-up table as a set of cartographic primitives, in the sense that each describes a graphic unit, a digital description of some geographic shape which might occur in any number of cartographic line types. (Recall for instance the visual similarities between a long fjord and a long reservoir, and how these two feature primitives were clustered in the same category.) In other words, as a line contains a particular combination of feature types, so a line's structure cannot be determined by the structure of a single feature.

Once again, the problem is raised of determining the scope of the line, and the number of features which must be identified. Again, the decision will vary according to the type of line structures being explored. However, consider that structural identification is required only at places along the line where the existing tolerance criteria must be changed. The low order parameters serve to specify these locations, as discussed in the previous section on line discrimination tasks. There is no reason why the same overlapping sections of the line cannot be utilized for both orders of structural measurements.

In order to identify features within any section of an unidentified line, the first task to be solved is how to locate them along the line. Recall from Chapter 5 that the context of high order structure provides for lines to be divided into trend lines and features, and that feature definitions are recursive. In the Chapter 5 definition, a feature is identified at places where a trend line simply bifurcates, as in the confluence of two rivers, or where a trend line can be seen to double back on itself. Features can also be identified along other features, using the above criteria, or by locating sudden changes in feature width.

One might consider the simple bifurcations and the 'doubling back' features in graphic space as a single line symbol and a parallel line symbol, respectively. But actually, the only difference between the two is in width: a simple bifurcation has a line width of one pixel, and the width of a 'doubling back' feature is greater than one pixel. For the identification algorithm, both kinds of features can be defined similarly, as a single line symbol having some thickness, that is, a resolvable and constant width. In concept, identification of any feature is reduced to a search for a single line (of some thickness) bifurcating from the trend.

To understand how the features are located, think for a moment about a line segment, one pixel wide, on the Apple screen. Any point along this line (any pixel) can be characterized as having at most two pixels of the eight possible adjacent to it which are also on the line. If more than two pixels which are adjacent are also part of the line, then one may identify a feature bifurcation.

Next, consider a box, of a size one pixel on a side, as a kind of moving frame, searching along any one of the sample cartographic lines. At each coordinate along the line, this adjacency test can be used to locate feature bifurcations which are one pixel wide. One might also envision larger searching frames, for example, 3 * 3 pixels, or 5 * 5 pixels, which could resolve bifurcations which are one to three pixels wide, or one to five pixels, respectively. Once a feature is located, the searching frame might be reduced in size, and then moved along within the feature to locate smaller features, according to the recursive definition set forth.

As each feature in the line is located, its structure can be measured for high order parameters, as described in Chapter 5. The actual assignment of a feature to a structure category involves matching the mean and variance of measured parameters with the mean and variance parameters for a known category of feature structure. Assignment to a particular feature category is based on maximum similarity in the largest number of parameters. It should be remembered that since each category has a unique standard deviation, it is possible that for certain parameters, large deviations from the cluster mean may still be acceptable for categorization. Weighting schemes can be assigned, to emphasize values for particular parameters. For example, the length of features is probably more important to feature identification than is some particular orientation angle. Different constraints will of course affect the category assignment in different ways.

After all features in a section of line have been categorized, the frequency of observed feature categories may be used to identify the structure of the line. The identification is also a matching process, but what is matched here is the pattern of feature categories observed for the unknown line, against patterns for known lines. In other words, a temporary column can be imagined for Table 5.9. In this temporary column is kept a running tally of features as they are categorized. The tallies are re-expressed as a percentage of all features measured for the unknown line, and the percentages are compared to the other columns in the table.

Comparison involves computing the squared discrepancy for each category, and summing these for all categories; in effect

$$\sum_{i=1}^{n} (p_i - m_i)^2$$

where p is the percentage for a known line
m is the percentage for the measured line
n is the number of categories being compared.

This is a simple least squares calculation: one might consider it as a distance computation in the cluster space of feature categories. In Figure 6.1, a hypothetical space is illustrated for three feature categories, and two known lines identified within it. (In Chapter 5, the cluster space was composed of nine feature categories and four known line types, but this would be difficult to illustrate.) The location of line types in the space is determined by the proportion of features in each catgory which compose them. Feature percentages are thus coordinates, and each category carries equal weight in the identification process (each dimension has the same scale in the cluster space).

The more closely located are two line types in cluster space, the more similar are the proportions of features which they contain, and the more similar are their high order structures. Identification of an unknown line

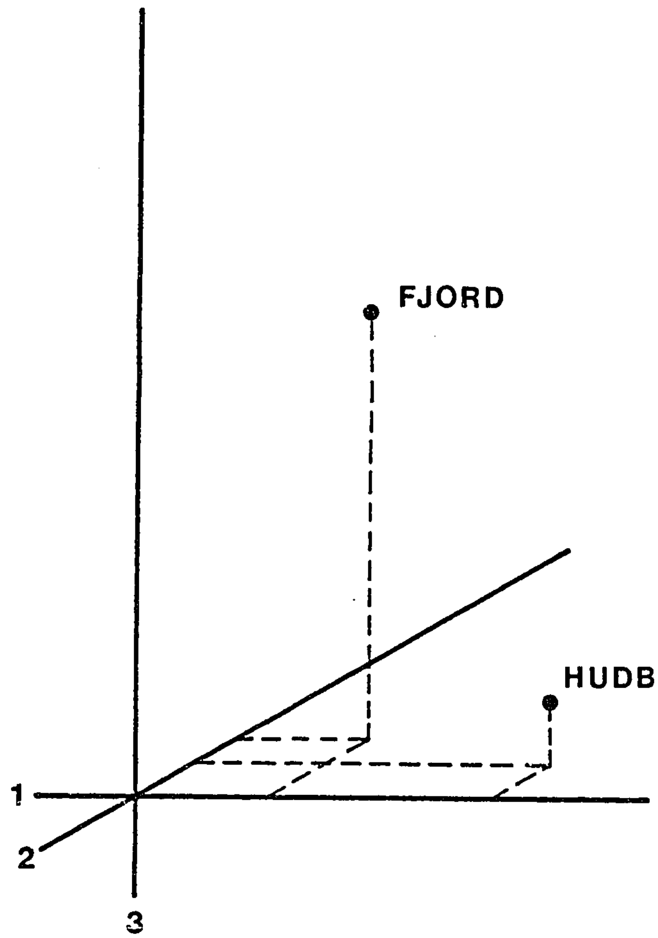


Figure 6.1
LOCATING A LINE TYPE IN CLUSTER SPACE

type will be based on whatever known line type lies closest to it in cluster space, computed as the smallest magnitude distance tally. Several authors (for example Tukey, 1979) recommend that in comparing proportions, one may use some multiple value of the standard error above or below the known values to express a range of confidence. For the line identification, a standard error might be computed as

$$\frac{\sum_{i=1}^{n} (p_i - m_i)^2}{n-2}$$

with variables defined as above.

To summarize, then, the line identification involves two matching tasks. First, the features contained in a line must be located, and measured for parameters of high order structure (including length, width, and orientation, as previously described). These parameters are matched with means and variance values for each of the nine feature categories, and each observed feature is assigned to a particular feature category. A running tally is kept of feature category assignments, and reexpressed in percentage form. The discrepancies between percentages of features in the observed line are compared to percentages for each known (previously identified) line, as a distance computation in cluster space; the measured line is assigned the structure of whichever known line lies closest to it in the cluster space.

Once matched with an existing line structure category, the newly identified line is assigned generalization tolerance values, based on its structural identification. One might also associate a confidence level to each identification, based on distances in the cluster space, to protect against erroneous assignments of deviant lines to a particular category.

Improving the Structural Identifications

There are several limitations to the algorithms previously described. In the case of the low order line discriminations, the accuracy of each discrimination remains a function of the resolution level of the strip tree. This has been demonstrated by the discriminant analysis in Chapter 4, wherein the line categories become less distinguishable as the size of the bandwidths decrease. This is not easily controlled for, except by a constant attention to the scale of the lines being distinguished.

In the case of high order line identification, the problem which is bound to occur is that a line will be introduced for identification which does not fit any of the look-up tables' existing categories. The algorithm which must be open-ended, and flexible during the final matching process. It must be designed in a way that each line presented for identification is not automatically forced into an existing structure category. No matter how exhaustive is the look-up table, there is always the chance that a particular kind of line structure has not been included; however, it is possible to design an algorithm which takes this into account.

What is called for is an identifier routine which modifies its own look-up table on the basis of line structures which are encountered. Two levels of this modification can be discussed, modification of the feature composition within line categories, and modification of the original feature categories. In either situation, the algorithm which is designed falls within the realm of Artificial Intelligence, or more accurately, within the realm of expert systems.

One of the principles of research in Artificial Intelligence (or AI) is to understand the principles which make inductive reasoning possible; however, this does not mean that the intention in applying AI techniques is to mimic human induction (Winston, 1977). For example, the algorithm for computer identification of lines is not presented as a reflection of the way in which a cartographer thinks or reasons. Its purpose is rather to provide an automated means of achieving the goals which a cartographer achieves inductively; as such, it can provide an instance in which AI programming can be applied to solve cartographic problems.

Many applications of AI programming relate to language analysis and to general problem solving; there are also AI systems which are designed to work with a single very specific domain. As these systems are designed to improve their performance over time, they are called expert systems. Mathematics was the domain of the earliest expert systems, as evidenced for example by the MACSYMA system produced at MIT, which solves formulas in integral calculus (McCorduck, 1979).

Other expert systems have since been developed which focus on a variety of tasks; examples include DENDRAL, which interprets the molecular structures displayed by chemical mass spectrograms, MYCIN, which is intended to aid physicians in the diagnosis of bacterial infections, and PROSPECTOR, a consultation system designed to help geologists with problems in mineral exploration. (Duda, 1981) The identifier algorithm discussed in this chapter may be redesigned as a cartographic expert system which improves its ability to identify line structures for generalization tasks.

The method by which performance is improved in an expert system includes appending discovered information to an existing data base, called a knowledge base. The digital look-up tables provide examples of a rudimentary knowledge base. Information can be added in the form of introducing additional details: for example, introducing a new feature category into the digital look-up table would provide new detail in the knowledge base.

Information may also be added in an expert system by introducing new relationships between existing details, as for example, finding a new combination of existing feature categories in order to identify a new line structure. As the knowledge base becomes more complete, the algorithm is able to provide interpretations based on increasingly thorough criteria. The resulting interpretations can be seen to become more refined over time.

Of course, one must anticipate that after several examples of some particular type of line structure have been identified, an update or readjustment of the knowledge base will prove useful. Readjustment implies modification of the means and standard deviation values for high order parameters, in order to 'polish' the summaries of feature categories, by clustering a larger sample of features. The larger sample should in theory result in tighter, more meaningful clusters. This modification may also point out deviant line structures which had been erroneously included within a particular line category.

Implementation of an intelligent identifier algorithm will not involve extensive redesign. A kind of training session is required, and a purposeful presentation of certain types of lines, to allow the knowledge base look-up tables to be further developed. In the jargon of AI, this sort of process is called learning, although it is easy to see that the kind of learning which is proposed here is not intended to mimic the training of a human cartographer. Winston (1977) writes that in training sessions, the system is presented with examples which are good (called 'hits') and those which are 'near-misses'. Each example is identified to the computer as, for example, a good example of a FJORD, or a near-miss example of TEXAS. The algorithm then proceeds through its identification task, locating features, measuring high order parameters, and comparing these with values in the knowledge base. Capabilities must be included to note whatever discrepancies occur, and to use the discrepancies to further modify its interpretive criteria for future line structure examples.

This extension of an identifier algorithm into the realm of AI programming is a very simplified discussion, but the reader should not infer that such an extension is a trivial programming accomplishment. Many aspects of the learning algorithm have not been included, for example the data structure of the knowledge base, or the detailed mechanics of searching through it during a matching task. The study of knowledge representation forms a significant part of AI research, and an obvious next step in the present research, to study and compare how various representation schemes may be applied to the learning of graphic information.

For the time being, these are preliminary remarks, intended to show how it is that intuitive tasks may be considered within a programming context. The dissertation has in effect served as a feasibility study, a pre-test; and this discussion is presented as a proposal that line identifications may be automated by means of the structure signatures which have been developed, and applied to computer generalization tasks. And while some gaps are evident in the procedures described here, it is suggested that AI programming techniques may provide a useful means by which many of the gaps may be filled in.

Summary

The purpose of this chapter has been to discuss several applications of structural measurements to cartographic generalization. One example illustrates the enhancement of line features, as stochastic models, and the concepts of probability in representing cartographic features; these concepts have been reiterated throughout the dissertation. A cartographic line is a probabilistic approximation of the geographic feature which it represents: as such, conventional assumptions about the accuracy of a generalized representation may be too stringent. Once these assumptions are relaxed, and graphic noise permitted to augment existing information, the representation can be enhanced as well as simplified. Generalization is expanded to a modelling process, and the stochastic line generation algorithm presented in this chapter provides one example of what can be accomplished. Future directions associated with research in stochastic models include design of perceptual tests of these representations, to evaluate map reader response to their inclusion in thematic maps.

The second application discussed in this chapter addresses a long-time obstacle to the automation of more conventional generalization tasks, namely the selection and modification of tolerance threshold criteria. The decisions made by a cartographer are often inductive, and based on some visual evaluation of shape and structure in the generalized line.

> In psychology, the study of intuitive perceptions of shape is entirely respectable, and has developed into the important branch known as Gestalt. In mathematics, on the contrary, to call a process intuitive is to condemn it as unanalyzed.
>
> Richardson, 1961, p.143

The purpose of the structural measurements made in this dissertation has not been to express some exhaustive typology of cartographic lines, but instead to demonstrate that structural measurements can be simply accomplished by automated means. The structural parameters can be applied to a variety of cartographic tasks, including line identification problems previously thought to be difficult to automate. Sketches of several algorithms have been presented by which identifications may be included in generalization routines, and also by which the identification capabilities may be improved.

In the final analysis, there are many cartographic problems which are considered 'not easily automated' due to their inductive nature. With the capabilities made available through research in Artificial Intelligence programming, many of these intuitive problems prviously abandoned by some computer scientists as 'not analyzeable' can be more easily addressed. One of the most difficult examples is found in the automation of many types of graphic displays, including the generalization of cartographic lines.

The most difficult part of the problem is not how to automate the intuitive aspects of cartography, but how to restate the intuitive aspects of the cartographic process so as to permit its automation. The work presented here has performed exactly this sort of restatement. Focusing on the aspects of the representations instead of on the recognition of a line has provided a means by which rudimentary line structures may be automated, and possibly imporved upon. It remains to be seen whether the same approach will be as fruitful with other problems commonly encountered in our field.

Conclusion

In this dissertation research, a procedure has been developed by which to derive digital definitions of graphic structure. The procedure has been applied to the generalization of cartographic lines, whose graphic structure has been shown to change with changes in the scale of their graphic representation. Traditional definitions of cartographic lines have been based on geomorphic process, or alternatively on perceptual response, that is, map reader impressions. The definitions derived here are based on changes in the geometry of a graphic representation as it is represented at a variety of map scales, and within a range of resolutions. It has been demonstrated that these geometric definitions lend themselves well to digital description tasks.

The purpose of the structural definitions is threefold. First, one may use them to distinguish between lines of differing visual characteristics. Secondly, the definitions may be used to identify particular line structues, or at least to categorize observed structures as being similar to, or different from, previously identified graphic structures. One may also apply the digital definitions to generate cartographic lines of a predictable graphic structure, in a stochastic modelling process.

The work has been discussed within the context of the generalization of cartographic lines. Here, the bottleneck to automation lies in the modification and selection of tolerance criteria. The development of consistent digital definitions of line character provides a means for tolerance modification by line distinction tasks, and a means for tolerance value assignments using line identification procedures. Algorithms for both aspects have been presented in the dissertation, and their application to aspects of automated generalization have been discussed.

The research discussed here does not provide a complete solution, by any means. The specific tolerance values which should be applied to a fjord or to a sandy beach in order to generalize its graphic representation are assumed to be given, for the purpose of this dissertation. This is a big assumption, as thorough guidelines have not yet been developed. The small sample of cartographic lines used in this study does not afford confident descriptions of the particular graphic structure of a fjord, or of a sandy beach. More appropriate description will require larger samples, and empiric evaluation of map reader impressions; the geometric procedures should not stand alone in defining cartographic structure.

On the other hand, this research does contribute a methodology by which one may arrive at digital descriptions of geometry as it changes with scale, and furthermore definitions which, once derived, may be expanded upon. This work lays the foundations for a generalization package which becomes more proficient over time, and is presented as a rudimentary application of Artifical Intelligence to a specific cartographic problem.

In the end, the aspects of this work which deal directly with the generalization of cartographic lines are not of primary importance: the application has provided a viable means by which to develop methods for structural description, and will additionally provide for evaluation of these methods. One must realize, however, that these methods may be extended from the realm of lines into two or three dimensions, for structural descriptions of terrain; and this extension may be applied to digital representations for flight simulators, navigational tasks, and other cartographic landscape displays.

Also, one might use these methods to analyze and compare structures of statistical surfaces and statistical solids, for example the data clusterings which result from factor analyses, scaling tasks, and discriminant functions. Here, too, the structure of the data will depend on the resolution at which it is collected, and analyzed; and the methods derived in this dissertation may provide a digital description by which the analyses and comparisons may be carried out.

What is of primary importance in this research is the development of a set of digital techniques which encompass scale-dependant changes in structure, and which are developed in a format which allows refinement of existing information, and implementation of new information as well. The cartographic application demonstrates only one example of the scale-dependant structures which can be analyzed. It is intended that the description techniques may be applied to the problems of other disciplines, as well.

BIBLIOGRAPHY

Attneave, F. (1954) "Some Informational Aspects of Visual Perception". Psychological Review vol.61, p.183-193.

Attneave, F. & M.P. Arnoult (1956) "Quantitative Study of Shape and Pattern Perception". Psychological Bulletin, vol.53(6), p.452-471.

Attneave, F. (1959) Applications of Information Theory to Psychology. New York: Holt, Rinehart, and Winston.

Ballard, D.H. (1981) "Strip-Trees: A Hierarchical Representation for Curves". Communications, Association for Computing Machinery, vol.24(5), p.310-321.

Barr, A. & E.A. Feigenbaum (1981) The Handbook of Artificial Intelligence. Volumes 1 and 2. Los Altos, California: William Kaufmann, Inc.

Barstow, D.R. (1979) Knowledge-Based Program Construction. New York: American-Elsevier.

Bartsch, H.J. (1974)(Trans. H. Liebscher) Handbook of Mathematical Formulas. New York: Academic Press.

Boyle, A.R. (1970) "The Quantized Line". Cartographic Journal, vol.7(2), p.91-94.

Boyle, A.R. (1981) "Concerns About the Present Applications of Computer-Assisted Cartography". Cartographica, vol.18(1), p.31-33.

Brophy, D.M. (1973) "An Automated Methodology for Linear Generalization in Thematic Cartography". Proceedings, American Congress on Surveying and Mapping, 33rd Annual Meeting, p.300-314.

Bunge, W. (1969) "Research Note on Simplicity". Geographical Analysis, vol.1(4), p.388-391.

Burrough, P.A. (1981) "Fractal Dimensions of Landscapes and Other Environmental Data". Nature, vol.294(5838), p.240-242.

Burrough, P.A. (1982a) "Problems of Superimposed Effects in the Statistical Study of the Spatial Varieties of Soil". Paper presented to the EGS Symposium on Spatial Variability of Soils, Leeds, England.

Burrough, P.A. (1982b) "Multiscale Sources of Spatial Variation in Soil: Part I - The Application of Fractal Concepts to Nested Levels of Soil Variation". Submitted to Journal of Soil Science (1983).

Carpenter, L.C. (1981) "Computer Rendering of Fractal Curves and Surfaces". Proceedings, ACM SIGGRAPH Conference, Seattle, Washington.

Chaikin, G.M. (1974) "An Algorithm for High Speed Curve Generation". Computer Graphics and Image Processing, vol.3, p346-349.

Chasen, S.H. (1978) Geometric Principles and Procedures for Computer Graphic Applications. Englewood, N.J.: Prentice Hall, Inc.

Chrisman, N.R. (1981) "Methods of Spatial Analysis Based on Maps of Categorical Coverages". Unpublished Ph.D. Thesis, Department of Geography, University of Bristol, England.

Church, M. & D.M. Mark (1980) "On Size and Scale in Geomorphology". Progress in Physical Geography, vol.4(3), p.342-390.

Davis, J.C. & M.McCullagh (eds.)(1981) Display and Analysis of Spatial Data. New York: John Wiley & Sons.

Davis, R. & D.B. Lenat (1982) Knowledge-Based Systems in Artificial Intelligence. New York: McGraw-Hill International.

Dent, B.D. (1972) "A Note on the Importance of Shape in Cartogram Communication." Journal of Geography, vol.71(7), p.393-401.

Dobson, M.W. (1980) "The Influence of the Amount of Graphic Information on Visual Matching". Cartographic Journal, vol.17(1), p.26-32.

Douglas, D.H. & T.K. Peucker (1973) "Algorithms for the Reduction of the Number of Points Required to Represent a Digitzied Line or its Caricature". Canadian Cartographer, vol.10(2), p.110-122.

Duda, R.O. (1981) "Knowledge-Based Expert Systems". BYTE Magazine, vol.6(9).

Duncan, J. (1981) "Directing Attention in the Visual Field". Perception and Psychophysics, vol.30(1), p.90-93.

Dutton, G.H. (1981) "Fractal Enhancement of Cartographic Line Detail". American Cartographer, vol.8(1), p.23-40.

Dyer, C.R. & A.Rosenfeld (1979) "Thinning Algorithms for Gray-Scale Pictures". IEEE Transactions in Pattern Analysis and Machine Intelligence, vol.1, p.88-89.

Fleischer, D. (1982) "Computer Animation: The Art of Hyper-Reality". Science Digest, vol.90(2), p.46-54.

Fournier, A. & D. Fussell (1980) "Stochastic Modelling in Computer Graphics". Proceedings, ACM SIGGRAPH Conference, Seattle, Washington.

Fournier, A., D.Fussell, and L.C.Carpenter (1982) "Computer Rendering of Stochastic Models". Communications, Association for Computing Machinery, vol.25(6), p.371-384.

Freund, J.E. (1971) Mathematical Statistics. New Jersey: Prentice-Hall, Inc.

Gardner, Martin (1978) "White and Brown Music, Fractal Curves and 1/f Fluctuations". Scientific American, Mathematical Games, vol.238(4), p.16-31.

Goodchild, M.F. (1980) "Fractals and the Accuracy of Geographic Measures". International Association for Mathematical Geology Journal, vol.12(2), p.85-98.

Gordon, W.J. & R.F. Riesenfeld (1974) "Bernstein-Bezier Methods for the Computer-Aided Design of Free-Form Curves and Surfaces". Journal of the Association for Computing Machinery, vol.21(2), p.293-310.

Gould, S.J. (1966) "Allometry and Size in Ontogeny and Phylogeny". Biological Reviews, vol.41(4), p.587-641.

Huxley, J.S. (1932) Problems of Relative Growth. London. As referenced in Gould, S.J. (1966) "Allometry and Size in Ontogeny and Phylogeny"; Biological Reviews, vol.41(4), p.587-641.

Jenks, G.F. (1963) "Generalization in Statistical Mapping". Annals, Association of American Geographers, vol.53(1), p.15-26.

Jenks, G.F. (1981) "Lines, Computers, and Human Frailties". Annals, Association of American Geographers, vol.71(1), p.1-10.

Jimenez, J. & J.L.Navalon (1982) "Some Experiments in Image Vectorization". IBM Journal of Resource Development, vol.26(6), p.724-734.

Johnston, R.J. (1978) Multivariate Statistical Analysis in Geography. London: Longman.

Kelley, P.S. (1977) "Information and Generalization in Cartographic Communication". Unpublished Ph.D. Thesis, Department of Geography, University of Washington.

Koeman, C. & F.L.T. Van der Weiden (1970) "The Application of Computation and Automatic Drawing Instruments to Structural Generalization". Cartographic Journal, vol7(1), p.47-49.

Krishnamurti, R. (1981) "The Construction of Shapes". Environment and Planning B, vol.8(1), p.5-40.

Lang, T. (1969) "Rules for Robot Draughtsmen". Geographical Magazine, vol.42(1), p.50-51.

Lavin, S.J. (1979) "Region Perception Variability on Choropleth Maps: Pattern Complexity Effects" Unpublished PhD Thesis, Department of Geography, University of Kansas.

Lenat, D.B. & J.McDermott (1977) "Less-Than-General Production System Architectures". Proceedings, International Joint Conferences on Artificial Intelligence, vol.5, p.928-932.

Lipkin, B.S. & A. Rosenfeld (1970) Picture Processing and Psychophysics. New York: Academic Press.

MacEachren, A.M. (1982) "Map Complexity: Comparison and Measurement". American Cartographer, vol.9(1), p.31-46.

Maling, D.H. (1968) "How Long is a Piece of String?". Cartographic Journal, vol.5(2), p.147-156.

Mandelbrot, B.B. (1967) "How Long is the Coast of Britain? Statistical Self-Similarity and Fractal Dimension". Science, vol.154, p.636-638.

Mandelbrot, B.B. (1977) Fractals: Form, Chance, and Dimension. San Francisco: W.H. Freeman Co.

Mandelbrot, B.B. (1982) The Fractal Geometry of Nature. San Francisco: W.H. Freeman Co.

Marino, J. (1979) "Identification of Characteristic Points Along Naturally Occurring Lines: An Empirical Study". Canadian Cartographer, vol.16(1), p.70-80.

Mark, D.M. (1975) "Geomorphometric Parameters: A Review and Evaluation". Geografiska Annaler, vol.57A, p.165-178.

McMaster, R. (1981) "Analysis of Error in Linear Simplification". Paper presented spring meetings Association of American Geographers, Los Angeles, California.

McMaster, R. (1982) "An Analysis of Line Simplification Algorithms". Paper presented spring meetings Association of American Geographers, San Antonio, Texas.

Monmonier, M.S. (1974) "Measures of Pattern Complexity for Choroplethic Maps". American Cartographer, vol.1(2), p.159-169.

Muehrcke, P.C. (1969) "Visual Pattern Analysis: A Look at Maps". Unpublished Phd Thesis, Department of Geography, University of Michigan.

Muller, J.C. (1979) "Perception of Continuosly-Shaded Maps". Annals, Association of American Geographers, vol.69(2), p.240-249.

Nilsson, N.J. (1980) Principles of Artificial Intelligence. Palo Alto, California: Tioga Publishing Company.

Nystuen, J.D. (1966) "Effects of Boundary Shape and the Concept of Local Convexity". Michigan Inter-University Community of Mathematical Geographers, Discussion Paper #10.

Pannekoek, A.J. (1962) "Generalization of Coastlines and Contours". International Yearbook of Cartography, vol.2, p.55-74.

Pavlidis, T. (1978) "A Review of Algorithms for Shape Analysis". Computer Graphics and Image Processing, vol.7, p.243-258.

Perkal, J. (1966) (Trans. R. Jackowski) "On the Length of Empirical Curves". Michigan Inter-University Community of Mathematical Geographers, Discussion Paper #10.

Perkal, J. (1966) (Trans. R. Jackowski) "An Attempt at Objective Line Generalization". Michigan Inter-University Community of Mathematical Geographers, Discussion Paper #10.

Peucker, T.K. (1975) "A Theory of the Cartographic Line". Proceedings, AUTO-CARTO II, Reston, Virginia, p.508-518.

Peucker, T.K. & N.Chrisman (1975) "Cartographic Data Structures". American Cartographer, vol.2(1), p.55-69.

Pfaltz, J.L. & A.Rosenfeld (1967) "Computer Representation of Planar Regions by Their Skeletons". Communicatons, Association for Computing Machinery, vol.10, p.119-122.

Poiker, T.K., R.Squirrell, & S.Xie (1981) "The Use of Computer Science and Artificial Intelligence in Cartographic Design". Unpublished manuscript, Simon Fraser University, Canada.

Prince, V. (1977) "Characteristics of Map Recall of a Single Regional Shape". Unpublished MA. Thesis, Department of Geography, University of Kansas.

Raisbeck, G. (1963) Information Theory: An Introduction for Scientists and Engineers. Cambridge: University Press.

Raisz, E.J. (1948) General Cartography. New York: McGraw-Hill, Inc.

Ramer, U. (1972) "An Iterative Procedure for the Polygonal Approximation of Plane Curves". Computer Graphics and Image Processing, vol.1(3).

Rechlin, A.T.M. (1972) "The Epsilon Generalizaton". Annals, Association of American Geographers, vol.62(4), p.578-581.

Rhind, D.W. (1973) "Generalisation and Realism Within Automated Cartographic Systems". Canadian Cartographer, vol.10(1), p.51-63.

Richardson, L.F. (1961) "The Problem of Contiguity: An Appendix to The Statistics of Deadly Quarrels". General Systems Yearbook, vol.6, p.139-187.

Riesenfeld, R.F. (1975) "On Chaikin's Algorithm". Computer Graphics and Image Processing, vol.4(3), p.304-310.

Robinson, A.H. & R.D.Sale (1969) Elements of Cartography. New York: Wiley & Sons, Inc. (3rd Ed.)

Rosenfeld, A. (1975) "A Characterization of Parallel Thinning Algorithms". Information Control, vol.29, p.286-291.

Rosenfeld, A. & A.C.Kak (1982) Digital Picture Processing. New York: Academic Press. (Volume 2)

Shannon, C.F. & W.Weaver (1949) The Mathematical Theory of Communication. Urbana: University of Illinois Press.

Spencer-Brown, G. (1979) Laws of Form. New York: E.P. Dutton.

Srnka, E. (1970) "The Analytical Solution of Regular Geneeralization in Cartography". International Yearbook of Cartography, vol.10, p.48-60.

Stefanelli, R. & A.Rosenfeld (1971) "Some Parallel Thinning Algorithms for Digital Pictures". Journal, Association for Computing Machinery, vol. 18(2), p.255-264.

Steinhaus, H. (1954) "Length, Shape, and Area". Colloquium Mathematica, vol.3.

Steinhaus, H. (1960) Mathematical Snapshots. Oxford:University Press.

Stevens, P.S. (1974) Patterns in Nature. Boston: Little, Brown, and Company.

Steward, H.J. (1974) "Cartographic Generalisation: Some Concepts and Explanation". Cartographica Monograph #10.

Strahler, A.N. & A.H. Strahler (1978) Modern Physical Geography. New York: John Wiley & Sons.

Sukhov, V.I. (1970) "Application of Information Theory in Generalization of Map Contents". International Yearbook of Cartography, vol.10, p41-46.

Thompson, D.W. (1961) On Growth and Form. Cambridge: Cambridge University Press.

Tobler, W.R. (1964) "An Experiment in the Computer Generalization of Maps". Technical Report #1, ONR Task 389-137, Contract NONR 1224(48).

Tobler, W.R. (1966) "Numerical Map Generalization". Michigan Inter-University Community of Mathematical Geographers, Discussion Paper Number 8.

Tobler, W.R. (1969) "Geographic Filters and Their Inverses". Geographic Analysis, vol.1(3), p.234-253.

Topfer, R. & W. Pillewizer (1966) "Principles of Selection". Cartographic Journal, vol.3(1), p.10-16.

Tukey, J. (1977) Exploratory Data Analysis. Cambridge: MIT Press.

Turner, E. (1977) "Use of Shape as a Nominal Variable on Multi-Pattern Dot Maps". Unpublished PhD. Thesis, Department of Geography, University of Washington.

Waterman, D.A. & Hayes-Roth (Eds.) (1978) Pattern-Directed Inference Systems. New York: Academic Press.

Whitted, T. (1982) "Special Issue on Computers and Electronics" Science, vol.215, p.767.

Williams, C.M. (1981) "Bounded Straight-Line Approximation of Digitized Planar Curves and Lines". Computer Graphics and Image Processing, vol.16(4), p.370-381.

Winston, P.H. (1977) Artificial Intelligence. Reading, Massachusetts: Addison-Wesley Publishing Company.

Zusne, L. (1970) Visual Perception of Form. New York: Academic Press.

APPENDIX A
INITIAL BANDS FOR EACH SAMPLE LINE

For each of the four cartographic lines, a set of eight points was chosen, to initialize the strip tree data structure, to bootstrap the bandwidth subdivision process, and to simulate a cartographer's lack of control in generalization about retention of those points critical to geographic accuracy. The procedure has been described in detail in Chapter 4; what follows here is an illustration of the bands and the geometric measurements which were made, and stored in the data structure.

The lines are presented in order, first SIOUX, then HUDBY, FJORD, and TEXAS. Each line has been subdivided into eight bands, and the heading of each band indicates which two geographic points bound the particular strip of the line (for example, STRIP 1 -- 2). Following this is the linear equation describing the band anchor line, its length, slope, and endpoints, and the number of further subdivisions to be made for the given band (see 'Levels Below'). Also tabulated are values for the five low order parameters. Please note that throughout this appendix, the segmentation parameter has been labelled as the bifurcation segment.

```
STRIP     1 --    2
     EQUATION FOR THIS STRIP IS    -14X +    -25Y +  4328.00
     LENGTH OF STRIP = -28.65                          STRIP SLOPE = -29.25
     ENDPOINTS   ( 252,   32)    ( 277,   18)     LEVELS BELOW =    0
     MONOTONICITY=  0.0909
     ERROR VARIANCE=  1.72                          NPTS IN STRIP=  11
     STRIP WIDTH=  0.16                        BIFURCATION SEGMENT=  0.72
     MAX DEVIATION= -0.12                      NEXT STRIP NODE AT RECORD      8
```

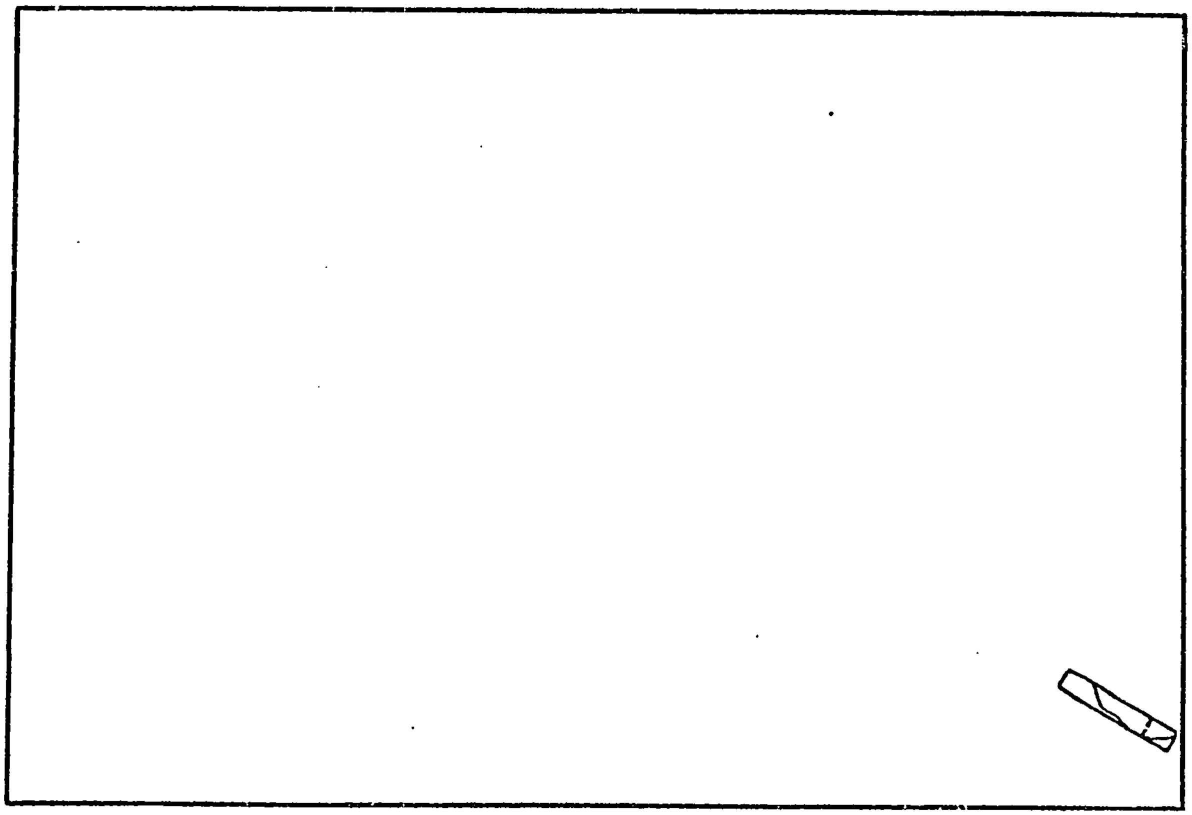


FIGURE A.1 a   SIOUX (Band 1)

```
STRIP     2 --    3
     EQUATION FOR THIS STRIP IS  -156X +    -83Y +  41968.0
     LENGTH OF STRIP = -176.71                        STRIP SLOPE = -61.98
     ENDPOINTS    ( 169, 188)     ( 252,   32)      LEVELS BELOW =    0
     MONOTONICITY=  0.0580
     ERROR VARIANCE=  14.98                          NPTS IN STRIP=  69
     STRIP WIDTH=  0.10                       BIFURCATION SEGMENT=  0.64
     MAX DEVIATION= -0.08                     NEXT STRIP NODE AT RECORD    45
```

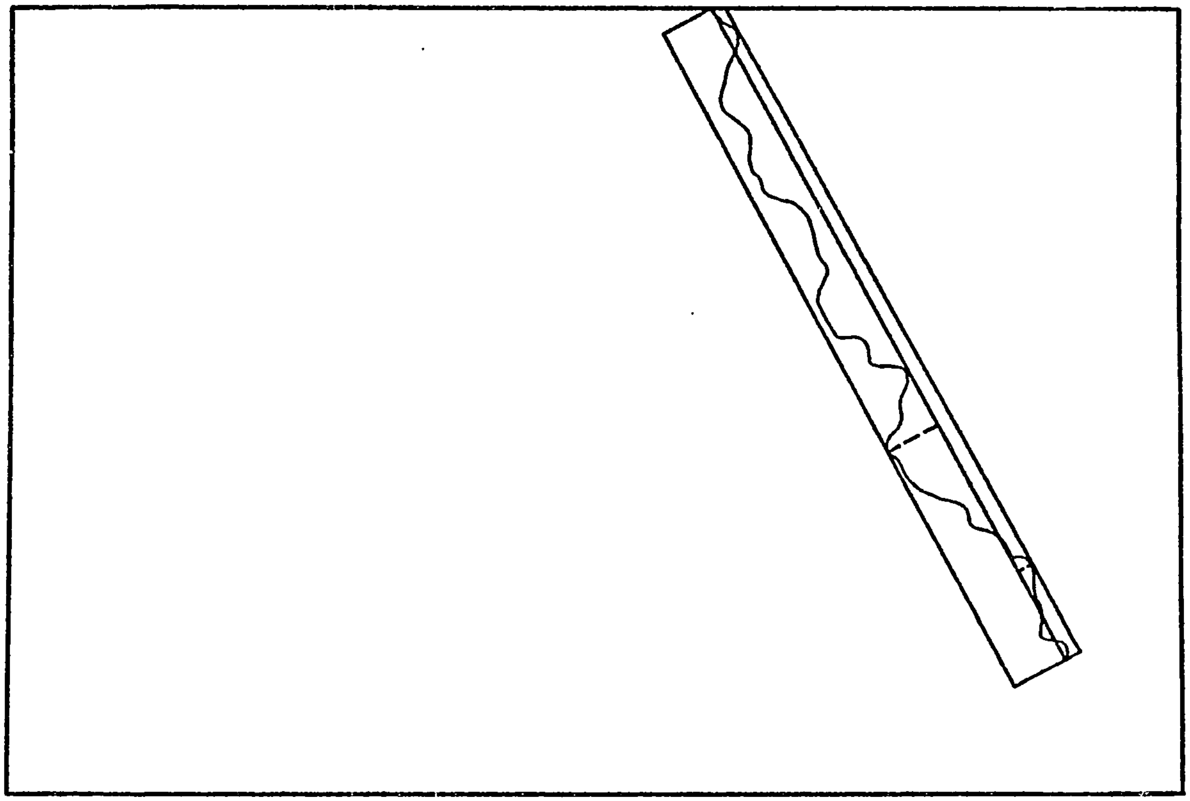


FIGURE A.1 b   SIOUX (Band 2)

```
STRIP    3 --   4
     EQUATION FOR THIS STRIP IS     14X +    -78Y + -1032.00
     LENGTH OF STRIP = -79.25                        STRIP SLOPE =  10.18
     ENDPOINTS    ( 174,   18)    ( 252,   32)     LEVELS BELOW =    0
     MONOTONICITY=  0.0714
     ERRCR VARIANCE=  8.45                          NPTS IN STRIP=  28
     STRIP WIDTH=  0.12                       BIFURCATION SEGMENT=  0.24
     MAX DEVIATION=  0.11                     NEXT STRIP NODE AT RECORD      9
```

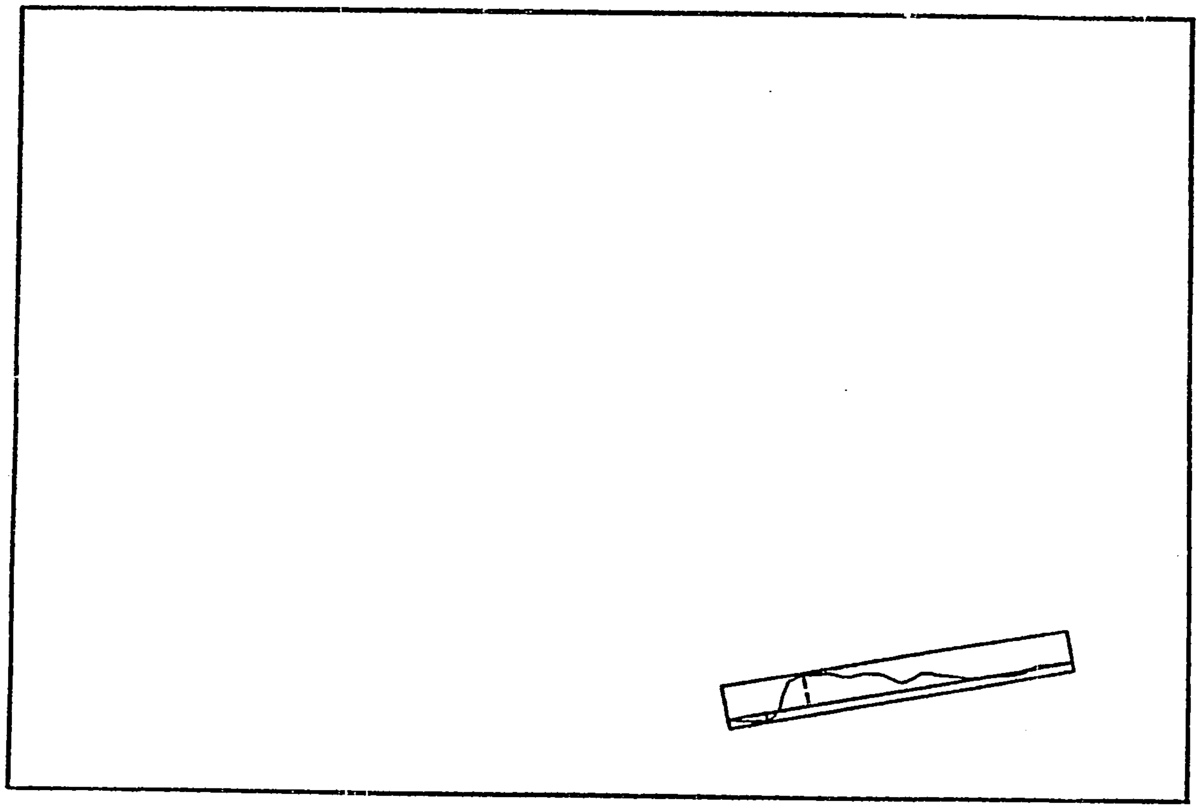


FIGURE A.1 c SIOUX (Band 3)

```
STRIP     4 --    5
     EQUATION FOR THIS STRIP IS     -9X +    172Y + -1530.00
     LENGTH OF STRIP =  172.24
                                                    STRIP SLOPE = -177.00
     ENDPOINTS    ( 174,  18)    (   2,   9)      LEVELS BELOW =   0
     MONOTONICITY=  0.0635
     ERROR VARIANCE=  42.04                          NPTS IN STRIP=  63
     STRIP WIDTH=  0.21                      BIFURCATION SEGMENT=  0.62
     MAX DEVIATION= -0.13                    NEXT STRIP NODE AT RECORD     40
```

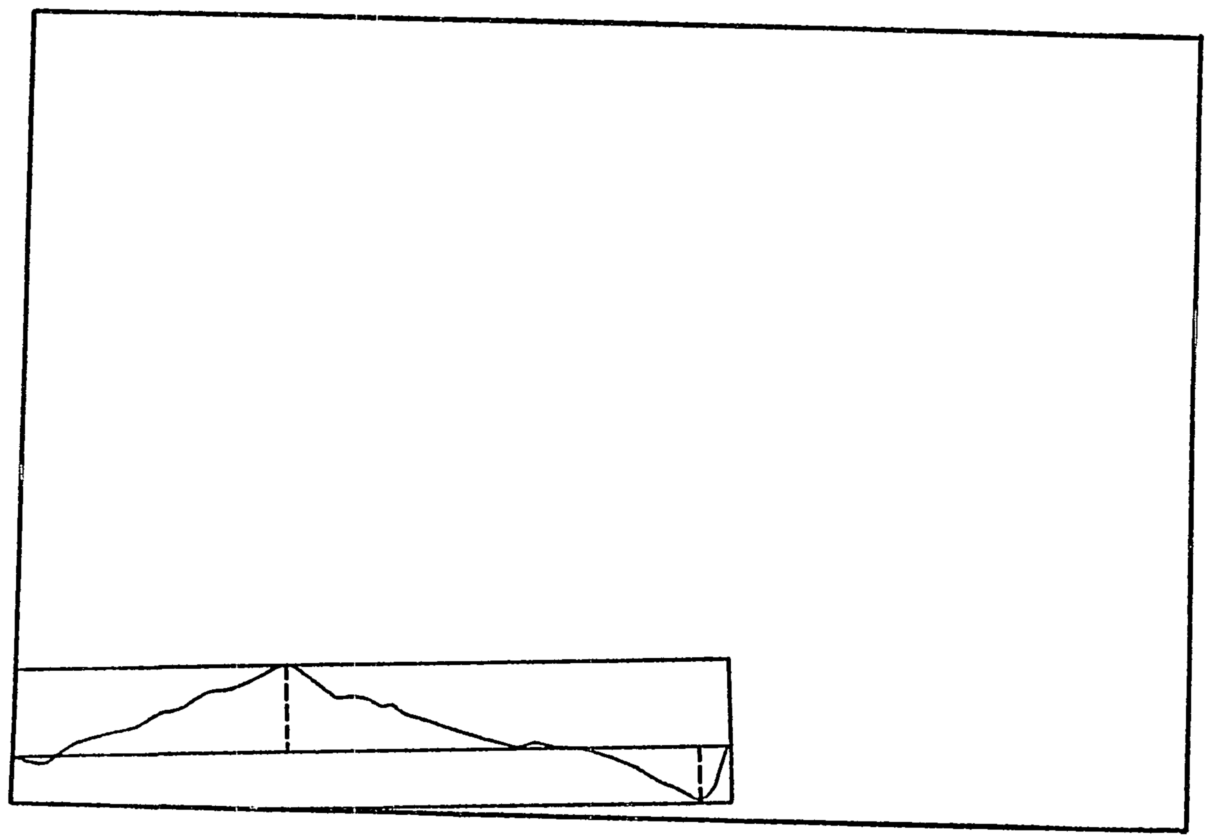


FIGURE A.1 d   SIOUX (Band 4)

```
STRIP     5 --    6
     EQUATION FOR THIS STRIP IS     40X +     60Y + -8040.00
     LENGTH OF STRIP =  72.11                        STRIP SLOPE =  146.31
     ENDPOINTS    ( 174,   18)    ( 114,   58)      LEVELS BELOW =    0
     MONOTONICITY=  0.0000
     ERROR VARIANCE=  3.43                         NPTS IN STRIP=  27
     STRIP WIDTH=  0.08                      BIFURCATION SEGMENT=  0.63
     MAX DEVIATION=  0.08                    NEXT STRIP NODE AT RECORD     17
```

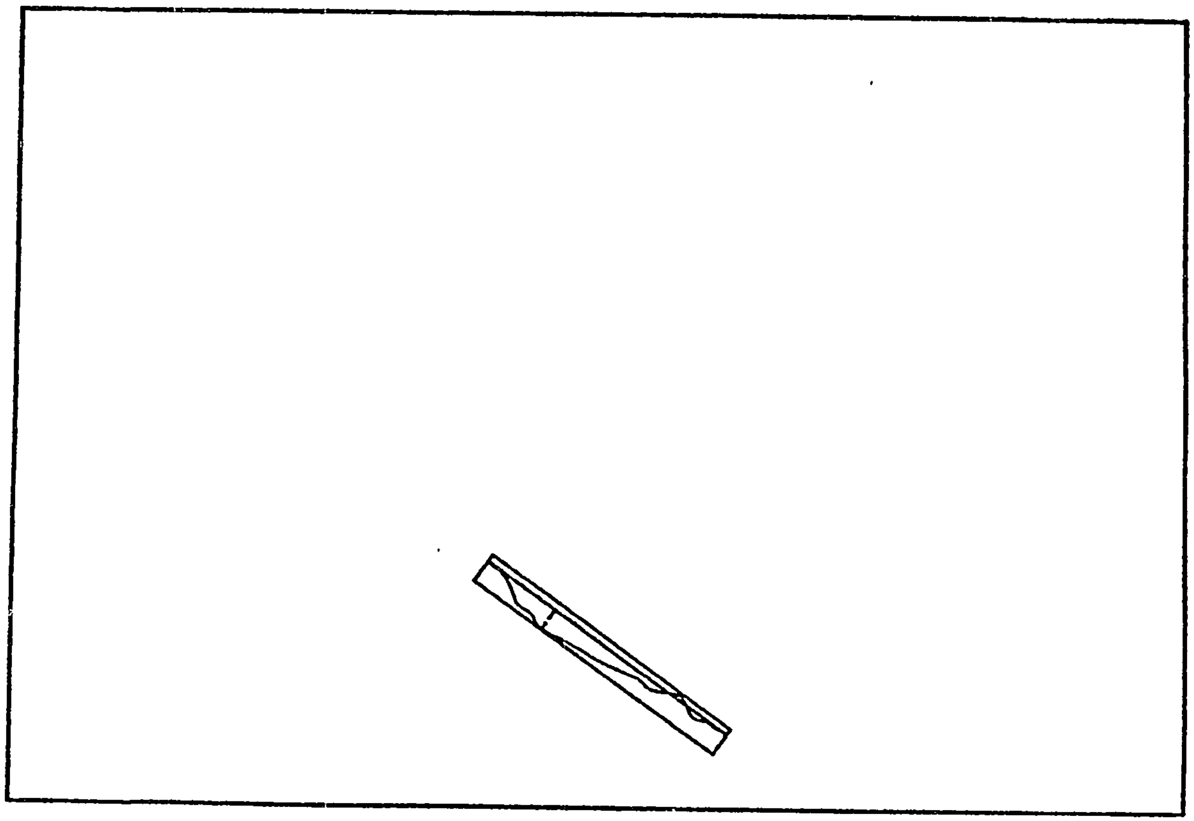


FIGURE A.1 e   SIOUX (Band 5)

```
STRIP     6 --    7
     EQUATION FOR THIS STRIP IS  -128X +   -89Y +  19754.0
     LENGTH OF STRIP = -155.90                      STRIP SLOPE = -55.19
     ENDPOINTS    (  25, 186)    ( 114,  58)     LEVELS BELOW =    0
     MONOTONICITY=  0.0606
     ERROR VARIANCE=  87.04                        NPTS IN STRIP=  66
     STRIP WIDTH=  0.21                      BIFURCATION SEGMENT=  0.40
     MAX DEVIATION= -0.20                    NEXT STRIP NODE AT RECORD     31
```

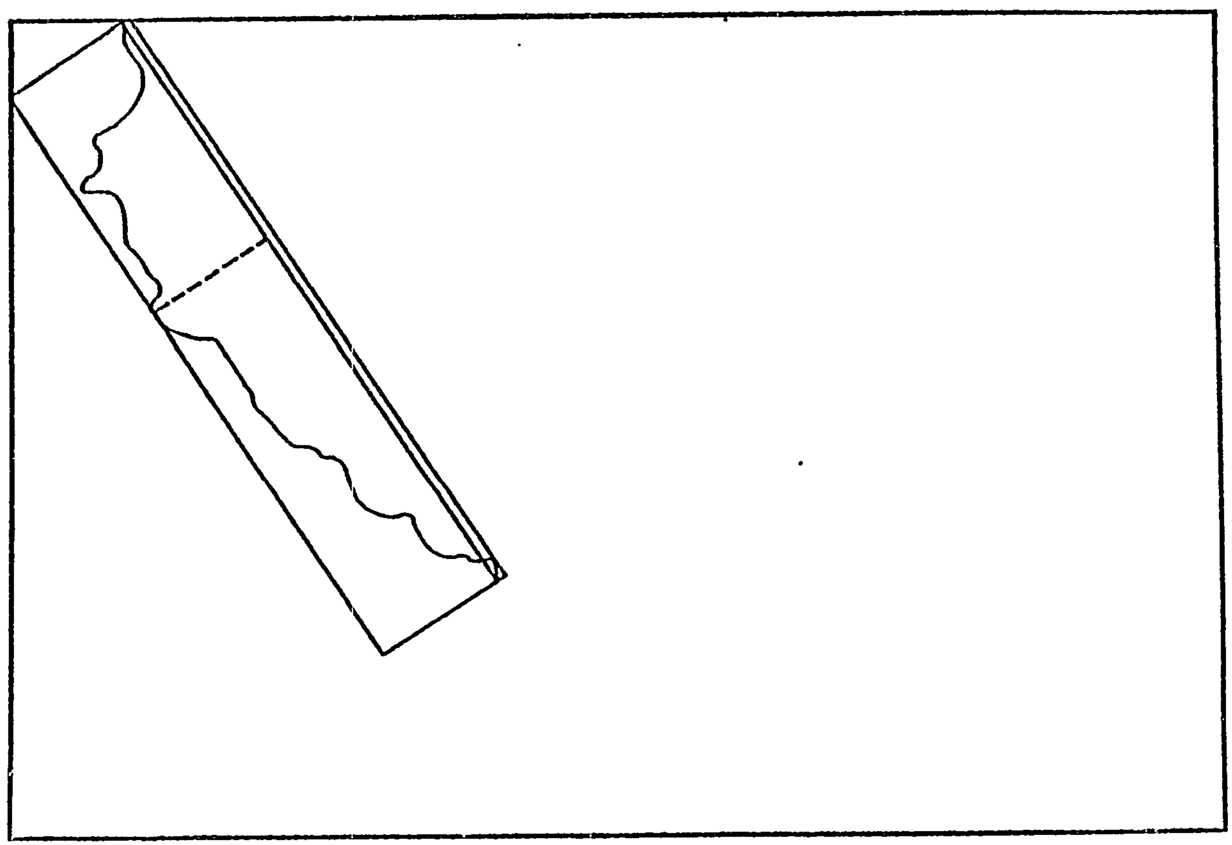


FIGURE A.1 f   SIOUX (Band 6)

```
STRIP     7 --    8
     EQUATION FOR THIS STRIP IS     53X +     69Y + -10044.0
     LENGTH OF STRIP =  87.01                      STRIP SLOPE =  142.47
     ENDPOINTS    ( 114,   58)     (  45, 111)     LEVELS BELOW =    0
     MONOTONICITY=  0.0000
     ERROR VARIANCE=  10.83                        NPTS IN STRIP=  33
     STRIP WIDTH=  0.13                     BIFURCATION SEGMENT=  0.17
     MAX DEVIATION=  0.13                   NEXT STRIP NODE AT RECORD      6
```

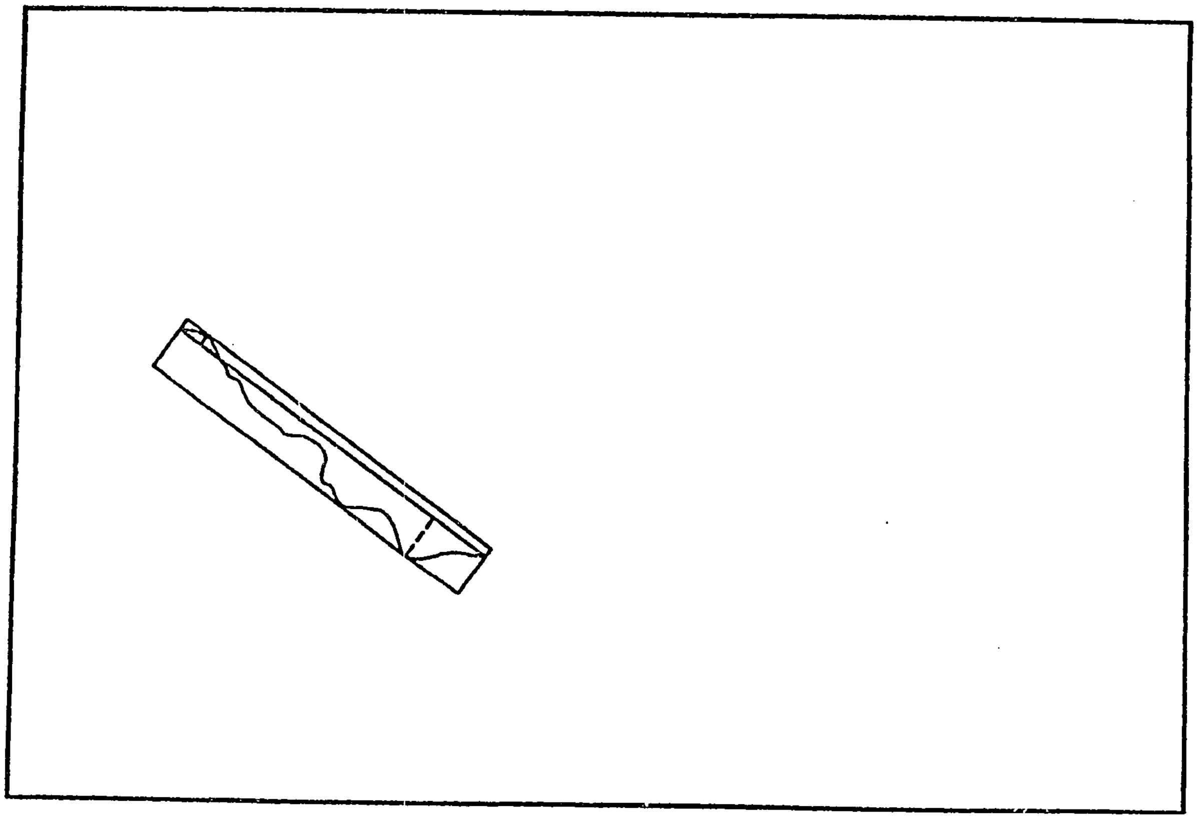


FIGURE A.1 g   SIOUX (Band 7)

```
STRIP     8 --    9
    EQUATION FOR THIS STRIP IS     75X +     20Y + -5595.00
    LENGTH OF STRIP =  77.62                        STRIP SLOPE =  104.93
    ENDPOINTS    (  45, 111)    (  25, 186)     LEVELS BELOW =    0
    MONOTONICITY=  0.0556
    ERROR VARIANCE=  41.56                          NPTS IN STRIP=  36
    STRIP WIDTH=  0.31                         BIFURCATION SEGMENT=  0.60
    MAX DEVIATION=  0.28                       NEXT STRIP NODE AT RECORD    20
```

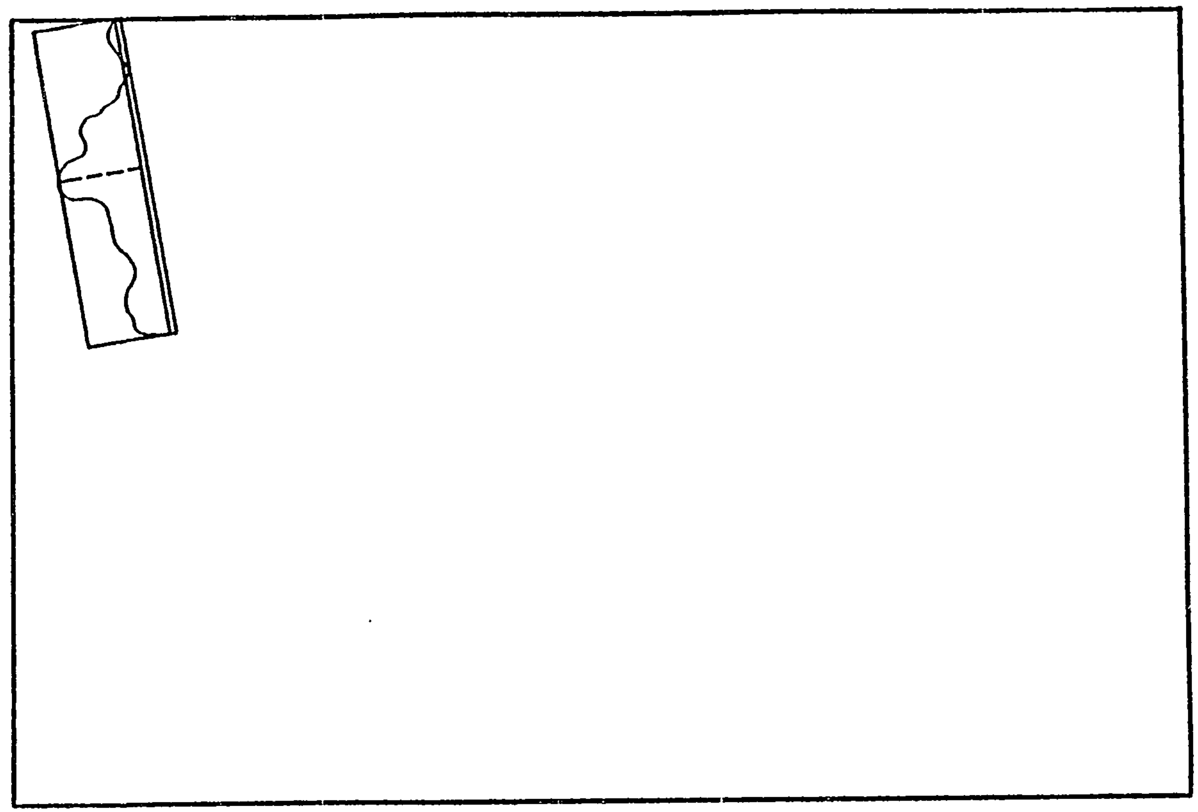


FIGURE A.1 h   SIOUX (Band 8)

```
STRIP     1 --    2
     EQUATION FOR THIS STRIP IS     67X +     48Y + -17650.0
     LENGTH OF STRIP =  82.42                          STRIP SLOPE =  125.62
     ENDPOINTS    ( 262,    2)    ( 214,   69)      LEVELS BELOW =    0
     MONOTONICITY=  0.0300
     ERROR VARIANCE=  68.16                          NPTS IN STRIP= 100
     STRIP WIDTH=  0.49                       BIFURCATION SEGMENT=  0.29
     MAX DEVIATION= -0.41                     NEXT STRIP NODE AT RECORD     38
```

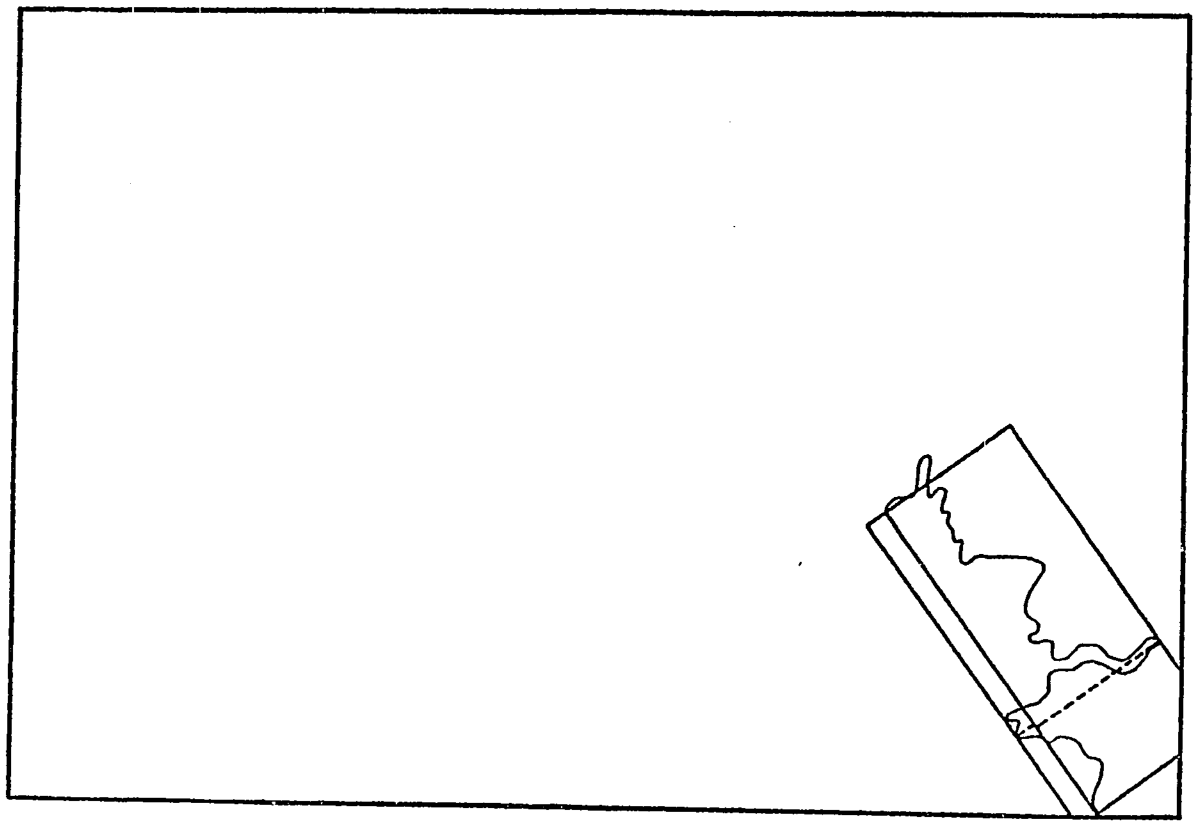


FIGURE A.2 a HUDBY (Band 1)

```
STRIP    2 --   3
     EQUATION FOR THIS STRIP IS     7X +    32Y + -3706.00
     LENGTH OF STRIP =  32.76                      STRIP SLOPE =  167.66
     ENDPOINTS   ( 214,  69)   ( 182,  76)     LEVELS BELOW =    0
     MONOTONICITY=  0.1379
     ERROR VARIANCE=  16.51                        NPTS IN STRIP=  29
     STRIP WIDTH=  0.58                      BIFURCATION SEGMENT=  0.25
     MAX DEVIATION=  0.43                    NEXT STRIP NODE AT RECORD     8
```

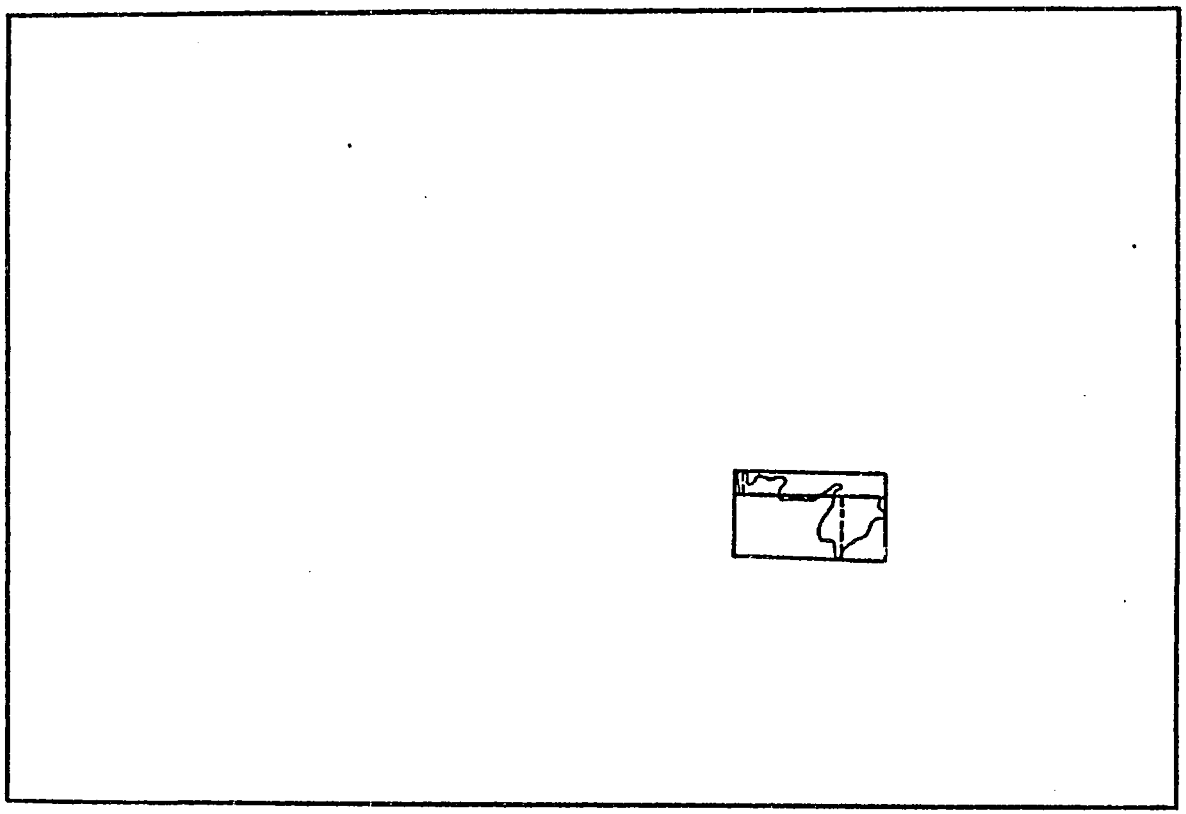


FIGURE A.2 b HUDBY (Band 2)

```
STRIP     3 --    4
     EQUATION FOR THIS STRIP IS     30X +     65Y + -10400.0
     LENGTH OF STRIP =  71.59                          STRIP SLOPE =  155.22
     ENDPOINTS    ( 182,   76)    ( 117, 106)      LEVELS BELOW =    0
     MONOTONICITY=  0.1000
     ERROR VARIANCE=  21.92                            NPTS IN STRIP=  50
     STRIP WIDTH=  0.42                        BIFURCATION SEGMENT=  0.74
     MAX DEVIATION=  0.21                      NEXT STRIP NODE AT RECORD     39
```

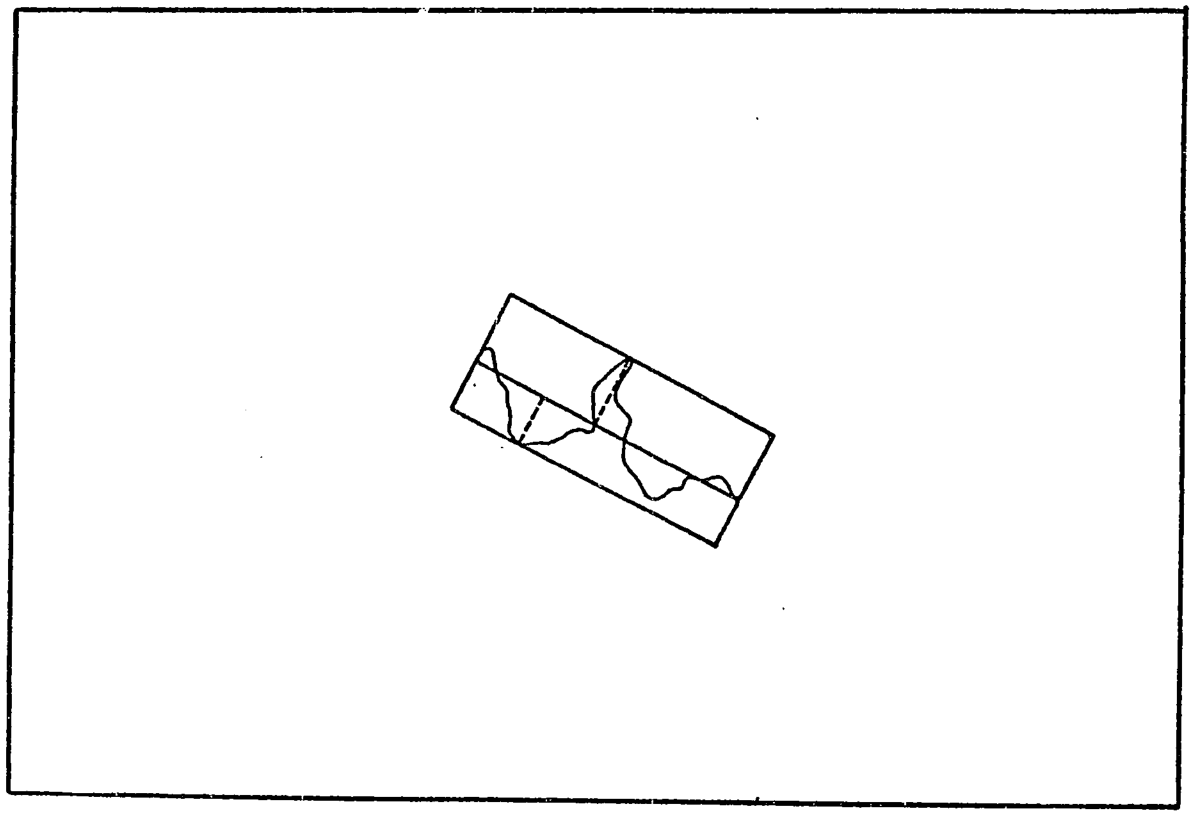


FIGURE A.2 c   HUDBY (Band 3)

```
STRIP     4 --    5
     EQUATION FOR THIS STRIP IS    -13X +      9Y +  567.00
     LENGTH OF STRIP =  15.81                          STRIP SLOPE = -124.70
     ENDPOINTS    ( 117, 106)    ( 108,   93)    LEVELS BELOW =    0
     MONOTONICITY=  0.0833
     ERROR VARIANCE=  8.69                        NPTS IN STRIP=  24
     STRIP WIDTH=  0.87                      BIFURCATION SEGMENT=  0.54
     MAX DEVIATION= -0.59                    NEXT STRIP NODE AT RECORD     12
```

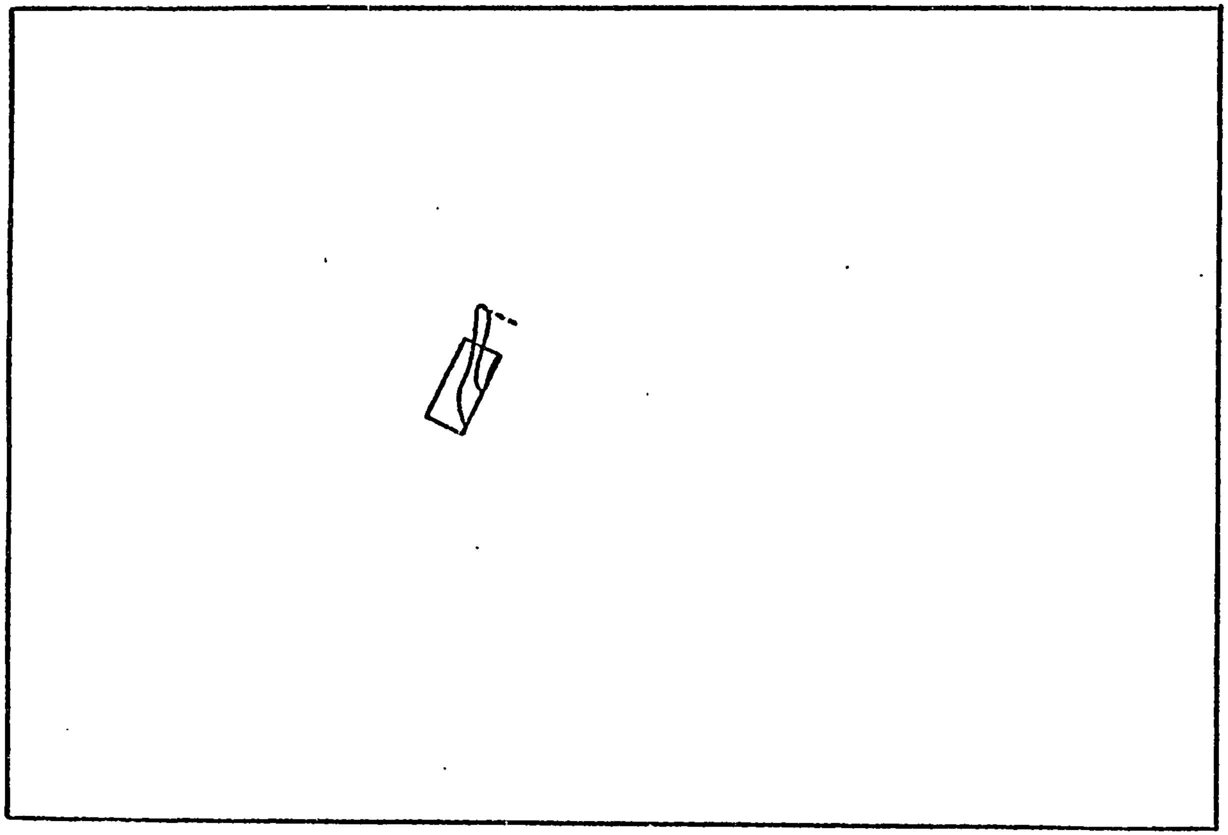


FIGURE A.2 d   HUDBY (Band 4)

```
STRIP    5 --   6
     EQUATION FOR THIS STRIP IS    -3X +    27Y + -2187.00
     LENGTH OF STRIP =  27.17                          STRIP SLOPE = -173.66
     ENDPOINTS    ( 108,   93)    (  81,   90)    LEVELS BELOW =    0
     MONOTONICITY=  0.0769
     ERROR VARIANCE=  40.44                          NPTS IN STRIP=  26
     STRIP WIDTH=  0.81                        BIFURCATION SEGMENT=  0.59
     MAX DEVIATION=  0.71                      NEXT STRIP NODE AT RECORD     17
```

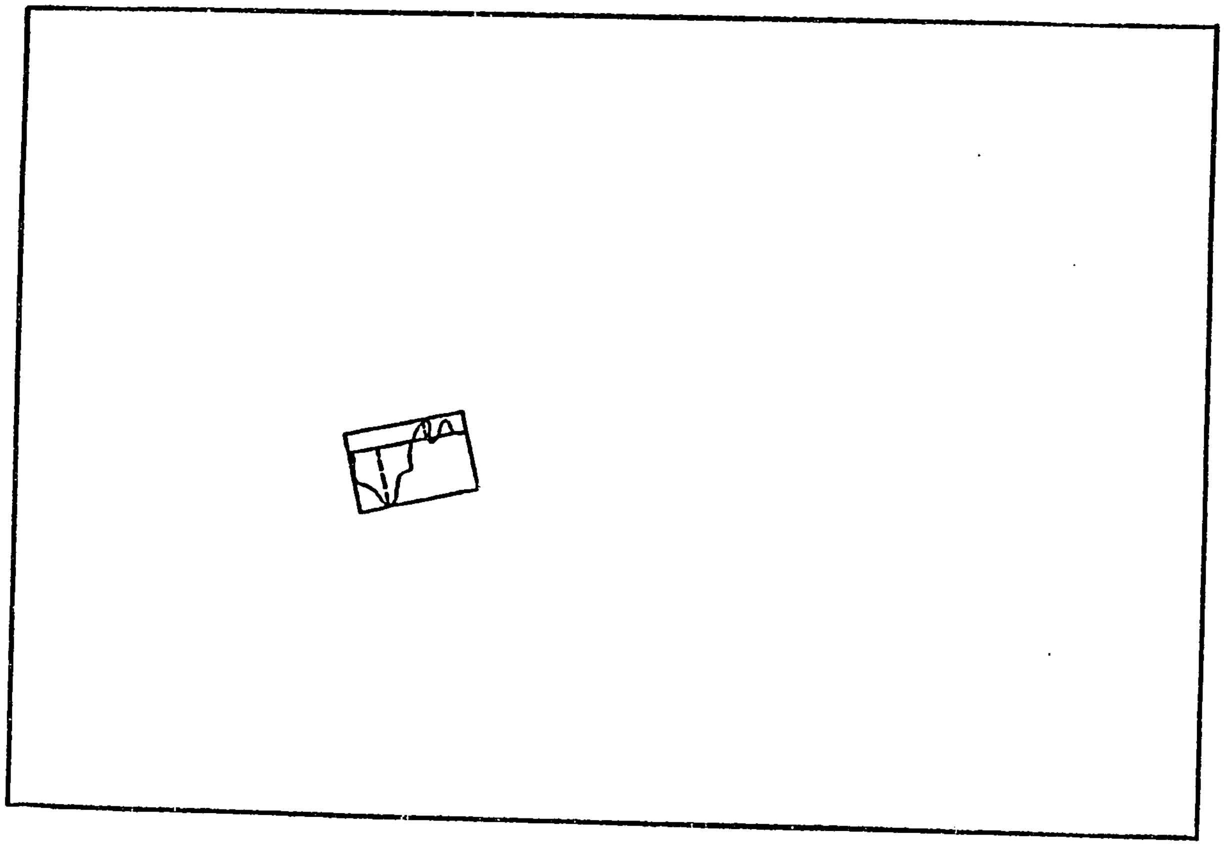


FIGURE A.2 e   HUDBY (Band 5)

```
STRIP     6 --    7
     EQUATION FOR THIS STRIP IS     58X +     52Y + -9378.00
     LENGTH OF STRIP =  77.90                        STRIP SLOPE =  131.88
     ENDPOINTS    (  81,   90)    (  29, 148)     LEVELS BELOW =    0
     MONOTONICITY=  0.0541
     ERROR VARIANCE=  22.31                          NPTS IN STRIP=  37
     STRIP WIDTH=  0.24                        BIFURCATION SEGMENT=  0.55
     MAX DEVIATION= -0.22                      NEXT STRIP NODE AT RECORD     21
```

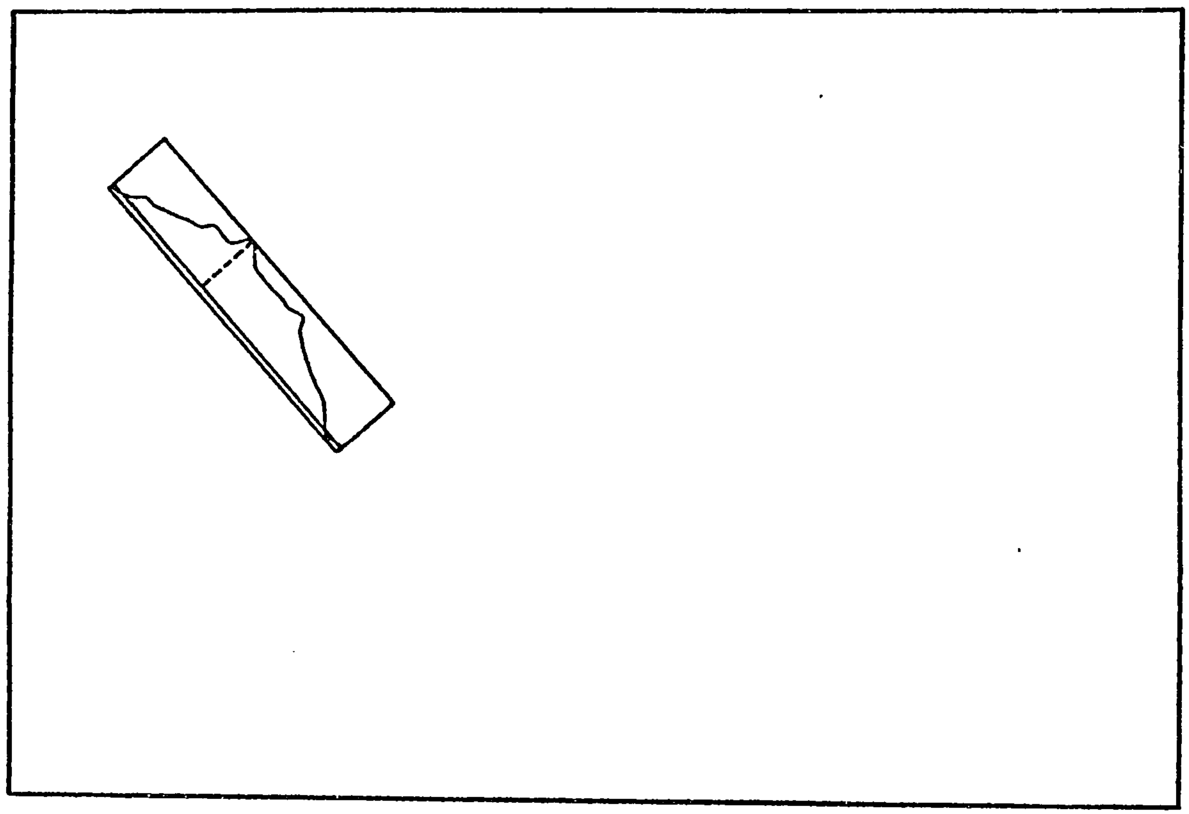


FIGURE A.2 f HUDBY (Band 6)

```
STRIP    7 --   8
    EQUATION FOR THIS STRIP IS     19X +     3Y + -995.00
    LENGTH OF STRIP =   19.24                          STRIP SLOPE =   98.97
    ENDPOINTS    (  29, 148)    (  26, 167)     LEVELS BELOW =    0
    MONOTONICITY=  0.1111
    ERROR VARIANCE=   7.94                            NPTS IN STRIP=  18
    STRIP WIDTH=  0.69                          BIFURCATION SEGMENT=  0.63
    MAX DEVIATION= -0.46                        NEXT STRIP NODE AT RECORD      8
```

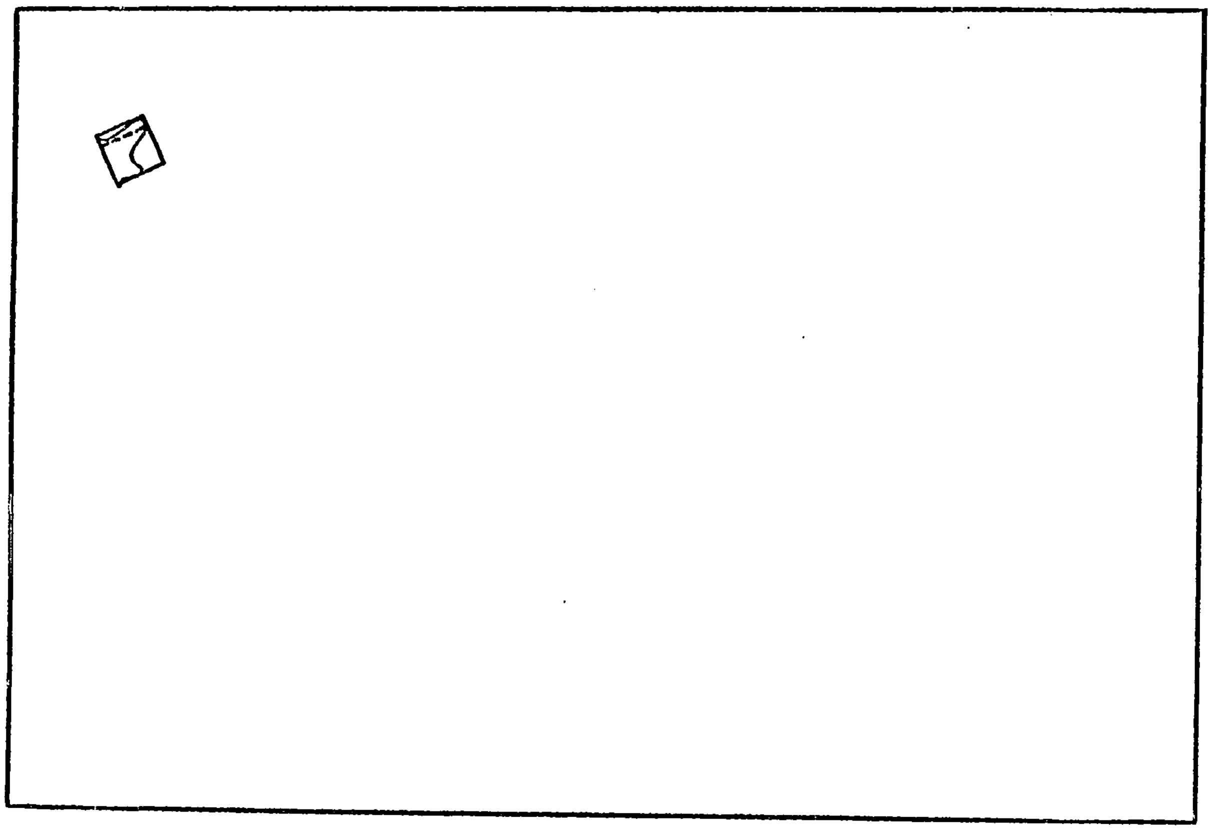


FIGURE A.2 g   HUDBY (Band 7)

STRIP 8 -- 9

EQUATION FOR THIS STRIP IS -4X + 25Y + -4071.00

LENGTH OF STRIP = 23.32 STRIP SLOPE = -170.91

ENDPOINTS (26, 167) (1, 163) LEVELS BELOW = 0

MONOTONICITY= 0.2222

ERROR VARIANCE= 2.70 NPTS IN STRIP= 18

STRIP WIDTH= 0.37 BIFURCATION SEGMENT= 0.16

MAX DEVIATION= -0.22 NEXT STRIP NODE AT RECORD 4

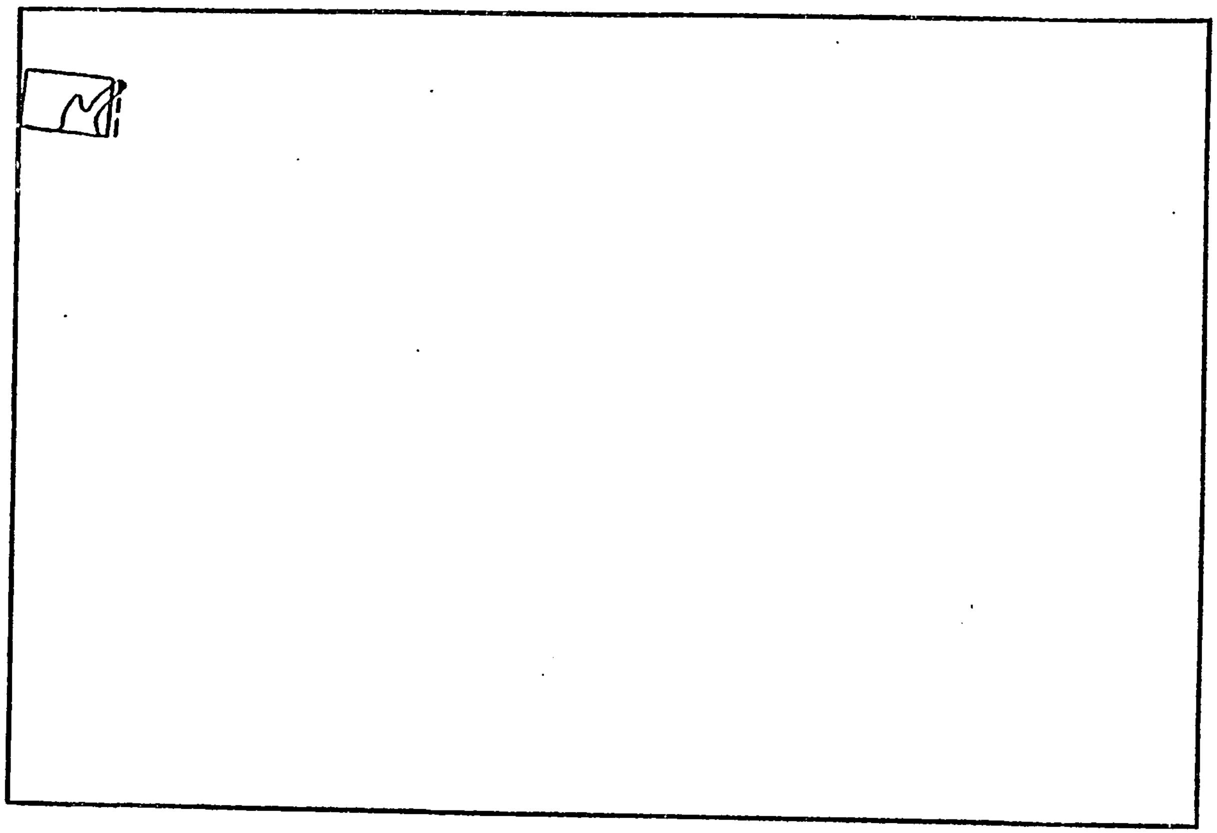


FIGURE A.2 h HUDBY (Band 8)

STRIP 1 -- 2
EQUATION FOR THIS STRIP IS 4X + -105Y + 15645.0
LENGTH OF STRIP = -105.08 STRIP SLOPE = 2.18
ENDPOINTS (0, 149) (105, 153) LEVELS BELOW = 0
MONOTONICITY= 0.1013
ERROR VARIANCE= 28.32 NPTS IN STRIP= 79
STRIP WIDTH= 0.35 BIFURCATION SEGMENT= 0.73
MAX DEVIATION= 0.19 NEXT STRIP NODE AT RECORD 62

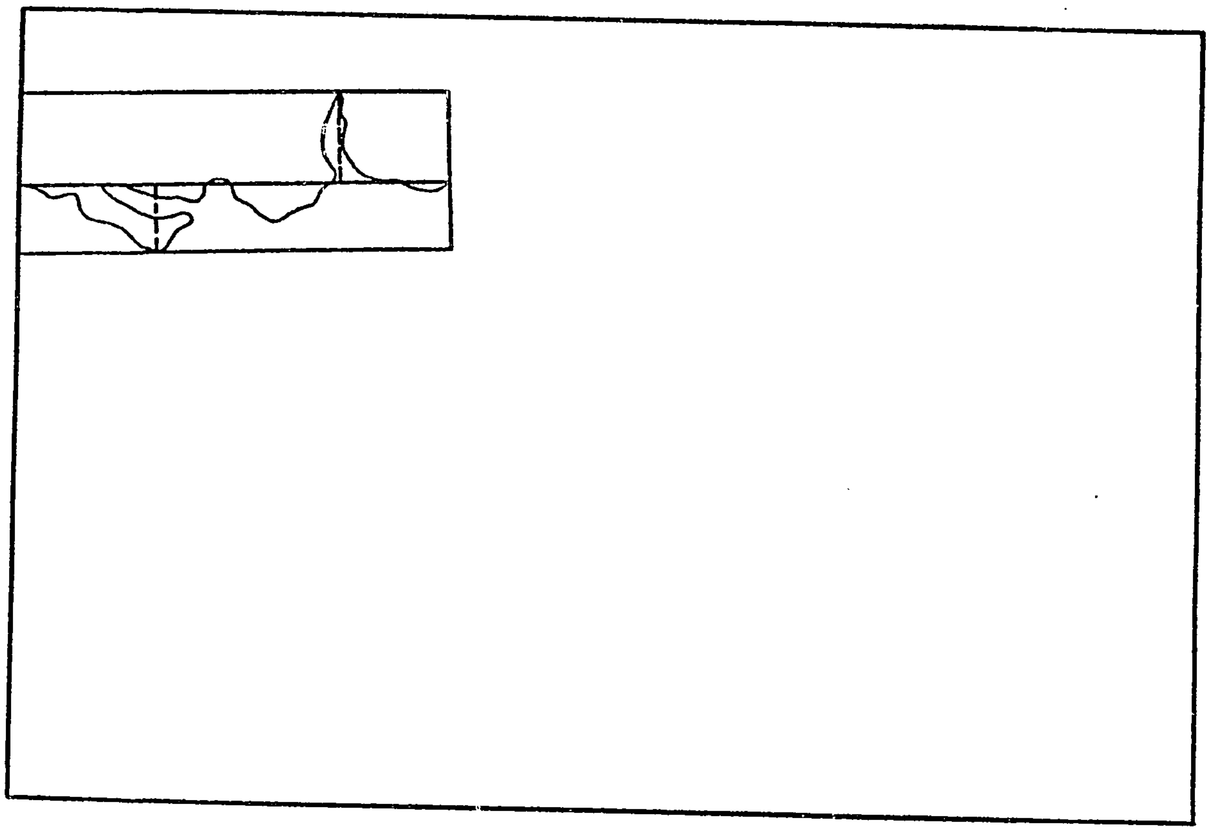


FIGURE A.3 a FJORD (Band 1)

```
STRIP     2 --    3
     EQUATION FOR THIS STRIP IS    -27X +      0Y +  2835.00
     LENGTH OF STRIP =  27.00                        STRIP SLOPE = -90.00
     ENDPOINTS    ( 105, 153)    ( 105, 126)     LEVELS BELOW =    0
     MONOTONICITY=  0.0270
     ERROR VARIANCE=  136.61                          NPTS IN STRIP=  37
     STRIP WIDTH=  1.26                          BIFURCATION SEGMENT=  0.59
     MAX DEVIATION= -1.26                        NEXT STRIP NODE AT RECORD     17
```

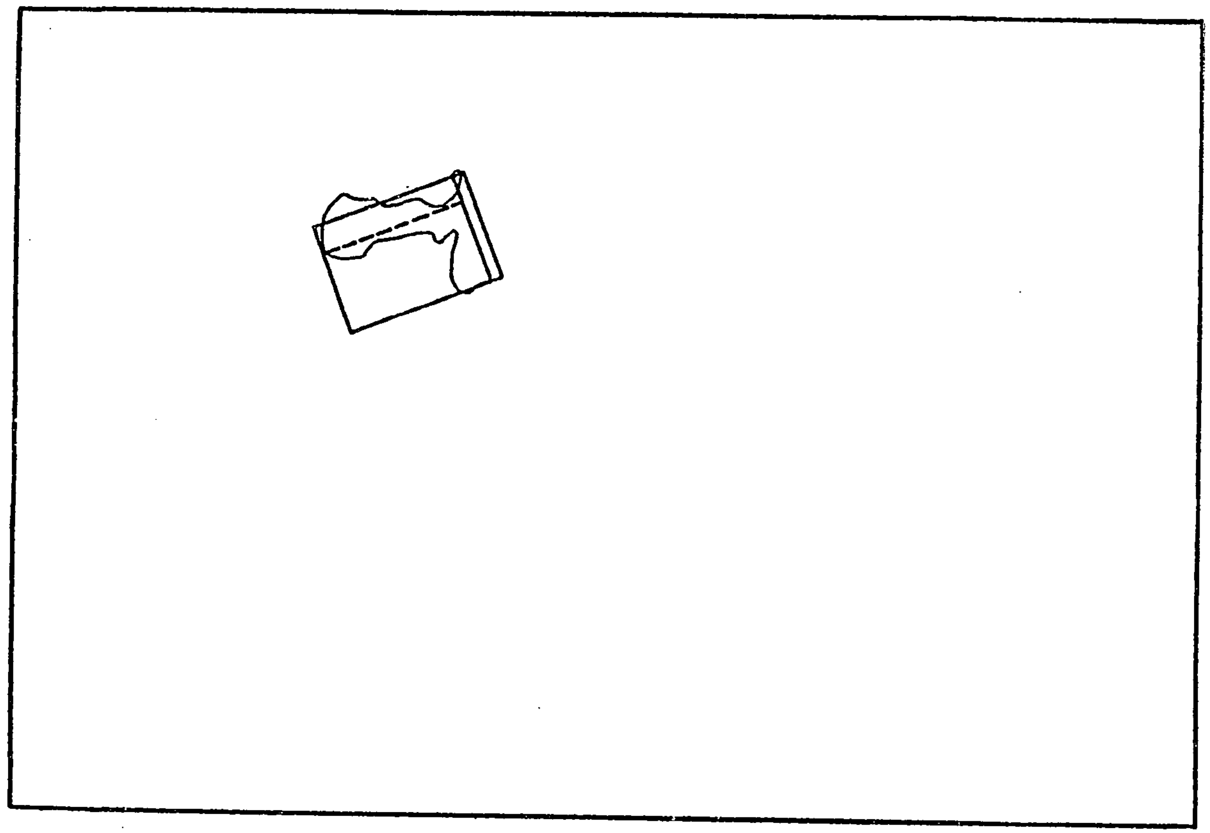


FIGURE A.3 b   FJORD (Band 2)

```
STRIP     3 --    4
     EQUATION FOR THIS STRIP IS     14X +    -48Y +  4578.00
     LENGTH OF STRIP = -50.00                      STRIP SLOPE =  16.26
     ENDPOINTS    ( 105, 126)    ( 153, 140)      LEVELS BELOW =   0
     MONOTONICITY=  0.0685
     ERROR VARIANCE=  87.61                        NPTS IN STRIP=  73
     STRIP WIDTH=  1.14                    BIFURCATION SEGMENT=  1.16
     MAX DEVIATION=  0.72                  NEXT STRIP NODE AT RECORD    57
```

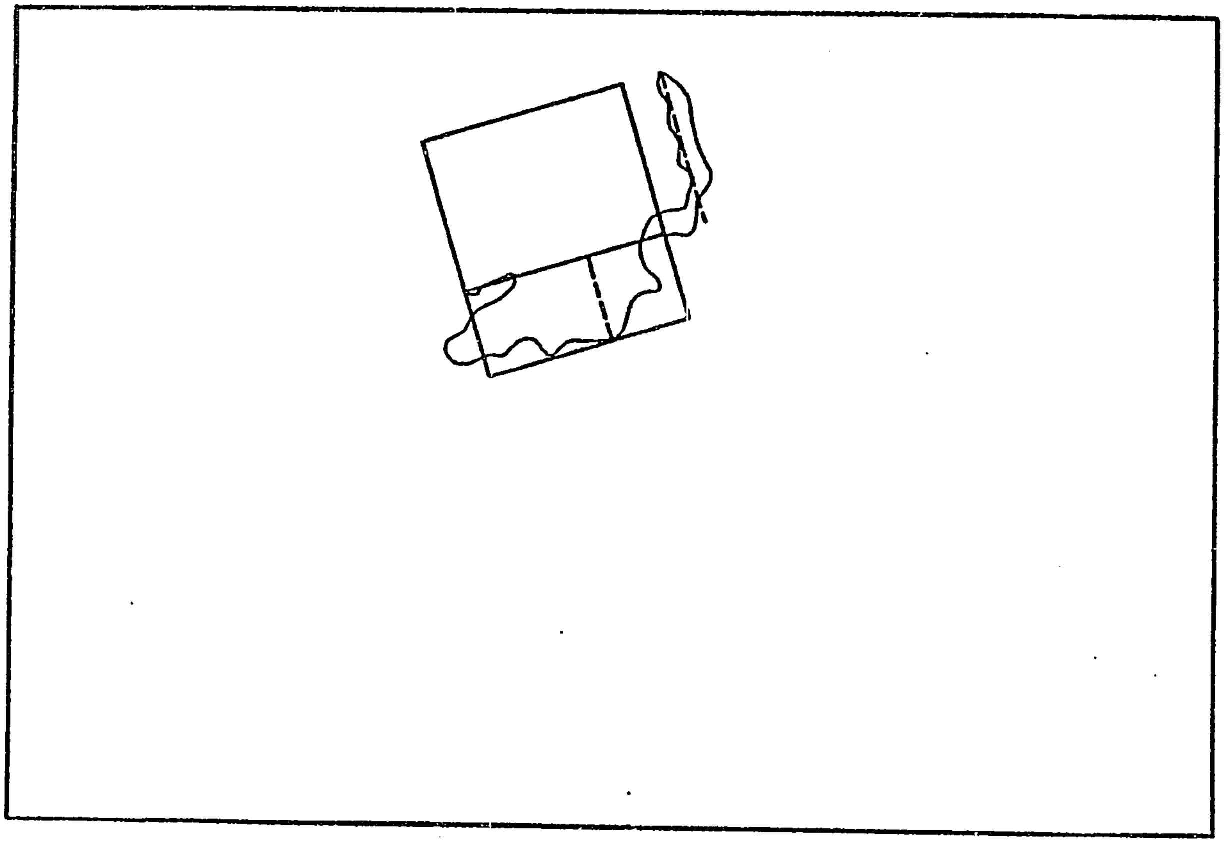


FIGURE A.3 c FJORD (Band 3)

```
STRIP     4 --    5
     EQUATION FOR THIS STRIP IS    -36X +     43Y + -512.00
     LENGTH OF STRIP =  56.08                          STRIP SLOPE = -140.06
     ENDPOINTS    ( 153, 140)     ( 110, 104)      LEVELS BELOW =    0
     MONOTONICITY=  0.0000
     ERROR VARIANCE=  33.32                            NPTS IN STRIP=  29
     STRIP WIDTH=  0.32                         BIFURCATION SEGMENT=  0.48
     MAX DEVIATION=  0.32                       NEXT STRIP NODE AT RECORD     13
```

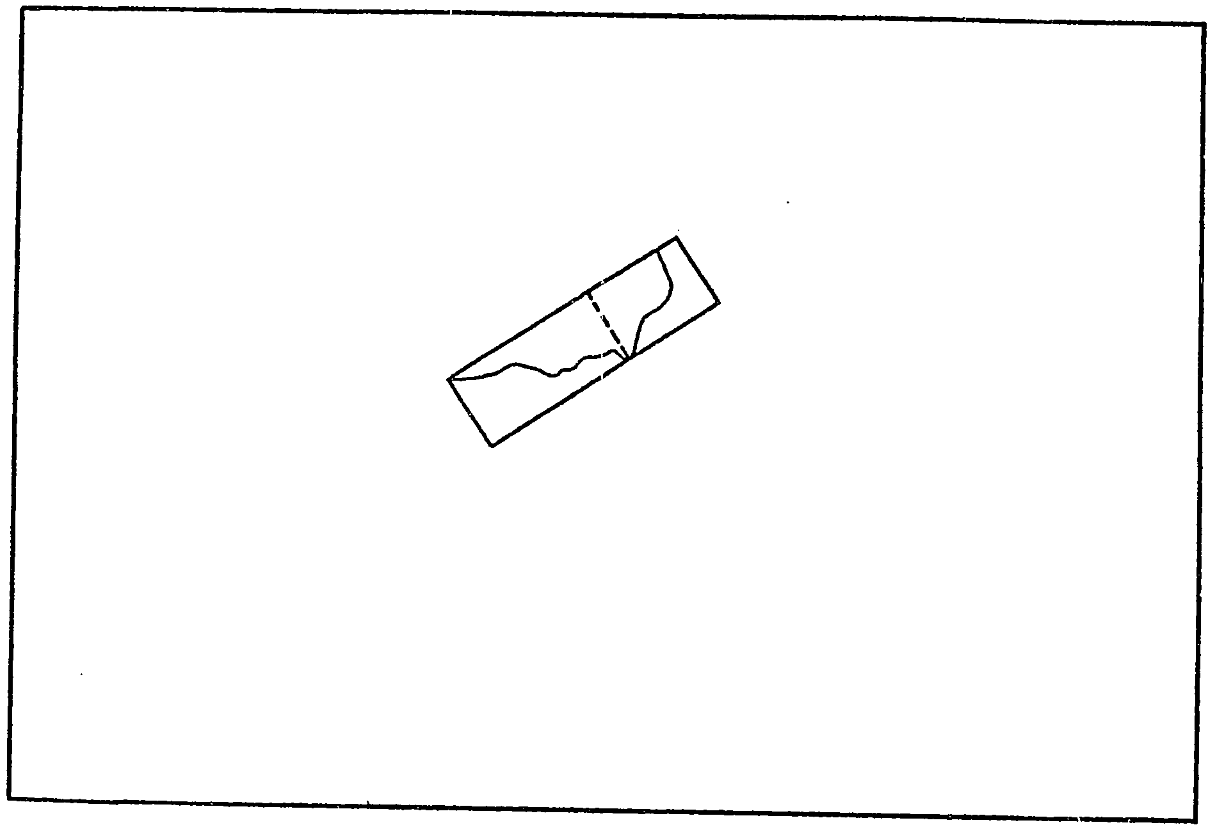


FIGURE A.3 d FJORD (Band 4)

```
STRIP     5 --    6
     EQUATION FOR THIS STRIP IS    -27X +     -4Y +  3386.00
     LENGTH OF STRIP = -27.29                   STRIP SLOPE = -81.57
     ENDPOINTS   ( 110, 104)    ( 114,  77)     LEVELS BELOW =    0
     MONOTONICITY=  0.0667
     ERROR VARIANCE=  21.99                     NPTS IN STRIP=  45
     STRIP WIDTH=  1.14                   BIFURCATION SEGMENT=  0.63
     MAX DEVIATION= -0.61                 NEXT STRIP NODE AT RECORD    37
```

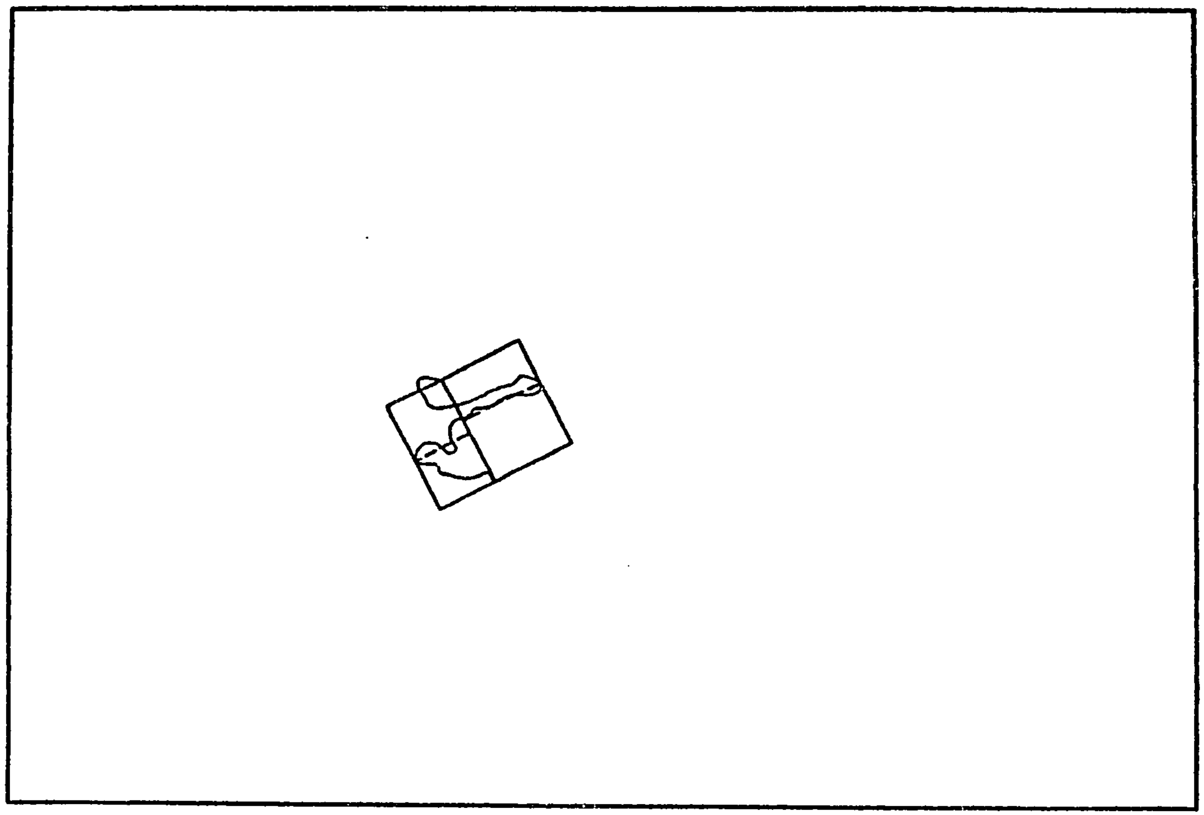


FIGURE A.3 e FJORD (Band 5)

```
STRIP    6 --    7
     EQUATION FOR THIS STRIP IS     20X +    -95Y +  5035.00
     LENGTH OF STRIP = -97.08                        STRIP SLOPE =  11.89
     ENDPOINTS    ( 114,   77)    ( 209,   97)      LEVELS BELOW =   0
     MONOTONICITY=  0.0549
     ERROR VARIANCE=  57.98                         NPTS IN STRIP=  91
     STRIP WIDTH=  0.54                        BIFURCATION SEGMENT=  0.61
     MAX DEVIATION=  0.28                      NEXT STRIP NODE AT RECORD    39
```

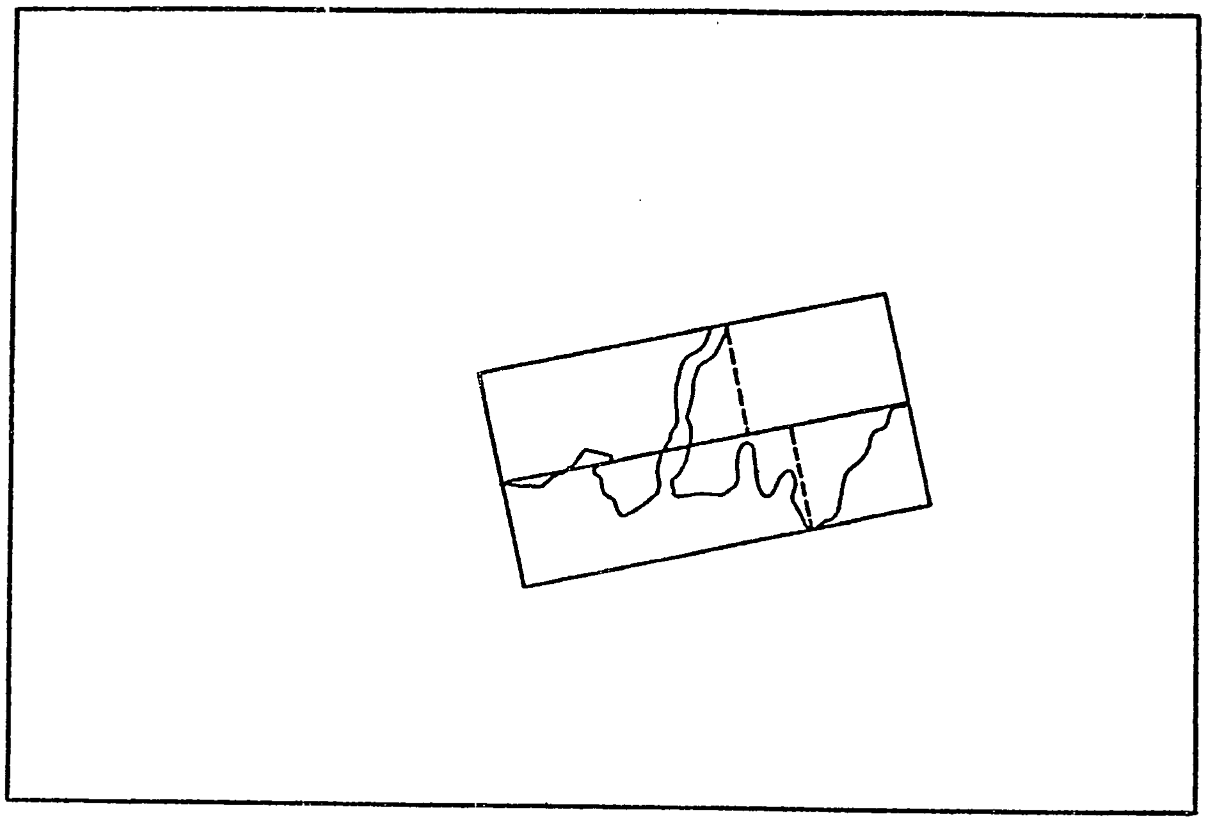


FIGURE A.3 f   FJORD (Band 6)

```
STRIP     7 --    8
     EQUATION FOR THIS STRIP IS    -29X +    -40Y +   9941.00
     LENGTH OF STRIP = -49.41                          STRIP SLOPE = -35.94
     ENDPOINTS    ( 209,   97)    ( 249,   68)      LEVELS BELOW =    0
     MONOTONICITY=  0.0374
     ERROR VARIANCE=  181.75                            NPTS IN STRIP= 107
     STRIP WIDTH=  1.86                           BIFURCATION SEGMENT=  0.43
     MAX DEVIATION=  1.01                         NEXT STRIP NODE AT RECORD     24
```

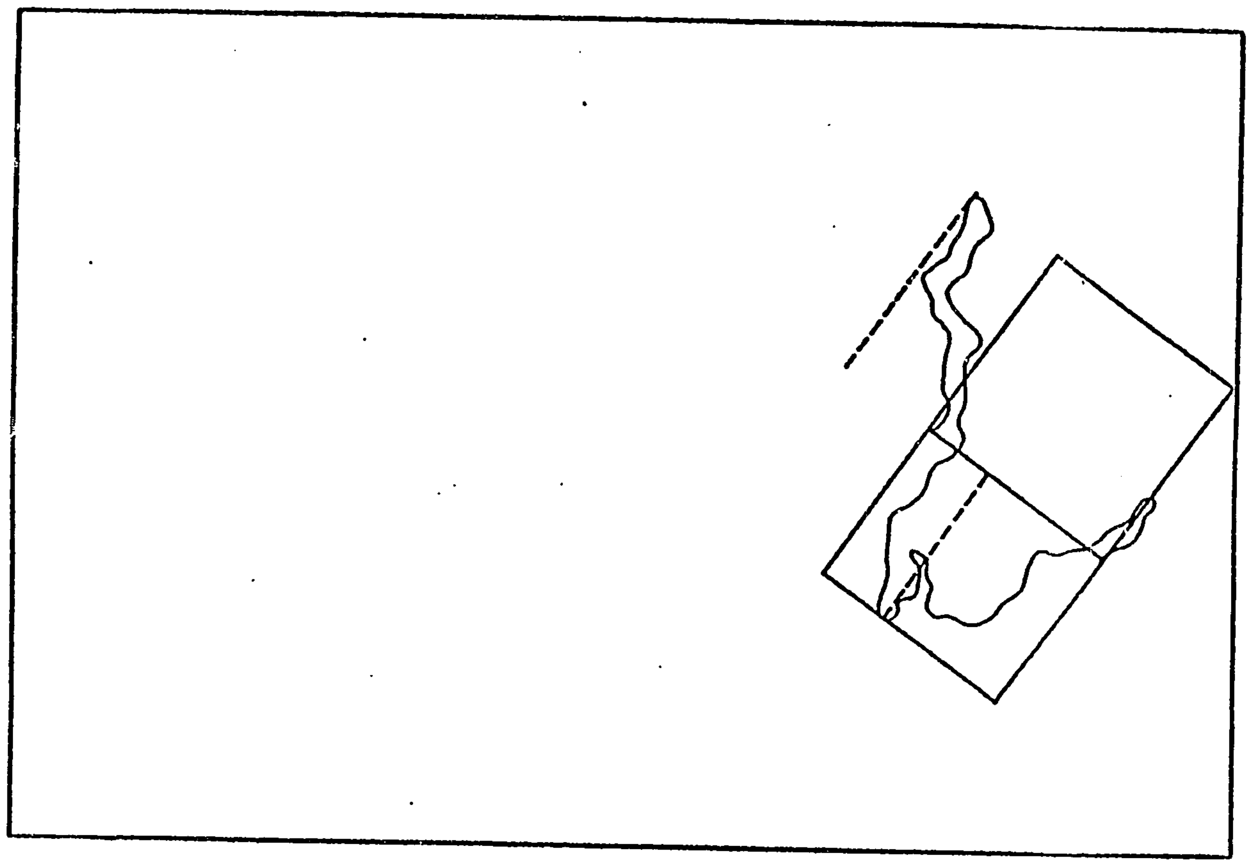


FIGURE A.3 g FJORD (Band 7)

STRIP 8 -- 9

EQUATION FOR THIS STRIP IS -68X + 25Y + 15232.0

LENGTH OF STRIP = 72.45 STRIP SLOPE = -110.19

ENDPOINTS (249, 68) (224, 0) LEVELS BELOW = 0

MONOTONICITY= 0.0652

ERROR VARIANCE= 13.93 NPTS IN STRIP= 46

STRIP WIDTH= 0.30 BIFURCATION SEGMENT= 0.27

MAX DEVIATION= -0.18 NEXT STRIP NODE AT RECORD 11

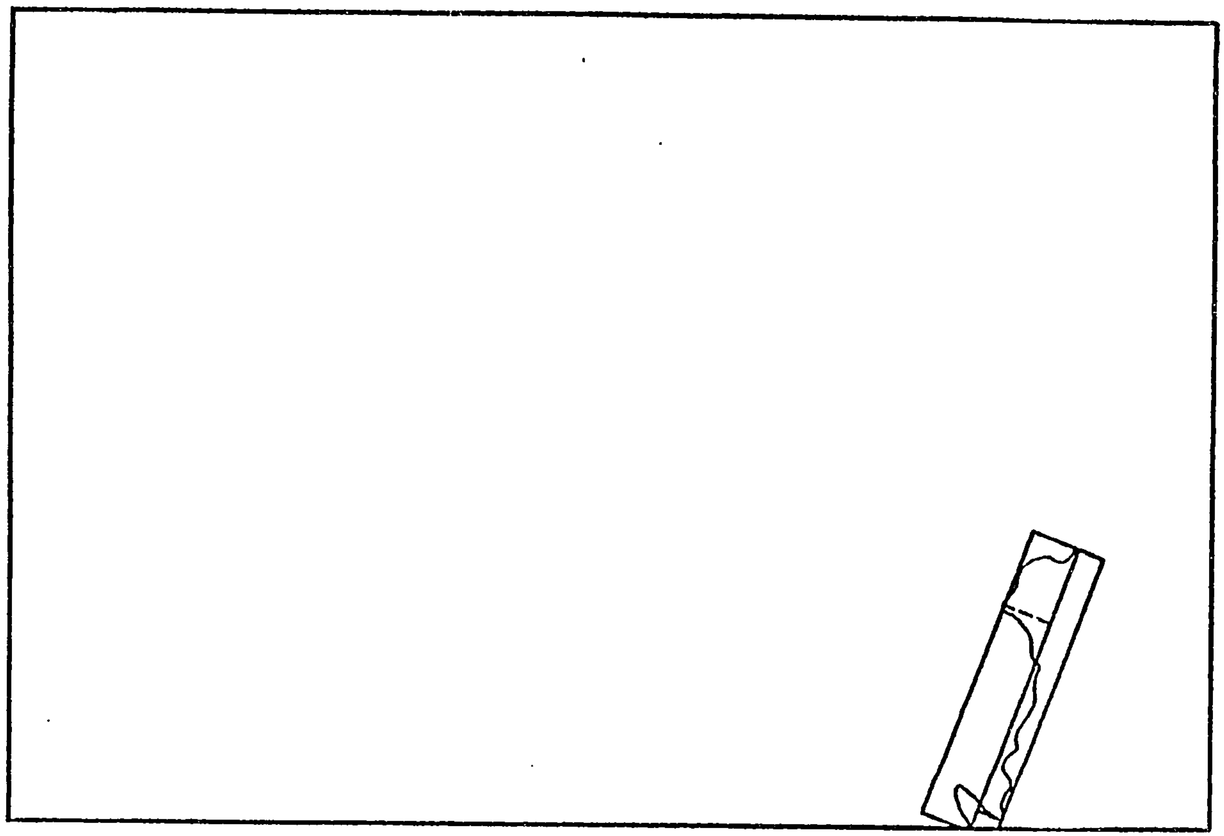


FIGURE A.3 h FJORD (Band 8)

```
STRIP    1 --   2
     EQUATION FOR THIS STRIP IS   -35X +    71Y + -2944.00
     LENGTH OF STRIP =  79.16                    STRIP SLOPE = -153.76
     ENDPOINTS   ( 279, 179)   ( 208, 144)     LEVELS BELOW =   0
     MONOTONICITY=  0.0909
     ERROR VARIANCE=  8.95                     NPTS IN STRIP=  55
     STRIP WIDTH=  0.15                  BIFURCATION SEGMENT=  0.60
     MAX DEVIATION= -0.13                NEXT STRIP NODE AT RECORD    18
```

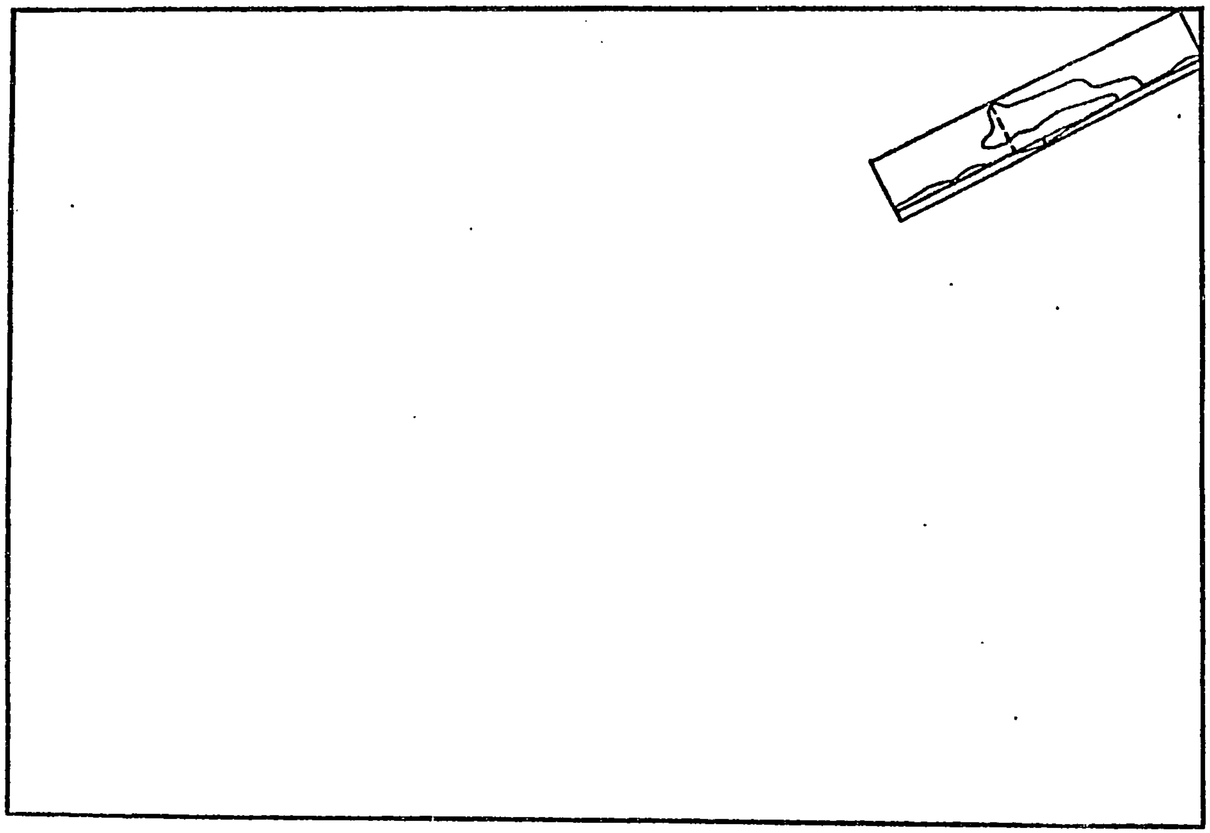


FIGURE A.4 a TEXAS (Band 1)

```
STRIP      2 --    3
     EQUATION FOR THIS STRIP IS     10X +      9Y + -3376.00
     LENGTH OF STRIP =  13.45                     STRIP SLOPE =  131.99
     ENDPOINTS    ( 208, 144)    ( 199, 154)     LEVELS BELOW =   0
     MONOTONICITY=  0.0364
     ERROR VARIANCE=  123.31                       NPTS IN STRIP=  55
     STRIP WIDTH=  4.66                      BIFURCATION SEGMENT=  0.69
     MAX DEVIATION=  2.88                    NEXT STRIP NODE AT RECORD     15
```

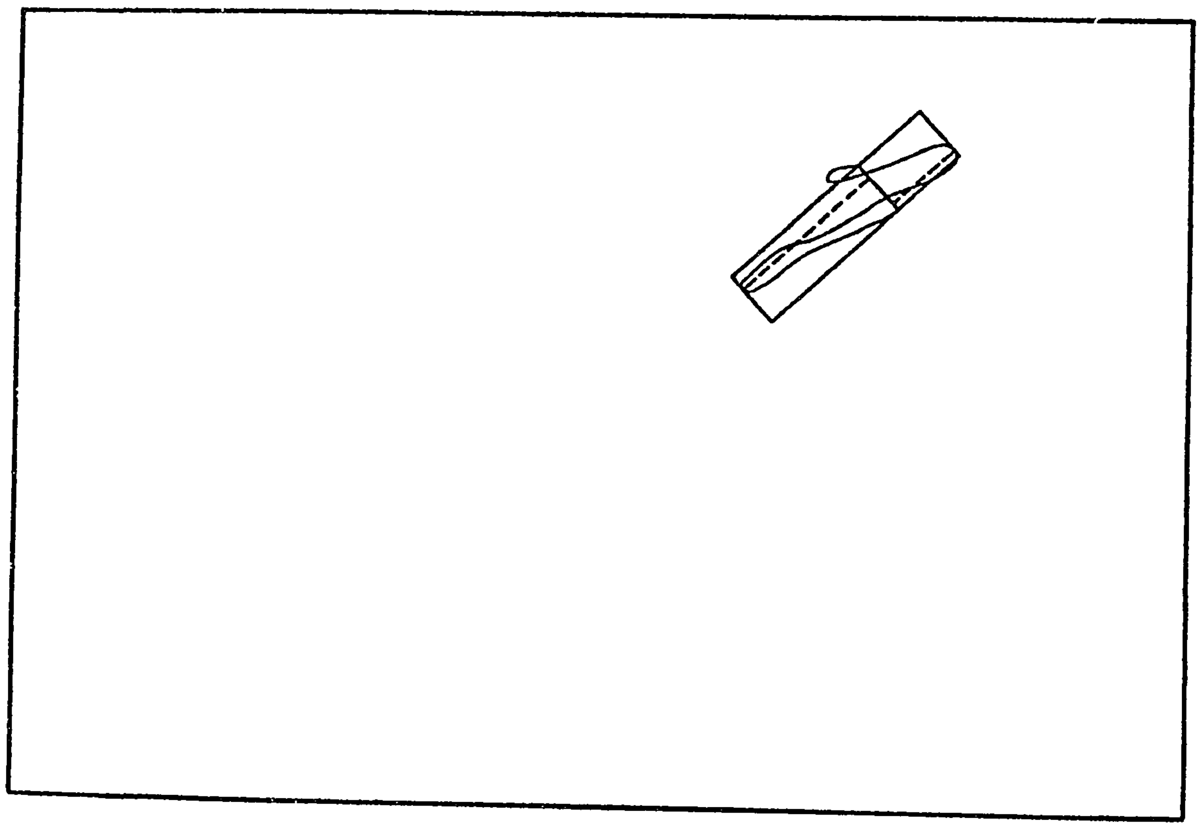


FIGURE A.4 b   TEXAS (Band 2)

```
STRIP     3 --    4
     EQUATION FOR THIS STRIP IS    -67X +    76Y +  1629.00
     LENGTH OF STRIP =  101.32                        STRIP SLOPE = -138.60
     ENDPOINTS   ( 199, 154)    ( 123,  87)     LEVELS BELOW =    0
     MONOTONICITY=  0.0261
     ERROR VARIANCE=  205.69                         NPTS IN STRIP= 115
     STRIP WIDTH=  0.62                        BIFURCATION SEGMENT=  0.34
     MAX DEVIATION= -0.53                      NEXT STRIP NODE AT RECORD     62
```

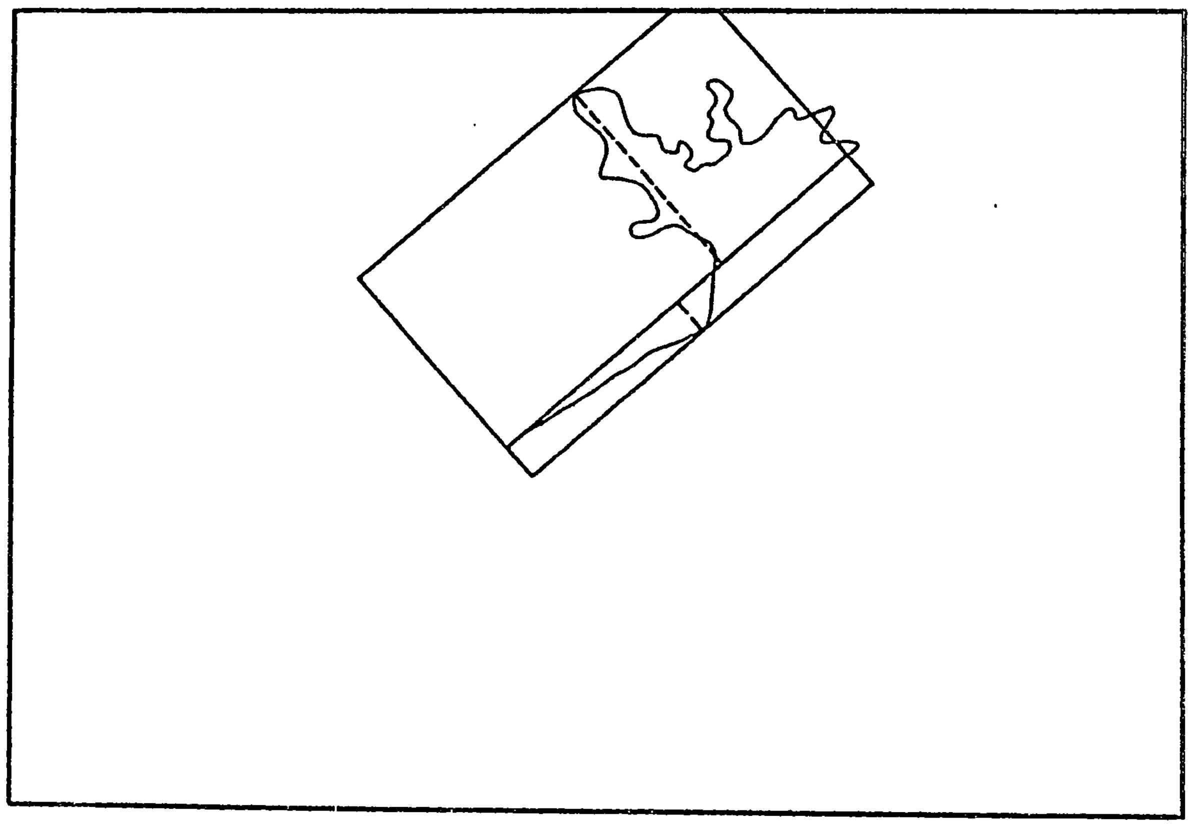


FIGURE A.4 c   TEXAS (Band 3)

```
STRIP     4 --    5
     EQUATION FOR THIS STRIP IS     41X +    -43Y + -1302.00
     LENGTH OF STRIP = -59.41                      STRIP SLOPE =  43.64
     ENDPOINTS    ( 123,   87)    ( 166, 128)     LEVELS BELOW =    0
     MONOTONICITY=  0.0959
     ERROR VARIANCE=  10.64                       NPTS IN STRIP=  73
     STRIP WIDTH=  0.31                     BIFURCATION SEGMENT=  0.20
     MAX DEVIATION=  0.17                   NEXT STRIP NODE AT RECORD    44
```

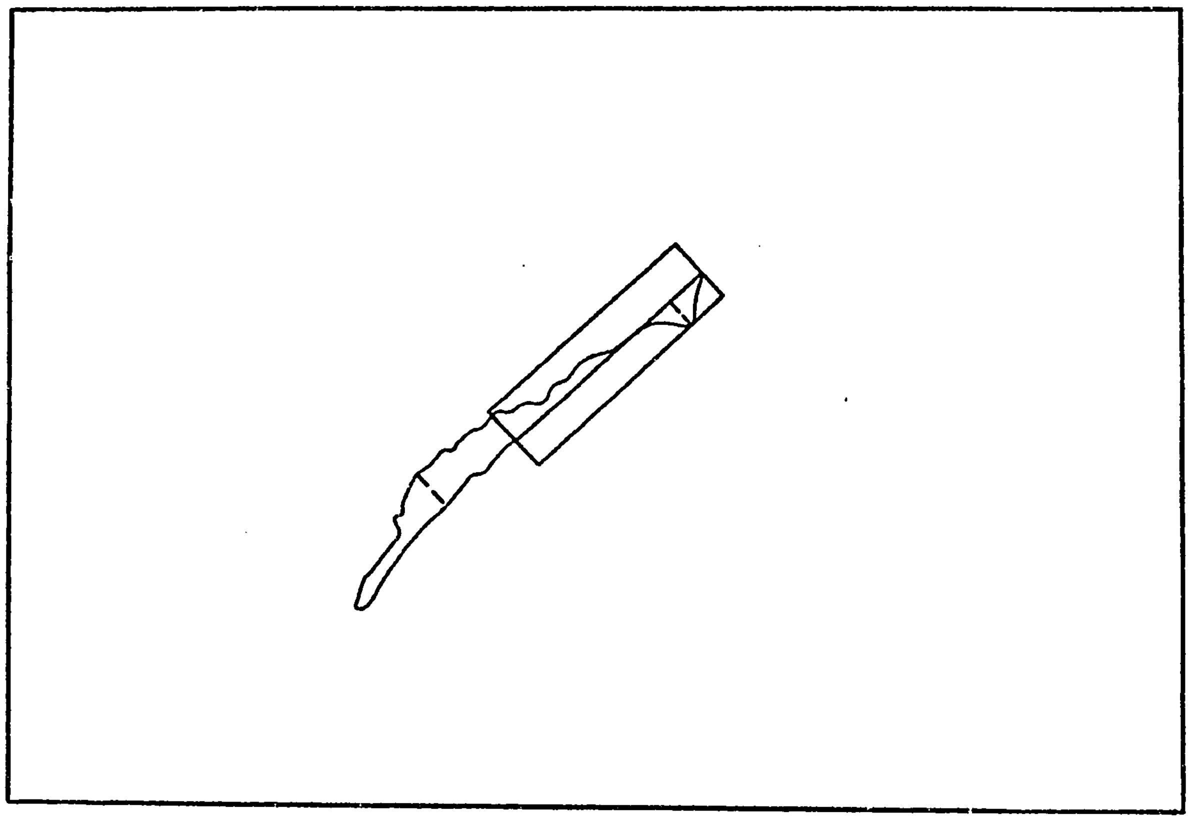


FIGURE A.4 d   TEXAS (Band 4)

```
STRIP    5 --   6
    EQUATION FOR THIS STRIP IS   -37X +    97Y + -6274.00
    LENGTH OF STRIP =  103.82                    STRIP SLOPE = -159.12
    ENDPOINTS   ( 166, 128)   (  69,  91)      LEVELS BELOW =   0
    MONOTONICITY=  0.0444
    ERROR VARIANCE=  43.53                       NPTS IN STRIP=  90
    STRIP WIDTH=  0.41                      BIFURCATION SEGMENT=  0.42
    MAX DEVIATION= -0.24                    NEXT STRIP NODE AT RECORD    25
```

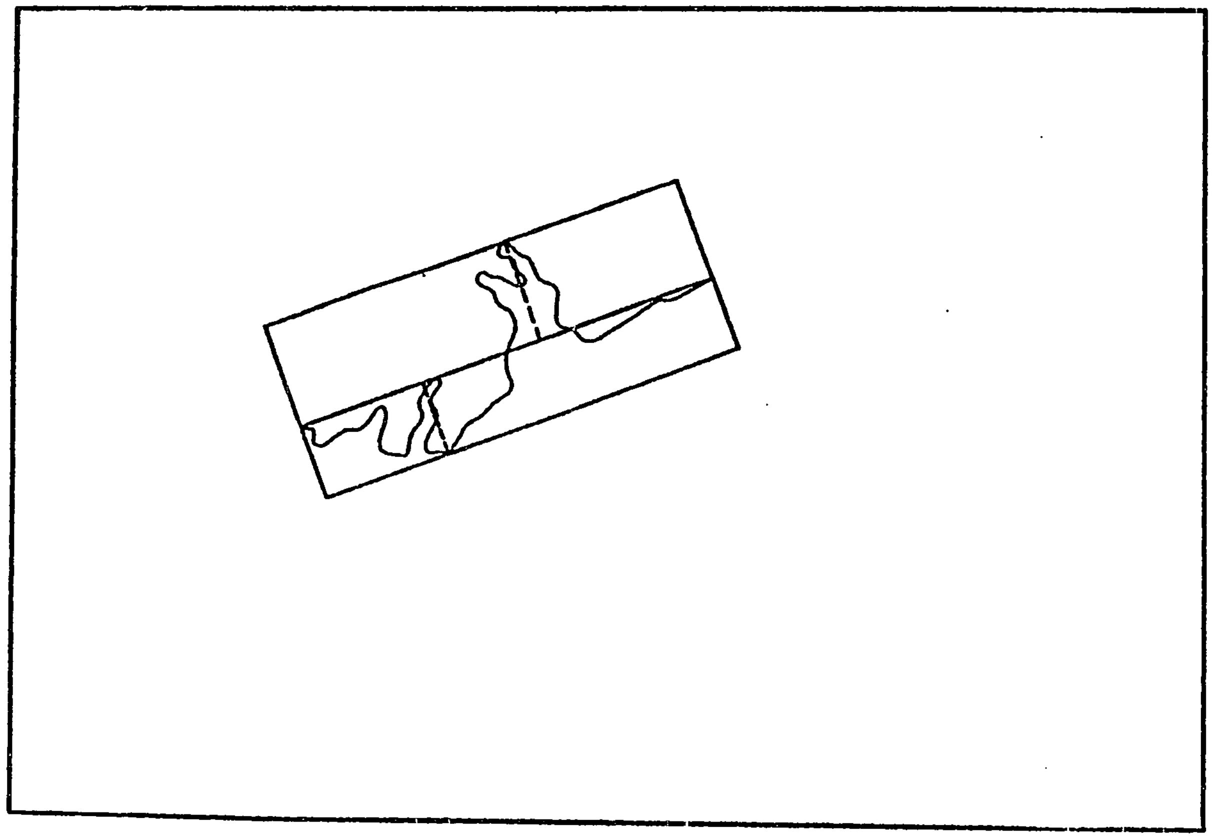


FIGURE A.4 e   TEXAS (Band 5)

```
STRIP     6 --    7
     EQUATION FOR THIS STRIP IS    -43X +    44Y + -1037.00
     LENGTH OF STRIP =  61.52                    STRIP SLOPE = -135.66
     ENDPOINTS    (  69,   91)    (  25,   48)   LEVELS BELOW =    0
     MONOTONICITY=  0.0286
     ERROR VARIANCE=  93.69                      NPTS IN STRIP=  70
     STRIP WIDTH=  0.55                   BIFURCATION SEGMENT=  0.63
     MAX DEVIATION=  0.52                 NEXT STRIP NODE AT RECORD     47
```

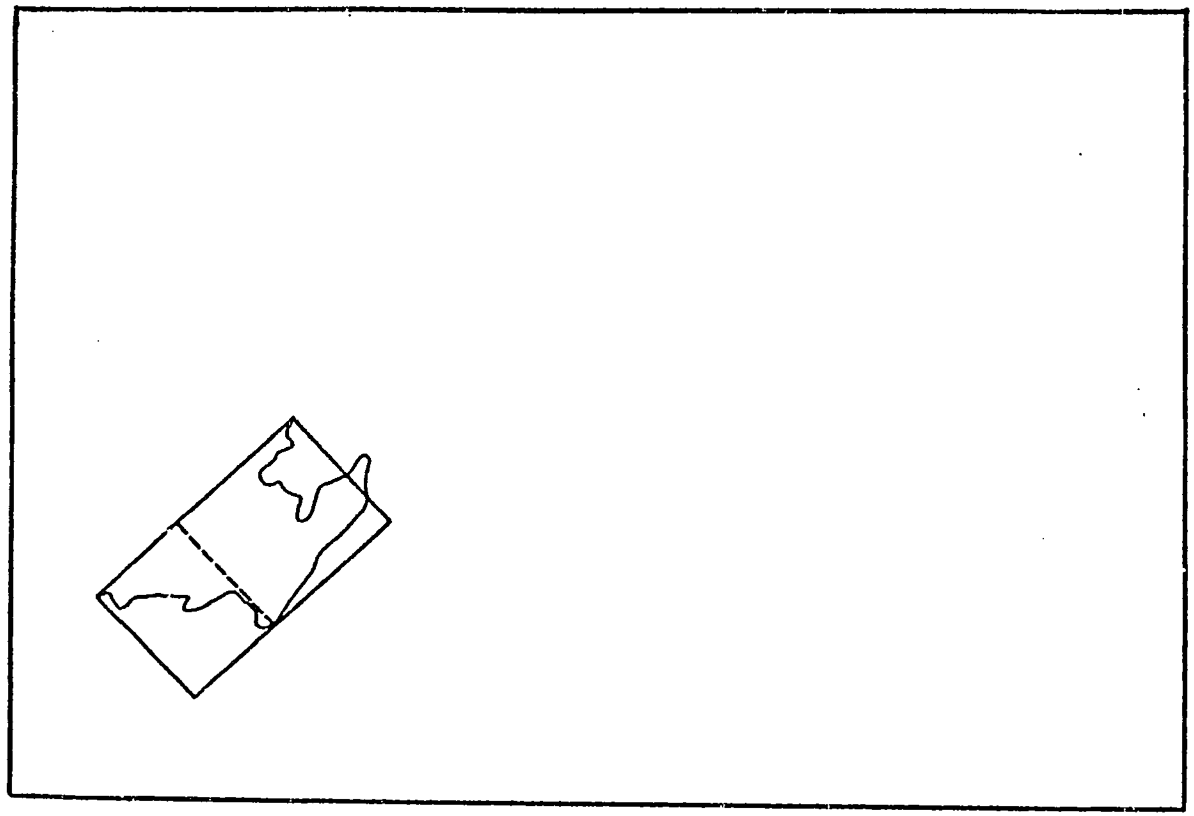


FIGURE A.4 f TEXAS (Band 6)

STRIP 7 -- 8

EQUATION FOR THIS STRIP IS -36X + -26Y + 2148.00

LENGTH OF STRIP = -44.41 STRIP SLOPE = -54.16

ENDPOINTS (25, 48) (51, 12) LEVELS BELOW = 0

MONOTONICITY= 0.0769

ERROR VARIANCE= 6.61 NPTS IN STRIP= 26

STRIP WIDTH= 0.34 BIFURCATION SEGMENT= 0.86

MAX DEVIATION= 0.22 NEXT STRIP NODE AT RECORD 22

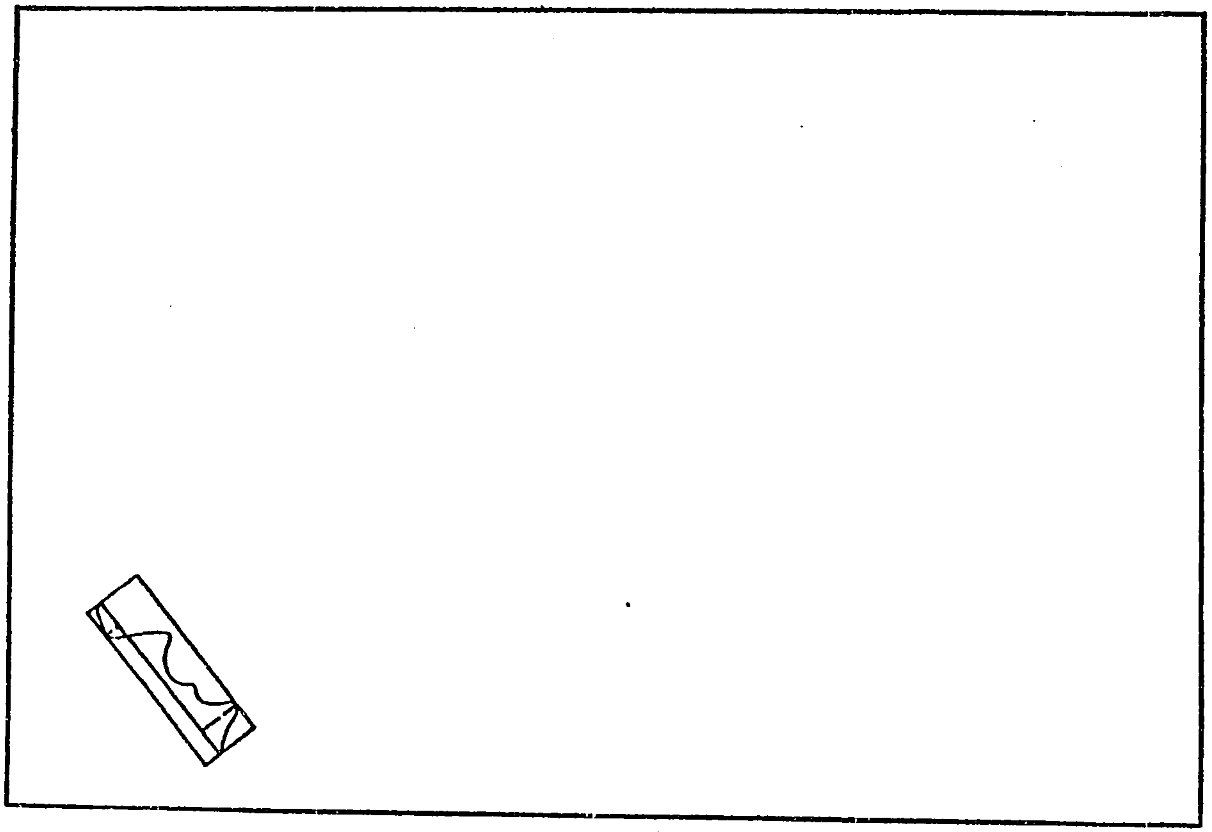


FIGURE A.4 g TEXAS (Band 7)

```
STRIP    8 --    9
     EQUATION FOR THIS STRIP IS    -12X +    -6Y +  684.00
     LENGTH OF STRIP = -13.42                          STRIP SLOPE = -63.43
     ENDPOINTS    (  51,   12)    (  57,    0)      LEVELS BELOW =    0
     MONOTONICITY=  0.0370
     ERROR VARIANCE=  200.66                           NPTS IN STRIP=  54
     STRIP WIDTH=  3.97                          BIFURCATION SEGMENT=  1.00
     MAX DEVIATION=  3.27                        NEXT STRIP NODE AT RECORD     35
```

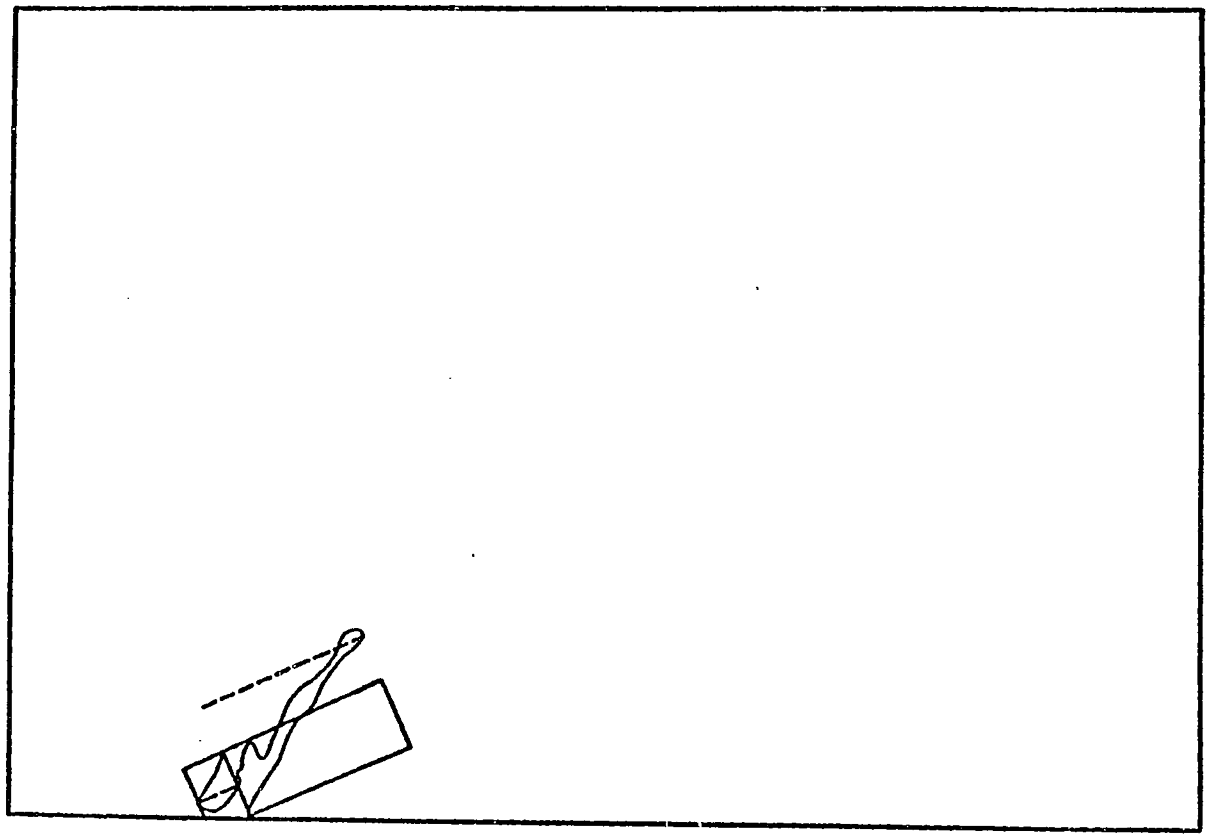


FIGURE A.4 h   TEXAS (Band 8)

APPENDIX B
LOW ORDER PARAMETERS FOR ALL BANDS

The purpose of the bandwidth subdivision algorithm is fourfold: first, to organize the structural measures for each line into a strip tree data structure; secondly, to subdivide the lines into bands; third, to perform the structural measurements as each subdivision is accomplished; and fourth, to store these measures for each level of resolution in a single pass through the coordinate line file.

The algorithm works through one level of subdivision at a time. Initially, the data structure contains only eight bands worth of parameters (these are the geographic bands, illustrated in Appendix A). As a band in this level is accessed, its coordinates are located within the origianl data file. This band is subdivided, according to the principle of maximum deviation; and the two bands which result are measured for low order structure. These newly measured parameters are entered into the strip tree at the next lower level. Thus the algorithm is building level 2 of the strip tree as it accesses level 1: when all bands in a level have been read in, the next lower level has been constructed, and can now be accessed for further band subdivisions.

What follows here is a listing of the low order structure parameters measured for each band of each line. Also listed are the number of points subsumed by a particular band, and the length of its anchor line. Each table is grouped by level of subdivision: thus the first 8 entries refer to the original eight bands, the second 16 entries refer to the sixteen bands formed when the first eight were divided, and so forth; and each subdivision level is separated in the table by a blank line. Only significant bands are included in this table, so the reader should be aware that the progression of 8, 16, 32, 64, etc. bands varies somewhat from line to line. The data in this table forms the basis for all discriminant analyses performed in the dissertation.

In the first column of the table is listed a four-digit identifier for each band. The first digit identifies from which of the eight original bands the particular band is subdivided. The second digit identifies at which level of subdivision the band has been resolved, and also which level in the strip tree data data structure contains the parameters for the band. Originally, each line has eight bands (2^3), and so the 'first' level of subdivision is called level 4 (16 bands = 2^4). Levels proceed on to level 7, where less than 10% of the bands retained significance (this was discussed in Chapter 4).

The last two digits in the strip identifier refer to the exact segment of the original band which is represented. For example, the first subdivision of any original band produces two bands, thus level '4' of the tree contains strip nodes 1401, 1402, 2401, 2402, etc., according to which half of which of the eight original bands is represented. The next level (level '5') contains not two but four subdivisions of each original band, and these strip identifiers range from 1501 - 1504, 2501 - 2504, etc.

This method of identification provides two capabilities. First, each node in the data structure is related to a specific band subdivided from each line; each set of digital low order measures can be visually related to a specific segment of each sample line. Secondly, the identifiers provide an indexing scheme for the data structure, because each identifier can be broken down, digit by digit, to point at the specific record in the disk file which contains the low order measures for a particular band. In this way, a system is achieved to allow direct cross-reference between segments of the cartographic line and respective measurements of low order structure.

TABLE B.1
LOW ORDER PARAMETERS

PRINTS:SIOUX.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
1	0.0909	1.719	0.72	-0.12	0.04	11	28.65
2	0.0580	14.978	0.64	-0.08	0.02	69	176.71
3	0.0714	8.448	0.24	-0.01	0.11	28	79.25
4	0.0635	42.042	0.62	-0.13	0.08	63	172.24
5	0.0000	3.434	0.63	0.00	0.08	27	72.11
6	0.0606	87.036	0.40	-0.20	0.01	66	155.90
7	0.0000	10.829	0.17	0.00	0.13	33	87.01
8	0.0556	41.558	0.60	-0.03	0.28	36	77.62
1401	0.2500	0.468	0.17	-0.01	0.08	8	20.62
1402	0.2500	0.134	0.22	-0.09	0.00	4	9.06
2401	0.0889	6.809	0.86	-0.05	0.09	45	114.59
2402	0.0000	8.750	0.63	0.00	0.14	25	64.54
3401	0.2222	2.192	0.38	-0.21	0.06	9	20.81
3402	0.1000	0.739	0.33	-0.04	0.00	20	61.03
4401	0.0250	15.405	0.06	-0.00	0.14	40	109.33
4402	0.0000	0.676	0.91	0.00	0.05	24	69.08
5401	0.3529	0.152	0.15	-0.02	0.03	17	45.62
5402	0.0000	0.279	0.51	0.00	0.06	11	27.59
6401	0.0968	14.695	0.54	-0.20	0.07	31	68.73
6402	0.2222	2.540	0.63	-0.06	0.04	36	99.20
7401	0.1667	0.267	0.63	-0.05	0.00	6	19.00
7402	0.1786	1.206	0.30	-0.05	0.05	28	72.86
8401	0.1000	6.005	0.28	-0.08	0.15	20	51.09
8402	0.0588	9.058	0.72	-0.24	0.00	17	38.28
1401	0.2500	0.468	0.17	-0.01	0.08	8	20.62
1402	0.2500	0.134	0.22	-0.09	0.00	4	9.06
2501	0.0256	9.187	0.76	-0.11	0.03	39	98.84
2502	0.0000	0.676	0.53	0.00	0.11	7	19.42
2503	0.3333	0.360	0.33	-0.05	0.04	15	41.40
2504	0.1818	0.798	0.70	-0.10	0.05	11	25.63
3501	0.2500	0.101	0.67	-0.07	0.00	4	9.06
3502	0.0000	1.227	0.63	0.00	0.21	6	13.60
3402	0.1000	0.739	0.33	-0.04	0.00	20	61.03
4501	0.2857	1.301	0.09	-0.16	0.05	7	16.12
4502	0.0882	1.508	0.85	-0.03	0.04	34	104.69
4402	0.0000	0.676	0.91	0.00	0.05	24	69.08
5401	0.3529	0.152	0.15	-0.02	0.03	17	45.62
5402	0.0000	0.279	0.51	0.00	0.06	11	27.59
6501	0.0556	9.304	0.28	-0.02	0.24	18	39.81
6502	0.0714	5.871	0.21	-0.00	0.22	14	34.13
6503	0.0870	3.884	0.87	-0.00	0.09	23	62.97
6504	0.1429	1.827	0.30	-0.03	0.12	14	37.12
7401	0.1667	0.267	0.63	-0.05	0.00	6	19.00

TABLE B.1
LOW ORDER PARAMETERS

PRINTS:SIOUX.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
7503	0.1111	3.705	0.60	-0.20	0.00	9	22.80
7504	0.1500	3.752	0.21	-0.12	0.02	20	50.45
8501	0.0000	0.157	0.67	0.00	0.06	6	16.16
8502	0.1333	8.276	0.58	-0.20	0.02	15	37.59
8503	0.3333	1.450	0.27	-0.13	0.09	12	29.21
8504	0.3333	0.651	0.84	-0.08	0.14	6	13.93
1401	0.2500	0.468	0.17	-0.01	0.08	8	20.62
1402	0.2500	0.134	0.22	-0.09	0.00	4	9.06
2601	0.0667	2.427	0.31	-0.07	0.07	30	76.22
2602	0.1000	0.747	0.30	-0.09	0.00	10	25.61
2502	0.0000	0.676	0.53	0.00	0.11	7	19.42
2503	0.3333	0.360	0.33	-0.05	0.04	15	41.40
2504	0.1818	0.798	0.70	-0.10	0.05	11	25.63
3501	0.2500	0.101	0.67	-0.07	0.00	4	9.06
3502	0.0000	1.227	0.63	0.00	0.21	6	13.60
3402	0.1000	0.739	0.33	-0.04	0.00	20	61.03
4501	0.2857	1.301	0.09	-0.16	0.05	7	16.12
4603	0.1071	2.091	0.35	-0.05	0.00	28	88.77
4604	0.1429	0.457	0.87	-0.10	0.05	7	16.64
4402	0.0000	0.676	0.91	0.00	0.05	24	69.08
5401	0.3529	0.152	0.15	-0.02	0.03	17	45.62
5402	0.0000	0.279	0.51	0.00	0.06	11	27.59
6601	0.0000	0.341	0.15	0.00	0.09	6	14.76
6602	0.2308	1.964	0.54	-0.15	0.06	13	30.02
6603	0.0000	0.163	0.40	0.00	0.08	4	10.20
6604	0.2727	1.261	0.74	-0.04	0.11	11	28.16
6605	0.1500	1.997	0.27	-0.06	0.07	20	55.00
6606	0.2500	0.077	0.31	-0.06	0.00	4	10.20
6607	0.2000	0.233	0.83	-0.10	0.02	5	12.04
6608	0.2000	2.431	0.30	-0.15	0.10	10	26.40
7401	0.1667	0.267	0.63	-0.05	0.00	6	19.00
7503	0.1111	3.705	0.60	-0.20	0.00	9	22.80
7607	0.5000	0.764	0.34	-0.05	0.19	6	12.17
7608	0.0000	3.013	0.23	0.00	0.11	15	40.36
8501	0.0000	0.157	0.67	0.00	0.06	6	16.16
8603	0.2222	1.702	0.35	-0.15	0.02	9	23.35
8604	0.2857	1.215	0.71	-0.03	0.17	7	17.20
8605	0.0000	0.450	0.75	0.00	0.15	4	8.94
8606	0.0000	3.949	0.35	0.00	0.24	9	21.47
8504	0.3333	0.651	0.84	-0.08	0.14	6	13.93
1401	0.2500	0.468	0.17	-0.01	0.08	8	20.62
1402	0.2500	0.134	0.22	-0.09	0.00	4	9.06
2701	0.0000	4.060	0.33	0.00	0.21	10	24.19

TABLE B.1
LOW ORDER PARAMETERS

PRINTS:SIOUX.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
2702	0.0952	6.093	0.63	-0.01	0.13	21	52.80
2602	0.1000	0.747	0.30	-0.09	0.00	10	25.61
2502	0.0000	0.676	0.53	0.00	0.11	7	19.42
2503	0.3333	0.360	0.33	-0.05	0.04	15	41.40
2504	0.1818	0.798	0.70	-0.10	0.05	11	25.63
3501	0.2500	0.101	0.67	-0.07	0.00	4	9.06
3502	0.0000	1.227	0.63	0.00	0.21	6	13.60
3402	0.1000	0.739	0.33	-0.04	0.00	20	61.03
4501	0.2857	1.301	0.09	-0.16	0.05	7	16.12
4603	0.1071	2.091	0.35	-0.05	0.00	28	88.77
4604	0.1429	0.457	0.87	-0.10	0.05	7	16.64
4402	0.0000	0.676	0.91	0.00	0.05	24	69.08
5401	0.3529	0.152	0.15	-0.02	0.03	17	45.62
5402	0.0000	0.279	0.51	0.00	0.06	11	27.59
6601	0.0000	0.341	0.15	0.00	0.09	6	14.76
6703	0.0000	0.476	0.50	0.00	0.08	7	16.97
6704	0.1429	2.403	0.58	-0.04	0.27	7	14.32
6603	0.0000	0.163	0.40	0.00	0.08	4	10.20
6604	0.2727	1.261	0.74	-0.04	0.11	11	28.16
6709	0.3333	0.253	0.29	-0.08	0.04	6	15.23
6710	0.0667	4.015	0.49	-0.13	0.00	15	40.36
6606	0.2500	0.077	0.31	-0.06	0.00	4	10.20
6607	0.2000	0.233	0.83	-0.10	0.02	5	12.04
6715	0.2500	0.183	0.75	-0.10	0.00	4	8.94
6716	0.2857	1.945	0.88	-0.03	0.17	7	18.79
7401	0.1667	0.267	0.63	-0.05	0.00	6	19.00
7503	0.1111	3.705	0.60	-0.20	0.00	9	22.80
7607	0.5000	0.764	0.34	-0.05	0.19	6	12.17
7608	0.0000	3.013	0.23	0.00	0.11	15	40.36
8501	0.0000	0.157	0.67	0.00	0.06	6	16.16
8603	0.2222	1.702	0.35	-0.15	0.02	9	23.35
8604	0.2857	1.215	0.71	-0.03	0.17	7	17.20
8605	0.0000	0.450	0.75	0.00	0.15	4	8.94
8711	0.0000	0.402	0.67	0.00	0.15	4	9.06
8712	0.1667	0.551	0.43	-0.12	0.01	6	14.87
8504	0.3333	0.651	0.84	-0.08	0.14	6	13.93

TABLE B.2
LOW ORDER PARAMETERS

PRINTS:HUDBY.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
1	0.0300	68.163	0.29	-0.41	0.07	100	82.42
2	0.1379	16.505	0.25	-0.15	0.43	29	32.76
3	0.1000	21.920	0.74	-0.20	0.21	50	71.59
4	0.0833	8.691	0.54	-0.59	0.28	24	15.81
5	0.0769	40.435	0.59	-0.10	0.71	26	27.17
6	0.0541	22.313	0.55	-0.22	0.03	37	77.90
7	0.1111	7.942	0.63	-0.46	0.22	18	19.24
8	0.2222	2.699	0.16	-0.22	0.15	18	25.32
1401	0.0000	85.830	0.28	0.00	0.64	38	41.44
1402	0.0476	22.649	0.99	-0.26	0.23	63	68.03
2401	0.0000	0.744	0.09	0.00	0.13	8	16.28
2402	0.0909	19.602	0.40	-0.54	0.03	22	28.32
3401	0.1282	51.809	0.59	-0.45	0.13	39	54.59
3402	0.0833	1.786	0.82	-0.15	0.07	12	24.60
4401	0.2500	3.105	0.49	-0.14	0.45	12	12.37
4402	0.1538	1.229	1.04	-0.01	0.13	13	25.71
5401	0.2353	5.316	0.09	-0.27	0.01	17	25.24
5402	0.0000	0.629	0.45	0.00	0.11	10	22.20
6401	0.0952	1.686	0.58	-0.10	0.07	21	45.88
6402	0.0000	1.269	0.17	0.00	0.10	17	39.00
7401	0.1250	4.150	0.24	-0.33	0.00	8	14.76
7402	0.1818	7.849	0.62	-0.10	0.66	11	11.66
8401	0.2500	1.117	1.00	-0.33	0.00	4	6.71
8402	0.2667	4.096	0.64	-0.01	0.20	15	29.73
1501	0.1765	8.987	0.24	-0.30	0.05	17	29.00
1502	0.1818	2.812	0.35	-0.08	0.16	22	39.82
1503	0.0000	33.152	0.34	0.00	0.29	56	69.66
1504	0.3750	1.225	0.43	-0.10	0.15	8	17.89
2401	0.0000	0.744	0.09	0.00	0.13	8	16.28
2503	0.2222	3.303	0.26	-0.04	0.24	9	19.10
2504	0.1429	3.363	0.65	-0.09	0.26	14	23.00
3501	0.0741	43.880	0.42	-0.02	0.50	27	40.31
3502	0.1538	5.182	0.51	-0.03	0.20	13	33.42
3503	0.0000	0.745	0.15	0.00	0.11	9	20.88
3504	0.2500	2.023	0.59	-0.55	0.00	4	5.39
4501	0.5000	1.040	0.72	-0.03	0.26	4	8.06
4502	0.1111	0.671	0.63	-0.12	0.00	9	19.03
4402	0.1538	1.229	1.04	-0.01	0.13	13	25.71
5501	0.0000	9.832	0.57	0.00	1.15	8	7.28
5502	0.2000	0.126	0.87	-0.00	0.04	10	24.04
5402	0.0000	0.629	0.45	0.00	0.11	10	22.20
6501	0.2308	0.318	0.48	-0.06	0.06	13	27.46
6502	0.0000	2.875	0.56	0.00	0.25	9	19.42

TABLE B.2
LOW ORDER PARAMETERS

PRINTS:HUDBY.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
6402	0.0000	1.269	0.17	0.00	0.10	17	39.00
7401	0.1250	4.150	0.24	-0.33	0.00	8	14.76
7503	0.1667	1.123	0.10	-0.24	0.00	6	10.44
7504	0.0000	2.567	0.78	0.00	0.44	6	9.00
8401	0.2500	1.117	1.00	-0.33	0.00	4	6.71
8503	0.2727	1.415	0.61	-0.20	0.07	11	20.00
8504	0.4000	0.073	0.74	-0.04	0.04	5	12.17
1601	0.2000	0.051	0.45	-0.05	0.02	5	11.05
1602	0.2308	0.919	0.87	-0.07	0.12	13	23.77
1603	0.1111	3.140	0.67	-0.28	0.01	9	15.00
1604	0.1429	3.914	0.65	-0.21	0.01	14	26.93
1605	0.1053	3.527	0.32	-0.00	0.18	19	31.02
1606	0.1053	10.433	0.29	-0.24	0.11	38	50.54
1504	0.3750	1.225	0.43	-0.10	0.15	8	17.89
2401	0.0000	0.744	0.09	0.00	0.13	8	16.28
2605	0.2500	2.819	0.56	-0.49	0.00	4	6.40
2606	0.3333	0.293	0.35	-0.03	0.09	6	15.23
2607	0.1429	0.281	0.47	-0.02	0.09	7	16.16
2608	0.2500	8.759	0.14	-0.74	0.10	8	10.00
3601	0.4000	1.346	0.12	-0.08	0.14	15	26.08
3602	0.0000	2.746	0.68	0.00	0.18	13	31.06
3603	0.1429	2.800	0.56	-0.24	0.00	7	18.36
3604	0.4286	0.237	0.71	-0.02	0.08	7	17.72
3503	0.0000	0.745	0.15	0.00	0.11	9	20.88
3504	0.2500	2.023	0.59	-0.55	0.00	4	5.39
4501	0.5000	1.040	0.72	-0.03	0.26	4	8.06
4502	0.1111	0.671	0.63	-0.12	0.00	9	19.03
4402	0.1538	1.229	1.04	-0.01	0.13	13	25.71
5601	0.5000	0.285	0.61	-0.09	0.11	4	9.22
5602	0.2000	0.624	0.78	-0.20	0.00	5	9.06
5502	0.2000	0.126	0.87	-0.00	0.04	10	24.04
5402	0.0000	0.629	0.45	0.00	0.11	10	22.20
6501	0.2308	0.318	0.48	-0.06	0.06	13	27.46
6502	0.0000	2.875	0.56	0.00	0.25	9	19.42
6402	0.0000	1.269	0.17	0.00	0.10	17	39.00
7401	0.1250	4.150	0.24	-0.33	0.00	8	14.76
7503	0.1667	1.123	0.10	-0.24	0.00	6	10.44
7504	0.0000	2.567	0.78	0.00	0.44	6	9.00
8401	0.2500	1.117	1.00	-0.33	0.00	4	6.71
8605	0.0000	4.498	1.00	0.00	0.43	8	12.65
8606	0.0000	0.183	0.75	0.00	0.10	4	8.94
8504	0.4000	0.073	0.74	-0.04	0.04	5	12.17
1601	0.2000	0.051	0.45	-0.05	0.02	5	11.05

TABLE B.2
LOW ORDER PARAMETERS

PRINTS:HUDBY.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
1602	0.2308	0.919	0.87	-0.07	0.12	13	23.77
1603	0.1111	3.140	0.67	-0.28	0.01	9	15.00
1707	0.3750	1.232	0.72	-0.05	0.15	8	18.03
1708	0.2857	0.286	0.38	-0.12	0.06	7	11.31
1709	0.3333	0.208	0.12	-0.09	0.06	6	11.66
1710	0.2857	0.916	0.62	-0.11	0.15	14	21.59
1711	0.2727	0.866	0.16	-0.12	0.12	11	19.42
1712	0.0714	16.198	0.34	-0.00	0.37	28	37.48
1504	0.3750	1.225	0.43	-0.10	0.15	8	17.89
2401	0.0000	0.744	0.09	0.00	0.13	8	16.28
2605	0.2500	2.819	0.56	-0.49	0.00	4	6.40
2606	0.3333	0.293	0.35	-0.03	0.09	6	15.23
2607	0.1429	0.281	0.47	-0.02	0.09	7	16.16
2715	0.0000	0.023	0.42	0.00	0.03	3	7.62
2716	0.3333	0.668	0.91	-0.11	0.19	6	11.05
3701	0.4000	0.720	0.28	-0.40	0.04	5	5.00
3702	0.1818	2.963	0.13	-0.20	0.07	11	23.09
3703	0.1111	0.853	0.76	-0.11	0.06	9	22.14
3704	0.2000	0.300	0.80	-0.12	0.04	5	11.18
3603	0.1429	2.800	0.56	-0.24	0.00	7	18.36
3604	0.4286	0.237	0.71	-0.02	0.08	7	17.72
3503	0.0000	0.745	0.15	0.00	0.11	9	20.88
3504	0.2500	2.023	0.59	-0.55	0.00	4	5.39
4501	0.5000	1.040	0.72	-0.03	0.26	4	8.06
4502	0.1111	0.671	0.63	-0.12	0.00	9	19.03
4402	0.1538	1.229	1.04	-0.01	0.13	13	25.71
5601	0.5000	0.285	0.61	-0.09	0.11	4	9.22
5602	0.2000	0.624	0.78	-0.20	0.00	5	9.06
5502	0.2000	0.126	0.87	-0.00	0.04	10	24.04
5402	0.0000	0.629	0.45	0.00	0.11	10	22.20
6501	0.2308	0.318	0.48	-0.06	0.06	13	27.46
6502	0.0000	2.875	0.56	0.00	0.25	9	19.42
6402	0.0000	1.269	0.17	0.00	0.10	17	39.00
7401	0.1250	4.150	0.24	-0.33	0.00	8	14.76
7503	0.1667	1.123	0.10	-0.24	0.00	6	10.44
7504	0.0000	2.567	0.78	0.00	0.44	6	9.00
8401	0.2500	1.117	1.00	-0.33	0.00	4	6.71
8709	0.4000	0.357	0.21	-0.09	0.04	5	13.45
8710	0.2500	2.034	0.83	-0.52	0.00	4	5.39
8606	0.0000	0.183	0.75	0.00	0.10	4	8.94
8504	0.4000	0.073	0.74	-0.04	0.04	5	12.17

TABLE B.3
LOW ORDER PARAMETERS

PRINTS:FJORD.TBL.FIN

STRIP	MONOTONICITY	ERROR VARIANCE	SEGMENTATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
1	0.1013	28.323	0.73	-0.16	0.19	79	105.08
2	0.0270	136.614	0.59	-1.26	0.00	37	27.00
3	0.0685	87.608	1.16	-0.42	0.72	73	50.00
4	0.0000	33.325	0.48	0.00	0.32	29	56.08
5	0.0667	21.994	0.63	-0.61	0.53	45	27.29
6	0.0549	57.983	0.61	-0.26	0.28	91	97.08
7	0.0374	181.746	0.43	-0.84	1.01	107	49.41
8	0.0652	13.926	0.27	-0.18	0.12	46	72.45
1401	0.0161	55.336	0.71	-0.33	0.00	62	79.40
1402	0.1111	23.405	0.42	-0.40	0.01	18	34.67
2401	0.0588	6.018	0.76	-0.21	0.12	17	37.58
2402	0.0952	17.773	0.85	-0.06	0.39	21	35.74
3401	0.0175	110.475	0.24	-0.50	0.00	57	68.68
3402	0.0000	8.572	0.65	0.00	0.26	17	37.05
4401	0.3077	3.044	0.38	-0.05	0.14	13	32.28
4402	0.1765	0.824	0.77	-0.08	0.04	17	34.37
5401	0.0541	13.618	0.31	-0.43	0.56	37	23.60
5402	0.2222	0.742	0.11	-0.12	0.08	9	19.70
6401	0.0256	63.279	0.42	-0.39	0.00	39	65.00
6402	0.0189	133.075	0.16	-0.88	0.01	53	47.01
7401	0.0833	7.305	0.60	-0.06	0.17	24	54.34
7402	0.0476	260.842	0.96	-0.61	0.17	84	86.65
8401	0.0909	3.972	0.42	-0.21	0.02	11	23.60
8402	0.0278	10.791	0.45	-0.10	0.20	36	54.33
1501	0.0833	9.279	0.53	-0.18	0.14	48	62.65
1502	0.1333	6.961	0.41	-0.20	0.01	15	34.93
1503	0.1429	0.472	0.74	-0.10	0.00	7	19.42
1504	0.1667	1.841	0.72	-0.16	0.04	12	25.00
2501	0.0000	4.702	0.17	0.00	0.21	13	29.43
2502	0.2000	0.186	0.19	-0.08	0.00	5	12.08
2503	0.0625	2.973	0.30	-0.16	0.00	16	33.24
2504	0.1667	0.700	0.87	-0.12	0.00	6	15.03
3501	0.0938	30.842	0.03	-0.49	0.17	32	37.64
3502	0.0769	12.620	0.61	-0.19	0.01	26	62.80
3503	0.3000	1.084	0.35	-0.04	0.13	10	25.50
3504	0.0000	3.576	0.92	0.00	0.32	8	16.64
4501	0.1667	1.891	0.38	-0.26	0.00	6	13.04
4502	0.2500	0.482	0.54	-0.08	0.06	8	20.59
4402	0.1765	0.824	0.77	-0.08	0.04	17	34.37
5501	0.1000	9.879	0.66	-0.62	0.11	20	15.13
5502	0.1667	2.028	0.80	-0.09	0.15	18	33.62
5402	0.2222	0.742	0.11	-0.12	0.08	9	19.70
6501	0.0435	15.224	0.57	-0.10	0.33	23	37.22

TABLE B.3
LOW ORDER PARAMETERS

PRINTS:FJORD.TBL.FIN

STRIP	MONOTONICITY	ERROR VARIANCE	SEGMENTATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
6502	0.1176	2.851	0.77	-0.01	0.11	17	45.54
6503	0.1875	0.953	0.00	-0.07	0.07	16	42.20
6504	0.0789	34.939	0.49	-0.36	0.09	38	57.38
7501	0.0667	4.247	0.68	-0.19	0.00	15	34.06
7502	0.4000	1.093	0.94	-0.06	0.13	10	23.60
7503	0.0750	7.690	0.54	-0.03	0.09	40	98.72
7504	0.0889	11.348	0.23	-0.16	0.25	45	52.89
8501	0.0000	1.589	0.54	0.00	0.27	6	11.18
8502	0.3333	0.177	0.40	-0.08	0.04	6	14.42
8503	0.0000	3.111	0.60	0.00	0.16	11	26.25
8504	0.1154	7.339	0.72	-0.27	0.25	26	32.20
1601	0.0667	1.609	0.32	-0.00	0.13	15	34.89
1602	0.0000	33.849	0.29	0.00	0.61	34	31.78
1603	0.1429	0.221	0.32	-0.03	0.07	7	15.62
1604	0.2222	0.779	0.68	-0.03	0.11	9	22.09
1503	0.1429	0.472	0.74	-0.10	0.00	7	19.42
1607	0.5000	0.075	0.28	-0.05	0.02	8	18.44
1[illegible]	0.4000	0.365	1.00	-0.08	0.17	5	8.06
2601	0.0000	0.353	0.72	0.00	0.15	4	8.06
2602	0.1000	0.355	0.80	-0.01	0.06	10	25.08
2502	0.2000	0.186	0.19	-0.08	0.00	5	12.08
2605	0.2000	1.096	0.27	-0.18	0.00	5	11.70
2606	0.0833	1.409	0.77	-0.14	0.12	12	23.41
2504	0.1667	0.700	0.87	-0.12	0.00	6	15.03
3601	0.0625	14.827	0.17	-0.07	0.64	16	18.44
3602	0.5294	1.049	0.55	-0.05	0.08	17	40.61
3603	0.2941	3.844	0.61	-0.09	0.14	17	40.25
3604	0.0000	0.713	0.19	0.00	0.10	10	27.20
3503	0.3000	1.084	0.35	-0.04	0.13	10	25.50
3607	0.3333	0.133	0.25	-0.03	0.06	6	16.49
3608	0.3333	0.011	0.42	-0.03	0.00	3	5.39
4501	0.1667	1.891	0.38	-0.26	0.00	6	13.04
4502	0.2500	0.482	0.54	-0.08	0.06	8	20.59
4402	0.1765	0.824	0.77	-0.08	0.04	17	34.37
5601	0.1111	7.881	0.74	-0.52	0.00	9	13.60
5602	0.1667	0.920	0.85	-0.05	0.12	12	26.68
5603	0.0714	5.590	0.77	-0.22	0.01	14	27.66
5604	0.2000	1.711	0.25	-0.34	0.00	5	8.06
5402	0.2222	0.742	0.11	-0.12	0.08	9	19.70
6601	0.0909	3.103	0.41	-0.19	0.00	11	24.17
6602	0.2308	6.049	0.86	-0.34	0.06	13	20.52
6603	0.3077	0.395	0.43	-0.05	0.03	13	35.69
6604	0.2000	0.150	0.25	-0.06	0.00	5	11.31

TABLE B.3
LOW ORDER PARAMETERS

PRINTS:FJORD.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
6605	0.0000	0.000	0.00	0.00	0.00	2	3.16
6606	0.2000	2.704	0.32	-0.12	0.01	15	42.06
6607	0.0000	25.478	0.41	0.00	0.48	25	34.93
6608	0.3571	0.624	0.47	-0.05	0.06	14	35.51
7601	0.1000	1.187	0.22	-0.13	0.00	10	23.54
7602	0.3333	0.216	0.45	-0.03	0.09	6	13.04
7502	0.4000	1.093	0.94	-0.06	0.13	10	23.60
7605	0.1304	3.391	0.39	-0.12	0.05	23	54.15
7606	0.0556	5.629	0.52	-0.17	0.01	18	46.10
7607	0.2000	2.307	0.37	-0.23	0.07	10	18.36
7608	0.1111	35.066	0.26	-0.44	0.35	36	42.05
8501	0.0000	1.589	0.54	0.00	0.27	6	11.18
8502	0.3333	0.177	0.40	-0.03	0.04	6	14.42
8503	0.0000	3.111	0.60	0.00	0.16	11	26.25
8607	0.0000	36.029	0.94	0.00	0.71	21	24.76
8608	0.3333	0.137	0.25	-0.07	0.03	6	12.65
1601	0.0667	1.609	0.32	-0.00	0.13	15	34.89
1703	0.1333	19.583	0.07	-0.63	0.02	15	21.40
1704	0.1000	5.760	0.61	-0.01	0.16	20	45.12
1603	0.1429	0.221	0.32	-0.03	0.07	7	15.62
1604	0.2222	0.779	0.68	-0.03	0.11	9	22.09
1503	0.1429	0.472	0.74	-0.10	0.00	7	19.42
1607	0.5000	0.075	0.28	-0.05	0.02	8	18.44
1608	0.4000	0.365	1.00	-0.08	0.17	5	8.06
2601	0.0000	0.353	0.72	0.00	0.15	4	8.06
2602	0.1000	0.355	0.80	-0.01	0.06	10	25.08
2502	0.2000	0.186	0.19	-0.08	0.00	5	12.08
2605	0.2000	1.096	0.27	-0.18	0.00	5	11.70
2711	0.0000	2.303	0.39	0.00	0.22	9	18.25
2712	0.2500	0.022	0.44	-0.05	0.00	4	6.40
2504	0.1667	0.700	0.87	-0.12	0.00	6	15.03
3701	0.2857	1.810	0.75	-0.08	0.25	7	12.00
3702	0.1000	1.469	0.82	-0.17	0.00	10	24.08
3703	0.2222	0.553	0.57	-0.10	0.00	9	22.14
3704	0.1111	2.694	0.16	-0.26	0.00	9	19.00
3705	0.0909	4.762	0.44	-0.25	0.00	11	25.71
3706	0.1429	0.838	0.69	-0.14	0.02	7	16.28
3604	0.0000	0.713	0.19	0.00	0.13	10	27.20
3503	0.3000	1.084	0.35	-0.04	0.13	10	25.50
3607	0.3333	0.133	0.25	-0.03	0.06	6	16.49
3608	0.3333	0.011	0.42	-0.03	0.00	3	5.39
4501	0.1667	1.891	0.38	-0.26	0.00	6	13.04
4502	0.2500	0.482	0.54	-0.08	0.06	8	20.59

TABLE B.3
LOW ORDER PARAMETERS

PRINTS:FJORD.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
4402	0.1765	0.824	0.77	-0.08	0.04	17	34.37
5701	0.0000	0.000	0.00	0.00	0.00	5	12.00
5702	0.2000	0.880	0.25	-0.28	0.00	5	8.06
5602	0.1667	0.920	0.85	-0.05	0.12	12	26.68
5705	0.1000	0.167	0.80	-0.05	0.05	10	21.93
5706	0.2000	0.155	0.11	-0.10	0.00	5	9.06
5604	0.2000	1.711	0.25	-0.34	0.00	5	8.06
5402	0.2222	0.742	0.11	-0.12	0.08	9	19.70
6601	0.0909	3.103	0.41	-0.19	0.00	11	24.17
6703	0.4000	1.460	0.24	-0.11	0.17	10	18.79
6704	0.5000	0.138	0.29	-0.10	0.03	4	7.62
6603	0.3077	0.395	0.43	-0.05	0.03	13	35.69
6604	0.2000	0.150	0.25	-0.06	0.00	5	11.31
6605	0.0000	0.000	0.00	0.00	0.00	2	3.16
6711	0.0000	0.302	0.45	0.00	0.10	6	14.21
6712	0.1000	0.652	0.58	-0.02	0.08	10	29.27
6713	0.0909	7.080	0.61	-0.32	0.00	11	22.20
6714	0.2000	2.428	0.43	-0.15	0.16	15	26.40
6608	0.3571	0.624	0.47	-0.05	0.06	14	35.51
7601	0.1000	1.187	0.22	-0.13	0.00	10	23.54
7602	0.3333	0.216	0.45	-0.03	0.09	6	13.04
7502	0.4000	1.093	0.94	-0.06	0.13	10	23.60
7709	0.0000	2.136	0.28	0.00	0.21	10	22.47
7710	0.2143	4.272	0.33	-0.02	0.17	14	33.24
7711	0.1818	0.983	0.58	-0.02	0.11	11	25.50
7712	0.1250	0.182	0.61	-0.02	0.05	8	23.09
7713	0.4000	1.520	0.12	-0.34	0.08	5	8.06
7714	0.3333	0.502	0.92	-0.05	0.15	6	12.17
7715	0.1111	3.253	0.69	-0.24	0.00	9	21.47
7716	0.1429	6.831	0.63	-0.04	0.26	28	36.24
8501	0.0000	1.589	0.54	0.00	0.27	6	11.18
8502	0.3333	0.177	0.40	-0.08	0.04	6	14.42
8503	0.0000	3.111	0.60	0.00	0.16	11	26.25
8713	0.0769	2.073	0.45	-0.17	0.00	13	29.43
8714	0.1111	0.297	0.08	-0.04	0.08	9	17.69
8608	0.3333	0.137	0.25	-0.07	0.03	6	12.65

TABLE B.4
LOW ORDER PARAMETERS

PRINTS:TEXAS.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
1	0.0909	8.947	0.60	-0.13	0.03	55	79.16
2	0.0364	123.311	0.69	-1.78	2.88	55	13.45
3	0.0261	205.694	0.34	-0.53	0.09	115	101.32
4	0.0959	10.639	0.20	-0.14	0.17	73	59.41
5	0.0444	43.533	0.42	-0.24	0.16	90	103.82
6	0.0286	93.688	0.63	-0.02	0.52	70	61.52
7	0.0769	6.611	0.86	-0.12	0.22	26	44.41
8	0.0370	200.660	1.00	-0.70	3.27	54	13.42
1401	0.0556	0.611	0.32	-0.05	0.05	18	48.51
1402	0.0526	29.512	0.66	-0.02	0.52	38	33.24
2401	0.1333	0.315	0.91	-0.04	0.05	15	39.82
2402	0.0488	33.203	1.49	-0.46	0.10	41	38.95
3401	0.0806	11.881	0.06	-0.19	0.18	62	64.29
3402	0.0370	154.394	0.58	-0.01	0.48	54	85.16
4401	0.0000	144.995	2.50	0.00	2.47	44	16.03
4402	0.1667	6.185	0.88	-0.13	0.01	30	72.42
5401	0.0000	45.267	0.56	0.00	0.41	25	50.64
5402	0.0000	67.507	0.35	0.00	0.43	66	65.07
6401	0.0426	39.210	0.16	-0.15	0.38	47	50.25
6402	0.0833	2.878	0.33	-0.15	0.07	24	39.62
7401	0.1364	3.116	0.11	-0.14	0.11	22	37.05
7402	0.2000	0.184	0.27	-0.04	0.09	5	11.70
8401	0.0571	5.859	0.12	-0.17	0.04	35	45.97
8402	0.0500	0.336	0.52	-0.04	0.02	20	51.62
1401	0.0556	0.611	0.32	-0.05	0.05	18	48.51
1503	0.1176	12.643	0.07	-0.35	0.01	17	28.16
1504	0.1364	0.474	0.35	-0.02	0.04	22	58.14
2401	0.1333	0.315	0.91	-0.04	0.05	15	39.82
2503	0.1250	0.455	0.25	-0.01	0.04	24	60.54
2504	0.1667	3.035	0.08	-0.20	0.14	18	26.08
3501	0.1667	4.651	0.31	-0.41	0.44	12	13.04
3502	0.0588	28.369	0.57	-0.08	0.28	51	61.13
3503	0.1143	5.091	0.78	-0.10	0.14	35	64.50
3504	0.0500	0.412	0.31	-0.03	0.00	20	53.85
4501	0.0476	1.000	0.60	-0.05	0.00	21	56.64
4502	0.0833	1.056	0.51	-0.01	0.08	24	46.53
4503	0.1600	2.183	0.83	-0.00	0.07	25	63.95
4504	0.1667	0.119	0.61	-0.06	0.00	6	13.15
5501	0.4615	0.083	0.81	-0.02	0.03	13	34.89
5502	0.2308	0.900	0.46	-0.08	0.07	13	30.61
5503	0.2727	4.032	0.31	-0.19	0.12	22	36.12
5504	0.1333	12.450	0.56	-0.14	0.22	45	50.80
6501	0.0690	48.183	0.61	-1.04	0.06	29	20.59

TABLE B.4
LOW ORDER PARAMETERS

PRINTS:TEXAS.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
6502	0.1053	1.697	0.27	-0.00	0.08	19	46.14
6503	0.1111	1.239	0.25	-0.10	0.19	9	14.42
6504	0.0000	1.431	0.26	0.00	0.14	16	27.02
7501	0.1667	4.000	0.29	-0.71	0.00	6	7.00
7502	0.0588	6.210	0.29	-0.08	0.24	17	34.99
7402	0.2000	0.184	0.27	-0.04	0.09	5	11.70
8501	0.1667	6.371	1.00	-0.60	0.00	6	10.05
8502	0.1333	1.641	0.26	-0.03	0.10	30	52.80
8402	0.0500	0.336	0.52	-0.04	0.02	20	51.62
1401	0.0556	0.611	0.32	-0.05	0.05	18	48.51
1605	0.1667	1.122	0.70	-0.23	0.00	6	10.05
1606	0.1667	0.310	0.58	-0.01	0.05	12	31.78
1504	0.1364	0.474	0.35	-0.02	0.04	22	58.14
2401	0.1333	0.315	0.91	-0.04	0.05	15	39.82
2503	0.1250	0.455	0.25	-0.01	0.04	24	60.54
2607	0.2500	0.770	0.83	-0.34	0.00	4	5.39
2608	0.2000	0.689	1.25	-0.06	0.10	15	25.00
3601	0.1250	6.087	0.31	-0 85	0.00	8	7.21
3602	0.2000	0.399	0.78	-0.14	0.01	5	10.30
3603	0.1176	30.927	0.59	-0.44	0.03	34	39.00
3604	0.2778	3.685	0.07	-0.19	0.03	18	31.40
3605	0.0323	12.375	0.70	-0.25	0.00	31	51.86
3606	0.2000	0.038	0.56	-0.03	0.00	5	16.03
3504	0.0500	0.412	0.31	-0.03	0.00	20	53.85
4501	0.0476	1.000	0.60	-0.05	0.00	21	56.64
4502	0.0833	1.056	0.51	-0.01	0.08	24	46.53
4503	0.1600	2.183	0.83	-0.00	0.07	25	63.95
4504	0.1667	0.119	0.61	-0.06	0.00	6	13.15
5501	0.4615	0.083	0.81	-0.02	0.03	13	34.89
5502	0.2308	0.900	0.46	-0.08	0.07	13	30.61
5605	0.2500	11.921	0.67	-0.10	0.73	12	13.42
5606	0.1818	4.506	0.42	-0.00	0.22	11	25.63
5607	0.1071	18.179	0.46	-0.44	0.12	28	30.53
5608	0.1111	10.992	0.38	-0.39	0.12	18	25.00
6601	0.1250	9.872	0.56	-0.38	0.01	16	25.02
6602	0.1429	1.866	0.53	-0.03	0.16	14	22.67
6502	0.1053	1.697	0.27	-0.00	0.08	19	46.14
6503	0.1111	1.239	0.25	-0.10	0.19	9	14.42
6504	0.0000	1.431	0.26	0.00	0.14	16	27.02
7601	0.3333	0.414	0.59	-0.21	0.00	3	5.39
7602	0.0000	0.125	0.32	0.00	0.10	4	7.07
7603	0.1667	0.099	0.39	-0.06	0.00	6	13.15
7604	0.0833	6.891	0.35	-0.27	0.00	12	26.25

TABLE B.4
LOW ORDER PARAMETERS

PRINTS:TEXAS.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
7402	0.2000	0.184	0.27	-0.04	0.09	5	11.70
8601	0.2000	0.029	0.26	-0.03	0.03	5	12.08
8602	0.0000	0.000	0.00	0.00	0.00	2	6.08
8603	0.1000	0.805	0.43	-0.17	0.00	10	14.56
8604	0.1905	4.081	0.04	-0.16	0.05	21	39.60
8402	0.0500	0.336	0.52	-0.04	0.02	20	51.62
1401	0.0556	0.611	0.32	-0.05	0.05	18	48.51
1605	0.1667	1.122	0.70	-0.23	0.00	6	10.05
1606	0.1667	0.310	0.58	-0.01	0.05	12	31.78
1504	0.1364	0.474	0.35	-0.02	0.04	22	58.14
2401	0.1333	0.315	0.91	-0.04	0.05	15	39.82
2503	0.1250	0.455	0.25	-0.01	0.04	24	60.54
2607	0.2500	0.770	0.83	-0.34	0.00	4	5.39
2608	0.2000	0.689	1.25	-0.06	0.10	15	25.00
3701	0.2500	0.136	1.00	-0.12	0.00	4	6.40
3702	0.0000	0.300	0.55	0.00	0.09	5	11.00
3602	0.2000	0.399	0.78	-0.14	0.01	5	10.30
3705	0.0000	21.354	0.68	0.00	0.49	19	28.86
3706	0.2500	8.199	0.59	-0.09	0.35	16	23.41
3707	0.2500	1.100	0.50	-0.35	0.00	4	6.32
3708	0.0667	2.654	0.15	-0.10	0.18	15	29.97
3709	0.1818	8.663	0.92	-0.06	0.25	22	38.28
3710	0.2000	3.104	0.35	-0.03	0.22	10	20.40
3606	0.2000	0.038	0.56	-0.03	0.00	5	16.03
3504	0.0500	0.412	0.31	-0.03	0.00	20	53.85
4501	0.0476	1.000	0.60	-0.05	0.00	21	56.64
4502	0.0833	1.056	0.51	-0.01	0.08	24	46.53
4503	0.1600	2.183	0.83	-0.00	0.07	25	63.95
4504	0.1667	0.119	0.61	-0.06	0.00	6	13.15
5501	0.4615	0.083	0.81	-0.02	0.03	13	34.89
5502	0.2308	0.900	0.46	-0.08	0.07	13	30.61
5709	0.2857	1.855	0.17	-0.23	0.09	7	13.00
5710	0.1667	2.967	0.73	-0.36	0.00	6	11.00
5711	0.4000	0.154	0.70	-0.03	0.08	5	12.04
5712	0.2857	1.667	0.50	-0.19	0.00	7	16.00
5713	0.0000	42.847	0.90	0.00	0.98	20	19.42
5714	0.4444	0.364	0.36	-0.07	0.06	9	21.19
5715	0.1111	2.439	0.33	-0.10	0.33	9	13.42
5716	0.0000	3.116	0.78	0.00	0.28	10	18.68
6701	0.3333	1.374	0.35	-0.12	0.18	9	16.64
6702	0.0000	2.371	0.72	0.00	0.27	8	14.87
6602	0.1429	1.866	0.53	-0.03	0.16	14	22.67
6502	0.1053	1.697	0.27	-0.00	0.08	19	46.14

TABLE B.4
LOW ORDER PARAMETERS

PRINTS:TEXAS.TBL.FIN

STRIP	MONOTO-NICITY	ERROR VARIANCE	SEGMEN-TATION	WIDTH (NEG)	WIDTH (POS)	NPTS	LENGTH
6503	0.1111	1.239	0.25	-0.10	0.19	9	14.42
6504	0.0000	1.431	0.26	0.00	0.14	16	27.02
7601	0.3333	0.414	0.59	-0.21	0.00	3	5.39
7602	0.0000	0.125	0.32	0.00	0.10	4	7.07
7603	0.1667	0.099	0.39	-0.06	0.00	6	13.15
7707	0.1667	0.542	0.67	-0.14	0.00	6	12.04
7708	0.1429	0.590	0.50	-0.12	0.00	7	17.89
7402	0.2000	0.184	0.27	-0.04	0.09	5	11.70
8601	0.2000	0.029	0.26	-0.03	0.03	5	12.08
8602	0.0000	0.000	0.00	0.00	0.00	2	6.08
8603	0.1000	0.805	0.43	-0.17	0.00	10	14.56
8707	0.0000	0.201	0.66	0.00	0.15	4	6.40
8708	0.1111	1.181	0.40	-0.01	0.08	18	39.41
8402	0.0500	0.336	0.52	-0.04	0.02	20	51.62

APPENDIX C
LOW ORDER PARAMETER LOOK-UP TABLES

Included in this Appendix are the means, standard deviations, and coefficients of variation for all parameters measured at each level of subdivision; these figures are computed as the summary for all measures within a given level of the strip tree data structure. These summary values are also stored digitally, as an in-core look-up table; their application to line discrimination tasks is described in detail in Chapter 4, and again in Chapter 6.

TABLE C.1
LOW ORDER SUMMARIES

SIOUX

MONOTONICITY	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.0500	0.0328	0.6560
	2	0.1320	0.1066	0.8076
	3	0.1573	0.1193	0.7584
	4	0.1571	0.1325	0.8434
	5	0.1548	0.1339	0.8650

ERROR VARIANCE	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	26.2560	29.2350	1.1135
	2	4.3400	5.2100	1.2005
	3	1.9100	2.5430	1.3314
	4	1.0870	1.0480	0.9641
	5	1.1430	1.3600	1.1899

SEGMENTATION	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.5000	0.2100	0.4200
	2	0.4600	0.2600	0.5652
	3	0.4900	0.2600	0.5306
	4	0.4700	0.2400	0.5106
	5	0.5100	0.2300	0.4510

NEGATIVE WIDTH	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	-0.0700	0.0700	-1.0000
	2	-0.0700	0.0800	-1.1429
	3	-0.0600	0.0600	-1.0000
	4	-0.0600	0.0600	-1.0000
	5	-0.0500	0.0500	-1.0000

POSITIVE WIDTH	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.0900	0.0900	1.0000
	2	0.0600	0.0500	0.8333
	3	0.0700	0.0700	1.0000
	4	0.0700	0.0600	0.8571
	5	0.0800	0.0700	0.8750

TABLE C.2
LOW ORDER SUMMARIES

HUDBY

MONOTONICITY	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.1019	0.0589	0.5780
	2	0.1192	0.0964	0.8087
	3	0.1533	0.1378	0.8989
	4	0.1802	0.1506	0.8357
	5	0.2096	0.1457	0.6951

ERROR VARIANCE	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	23.5800	21.4800	0.9109
	2	13.3000	23.4700	1.7647
	3	5.1500	10.2300	1.9864
	4	1.9700	2.3100	1.1726
	5	1.4600	2.5400	1.7397

SEGMENTATION	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.4700	0.2100	0.4468
	2	0.5300	0.3200	0.6038
	3	0.5000	0.2600	0.5200
	4	0.5400	0.2700	0.5000
	5	0.5400	0.2700	0.5000

NEGATIVE WIDTH	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	-0.2900	0.1700	-0.5862
	2	-0.1700	0.1800	-1.0588
	3	-0.1000	0.1400	-1.4000
	4	-0.1300	0.1800	-1.3846
	5	-0.1300	0.1500	-1.1538

POSITIVE WIDTH	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.2600	0.2200	0.8462
	2	0.1900	0.2100	1.1053
	3	0.1800	0.2400	1.3333
	4	0.1000	0.1100	1.1000
	5	0.0900	0.1000	1.1111

TABLE C.3
LOW ORDER SUMMARIES

FJORD

MONOTONICITY	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.0526	0.0308	0.5856
	2	0.0846	0.0850	1.0047
	3	0.1338	0.0983	0.7347
	4	0.1839	0.1372	0.7461
	5	0.1852	0.1291	0.6971

ERROR VARIANCE	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	70.1900	60.8390	0.8668
	2	44.9420	70.2650	1.5635
	3	6.0550	8.3790	1.3838
	4	4.4080	9.0560	2.0544
	5	1.6130	2.7920	1.7309

SEGMENTATION	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.6100	0.2600	0.4262
	2	0.5100	0.2500	0.4902
	3	0.5200	0.2600	0.5000
	4	0.4800	0.2500	0.5208
	5	0.4600	0.2500	0.5435

NEGATIVE WIDTH	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	-0.4700	0.4200	-0.8936
	2	-0.2800	0.2500	-0.8929
	3	-0.1400	0.1400	-1.0000
	4	-0.1000	0.1200	-1.2000
	5	-0.1000	0.1100	-1.1000

POSITIVE WIDTH	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.4000	0.3400	0.8500
	2	0.1400	0.1600	1.1429
	3	0.1100	0.1000	0.9091
	4	0.1100	0.1700	1.5455
	5	0.0700	0.0800	1.1429

TABLE C.4
LOW ORDER SUMMARIES

TEXAS

MONOTONICITY	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.0545	0.0287	0.5266
	2	0.0715	0.0594	0.8308
	3	0.1336	0.0886	0.6632
	4	0.1461	0.0922	0.6311
	5	0.1547	0.1164	0.7524

ERROR VARIANCE	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	86.6400	83.3600	0.9621
	2	34.0900	49.4100	1.4494
	3	5.3300	10.2400	1.9212
	4	3.4700	6.1500	1.7723
	5	2.4800	6.6400	2.6774

SEGMENTATION	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.5900	0.2700	0.4576
	2	0.6100	0.6300	1.0328
	3	0.4500	0.2500	0.5556
	4	0.4800	0.2500	0.5208
	5	0.5300	0.2500	0.4717

NEGATIVE WIDTH	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	-0.4600	0.5800	-1.2609
	2	-0.1600	0.1200	-0.7500
	3	-0.1600	0.2500	-1.5625
	4	-0.1300	0.1800	-1.3846
	5	-0.0800	0.0900	-1.1250

POSITIVE WIDTH	LEVEL	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION
	1	0.9200	1.3400	1.4565
	2	0.3400	0.6000	1.7647
	3	0.1000	0.1000	1.0000
	4	0.0700	0.1200	1.7143
	5	0.1100	0.1700	1.5455

APPENDIX D
BAND WIDTH APPROXIMATIONS FOR THE SAMPLE LINES

By subdividing each cartographic line into smaller and smaller bands or segments, and then representing the lines using only the anchor lines for each band, a series of progressively more detailed approximations of each sample line can be achieved, in a kind of inverse generalization. Presumably, one could continue subdividing, and produce increasing complexities of graphic approximation: this is similar to the concept by which fractal geometries of geographic lines and landscapes are justified. Only five approximations are included here for each line, however, to represent the five levels of subdivision contained in each strip tree.

One may notice in looking at these approximations that the adequacy of the representations varies from one sample line to the next. That is to say, the anchor line approxiamtions for SIOUX seem to produce a much more reliable visual representation at all levels of subdivision than do the approximations for TEXAS, for example. It is hypothesized in the text of the dissertation that these variations may be accounted for by computing a fractal dimension for each line, or that the complexity of a line may be related to the quality of its generalized representations. Verification of this hypothesis must include a larger sample of cartographic lines, and also be based on empiric testing of map reader reactions. For the time being, then, the illustrations in this Appendix stand as illustration of the concept, rather than as a validation. Each anchor line approximation is labelled with its respective level of subdivision.

TABLE D.1
ANCHOR LINE APPROXIMATIONS FOR SIOUX

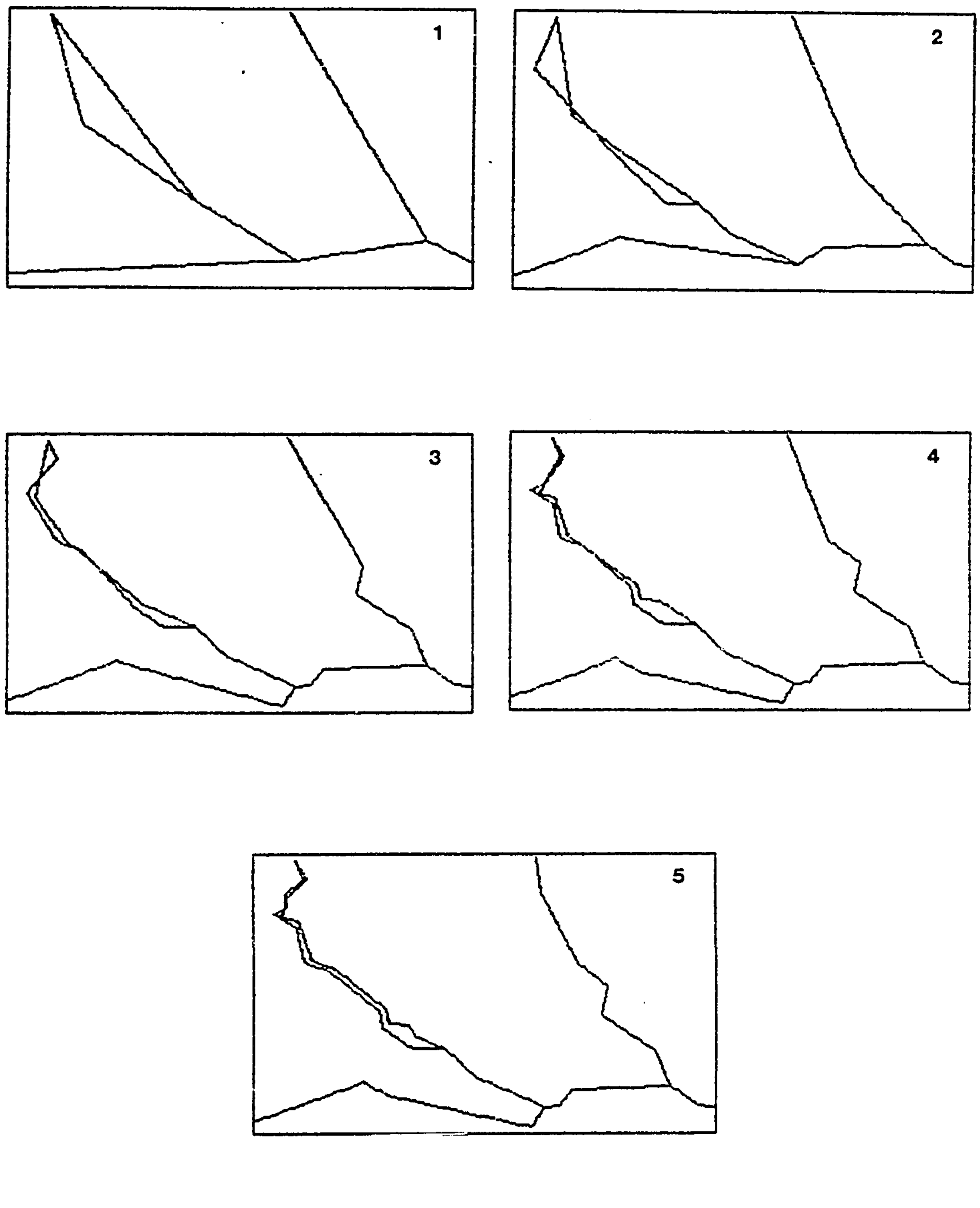


TABLE D.2
ANCHOR LINE APPROXIMATIONS FOR HUDBY

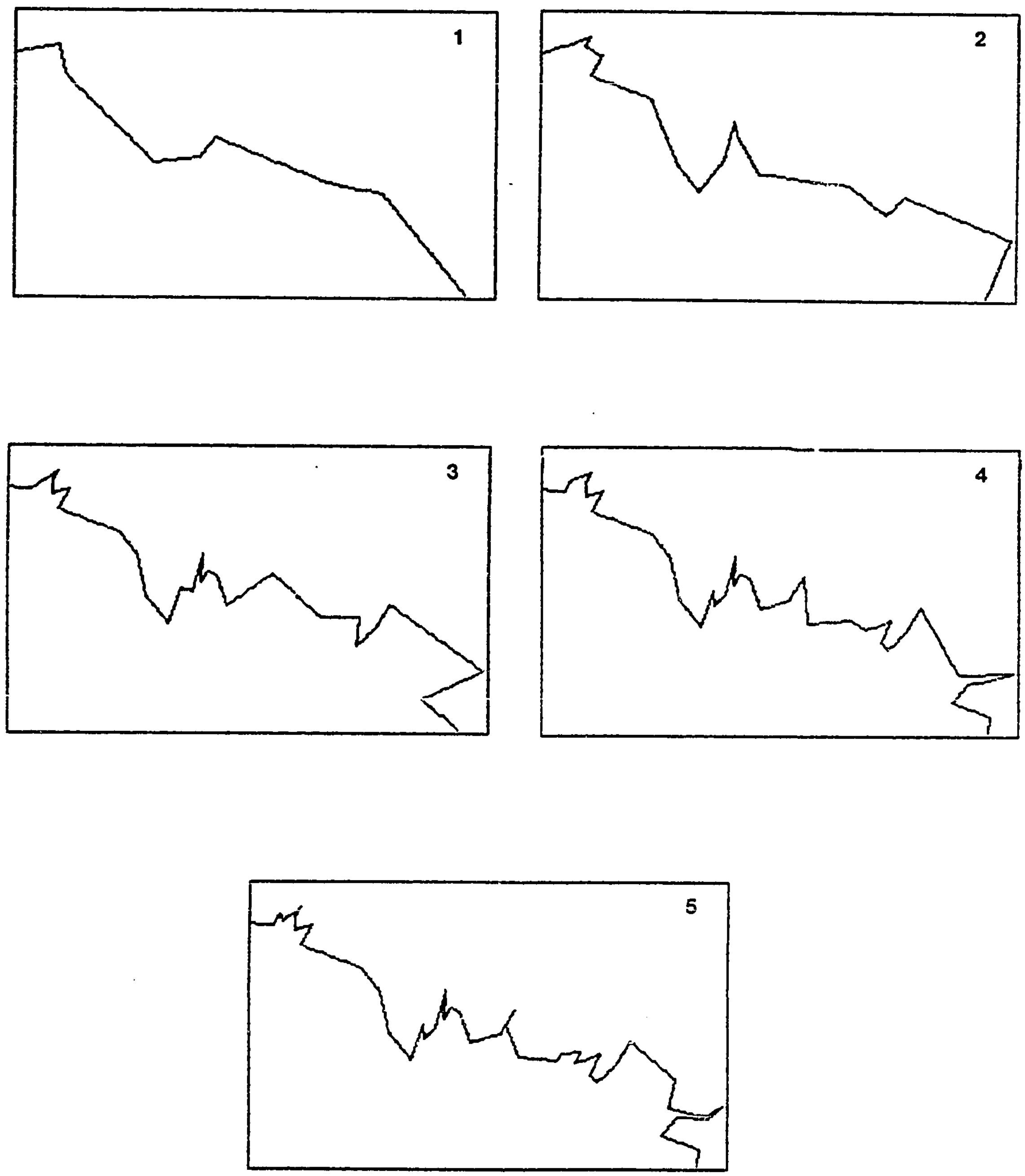


TABLE D.3
ANCHOR LINE APPROXIMATIONS FOR FJORD

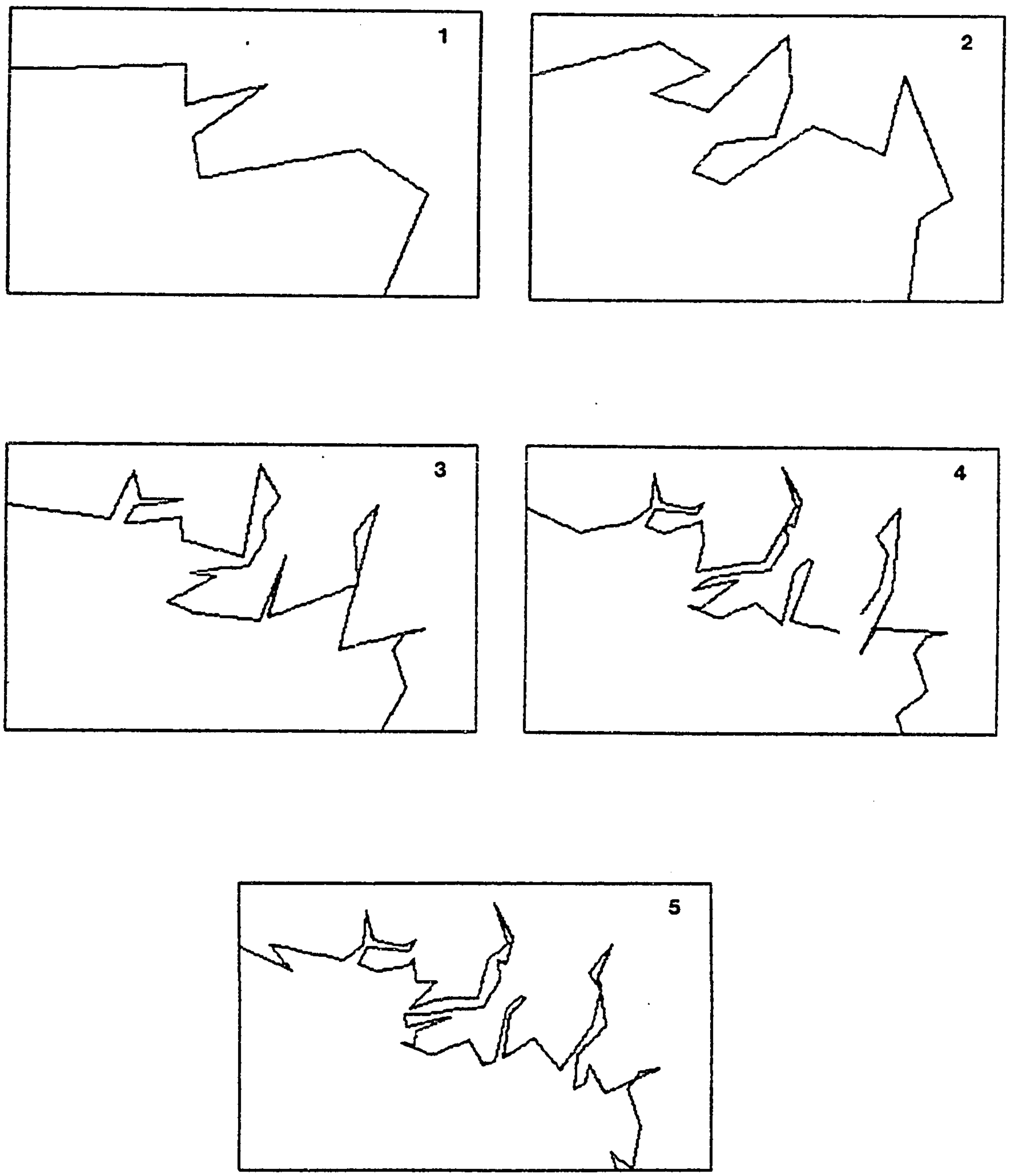
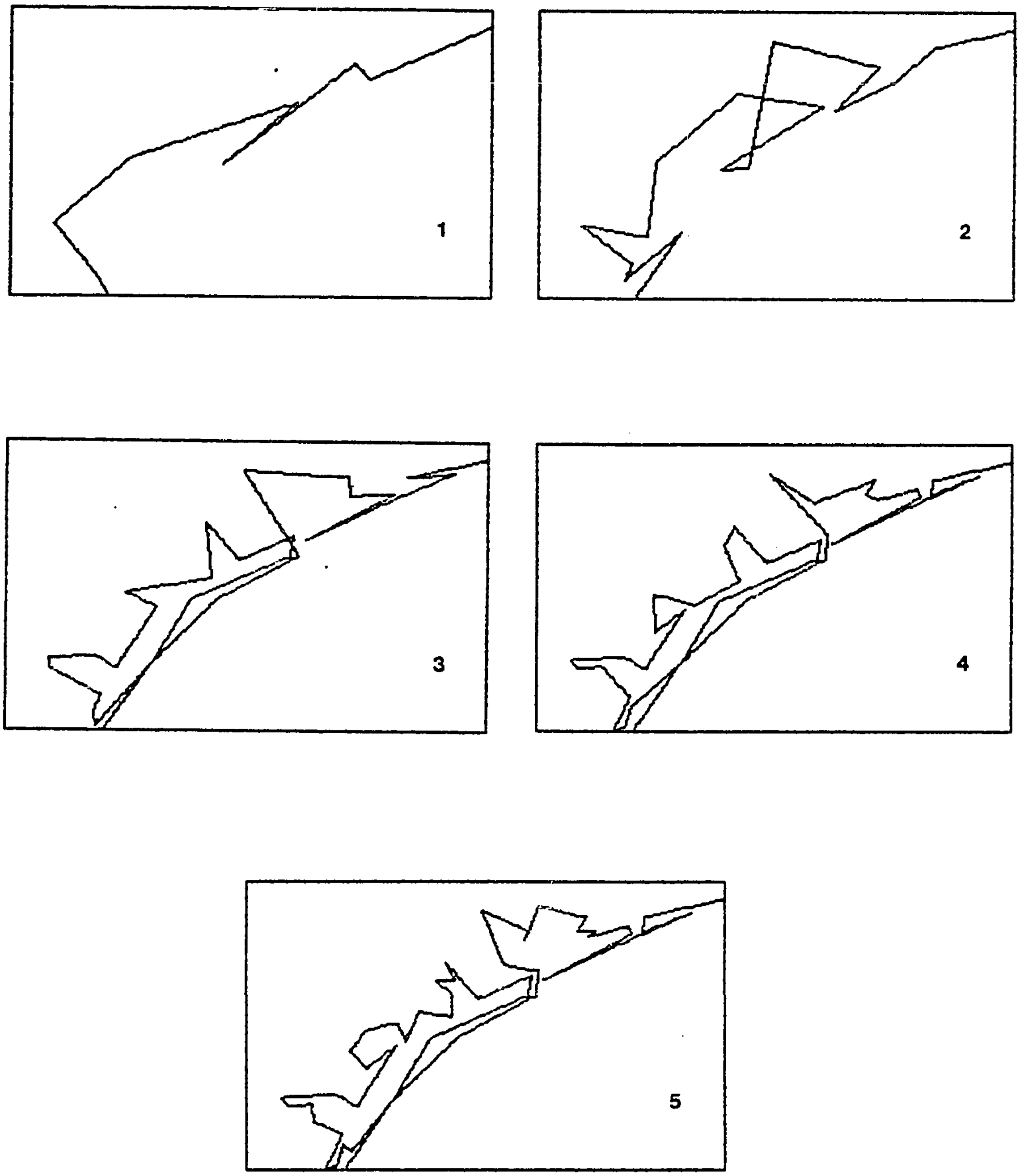


TABLE D.4
ANCHOR LINE APPROXIMATIONS FOR TEXAS



APPENDIX E
STATISTICAL COMPUTATIONS FOR FRACTAL DIMENSION

As described in Chapter 5, the computation of fractal dimensions for the four sample lines has been accomplished indirectly, by means of linear regression. The length of all anchor lines has been summed for each level of the strip tree, to measure the total length of each line at each level of subdivision. This total length is next divided by the number of (significant) bands occurring at that level, to provide an average strip length. Logarithms are computed for the total and average length measures, as displayed in the table below. Chapter 5 includes a discussion of the statistical analysis of this data.

TABLE E.1
LENGTH MEASURES FOR THE SAMPLE LINES
(AFTER RICHARDSON, 1961)

	LEVEL	TOTAL LENGTH	NUMBER OF STRIPS	AVERAGE LENGTH	ln(TOTAL)	ln(AVERAGE)
SIOUX	3	849.49	8	106.19	6.74	4.67
	4	891.43	16	55.71	6.79	4.02
	5	924.93	25	37.00	6.83	3.61
	6	947.63	34	27.87	6.85	3.33
	7	954.05	39	24.46	6.86	3.20
HUDBY	3	352.21	8	44.03	5.86	3.78
	4	466.52	16	29.16	6.15	3.37
	5	580.03	26	22.31	6.36	3.11
	6	637.63	35	18.22	6.46	2.90
	7	667.76	42	15.90	6.50	2.77
FJORD	3	484.39	8	60.55	6.18	4.10
	4	734.00	16	45.88	6.60	3.83
	5	958.71	30	31.96	6.87	3.46
	6	1000.69	48	22.93	7.00	3.13
	7	1196.40	63	19.30	7.09	2.96
TEXAS	3	476.51	8	59.56	6.17	4.09
	4	752.34	16	47.02	6.62	3.85
	5	1062.79	28	37.96	6.97	3.64
	6	1152.00	40	28.80	7.05	3.36
	7	1233.72	51	24.19	7.12	3.19

APPENDIX F
HIGH ORDER PARAMETERS FOR ALL FEATURES

This appendix contains listings of high order parameters for the four sample cartographic lines. In contrast to the low order parameteres, which are organized as binary subdivisions of line segments, the high order parameters are measured on a 'feature-by-feature' basis, as discussed in Chapter 5. The definition of a feature is recursive, in that a feature may be broken into a trend line and a set of sub-features which birfurcate from that trend.

At least two levels of feature bifurcations have been delineated for each line. TEXAS is the exception to this, having three levels of bifurcation. As with the low order parameters, a four-digit indexing scheme has been developed to identify each feature. The first digit is always assigned a value of 9, to identify a high order parameter. (Recall that for all low order parameters, the first digits for the indices range from 1 to 8.) The second digit ranges from 1 to 3, depending on the level of feature bifurcation. The final two digits identify the sequence in which features in a particular level have occurred.

TABLE F.1
HIGH ORDER PARAMETERS

SIOUX

FEATURE NUMBER: 9101
MIDPOINT: X= 252 Y= 32 ENDPOINT: X= 169 Y= 188
LENGTH OF AXIS: 204.399 (72 POINTS)
WIDTH AT MIDPOINT: 0.0000
ANGLE OF BIFURCATION: -0.7378 (-42.2737 DEGREES)

FEATURE NUMBER: 9102
MIDPOINT: X= 174 Y= 18 ENDPOINT: X= 114 Y= 58
LENGTH OF AXIS: 74.9664 (164 POINTS)
WIDTH AT MIDPOINT: 0.0000
ANGLE OF BIFURCATION: -0.6209 (-35.5736 DEGREES)

FEATURE NUMBER: 9201
MIDPOINT: X= 114 Y= 58 ENDPOINT: X= 27 Y= 188
LENGTH OF AXIS: 213.978 (135 POINTS)
WIDTH AT MIDPOINT: 6.4000
ANGLE OF BIFURCATION: -0.5834 (-33.4248 DEGREES)

TABLE F.2
HIGH ORDER PARAMETERS

HUDBY

FEATURE NUMBER: 9101
MIDPOINT: X= 245 Y= 33 ENDPOINT: X= 251 Y= 34
LENGTH OF AXIS: 11.5765 (39 POINTS)
WIDTH AT MIDPOINT: 10.3723
ANGLE OF BIFURCATION: -0.7854 (-45.0000 DEGREES)

FEATURE NUMBER: 9102
MIDPOINT: X= 237 Y= 53 ENDPOINT: X= 249 Y= 59
LENGTH OF AXIS: 16.0395 (13 POINTS)
WIDTH AT MIDPOINT: 9.8229
ANGLE OF BIFURCATION: -2.5830 (-147.995 DEGREES)

FEATURE NUMBER: 9103
MIDPOINT: X= 216 Y= 58 ENDPOINT: X= 221 Y= 75
LENGTH OF AXIS: 35.0864 (33 POINTS)
WIDTH AT MIDPOINT: 11.7777
ANGLE OF BIFURCATION: -1.6476 (-94.3987 DEGREES)

FEATURE NUMBER: 9104
MIDPOINT: X= 196 Y= 67 ENDPOINT: X= 205 Y= 76
LENGTH OF AXIS: 14.3891 (11 POINTS)
WIDTH AT MIDPOINT: 8.6221
ANGLE OF BIFURCATION: -1.3258 (-75.9638 DEGREES)

FEATURE NUMBER: 9105
MIDPOINT: X= 184 Y= 71 ENDPOINT: X= 193 Y= 77
LENGTH OF AXIS: 14.8323 (12 POINTS)
WIDTH AT MIDPOINT: 9.7793
ANGLE OF BIFURCATION: -2.9764 (-170.538 DEGREES)

FEATURE NUMBER: 9106
MIDPOINT: X= 148 Y= 85 ENDPOINT: X= 147 Y= 95
LENGTH OF AXIS: 12.1635 (16 POINTS)
WIDTH AT MIDPOINT: 9.1575
ANGLE OF BIFURCATION: -0.6823 (-39.0939 DEGREES)

FEATURE NUMBER: 9107
MIDPOINT: X= 118 Y= 88 ENDPOINT: X= 120 Y= 108
LENGTH OF AXIS: 23.3651 (31 POINTS)
WIDTH AT MIDPOINT: 11.1488
ANGLE OF BIFURCATION: -1.2925 (-74.0546 DEGREES)

FEATURE NUMBER: 9108
MIDPOINT: X= 107 Y= 88 ENDPOINT: X= 106 Y= 92
LENGTH OF AXIS: 6.1231 (5 POINTS)
WIDTH AT MIDPOINT: 7.0000
ANGLE OF BIFURCATION: 1.5708 (90.0000 DEGREES)

TABLE F.2
HIGH ORDER PARAMETERS

HUDBY

FEATURE NUMBER: 9109
MIDPOINT: X= 99 Y= 80 ENDPOINT: X= 101 Y= 95
LENGTH OF AXIS: 15.3351 (12 POINTS)
WIDTH AT MIDPOINT: 10.8003
ANGLE OF BIFURCATION: -2.6779 (-153.435 DEGREES)

FEATURE NUMBER: 9110
MIDPOINT: X= 60 Y= 130 ENDPOINT: X= 65 Y= 133
LENGTH OF AXIS: 5.8416 (6 POINTS)
WIDTH AT MIDPOINT: 11.1803
ANGLE OF BIFURCATION: -2.0344 (-116.565 DEGREES)

FEATURE NUMBER: 9111
MIDPOINT: X= 27 Y= 152 ENDPOINT: X= 35 Y= 163
LENGTH OF AXIS: 17.4181 (14 POINTS)
WIDTH AT MIDPOINT: 11.6444
ANGLE OF BIFURCATION: -0.6288 (-36.0274 DEGREES)

FEATURE NUMBER: 9112
MIDPOINT: X= 18 Y= 160 ENDPOINT: X= 17 Y= 169
LENGTH OF AXIS: 15.0486 (18 POINTS)
WIDTH AT MIDPOINT: 7.3259
ANGLE OF BIFURCATION: -0.7854 (-45.0000 DEGREES)

TABLE F.2
HIGH ORDER PARAMETERS

HUDBY

FEATURE NUMBER: 9201
MIDPOINT: X= 251 Y= 34 ENDPOINT: X= 276 Y= 41
LENGTH OF AXIS: 31.0219 (26 POINTS)
WIDTH AT MIDPOINT: 14.2413
ANGLE OF BIFURCATION: -0.4637 (-26.5700 DEGREES)

FEATURE NUMBER: 9202
MIDPOINT: X= 221 Y= 75 ENDPOINT: X= 222 Y= 85
LENGTH OF AXIS: 10.1623 (8 POINTS)
WIDTH AT MIDPOINT: 5.2942
ANGLE OF BIFURCATION: 0.2179 (12.4857 DEGREES)

FEATURE NUMBER: 9203
MIDPOINT: X= 147 Y= 95 ENDPOINT: X= 154 Y= 105
LENGTH OF AXIS: 12.9061 (9 POINTS)
WIDTH AT MIDPOINT: 4.7370
ANGLE OF BIFURCATION: -0.8803 (-50.4350 DEGREES)

FEATURE NUMBER: 9204
MIDPOINT: X= 113 Y= 99 ENDPOINT: X= 115 Y= 119
LENGTH OF AXIS: 20.6491 (16 POINTS)
WIDTH AT MIDPOINT: 6.4491
ANGLE OF BIFURCATION: -1.8925 (-108.435 DEGREES)

FEATURE NUMBER: 9205
MIDPOINT: X= 23 Y= 165 ENDPOINT: X= 31 Y= 172
LENGTH OF AXIS: 10.9907 (8 POINTS)
WIDTH AT MIDPOINT: 5.6796
ANGLE OF BIFURCATION: -2.6779 (-153.435 DEGREES)

TABLE F.3
HIGH ORDER PARAMETERS

FJORD

FEATURE NUMBER: 9101
MIDPOINT: X= 42 Y= 144 ENDPOINT: X= 18 Y= 150
LENGTH OF AXIS: 25.7641 (20 POINTS)
WIDTH AT MIDPOINT: 4.5594
ANGLE OF BIFURCATION: 2.0344 (116.565 DEGREES)

FEATURE NUMBER: 9102
MIDPOINT: X= 67 Y= 137 ENDPOINT: X= 104 Y= 154
LENGTH OF AXIS: 45.9662 (49 POINTS)
WIDTH AT MIDPOINT: 10.7473
ANGLE OF BIFURCATION: 1.3258 (75.9638 DEGREES)

FEATURE NUMBER: 9103
MIDPOINT: X= 92 Y= 130 ENDPOINT: X= 104 Y= 141
LENGTH OF AXIS: 22.9508 (16 POINTS)
WIDTH AT MIDPOINT: 12.7434
ANGLE OF BIFURCATION: 0.8795 (50.3893 DEGREES)

FEATURE NUMBER: 9104
MIDPOINT: X= 103 Y= 120 ENDPOINT: X= 117 Y= 129
LENGTH OF AXIS: 18.2521 (14 POINTS)
WIDTH AT MIDPOINT: 8.8392
ANGLE OF BIFURCATION: 2.3562 (135.000 DEGREES)

FEATURE NUMBER: 9105
MIDPOINT: X= 99 Y= 106 ENDPOINT: X= 151 Y= 177
LENGTH OF AXIS: 131.083 (90 POINTS)
WIDTH AT MIDPOINT: 22.1413
ANGLE OF BIFURCATION: 2.3562 (135.000 DEGREES)

FEATURE NUMBER: 9106
MIDPOINT: X= 98 Y= 91 ENDPOINT: X= 106 Y= 95
LENGTH OF AXIS: 13.3006 (28 POINTS)
WIDTH AT MIDPOINT: 9.3425
ANGLE OF BIFURCATION: 2.2318 (127.875 DEGREES)

FEATURE NUMBER: 9107
MIDPOINT: X= 135 Y= 74 ENDPOINT: X= 134 Y= 80
LENGTH OF AXIS: 11.6503 (15 POINTS)
WIDTH AT MIDPOINT: 11.7522
ANGLE OF BIFURCATION: 2.2962 (131.565 DEGREES)

FEATURE NUMBER: 9108
MIDPOINT: X= 154 Y= 75 ENDPOINT: X= 169 Y= 115
LENGTH OF AXIS: 45.5068 (31 POINTS)
WIDTH AT MIDPOINT: 7.1047
ANGLE OF BIFURCATION: 1.5708 (90.0000 DEGREES)

TABLE F.3
HIGH ORDER PARAMETERS

FJORD

FEATURE NUMBER: 9109
MIDPOINT: X= 174 Y= 76 ENDPOINT: X= 174 Y= 88
LENGTH OF AXIS: 12.6503 (10 POINTS)
WIDTH AT MIDPOINT: 6.6881
ANGLE OF BIFURCATION: 1.8925 (108.435 DEGREES)

FEATURE NUMBER: 9110
MIDPOINT: X= 182 Y= 75 ENDPOINT: X= 184 Y= 82
LENGTH OF AXIS: 7.5765 (7 POINTS)
WIDTH AT MIDPOINT: 6.4142
ANGLE OF BIFURCATION: 1.1071 (63.4349 DEGREES)

FEATURE NUMBER: 9111
MIDPOINT: X= 194 Y= 60 ENDPOINT: X= 218 Y= 151
LENGTH OF AXIS: 106.341 (76 POINTS)
WIDTH AT MIDPOINT: 10.8251
ANGLE OF BIFURCATION: 2.3775 (136.219 DEGREES)

FEATURE NUMBER: 9112
MIDPOINT: X= 205 Y= 56 ENDPOINT: X= 207 Y= 70
LENGTH OF AXIS: 15.1820 (12 POINTS)
WIDTH AT MIDPOINT: 5.6152
ANGLE OF BIFURCATION: 1.8925 (108.435 DEGREES)

FEATURE NUMBER: 9113
MIDPOINT: X= 226 Y= 54 ENDPOINT: X= 260 Y= 82
LENGTH OF AXIS: 48.0653 (35 POINTS)
WIDTH AT MIDPOINT: 7.6150
ANGLE OF BIFURCATION: 1.1071 (63.4350 DEGREES)

FEATURE NUMBER: 9114
MIDPOINT: X= 225 Y= 14 ENDPOINT: X= 233 Y= 0
LENGTH OF AXIS: 18.5963 (16 POINTS)
WIDTH AT MIDPOINT: 6.0267
ANGLE OF BIFURCATION: 1.3045 (74.7449 DEGREES)

TABLE F.3
HIGH ORDER PARAMETERS

FJORD

FEATURE NUMBER: 9201
MIDPOINT: X= 77 Y= 153 ENDPOINT: X= 76 Y= 172
LENGTH OF AXIS: 20.2109 (14 POINTS)
WIDTH AT MIDPOINT: 5.0248
ANGLE OF BIFURCATION: 1.8158 (104.036 DEGREES)

FEATURE NUMBER: 9202
MIDPOINT: X= 106 Y= 95 ENDPOINT: X= 125 Y= 102
LENGTH OF AXIS: 20.5645 (18 POINTS)
WIDTH AT MIDPOINT: 4.7902
ANGLE OF BIFURCATION: 2.1217 (121.565 DEGREES)

FEATURE NUMBER: 9203
MIDPOINT: X= 134 Y= 80 ENDPOINT: X= 138 Y= 85
LENGTH OF AXIS: 8.6569 (9 POINTS)
WIDTH AT MIDPOINT: 7.9941
ANGLE OF BIFURCATION: 1.4289 (81.8699 DEGREES)

TABLE F.4
HIGH ORDER PARAMETERS

TEXAS

FEATURE NUMBER: 9101
MIDPOINT: X= 227 Y= 163 ENDPOINT: X= 226 Y= 155
LENGTH OF AXIS: 12.1876 (76 POINTS)
WIDTH AT MIDPOINT: 5.6154
ANGLE OF BIFURCATION: 0.5880 (33.6900 DEGREES)

FEATURE NUMBER: 9102
MIDPOINT: X= 179 Y= 142 ENDPOINT: X= 159 Y= 147
LENGTH OF AXIS: 29.1488 (94 POINTS)
WIDTH AT MIDPOINT: 28.6279
ANGLE OF BIFURCATION: -1.4728 (-84.3824 DEGREES)

FEATURE NUMBER: 9103
MIDPOINT: X= 167 Y= 130 ENDPOINT: X= 166 Y= 115
LENGTH OF AXIS: 15.2361 (97 POINTS)
WIDTH AT MIDPOINT: 6.7623
ANGLE OF BIFURCATION: 0.4734 (27.1250 DEGREES)

FEATURE NUMBER: 9104
MIDPOINT: X= 127 Y= 106 ENDPOINT: X= 123 Y= 126
LENGTH OF AXIS: 24.9127 (34 POINTS)
WIDTH AT MIDPOINT: 12.4510
ANGLE OF BIFURCATION: -1.9693 (-112.834 DEGREES)

FEATURE NUMBER: 9105
MIDPOINT: X= 93 Y= 79 ENDPOINT: X= 86 Y= 84
LENGTH OF AXIS: 14.3071 (67 POINTS)
WIDTH AT MIDPOINT: 9.6601
ANGLE OF BIFURCATION: 1.2424 (71.1859 DEGREES)

FEATURE NUMBER: 9106
MIDPOINT: X= 58 Y= 31 ENDPOINT: X= 40 Y= 44
LENGTH OF AXIS: 28.0733 (44 POINTS)
WIDTH AT MIDPOINT: 17.0151
ANGLE OF BIFURCATION: -1.7127 (-98.1301 DEGREES)

FEATURE NUMBER: 9107
MIDPOINT: X= 51 Y= 0 ENDPOINT: X= 55 Y= 2
LENGTH OF AXIS: 4.8284 (50 POINTS)
WIDTH AT MIDPOINT: 14.0454
ANGLE OF BIFURCATION: 0.3719 (21.3099 DEGREES)

TABLE F.4
HIGH ORDER PARAMETERS

TEXAS

FEATURE NUMBER: 9209
MIDPOINT: X= 123 Y= 126 ENDPOINT: X= 116 Y= 136
LENGTH OF AXIS: 13.8836 (11 POINTS)
WIDTH AT MIDPOINT: 6.3911
ANGLE OF BIFURCATION: 0.6708 (38.4350 DEGREES)

FEATURE NUMBER: 9210
MIDPOINT: X= 120 Y= 120 ENDPOINT: X= 110 Y= 127
LENGTH OF AXIS: 12.3651 (10 POINTS)
WIDTH AT MIDPOINT: 8.2164
ANGLE OF BIFURCATION: -1.7895 (-102.529 DEGREES)

FEATURE NUMBER: 9211
MIDPOINT: X= 96 Y= 85 ENDPOINT: X= 100 Y= 104
LENGTH OF AXIS: 20.5859 (17 POINTS)
WIDTH AT MIDPOINT: 5.9429
ANGLE OF BIFURCATION: -1.8158 (-104.036 DEGREES)

FEATURE NUMBER: 9212
MIDPOINT: X= 86 Y= 84 ENDPOINT: X= 60 Y= 77
LENGTH OF AXIS: 32.6629 (41 POINTS)
WIDTH AT MIDPOINT: 11.8872
ANGLE OF BIFURCATION: 1.2861 (73.6900 DEGREES)

FEATURE NUMBER: 9213
MIDPOINT: X= 40 Y= 44 ENDPOINT: X= 20 Y= 46
LENGTH OF AXIS: 22.3071 (22 POINTS)
WIDTH AT MIDPOINT: 9.4450
ANGLE OF BIFURCATION: 0.9326 (53.4349 DEGREES)

FEATURE NUMBER: 9214
MIDPOINT: X= 55 Y= 2 ENDPOINT: X= 62 Y= 11
LENGTH OF AXIS: 17.5432 (48 POINTS)
WIDTH AT MIDPOINT: 6.8462
ANGLE OF BIFURCATION: 1.3258 (75.9638 DEGREES)

TABLE F.4
HIGH ORDER PARAMETERS

TEXAS

FEATURE NUMBER: 9201
MIDPOINT: X= 230 Y= 156 ENDPOINT: X= 260 Y= 170
LENGTH OF AXIS: 34.2773 (24 POINTS)
WIDTH AT MIDPOINT: 6.0182
ANGLE OF BIFURCATION: 1.9030 (109.036 DEGREES)

FEATURE NUMBER: 9202
MIDPOINT: X= 226 Y= 155 ENDPOINT: X= 174 Y= 125
LENGTH OF AXIS: 62.9254 (46 POINTS)
WIDTH AT MIDPOINT: 3.5535
ANGLE OF BIFURCATION: -1.3258 (-75.9638 DEGREES)

FEATURE NUMBER: 9203
MIDPOINT: X= 194 Y= 155 ENDPOINT: X= 204 Y= 158
LENGTH OF AXIS: 11.3592 (9 POINTS)
WIDTH AT MIDPOINT: 4.9492
ANGLE OF BIFURCATION: -1.2490 (-71.5651 DEGREES)

FEATURE NUMBER: 9204
MIDPOINT: X= 186 Y= 158 ENDPOINT: X= 198 Y= 167
LENGTH OF AXIS: 19.1197 (15 POINTS)
WIDTH AT MIDPOINT: 11.2522
ANGLE OF BIFURCATION: -2.0344 (-116.565 DEGREES)

FEATURE NUMBER: 9205
MIDPOINT: X= 174 Y= 158 ENDPOINT: X= 169 Y= 166
LENGTH OF AXIS: 14.0711 (18 POINTS)
WIDTH AT MIDPOINT: 6.3570
ANGLE OF BIFURCATION: -1.2490 (-71.5651 DEGREES)

FEATURE NUMBER: 9206
MIDPOINT: X= 159 Y= 147 ENDPOINT: X= 137 Y= 171
LENGTH OF AXIS: 41.1881 (34 POINTS)
WIDTH AT MIDPOINT: 12.5029
ANGLE OF BIFURCATION: -0.1206 (-6.9112 DEGREES)

FEATURE NUMBER: 9207
MIDPOINT: X= 158 Y= 139 ENDPOINT: X= 150 Y= 135
LENGTH OF AXIS: 10.5963 (9 POINTS)
WIDTH AT MIDPOINT: 4.9767
ANGLE OF BIFURCATION: -0.9828 (-56.3099 DEGREES)

FEATURE NUMBER: 9208
MIDPOINT: X= 166 Y= 115 ENDPOINT: X= 97 Y= 64
LENGTH OF AXIS: 93.4674 (87 POINTS)
WIDTH AT MIDPOINT: 8.6890
ANGLE OF BIFURCATION: -1.2532 (-71.8014 DEGREES)

TABLE F.4
HIGH ORDER PARAMETERS

TEXAS

FEATURE NUMBER: 9301
MIDPOINT: X= 169 Y= 166 ENDPOINT: X= 170 Y= 174
LENGTH OF AXIS: 10.9061 (9 POINTS)
WIDTH AT MIDPOINT: 4.3346
ANGLE OF BIFURCATION: -0.2618 (-15.0000 DEGREES)

FEATURE NUMBER: 9302
MIDPOINT: X= 97 Y= 64 ENDPOINT: X= 84 Y= 45
LENGTH OF AXIS: 23.5246 (20 POINTS)
WIDTH AT MIDPOINT: 3.4403
ANGLE OF BIFURCATION: -0.2411 (-13.8134 DEGREES)

FEATURE NUMBER: 9303
MIDPOINT: X= 83 Y= 90 ENDPOINT: X= 87 Y= 96
LENGTH OF AXIS: 7.8416 (8 POINTS)
WIDTH AT MIDPOINT: 8.1618
ANGLE OF BIFURCATION: -2.0344 (-116.565 DEGREES)

FEATURE NUMBER: 9304
MIDPOINT: X= 70 Y= 85 ENDPOINT: X= 66 Y= 90
LENGTH OF AXIS: 6.9907 (7 POINTS)
WIDTH AT MIDPOINT: 7.4352
ANGLE OF BIFURCATION: -1.1071 (-63.4350 DEGREES)

FEATURE NUMBER: 9305
MIDPOINT: X= 71 Y= 72 ENDPOINT: X= 70 Y= 66
LENGTH OF AXIS: 6.4142 (7 POINTS)
WIDTH AT MIDPOINT: 4.8028
ANGLE OF BIFURCATION: -1.8925 (-108.435 DEGREES)

FEATURE NUMBER: 9306
MIDPOINT: X= 62 Y= 11 ENDPOINT: X= 84 Y= 44
LENGTH OF AXIS: 41.6276 (33 POINTS)
WIDTH AT MIDPOINT: 5.3446
ANGLE OF BIFURCATION: -0.1821 (-10.4350 DEGREES)

VITA

Barbara Pfeil Buttenfield was born 14 August, 1952, in Pittsburgh, Pennsylvania. She attended The Ellis School in preparation for college, and graduated in 1969. Her college degrees include a Bachelor of Arts (1974) in Geography from Clark University, in Worcester, Massachusetts, and a Master of Arts (1979) in Geography from the University of Kansas at Lawrence. She has pursued the doctorate through the Geography Department at the University of Washington in Seattle.