INFORMATION TO USERS

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

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

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

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

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

UMI

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

Simulation of bidirectional reflectance, modulation transfer, and spatial interaction for the probabilistic classification of Northwest forest structures using Landsat data

by

Jeffrey Lee Moffett

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

Doctor of Philosophy

University of Washington

1998

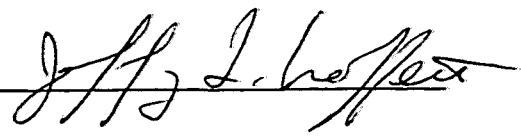
Approved by Julian Berry
(Co-Chairperson of Supervisory Committee)

Gerard F Schmidt
(Co-Chairperson of Supervisory Committee)

Program Authorized
to Offer Degree College of Forest Resources

Date October 22, 1998

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

Signature 

Date October 22, 1998

University of Washington

Abstract

Simulation of bidirectional reflectance, modulation transfer, and spatial interaction for the probabilistic classification of Northwest forest structures using Landsat data

by Jeffrey Lee Moffett

Co-Chairpersons of Supervisory Committee

Professor Julian E. Besag

Department of Statistics

Professor Gerard F. Schreuder

College of Forest Resources

Satellite remote sensors sample the upwelling reflected radiance in the form of digital images. A hierarchical model linking several sub-models of the image acquisition process and the spatial interaction of the classes has been developed to infer forest structure classes from these samples. Forest classifications exhibit large-scale spatial patterns. A fifth-order Markov random field (MRF) is used for modelling spatial interaction consistent with forest classifications at both the local and global levels. Forests scatter irradiance anisotropically. An unobserved image of the reflected radiance field is determined by a bidirectional reflectance factor and a normalized mean reflectance value for each class. The effective system resolution is partially dependent upon the atmospheric characteristics at the time of acquisition. A point spread function models the observed image data. The joint posterior distribution for the hierarchical model

is constructed using Bayes theorem to link each of these sub-models. Exact optimization and direct simulation of the resulting MRF are infeasible due to its high dimensionality. Estimates of image attributes are obtained via Markov chain Monte Carlo simulation. The marginal posterior modes estimate (MPM) minimizes the expected number of misclassifications and the posterior probability estimates provide spatially explicit information about the certainty of the MPM estimate. Parameter estimation is a formidable task. A radiosity model is used to simulate the bidirectional reflectance of each forest structure over a wide range of surface orientations. Wavelength dependent separable system point spread functions are estimated for each image classified. The class means effectively supervise the classification and several approaches for their estimation are evaluated. Prior distributions are specified for each of the covariance parameters. Multiple images are classified to assess the consistency of the solutions under varying illumination geometries. The estimated solutions are primarily sensitive to the selection of training data and the MRF parameters. The improvements achieved by detailed modelling of bidirectional reflectance remain subject to sources of variation not accounted for by this approach.

The original aspects of this dissertation include the development of a hierarchical framework for classifying remotely sensed data that overcomes the assumption of class conditional independence. Classifications are based on the effective resolution of the image data. Simulating multiple slope and class dependent bidirectional reflectance distribution functions for the purpose of normalizing projected area variation in rugged terrain and between multiple images is an important and original feature of this research. The application of a fifth-order random field to forest remote sensing is also unique.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank all of the members of my supervisory committee for the many insightful discussions and research reviews. I am especially grateful to my co-chairs. Gerard Schreuder pointed me in the direction of remote sensing and statistics, while Julian Besag encouraged me to endeavor and persevere in the face of the complexity of this research. I feel very fortunate to have received so much support from Chadwick Oliver through the Landscape Management Project sponsored by the USDA, Forest Service. As a research assistant in Chad's silviculture lab I have been able to learn a lot of valuable skills for this and other research projects. At times I have worked on projects for Bruce Lippke and I am grateful that those opportunities were available when necessary. Milton Smith provided the image data for this project. The Geology Remote Sensing Lab, with which Milton is affiliated, provided a tremendous amount of cpu for this project. Steve Cothern and Bill Gustafson have done a great job of maintaining those systems. Additional cpu was available from the College of Forest Resources GIS Lab and Phil Hurvitz did a great job of maintaining those systems. I have worked closely with Simon Byers and Wen-Hao Li and really appreciate the efforts they have contributed to this project.

Closer to home, the Silviculture Lab has been a great second home. All of the students in the lab have been great people to work with, great friends, and very accommodating in the past few months as I wrote this dissertation. Jim McCarter deserves a special thank you for keeping all of the computer systems running and stocking the soda fridge. The College of Forest Resources as a whole has been a wonderful place to be a student. The staff members have never failed to solve any

problem quickly. Teresa, Carrie, Lynn, Linda, Cherrie, Sam, Noel, Rosemary, Joyce, Jeff, Marlene, and all the other staff members have been great.

Off campus, I am very fortunate and grateful to my parents, Jane and Read Moffett, for emphasizing and providing me with a great education to pursue this research. All of my friends in Seattle, many of whom I met years ago working at REI, have been very supportive as well as providing a reality check on the balance of life. Jim McCaa not only fell into this category, but also helped out with many programming problems back on campus. My housemates in recent months, Tamsin, Lisa, Todd, and Carrie did a great job of understanding my lack of attention to details around the house as I wrote. Finally, Gavin and Marisha have offered a big incentive to actually bring this research to a close and I look forward to embarking on our voyage south.

DEDICATION

To the snow and the wind

Chapter 1

INTRODUCTION

Satellite remote sensing offers a unique source of information for understanding and managing the earth's ecosystems and natural resources. However, public and private organizations are not using satellite imagery for these purposes on a routine and cost effective basis. The experiences of the past few decades indicate that technical barriers remain unsolved, but also provide a base of knowledge for approaching the problems. Operational remote sensing systems will require reliable and extendible automated information extraction techniques. A historical perspective suggests that operational systems can be developed.

1.1 Background

The origins of modern remote sensing can be traced back for over a hundred years. In 1860, James Black recorded the first aerial photograph over America from a balloon above Boston. In 1903, Julius Neubronner designed tiny cameras with automatic timers to be mounted on carrier pigeons (Strain and Engle 1992). Throughout this century technological advances have continued at rate that by the 1980's the Voyager spacecrafts were returning fantastic images of the outer planets to Earth.

Over thirty years ago the value of multispectral aerial photography was recognized for many purposes ranging from distinguishing cement runways from surrounding grassy fields and identifying diseased trees in orange groves. During the late 1950's, Robert Colwell presented landmark research using infrared aerial photography for

detecting plant diseases, pointing out that the ability to remotely detect various surface features was dependent upon careful spectral analysis and proper film and filter selection (Colwell 1961). He also observed that early achievements created expectations for future remote sensing and prompted some photo interpreters to conclude that through the use of "integrated sensors," the concept of multispectral spectral reconnaissance could lead to automatic interpretation by electronic means of almost any object or condition on the face of the earth (Colwell 1961).

Colwell himself felt that extensive terrain analysis was a complex and formidable task. With the development of satellites, forest mapping was seen as a potential application of space photography, but initially not at a *scale* that would replace aerial photography (Lowman 1965). However by 1970, early successes in agricultural remote sensing created an optimism that aerial photography based forest mapping tasks could be extrapolated to space. This vision was published in an impressive National Academy of Sciences (NAS) report on the status and research needs for agriculture and forestry remote sensing (National Academy of Sciences 1970).

The NAS report resembles contemporary textbooks by reviewing conventional classification methods and outlines many of the problems still facing the remote sensing community. The report explains that the spectral signatures of land features, having means and variances, are statistical in character. The report also emphasizes the need for improved automated classification techniques. The NAS specifically identified the multivariate Gaussian assumption for maximum likelihood classification (ML) and texture algorithms as research priorities. Despite these developments the authors concluded that, "the present capabilities of sensor technology appear to exceed our knowledge of how to employ the sensor outputs (National Academy of Sciences 1970)." Arguably, this finding still applies today.

In 1972, Landsat-1 [originally named Earth Resources Technology Satellite-1, (ERTS-1)], was deployed and multispectral scanner (MSS) data became available on a regular basis. The most significant effort to develop an operational use of Land-

sat data was the Large Area Crop Inventory Experiment (LACIE) started in 1973, designed to monitor wheat production for various parts of the world. The justification for LACIE arose from the 1972 U.S.-Russian wheat sales (Slater 1980; Campbell 1987), when the U.S. was unprepared for large Russian purchases of wheat. The idea behind LACIE was that advance knowledge of global wheat production would allow the U.S. to develop more profitable pricing strategies.

While LACIE demonstrated some promising results for a few years, political pressures to validate results and demonstrate uses, and debates over methodologies and accuracy plagued the program (Mack 1990). The fate of LACIE was determined in 1977 when the USDA decided not to adopt the methodology for measuring planted crop areas. A policy review of the LACIE experience observed the failure to make the technology transfer from a large research and development system under NASA to an efficient operation under USDA (Mack 1990). The lack of leadership directing NASA and USDA in the development of the Landsat system may stem from the questionable political and economic benefits arising from resource remote sensing. Research and development costs are enormous. Large potential economic gains such as the foreign wheat sales have been the exception not the rule. The environmental benefits of long term resource monitoring are more likely to aid future generations and thus have limited short term political returns.

The limitations of using MSS data for classifying wildland vegetation as opposed to agricultural croplands were realized as early as 1973 and attributed to the non-homogeneous nature of the wildland vegetation (Crea 1973). One disappointing study of Minnesota forests (Eller and Ulliman 1974) concluded that, "there are factors other than forest species and condition classes influencing reflectance from vegetation. Any future studies should attempt to determine what those factors are before developing a classification scheme and conducting interpreter evaluations." By 1976, the NAS had become "acutely aware" of the limited number of repeatable experimental results in the technical literature (National Academy of Sciences 1976). In 1980, Philip Slater

(1980) argued that the Landsat program was experimental and not operational, outlining four areas requiring improvement before a remote sensing program would be operational: data management and compression; automated information extraction methods; predictive system modelling; and performance of data collection systems. By the 1980's, "better data" was seen as the key to the operational success of remote sensing technology and Landsats 4 and 5 were deployed with Thematic Mapper (TM) sensor. The TM was designed to have finer spectral, spatial, and radiometric *resolution* than the MSS with the aim of improved agricultural monitoring (Slater 1980).

Many remote sensing studies have focused on applications for the data given the technology, rather than on improving the technology. There are many examples, including Best (1982), Craighead et al. (1985), Hill and Kelly (1987), Iverson et al. (1989). Fiorella and Ripple (1993). The economic development of the GIS industry allowed virtually any group to afford and apply digital image processing (Philipson 1986). The availability of existing technology has made it possible for an increasing number of professional and special interest organizations to become involved in resource allocation and forest planning efforts. Remotely sensed data and GIS technology became tools for organizations trying to influence policy and defend positions; see, for example, Norheim (1996). Given the number of assumptions implicit in the process, it is not surprising that mapping results reflect the intentions of the analyzing organization (see Section 1.3.1). Consultants acting under time and budget constraints have little time for research, accuracy assessment, and use existing data whenever possible (Green et al. 1993). Clearly trade-offs exist between cost and accuracy. Assumptions made to produce a map quickly and for minimum cost will result in a less accurate map.

Many GIS and remote sensing software vendors have emphasized user-friendly systems and compete on the basis of interfaces. Such systems can be an integral part of the technology transfer process and play an important role in resource management

if developed appropriately. Commercial software has allowed many people to process image data without knowledge of the physical science or statistical assumptions inherent in the procedures; see, for example, Croft and Kessler (1996). The trade-off is that simple is restrictive. While software packages are becoming easier to use, they have limited analysts to employing techniques with known deficiencies rather than facilitating innovation and development of new tools.

The commercialization of Landsat data in 1984 under the Land Remote-Sensing Commercialization Act of 1984 led to an oversimplification and overselling of remote sensing applications especially in forestry. Schwaller and Dealy (1986), writing for the Earth Observation Satellite Company, provide a perfect example of the oversimplification of using Landsat data for forest change detection, "since two scenes are referenced to the same coordinate system, a simple check of the classifications can identify where changes have occurred, and what type of changes have taken place." Idealizing change detection in this manner probably helped to sell images, but registration, calibration, classification, and other errors have impeded the operational use of image data for monitoring forest changes. In light of such trends, Chrisman (1987) observed that, "there is a danger that the tough issues, temporarily swept under the rug, will reemerge, perhaps to discredit the whole process." In 1987, Colwell observed that "the overselling, overkilling, undertraining, underinvolvement, spurious evaluation, misapplication, timidity, inadequate infrastructure, inadequate understanding, and inordinate distrust" acted as institutional barriers to the adoption of modern remote sensing technology (Lauer et al. 1991). Other authors have also observed the overselling of remote sensing technology for resource management (Bondi 1989; Mack 1990; Congalton 1991). Arguably in a few cases, the technology continues to be oversold; see, for example, Sample (1994).

The privatization of Landsat data also increased prices and reduced the availability of data to the academic research community (Eisenbeis 1995). By hindering the research necessary to produce a truly operational system that could legitimately be

privatized, the Commercialization Act of 1984 was overly premature. By 1992, this had become evident and President Bush signed the Land Remote Sensing Policy Act of 1992 repealing the 1984 law. The hope inspired by the 1992 Act was followed a year later by a significant setback when Landsat-F, with its Enhanced Thematic Mapper, failed to reach orbit.

The social structure in place for environmental remote sensing requires leadership and cooperation as the data providers, analysts and users are separate parties each with their own objectives. NASA is primarily focused on developing new sensor technology, the data providers want to sell data, the analysts want to win contracts and make a profit, and the managing agencies are guided by legislation. Conceptual paradigms for image interpretation have not been well developed (Ryerson 1989). Land managers and ecologists are inclined to map and study different classes and processes. The development of remote sensing technology has been impeded by the lack of a clear and driving application.

Remote sensing is often seen as an important tool for global ecosystem and tropical forest monitoring; see, for example, Iverson et al. (1989), Sader et al. (1990), Lillesand (1993), and Lund et al. (1992). Without specific directives for *what* is to be monitored, *how*, and by *who*, it is not clear that remote sensing will meet this need. Operational development at the global scale is hindered by the lack of an organization with a mandate, as prescribed by Chrisman (1987), for monitoring ecosystem change. In contrast, remote sensing has been very successfully applied to defense objectives in which case the provider of the satellite equals the operator of the satellite equals the user of the information (Bondi 1989). Indeed, as early as the 1960's with the CORONA project, the military successfully deployed and used spaceborne reconnaissance systems (McDonald 1995). Learning from failures was a key to the success of CORONA.

Perhaps the most significant challenges facing remote sensing are physical in nature, not socio-political. Atmospheric attenuation and topographic variability present

obstacles for all users. Researching solutions to these problems without regard to a specific application is paramount to the development of operational remote sensing applications.

1.2 Forest classification

Forest characteristics vary continuously at all landscape scales depending on numerous geographic factors as well as previous disturbance patterns (Oliver and Larson 1990). For a given scale, classification provides a method for organizing information by grouping it into categories (MacEachren 1994). For example, forest structure classifications reduce the number of structural features required to describe each stand condition (Oliver et al. 1996). The notion of classification has a long history. Holmgren (1995) traces the concept of forest stand mapping back to 1829. The need to organize and reduce forest information is particularly acute in Canada, a country with a small population relative to its vast forests. This has led to the development of elaborate forest classification systems in Canada; see, for example, Rowe and Sheard (1981), Burger and Pierpoint (1990), Sims and Uhlig (1992). Latham (1997) is an example of a recently developed classification system for the forests of the Upper Columbia River basin.

Classification systems reflect the purpose for using the information. Rowe (1971) argued for the need to have multiple non-coincidental classifications in order to account for all forest functions. The appropriate number of classes and their definitions are determined by the mapping objective. Switzer (1980) refers to the categories of land cover classifications as being "contemplated," emphasizing the subjective nature of classification.

Classes are extremely scale dependent (Adams and Smith 1995). For example, at a fine scale individual trees may be grouped by age class, but at larger scale whole forests may be classified as either conifer or deciduous. Forest classification schemes

have been proposed for a variety of scales; see, for example, O'Hara et al. (1996). Anderson et al. (1976) attempted to develop a standardized land use classification scheme for remotely sensed data of the entire United States. While a consistent set of classes will have its merits, classification schemes will be more useful when designed as needed at appropriate scales. Classification schemes may fall into a hierarchical framework where sets of classes at one scale define a single class at another scale.

For classifying spatial information, class definitions imply boundaries. The topological characteristics of modern GIS technology allow for flexibility in classification process; see, for example, Davis and Dozier (1990) and Holmgren (1995). However, this flexibility is only an advantage if used properly as a classification system for one objective may not be appropriate for an objective at another scale and/or another purpose.

1.2.1 Classification of multispectral data

Remote sensors indirectly observe the continuous nature of forest distribution patterns by sampling the *upwelling reflected radiance* at a specific spatial and temporal resolution. The objective of image interpretation is to reduce the large amounts of *data* collected into the *information* needed by end-users (Hoffer 1986). Forest classification of multispectral image data is accomplished by inferring forest stand types and attributes from samples of reflected radiance. For example, the TM sensor measures radiance for six reflective bandpasses across the *electromagnetic spectrum* with a nominal spatial resolution of 30 m, recording the measurements with eight bit radiometric resolution. However, typical forest map attributes are species, age, and stand structure class for homogeneous vegetative areas or stands which are usually much larger than the TM spatial resolution. Thus, the process of *inferring* forest classes from remotely sensed data is one of "bridging the gap" between what is measured and what is to be mapped. In other words, classification transforms the measurement framework (Chrisman 1997). As an inference, classification should reflect the

uncertainty inherent in the process.

1.2.2 Signature extension

From the outset of multispectral remote sensing *signature extension* has been considered the primary mechanism for automating the information extraction process. Signature extension is a *supervised* classification technique whereby the areas of unknown surface cover within an image are classified by comparing their radiance measurements to the observations of known surface features. Signature extension relies on the assumption that the measurements of known "training sites" can be extrapolated over space and/or time.

Early researchers had difficulties implementing this technique, but remained optimistic that the problems they encountered could be solved:

> Some of the aspects of the discrimination problem require further research. It is readily apparent that much additional signature data must be obtained and the manners and amounts of their variances studied and analyzed in order to derive fully satisfactory decision procedures. The results to date, however, give strong evidence that satisfactory methods can be developed with a realistic amount of research (National Academy of Sciences 1970).

With the advent of the LACIE project, NASA scientists began to focus on the signature extension problem. In 1972, the essence of the problem was described as follows, "a computer program used to tell wheat from corn in one place would not necessarily work a few hundred miles away or on pictures taken a few days later." Substantial variations in reflected radiance resulted from changes in soil, water supply,the crop variety the transparency of the atmosphere, and the growth of the crop (Mack 1990). Indeed, there are many factors that influence the reflectance of vegetation,

resulting in a high degree of variation. Forest canopies and structures vary continuously over space and time. The reflectance from forests on rugged terrain also varies with topography. The class definitions determine which variation is useful, leaving the undesired variation to complicate the extension process.

By 1974, signature extension was marked as the key problem for the LACIE project to solve. However, in 1975 efforts to implement signature extension at the proto-operation level were ceased as successful applications occurred only 20 percent of the time and in random patterns (Mack 1990). The result was that more data had to be processed by human analysts.

Maximum likelihood assuming a multivariate Gaussian distribution is perhaps the most common method of signature extension, although the minimum distance, parallelepiped, and Mahalanobis classifiers are used as well (Richards 1986; Lillesand and Kieffer 1994). The limitations of the ML approach have been since at least 1970 (National Academy of Sciences 1970). However, ML is often chosen "because of its utility and generally acceptable results and because no better general scheme has yet been found (Harding and Scott 1978)."

Unsupervised classification is provides an alternative to supervised classification. In this approach the classes are defined in terms of their spectral spectral separability as opposed to their land cover characteristics. The challenge for unsupervised classification then becomes interpreting the land cover significance of the defined classes. In contrast supervised classification clearly distinguishes the land cover first, only to face problems with spectral separability. Unsupervised classes can be identified with various clustering algorithms (Richards 1986).

In general however, the remote sensing community believes that better results are to come from increased resolution data rather than improved processing techniques. Effective digital analysis tools that reliably produce needed information will be an integral part of an operational remote sensing system.

1.2.3 Contextual information

Knowledge of the size and shape of the class areas provides information about the spatial configuration of the image elements (pixels). *Contextual classification* methods consider the spatial context of a pixel in determining its class. The 1970 NAS report identified contextual classification as a research priority, specifically mentioning texture algorithms. Many approaches for contextual classification have been presented in the literature. In general spatial information is incorporated into spectral classifiers, and almost always results improve at the expense of computational time and complexity. While one could argue that ML has become the predominant signature extension method, no single contextual classification method has risen above the rest in remote sensing.

Texture algorithms seek to identify unique spatial patterns within an image and classify each pixel based on its pattern. (Haralick 1979) reviews texture models developed through the 1970's. More recent applications to satellite remote sensing include Wang and He (1990), Lee and Philpot (1991), Barber and LeDrew (1991), and Peddle and Franklin (1991). References for texture algorithms based on Markov random fields are provided in Section 2.2.2. Across the board these techniques have improved results relative to non-contextual classifiers. However, as yet there is no unambiguous definition of texture (Lee and Philpot 1991) and unlike size and shape, texture is a more esoteric attribute of the scene objects being classified. While air photo interpreters implicitly rely on texture for stratification, at satellite image scales there is less reason to assume classes should exhibit different textures or that texture would be consistent between all objects of a particular class.

Another contextual classification approach considers the spatial homogeneity of the regions being classified. Kettig and Landgrebe (1976) partitioned the image data into regions of similar values and then classified the regions via signature extension. While this takes spatial information into account, the data observed at each pixel is

not considered as classes are assigned. This is also true of filters that smooth classified maps and iterative relaxation procedures; see, for example, Therrien (1983).

Bayes decision rule has become a popular method for incorporating knowledge of spatial continuity into a signature extension framework and thereby producing spatially homogeneous classifications. Switzer (1980) proposed several extensions to standard linear discriminant analysis based on the observation that surface classes tend to be larger than the corresponding size of the pixels. This led to methods for basing the prior probability within the optimal decision rule on the configuration of class values within the neighborhood of the pixel being classified (Switzer 1983). In a published discussion of the paper by Switzer (1983), Besag (1983) proposed what later became known as iterated conditional modes (ICM) for Bayesian image analysis (Besag 1986). Swain et al. (1981) and Tilton et al. (1982) developed a context function for computing prior probabilities of class occurrences. Yu and Fu (1983) considered the joint density of a pixel and its neighbors. Owen (1984) estimated prior probabilities of particular pixel patterns using a geometric probability model. The research of Sæbø et al. (1985) and Mohn et al. (1987) incorporated spatial correction factors into Bayes rule based on neighboring pixel configurations. An iterative approach that estimates prior probabilities based on configurations found in the previous iteration has been presented by Klein and Press (1992).

These contextual classifiers use Bayes rule in some form and induce spatial homogeneity by considering small configurations of pixels. However, these techniques do not model a probability distribution of the pixel configurations. Indeed, prior distributions for each pixel and whole classifications can be constructed from basic local knowledge of the scene being classified (Besag 1974). These distributions are based on the conditional probability of a class occurring at a pixel given its neighbors and define *Markov random fields*. Modelling prior distributions provides a more intuitively satisfying approach and makes full use of the Bayesian paradigm.

1.2.4 Probabilistic classification

Thematic land cover classes are categorized abstractions of varying physical phenomena and therefore cannot be observed directly by satellite remote sensors. The spectral similarity of the classes and the physical variability of surface reflectance make classification an uncertain process even under the best of circumstances. Confusion also arises when the area of measurement falls on the boundary between two classes. Probabilistic classifications are the logical result of the inference required to map surface cover from samples of reflected radiance. Bayesian image analysis methods are inherently probabilistic and have applications ranging from tumor detection to mineral prospecting to land classification (Besag 1989).

Decision theory considers uncertainty in the calculation of expected values. This approach extends to mapped variables and spatial allocation problems. Bernknopf et al. (1988) used probabilistic maps of landslide susceptibility to estimate and map expected property value losses. Such estimates aid in the evaluation of landslide hazard mitigation regulations. Bonham-Carter et al. (1988) used Bayes theorem to estimate maps of the posterior probability of gold mineralization for exploration in Nova Scotia.

In the fields of forestry and resource planning Lowell et al. (1992) argue that each thematic map in a forest management GIS should have an associated uncertainty map representing estimates of error. They further contend that spatial analysis should be conducted with the "certainty surfaces" to produce probabilistic results and caution that understanding and using such maps will require considerable effort. Berry (1994) gives a hypothetical example of the utility of "shadow maps of uncertainty" in an overlay analysis of forest types and soils. Berry suggests that uncertainty maps will become commonplace in future GIS development. Goodchild et al. (1992) developed an error model for class area measurements and overlay. MacEachren (1994) outlines various methods for representing uncertainty.

Probability surfaces have also been analyzed in the context of error propagation through the buffer operation, where it was shown that errors in the estimation of the probabilities propagate as do conventional thematic map errors (Veregin 1996). The point here is that while probabilistic classifications capture the inherent uncertainty of the process, they themselves are subject to estimation error. Like a weather forecast predicting a twenty percent chance of rain, a probabilistic estimate of the occurrence of a land cover class is difficult to verify. Faith in such estimates will depend on the validity of the assumptions involved.

1.3 Objectives and potential applications

Whether or not remote sensing can be a useful tool on an operational basis depends on the specific needs and objectives of ecosystem scientists and resource managers. While there have been a few operational successes in the agricultural sector at both the regional (Corbley 1994) and continental scales (Hutchinson 1991; Prince and Justice 1991), such examples are hard to find. Bartholome (1991) points out that scientific research results have not led to the operational monitoring of crop production because the results have not been relevant for end-user needs.

The 1976 NAS report observed that most of the published studies in remote sensing were performed by technologists with only scattered participation by operational managers (National Academy of Sciences 1976). More recently, Meyer and Werth (1990) claimed that remote sensing investigators failed to define specific management needs and have at best only a vague and simplistic view of operational functions. However, the operational use of remote sensing technology has been limited to the extent that it was originally perceived to replace aerial photography. While it is generally accepted that photography performs better than image data for certain tasks, air photos and satellite imagery provide different information at different scales. Field foresters use aerial photography on a regular basis, but satellite remote sensing pro-

duces regional global information. Thus, the information extracted from imagery is of little interest to foresters at the local level, while the information needed by local foresters cannot be accurately recovered from image data.

One of the most dangerous misconceptions associated with GIS is that it enables the modification of scale and resolution thereby reducing the significance of these issues. In fact the opposite is true. Scale and resolution have become extremely more important and difficult to deal with in the digital environment (Evanisko 1990). Successful image classification requires understanding needs and identifying objectives at appropriate scales. Once the scale of the objective is known, then the appropriate data can be selected.

1.3.1 Regional landscape planning

The efforts to map the old growth ecosystems of the Pacific Northwest provide examples of the usefulness of remote sensing when applied at an appropriate scale to forest classification, as well as the need for further research and development. In 1989, both the Wilderness Society (Morrison 1992) and the Pacific Northwest Region of the U.S. Forest Service (Teply and Green 1991; Congalton et al. 1993) initiated programs to map Northwest forest types from remotely sensed data. See Norheim (1996) for a thorough review and comparison of these efforts. The demands on Northwest forests range from habitat provisions for the spotted owl and Pacific yew to a base for commodity production. By identifying the extent and distribution of various forest types over extensive land areas, regional forest classifications are an important tool for designing conservation and allocation strategies. See Leversee (1993) for a similar study conducted in British Columbia.

Each of these studies used the same commercial GIS and image processing software and similar data to map the same region. Neither study included more than a limited research component (Cohen 1994), although the Wilderness Society study did investigate various texture algorithms for classification. In the end, the Wilderness

Society concluded that only ten percent of the old growth was left, while the Forest Service found that a third remained.

Clearly the definition of old growth adopted by each organization will influence how much of it they map, respectively. However, the subjective selection of the image processing techniques also has an effect on the results. As an example consider the approach taken by each party for normalizing the topographic variation. The contractor for the Forest Service study did not employ a systematic correction technique, although one was used in a similar study (Green et al. 1993). The Wilderness Society used either of two normalization techniques for most of their analysis (Norheim 1996), including the method of Civco (1989).

Normalization techniques are designed to compensate for the effect of projecting radiance onto topographic surfaces. In appearance, normalization removes "topographic shade." Old growth canopies are structurally diverse and exhibit a higher proportion of tree shadows, which are spectrally equivalent to projecting radiance onto larger areas. By not compensating for topography, there is a greater likelihood in the Forest Service study to have mapped second growth forests on north facing slopes as old growth. Likewise, it would have been possible for the Wilderness Society to have overcorrected for shaded slopes and mapped old growth forests on north facing slopes as second growth.

The ability of various interpreters to subjectively select image processing techniques in this manner indicates the lack of standardized methods. For example, where multiple images overlapped, the Forest Service study processed only one image, neglecting the cross check classifying a second would have provided (Norheim 1996). Public confidence in the technology will be limited as long as it is apparent that the user can influence the results. A controversy over the use of MSS data to map the deforestation of Puget Sound has reached the front pages of Seattle's newspapers as this is being written. Technological improvements will not resolve the dispute over the definition of old growth, but remote sensing technology should be developed to

the point where discrepancies in class areas depend on class definitions and not the image processing techniques themselves.

1.3.2 Forest inventory

A multi-stage sampling scheme provides an efficient method for conducting inventories of extensive forest areas. Kent et al. (1979) and Kleinn (1996) review the statistical theory behind this approach in the context of forest inventory and monitoring. In the first stage of a three-stage design, satellite remote sensing data is useful for stratifying land cover into broad cover classes. In the second stage, medium scale photography further stratifies the forest by type. Ground-based measurements are then collected in the third stage. Indeed, multi-stage sampling has been successfully applied many times for forest inventory; see Duggin (1990) for additional references.

Used in this manner satellite imagery is a complementary tool instead of a replacement for low altitude aerial photography. The scale of imagery is most suited for mapping land cover extending over tens or hundreds of kilometers, such as stratifying a large area forest into a few broad forest classes. Medium scale aerial photographs provide stand specific information. It requires 3,744 photos (1:24,000, stereo) to cover the same area as a single TM scene (Hoffer 1994).

Harding and Scott (1978) and Franklin et al. (1986) used Landsat MSS data in this manner for conducting large area inventories in Washington and California, respectively. Their results indicated that for comparable accuracy, the image classification approach reduced costs. Hahn et al. (1992) proposed using TM data to stratify an array of 45,000 Forest Inventory and Assessment plots in the north central United States. Woodcock et al. (1994) and Woodcock et al. (1997) indicate that classified TM imagery is useful for stratification, although unreliable for obtaining tree size estimates directly. Gemmell (1995) also obtained marginal results trying to estimate volume directly from TM data. In Switzerland, image classification was not considered accurate enough for the stratification phase of a two stage design, but

may be used as the first phase in a three stage design Keller and Köhl (1996). Of course, results derived from image stratification will depend on the accuracy and appropriateness of the classification. Further, stratified sampling requires stratification into homogeneous regions. Contextual classification techniques are ideally suited for producing the right amount of homogeneity.

1.4 The scope of this study

In several ways this study pursues the research priorities of the 1970 NAS report by seeking to develop a classification method that does not require the assumption of class conditional independence and incorporates contextual information. The statistical classification of image data requires a number of assumptions. While this cannot be avoided, the intent of this study is to consider the most realistic and feasible assumptions. Although many remote sensing studies have not resulted in operational systems they nevertheless constitute a wealth of information and lessons learned that did not exist 30 years ago. Remote sensing research must build on detailed examinations of previous studies. Understanding *why* experimental systems have not become operational is more important than asking *what* can be accomplished using the data.

In mountainous areas topographic variation severely complicates forest classification. The atmosphere and sensor characteristics decrease the effective resolution beyond the nominal value. For this reason image pixels are not independently distributed random variables. A primary goal of this thesis is to explicitly model these processes, thus accounting for their effects in the eventual classification. Overcoming the obstacles presented by these forms of degradation will benefit all classification objectives.

The classifier developed for this thesis is contextual and probabilistic. A prior probability distribution models the spatial configuration of forest classes. Although a point estimate of the classification minimizing the expected number of misclassified

pixels is given, it represents the mode of a discrete probability distribution estimated for each pixel. Thus, a probability surface is given for each class, capturing the uncertainty inherent in such an inference process. Although these surfaces are not developed for a specific application, the literature reviewed in Section 1.2.4 calls for the development of probabilistic mapping techniques.

Insufficient spatial resolution is a common criticism of multispectral image data. This may stem from an expectation that image data perform the same tasks as aerial photography. Many authors imply that increased spatial resolution can improve remote sensing results across the board; see, for example, Cohen et al. (1995) and Croft and Kessler (1996). However, finer spatial resolution alone will not lead to increased per pixel accuracies (see Section 4.1.7). Alternatively, this study focuses on improving results by adequately modelling the image acquisition process. Although there is no rule for determining the appropriate classification scale, previous studies suggest several forest classes are appropriate for Landsat data.

The development of remote sensing technology for resource management is inherently a multidisciplinary prospect, which unfortunately faces a number institutional obstacles in American universities (Lillesand 1993). The classifier developed for this thesis integrates research conducted simultaneously within the College of Forest Resources, Statistics Department, and the Geology Remote Sensing Laboratory at the University of Washington. This informal cooperation represents an achievement in interdisciplinary research.

Although the techniques developed here are applied to TM data, they should apply equally well for MSS and other multispectral data for an appropriate set of classes. Unless specified otherwise, throughout this thesis descriptions of specific data attributes refer to TM data.

This thesis is organized as follows. Given the multidisciplinary nature of this research the relevant literature is reviewed where necessary in several Chapters. Chapter 2 reviews the theory for modelling spatial context using Markov random fields

and incorporating that information into a hierarchical framework. The relevant literature applying random fields to image classification problems is reviewed. Chapter 3 reviews theoretical and empirical studies of forest reflectance. A model of reflected radiance from rugged terrain is presented. Many approaches exist for modelling forest reflectance. The pertinent literature is reviewed and the method adopted for this study is outlined in detail. The TM and relevant sensor literature is reviewed in Chapter 4 along with a literature review of data models. The approach chosen for this study is appropriate for use with the reflected radiance model. A few methods exist for estimating the effective system resolution. These are reviewed as well. Prior distributions for hyperparameters are presented in Chapter 5 along with approaches for training the classifier. Chapter 6 discusses the application of stochastic simulation for producing particular classification estimates. This classifier is evaluated using TM data in Chapter 7. The first part of the chapter outlines the assumptions required to apply the model. Then various parameterizations for single and multiple images are compared to assess the stability of the solutions. The application to specific objectives and formal methods for assessing the accuracy of probabilistic estimates are beyond the scope of this research. Chapter 8 concludes this thesis with a further discussion of the results.

In many ways remote sensing remains a young subject (Mather 1989) with many technical challenges left to face. There is plenty of room for innovation. Although the limitations of forest remote sensing are become more familiar, "methods to make use of such imprecise, but objective information have, however, barely started to evolve (Holmgren 1995)."

Chapter 2

BAYESIAN IMAGE ANALYSIS METHODOLOGY

2.1 Introduction

2.1.1 Some definitions

Consider a two-dimensional rectangular array of *pixels* partitioning a finite region S with each pixel identified by an index $i \in \mathcal{N} = \{1, 2, \ldots, n\}$. For each pixel in S, and for each bandpass, a remote sensor records a sample of the upwelling reflected radiance from the earth's surface. These observations form the remotely sensed image data $y = \{y_i : i \in S\}$. For the reflective bands of the Landsat Thematic Mapper sensor, each y_i has six components, corresponding to bands 1-5 and 7. More generally, the components of y_i are indexed by $p, q = 1, \ldots, r$.

Also consider a corresponding two-dimensional rectangular array of *sites* over the region S with each site also indexed $i \in \mathcal{N}$. Let $X = \{X_i : i \in S\}$ denote a random vector that classifies y over region S. A univariate X_i located at each site takes on values among a finite set of unordered classes $\Lambda = \{1, \ldots, m\}$. The set of classes Λ is defined to be the *state space* of X_i, while the state space of X is $\Omega = \Lambda^n$ such that $x \in \Omega$. The *minimal sample space* is the set of all *possible* classifications; classifications having probability greater than zero. For the purpose of review, the minimal sample space is considered equivalent to Ω throughout this Chapter, while in practice the minimal sample space may be a subset of Ω. The primary objective of classification is to estimate the desired or "true" classification x^* for $x^* \in \Omega$.

An arbitrary realization of X is denoted as x, likewise x_i is a particular realization of X_i. Unless otherwise shown, π is used to denote probability distributions

and densities. Also note that a pixel is distinguished from a site. A pixel corresponds to a remotely sensed sample, whereas a site is a component of X. For this research the pixel and site dimensions are identical, but this is not a requirement of the methodology (Besag 1989).

2.1.2 Bayes theorem

For a given scale, a classification of S can be represented as x having an associated joint probability distribution $\pi(x)$. By adopting a Bayesian approach this probability distribution can be combined with a joint probability density for the data y given x. The conditional density for $l(y \mid x)$ is a likelihood function of x for the observed image y. Bayes theorem can be used to construct a conditional posterior probability distribution for the classification given the image data as follows

$$\pi(x \mid y) = \frac{\pi(x, y)}{\pi(y)} = \frac{l(y \mid x)\pi(x)}{\pi(y)} \propto l(y \mid x)\pi(x), \tag{2.1}$$

where $\pi(x, y)$ is the joint density of the classification and the data, $\pi(y)$ behaves as a normalizing constant. Typically in imaging applications the normalizing constant cannot be evaluated directly because $\pi(y)$ requires an intractable n-dimensional summation over x. In this case the posterior distribution $\pi(x \mid y)$ is known only up to scale. Fortunately, the indirect estimation procedures described in Section 6.3 only require the posterior up to scale. The simplest likelihood function assumes class conditionally independence

$$l(y \mid x) = \prod_{i \in S} f(y_i | x_i). \tag{2.2}$$

Statistical inferences about x^* are based on the posterior distribution $\pi(x \mid y)$.

This introduction demonstrates how Bayes theorem is used to integrate what is known *a priori* about the spatial characteristics of the desired classification itself with the information contained in the data. Each site is then classified on the basis of its spatial context and spectral observations. As Besag (1989) explains, "faith in the

data is balanced against regularities supported by the prior." The advantage of this approach is that spatial information is incorporated without losing fidelity to the data, meaning that the data is considered at all stages of classification. For example, Moffett and Besag (1996) compared Bayesian and ML algorithms for classifying a TM image of a small lake surrounded by forest. Both the Bayesian and ML algorithms classified the lake as a single site surrounded by forest. Because of the unique spectral observation at the center of the lake, the Bayesian classifier maintained the single site as lake in an otherwise homogeneous classification. In contrast, the ML classification would have required a post-classification smoothing algorithm, such as a majority filter, to achieve the same degree of homogeneity. However, such a filter would have completely smoothed over a single unique site leaving no indication of the lake in the final map. The degree to which such distinct sites are preserved by a Bayesian classifier depends upon the objective.

The next section describes the derivation and various specifications of $\pi(x)$ in detail. A fully Bayesian approach models a distribution for the probability of x. Not all "Bayesian" classification methods do so; see, for example, Gorte and Stein (1998). In the conventional "Bayes classifier" commonly found in commercial software the prior probability of a particular class is a user-defined singular value. Section 2.3 extends Equation (2.1) to include several likelihood functions. The chapter concludes with a brief summary. In each section the relevant literature is reviewed as necessary, with particular emphasis on the application of Bayesian methods to remote sensing problems.

2.2 Markov random fields

The components of X interact spatially to form a classification. As a collection of interacting random variables, X is a stochastic system. For dependent random variables it is more convenient to consider the conditional distribution of the random

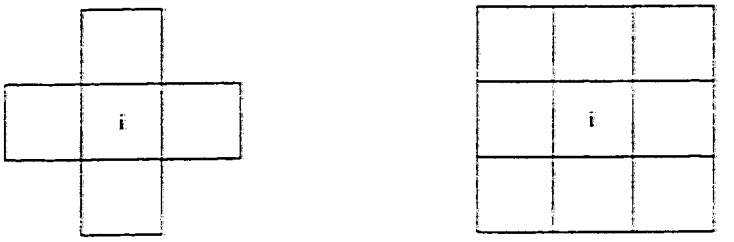
variable X_i at each site, rather than the joint distribution of the set X. However, unlike time dependent variables, there is not a natural ordering for spatially distributed variables. For spatially dependent variables the only logical conditioning set for X_i is the set of all other site variables in X denoted as X_{-i}. Thus, $\pi(x_i|x_{-i})$ becomes the *full conditional distribution* of X_i, otherwise known as the local characteristic of X_i in spatial statistics.

At issue is whether the n full conditionals determine the joint distribution $\pi(x)$ for $x \in \Omega$, which is required in order to adopt the conditional probability perspective. Further, if this requirement is met, what are the available functional forms of the full conditionals. Besag (1974) used the Brook expansion to derive $\pi(x)$ from the full conditionals confirming that the full conditionals determine the joint distribution provided that the system satisfies the *positivity* condition. Positivity states that if $\pi(x_i) > 0$ for each site, then $\pi(x) > 0$, ensuring that there are no deterministic exclusions. In other words, a system satisfies positivity if the minimal sample space is the Cartesian product of the minimal sample spaces of each site. Positivity is assumed throughout this research.

The result of this expansion implies hidden constraints on the form of the full conditionals in order to render valid spatial models with conditional probabilities and a mathematically consistent joint probability distribution. The Hammersley-Clifford theorem gives the most general form of $\pi(x)$, by first defining a Markov random field (MRF) as a system of n sites, each having a specific set of neighbors (Besag 1974).

An MRF is constructed by defining a *set of neighbors* $x_{\partial i}$ for each site $i \in S$. Site $j \neq i$ is defined to be a neighbor of site i if x_j contributes to the full conditional of x_i. From this definition it is clear that adjacency is not required for two sites to be neighbors, although adjacency and close proximity are logical considerations. For a rectangular array of sites, a hierarchy of models is formed by progressively expanding the size of the neighborhoods (Besag 1974; Besag 1986). The hierarchy begins with a first-order MRF which, for an individual site, is defined as the four adjacent sites

in each of the horizontal and vertical directions. Adding the diagonally adjacent sites forms a second-order random field where each interior site has eight neighbors. Figure 2.1 shows the first- and second-order neighborhoods for a rectangular array. The neighborhoods are also assumed translation invariant. In principle, the hierarchy can be continued in this manner until $x_{\partial i} = x_{-i}$, but the key to developing a useful MRF is defining a neighborhood structure that is elaborate enough for the task at hand, yet manageable.



(a) first-order neighborhood

(b) second-order neighborhood

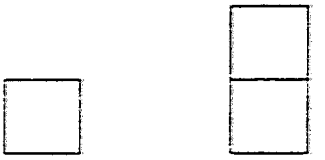
Figure 2.1: First- and second-order neighborhoods.

For a given neighborhood structure a *clique* is defined as a set of sites, single or otherwise, in which every site is a neighbor of every other site in the set. A clique can also be described as a maximal set of pixels, all distinct pairs of which are neighbors of each other. Figure 2.2 shows the possible clique types for first- and second-order neighborhoods up to rotation and reflection.

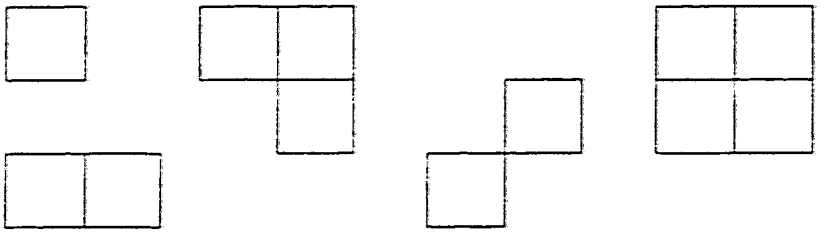
Either definition of a clique is necessary for identifying valid forms of $\pi(x)$ via the Hammersley-Clifford theorem. Given finite Λ and under positivity, the theorem states that for a given neighborhood structure the most general form that $\pi(x)$ must satisfy is

$$\pi(x) \quad \propto \quad \exp\left\{-\sum_{C} V_C(x_C)\right\}, \quad x \in \Omega, \tag{2.3}$$

where the V_Cs are arbitrary finite *potential functions* defined on the configurations x_C of the cliques C of the neighborhood system. The summation is over all cliques



(a) first-order clique types



(b) second-order clique types

Figure 2.2: First- and second-order MRF clique types.

C. Positivity ensures that the potential functions are positive for all $x \in \Omega$. Note that this form is especially restrictive for small neighborhoods (Besag 1986). The Hammersley-Clifford theorem was originally proved using a circuitous method in 1971 and later published in 1990. The theorem was first published in 1974 on the basis of an alternative proof (Besag 1974).

Equation (2.3) states the theorem using statistical physics terminology, whereby $\pi(x)$ is a Gibbs distribution with respect to the specified neighborhood structure (Geman and Geman 1984). The equivalence between Markov random fields and Gibbs distributions allows $\pi(x)$ to be described in terms of the full conditional distributions or by Gibbs potential functions and provides an explicit formula for the jointly distributed random field[1].

[1]The analogy to physical systems can be helpful for understanding MRF models. The Gibbs model describes the interaction of a macroscopic system in thermal equilibrium in the same way the spatial Markov models describe local dependence (Geman and Geman 1984; Zhang et al.

As an additional consequence of the theorem, the neighborhood specification must be symmetric, in that if i is a neighbor of j, then j must be a neighbor of i for all $i, j \in S$. Further, for a given neighborhood system the Markov property states

$$\pi(x_i|x_{-i}) \equiv \pi(x_i|x_{\partial i}), \quad x \in \Omega. \tag{2.4}$$

The Markov property allows the full conditional distribution of x_i to be specified in terms of its neighbors. This follows from the Hammersley-Clifford theorem which establishes, under positivity, that for any given MRF the local and global Markovian properties are equivalent (Besag 1974). Thus, $\pi(x)$ is referred to as a *locally dependent* MRF. A model of local characteristics may have undesirable global properties. Section 2.2.3 discusses the use of higher-order random fields for overcoming this inconsistency. Further, $\pi(x)$ does not model causality.

In terms of potential functions the conditional probability of x_i can be written as

$$\pi(x_i|x_{-i}) \quad \propto \quad \exp\left\{-\sum_{C:i\in C} V_C(x_C)\right\}, \quad x \in \Omega. \tag{2.5}$$

Specification of the V_Cs depends knowledge of the system being modelled. Any choice of potential functions determines the conditional distributions for all x_i. Specifying the full conditionals in terms of the neighboring sites drastically reduces the computational burden of simulating the system; necessary for indirect estimation of $\pi(x)$. Note that $\pi(x)$ may contain unknown parameters.

2.2.1 Prior information

Within the Bayesian framework, the joint probability distribution $\pi(x)$ is viewed as the prior distribution of x, whereby $\pi(x)$ assigns classes to S. The desired classification x^* is assumed to be a realization of a locally dependent MRF. Prior knowledge of

---

1990). In thermal equilibrium the most probable molecular configurations occur at the lowest energies/coolest temperatures. Likewise, the most probable classification will occur when the potential functions are minimized.

local properties provides an imperfect source of information about x^*, which can be quantified probabilistically as an MRF in terms of the potential functions shown in Equation (2.3).

As a prior distribution $\pi(x)$ represents *a priori* knowledge of the actual forest structure distribution patterns being classified. These characteristics are inherently spatial in nature, with the probability of particular spatial patterns analogous to levels of energy interaction. In principle, this information has absolutely nothing to do with in satellite image data! However, the local properties being modelled will depend on the scale of the classification, and the classification scale may be chosen to correspond to the resolution of the data; although, as mentioned above, this is not necessary.

When forest class areas are known to be larger than individual sites adjacent sites will tend to belong to the same class. Exploiting the spatial continuity of the surface cover at the scale of the image resolution implies that a degree of homogeneity in the classification is expected and desirable.

Depending on the mapping scale and class descriptions, two forest types may not be found adjacent to one another if one disperses allelopathic chemicals preventing the other's germination. Some classes may exhibit particular arrangements, such as objects having closed boundaries on a background. Other classes may occur as linear features such as rivers and roads for which specific MRFs can be developed.

At higher-orders of interaction the edges between two classes, clearcut and old-growth for example, may be assigned higher probability than the prospect of finding an old-growth site between two clearcut sites. This emphasizes the degree to which an MRF is defined for a particular scale, because at one scale this pattern may be entirely reasonable and at another scale highly unlikely. Modelling spatial coherence in this manner suppresses overly heterogeneous patterns.

The prior distribution may also include what is known about the classes relative to information other than the data being classified. For example, finding particular species, such as red alder, within a certain proximity to a stream would require a

digital stream map. Certain species may only be found above or below a particular elevation, or within a certain range, in which case the prior incorporates knowledge about the distribution of tree species with respect to elevation. Green et al. (1994) integrated road network information into an MRF prior for classifying TM data on the basis of proximity to urban development.

Clearly a wide range of prior distributions are feasible depending on the scale. The ability to incorporate all that is known about a particular objective gives Bayesian image analysis a high degree of flexibility. Further the *a priori* knowledge is modelled probabilistically which has an intuitive appeal given its nature. To date only a few of the many possibilities for forest related prior distributions have been explored, which calls for greater creativity.

2.2.2 Pairwise interaction Markov random fields

Many of the Bayesian image analysis applications in remote sensing classification have been based on the interaction between sites within pairwise cliques (see Figure 2.2). MRF models of *pairwise interaction* induce spatial homogeneity in the desired classification. Most of these models define second-order neighborhoods, although Haslett (1985) applied a first-order neighborhood model to synthetic data. Kittler and Föglein (1984) defined a first-order neighborhood to classify synthetic and MSS data. Kittler and Pairman (1985) classified clouds in AVHRR data also using a first-order neighborhood. More recently, Jeon and Landgrebe (1990) modelled first-order neighbors in both time and space for classifying multiple TM images. All of these studies showed improvements over non-contextual classifiers in varying degrees.

Equation (2.3) provides the most general form of $\pi(x)$. A second-order pairwise interaction MRF (Besag 1986), for which the neighbors of each (interior) site are defined to be the eight sites surrounding it, is given by

$$\pi(x) \propto \exp\left\{ \sum_{1 \le k \le m} \alpha_k n_k - \sum_{1 \le k < l \le m} \beta_{kl} n_{kl} \right\}, \tag{2.6}$$

where n_k is the number of sites in class k and n_{kl} is the number of neighboring pairs of sites having class (k, l). The α_k's and the β_{kl}'s govern the percentage of pixels in each class and the strength of interaction between different classes; for $\beta_{kl} > 0$ classes k and l are discouraged from being neighbors. Often the model is reduced to the Potts form in which $\alpha_k = 0$ for all k and $\beta_{kl} = \beta$ for all $k < l$, so that classes are treated as exchangeable (Potts 1952). The sites at the edge of the array may have fewer neighbors.

The Potts MRF, with a particular choice or estimate of β, is frequently adopted as the prior distribution for the Bayesian approach to classifying synthetic aperture radar (SAR) data (Derin and Elliott 1987; Kelly et al. 1988), MSS data (Zhang et al. 1990), TM data (Green et al. 1994; Moffett and Besag 1996), TM and SAR data (Solberg and Taxt 1994) to name a few examples. In each case the inclusion of spatial context using the pairwise interaction MRF improved results over non-contextual methods.

Parameter estimation remains a problem. Kittler and Pairman (1985) implicitly estimated β_{kl} for a first-order MRF from an ML classification of an AVHRR image, thus uniquely characterizing the interaction between different cloud types. Better estimates would have been achieved by using training data but this was not possible. They concluded that parameter estimation was a primary limitation of the methodology. Likewise, Zhang et al. (1990) used maximum likelihood to estimate neighborhood conditional transition probabilities. As an alternative to maximum likelihood estimates, Besag (1974) introduced a "coding method." A maximum "pseudo-likelihood" estimate is also available before and during classification (Besag 1986). Whether estimated or arbitrarily chosen, β may vary during image estimation. Derin and Elliott (1987) proposed a least square error method for parameter estimation. Finally, a fully Bayesian approach would specify a prior distribution, $\pi(\beta)$, for β. However this requires a normalizing constant that in most situations is unavailable in closed form (Besag 1989).

The success of pairwise interaction MRFs is attributable to the fact that at the

local level the MRF model is consistent with the characteristics of the desired classification. Forest scenes are often homogeneous at the scale of the image data, which also acts as an external magnetic field to prevent long-range order (Besag 1989). Indeed, in choosing to model spatial dependency with a locally dependent field merely requires that the local properties be consistent. However, the global characteristics of pairwise MRFs bear no relationship to typical forest classifications. Figure 2.3 shows a realization from a first-order Potts model with $\beta = 1.0$, while Figure 2.4 shows a realization corresponding to $\beta = 1.5$. In each realization, although adjacent pixels are alike, the long-range properties are rather undesirable. Local pairwise interaction models do not describe global interaction well.

An MRF model of local dependence will exhibit positive correlations, hence long-range order, over significant distances. This is best demonstrated by the Ising model of statistical physics which corresponds to the binary, first-order version of Equation 2.6 on the infinite rectangular lattice. This auto-logistic model was originally developed by Ising in 1925 to describe spontaneous magnetism of some materials at low temperatures. See Kindermann and Snell (1980) for a further discussion of the Ising model. Even for moderate values of β the Ising model produces positive correlations between site values infinitely far apart (Besag 1986). This tendency for infinite single color patches is evident on the finite lattice as well as shown in Figure 2.4.

Simulations produce primarily single class realizations for values of β greater than the critical value at which the transition takes place. For the Potts MRF, this value is given by $\ln(1+\sqrt{m})$ (Potts 1952). Thus, for $m = 6$, the critical value is approximately 1.24. Figure 2.5 shows a realization from the first-order model with $\beta = \ln(1+\sqrt{6}) \approx 1.24$. Although larger patches of each class appear, the realization remains rather "noisy." These properties hold for second-order pairwise interaction.

In classifying image data using the Potts model the limitation of pairwise interaction models becomes apparent in choosing an appropriate value for β. Larger values induce long-range order in the form of greater homogeneity. However, a value large

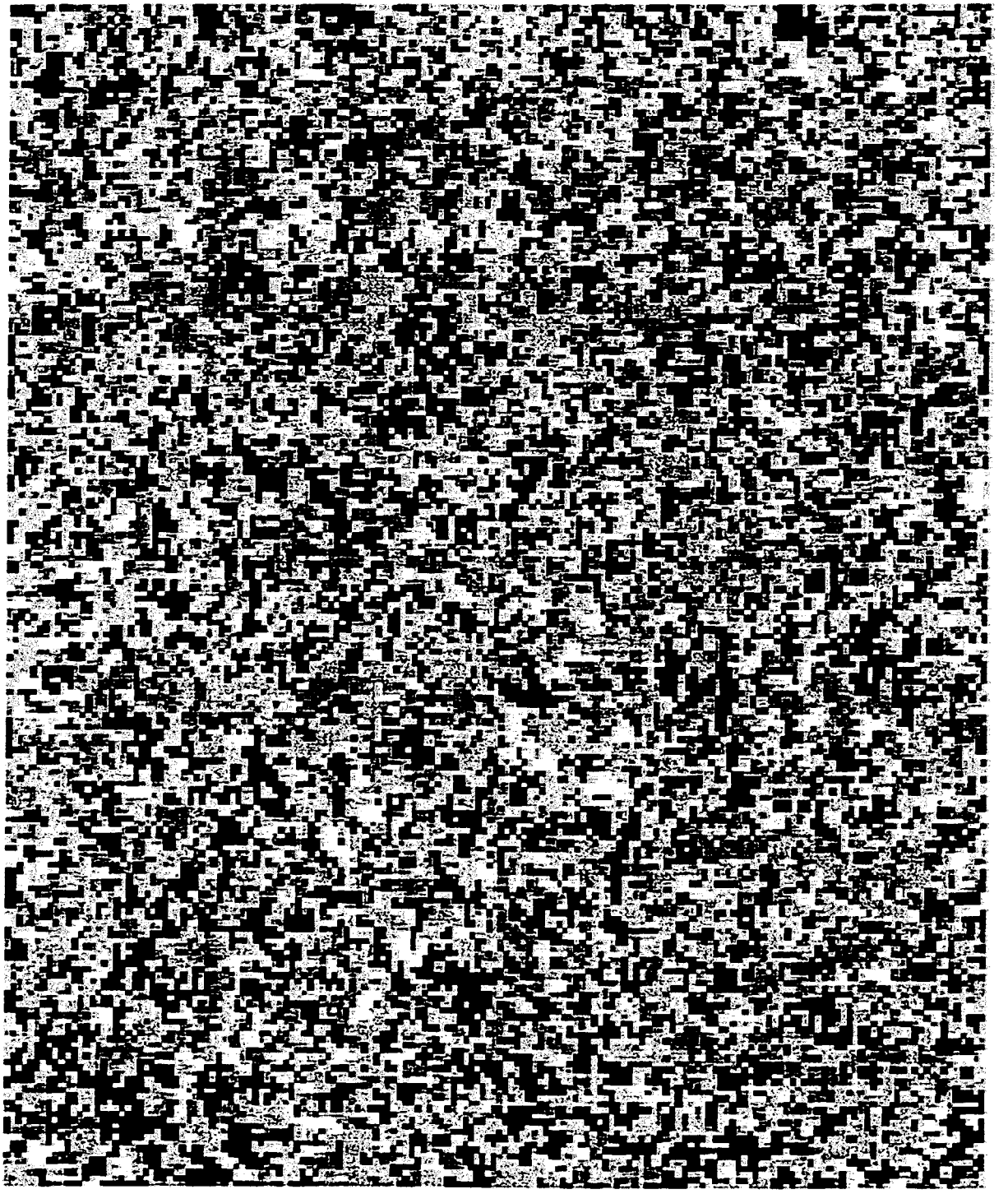


Figure 2.3: A realization from a first-order Potts prior distribution with $\beta = 1.0$ and $m = 6$.

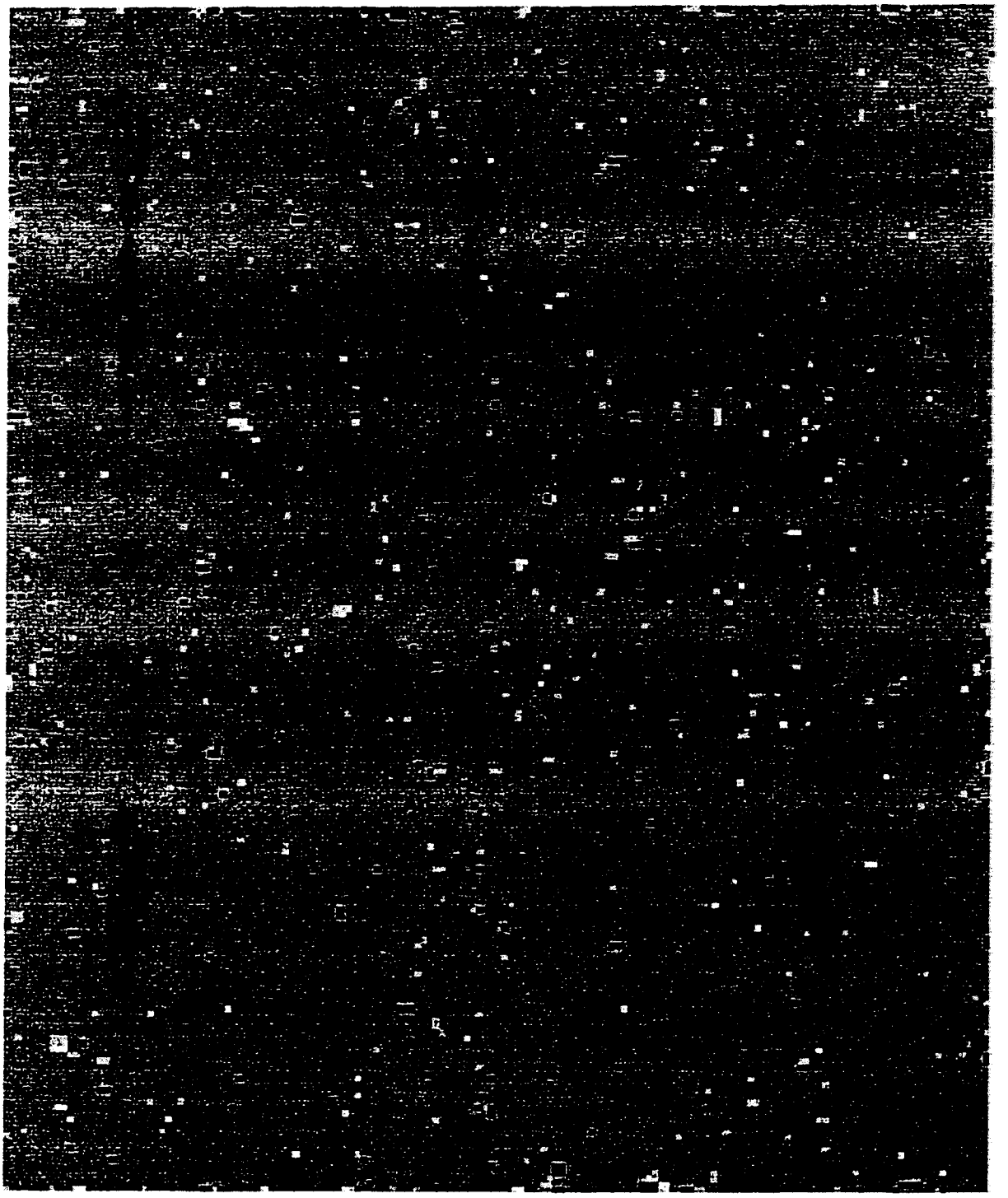


Figure 2.4: A realization from a first-order Potts prior distribution with $\beta = 1.5$ and $m = 6$.

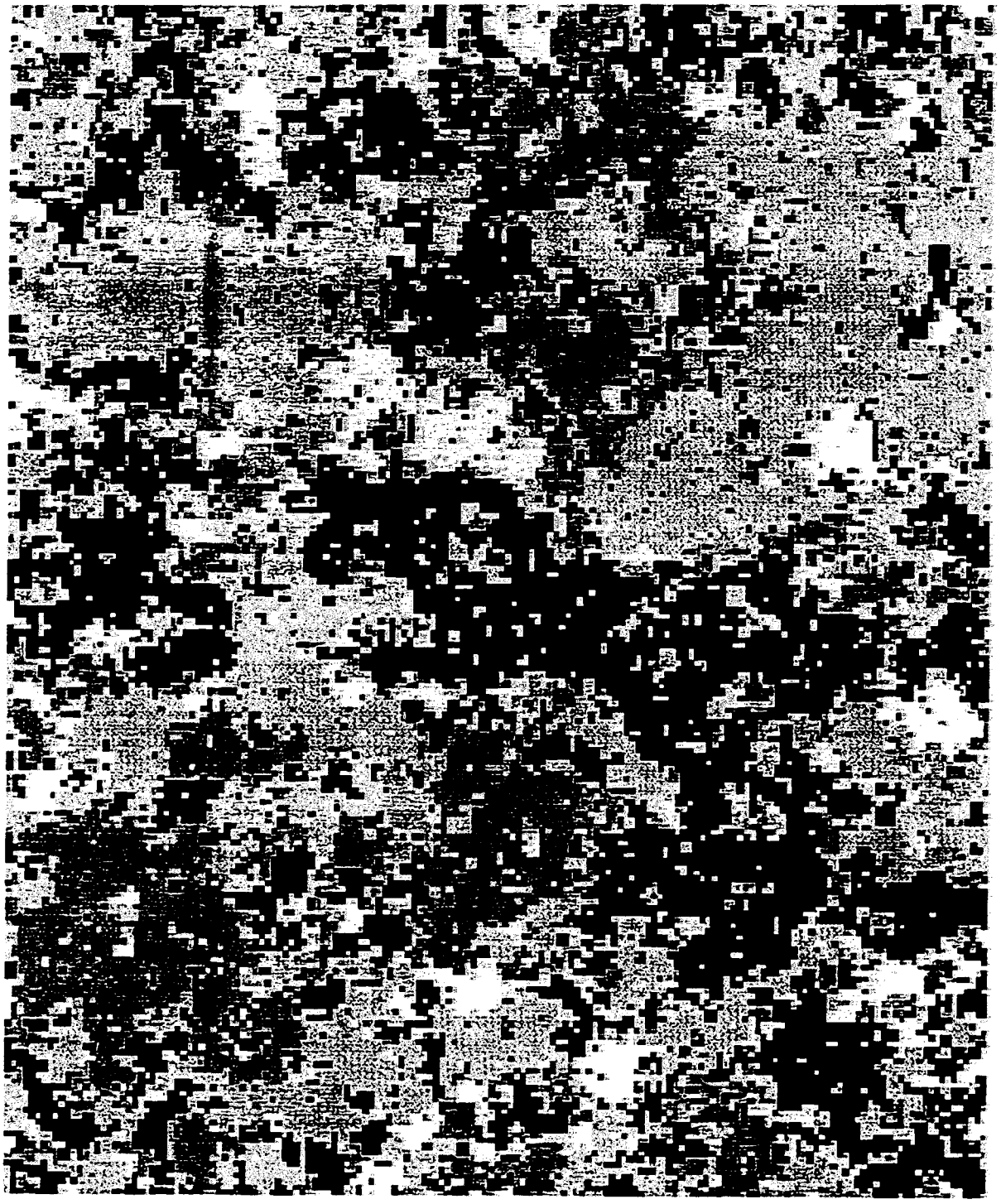


Figure 2.5: A realization from a first-order Potts prior distribution with $\beta \approx 1.24$ and $m = 6$.

enough to eliminate "noisy" site patterns throughout a classification often leads to excessive homogeneity in the form of the distortion and/or elimination of individual forest class areas when applied to actual data. Descombes et al. (1995) demonstrate that the implicit objective of Ising and Potts models is to minimize the length of edges. Further, even when restoration appears to work well, the use of pairwise interaction parameters may lead to poor estimates of classification parameters such as class area and perimeter length (Tjelmeland and Besag 1998).

In practice, β is often chosen arbitrarily depending on the classification objective. For a desired level of homogeneity the appropriate choice of β will depend in part on the spatial variation within the image data, subject to the limitations of pairwise models discussed below. Using a second-order model, Kato et al. (1992) specified 2.0 for a synthetic image and 0.5 for a SPOT image, both relative to a factor of 0.2. Again for a second-order MRF, preliminary research for this study suggests that 0.5 is more desirable for TM data than 1.0. Settle (1989) proposed allowing β to vary spatially across a lattice. This may be reasonable when the image data variation is not spatially uniform as the appropriate choice of β depends on interplay between the classification objective and information contained in the data.

Within the confines of pairwise interaction several modifications have been proposed to adequately address long-range order. Geman and Geman (1984) augmented a second-order pairwise model with an unobserved line process model defined on a dual lattice. The line process is designed to restrict long-range order by inducing object boundaries. Adding the line process model was shown to improve results for fairly homogeneous scenes. However, this design does not preserve linear features themselves nor small objects (Descombes et al. 1995). Bouman and Shapiro (1994) constructed a hierarchical pyramid of dependent pairwise interaction MRFs at different scales to maintain consistent properties at both the local and global scales. However, when applied to SPOT data this method did not show an improvement over other contextual classifiers.

Texture models and boundary detection algorithms represent another approach to image segmentation that more specifically address global characteristics. Cross and Jain (1983) model textures as auto-binomial processes (Besag 1974) using first- through fourth-order neighborhoods. Several texture segmentation methods have been presented using second-order random fields including Derin and Elliott (1987), Bouman and Liu (1991), and Yamazaki and Gingras (1995). Bouman and Liu (1991) assume second-order neighborhoods at multiple scales. Building on the notion of an unobserved line process (Geman and Geman 1984), Geman (1987) and Geman et al. (1990) model the joint probability of image intensity values and boundary indicators to segment images on the basis of homogeneous textures. Although modelling textures may lead to improved forest classification methods, it is outside the scope of this research.

2.2.3 Higher-order interaction MRFs

Markov random fields of higher-order spatial interaction provide a technique for modelling realistic forest patterns whereby the model and the desired classification are consistent at both the local and global levels. Although pairwise interaction models induce spatial continuity, they do not encourage region shape. By recognizing the shapes of objects, higher-order interaction MRFs exhibit interesting large-scale characteristics. Triple cliques can be designed to simulate convex and concave boundaries. Triple and quadruple site interaction can be modelled by second-order neighborhoods (Besag 1974). Quadruples allow for homogeneous object interiors. Higher-order MRFs have the drawback of being computationally expensive.

On the hexagonal lattice triple cliques are possible with a first-order random field. However, to adequately model the large-scale properties of object on background images, Tjelmeland and Besag (1998) develop second- and fourth-order MRFs. Hexagonal arrays offer more appealing neighborhood geometries. In addition to higher-order MRFs other methods exist for modelling the global characteristics of images; not all

are lattice-based; see Tjelmeland and Besag (1998) for a brief review.

Returning to the finite rectangular lattice, Descombes et al. (1995) have introduced a class of MRFs designed to preserve fine structures. Realistic classification features such as lines and convex object boundaries are represented as low energy/high probability clique configurations via specification of the clique potential functions. Noisy patterns are considered high energy states. The problem reduces to developing a method for coherently assigning an energy value to each possible clique configuration. Originally, the Descombes et al. model was proposed for binary images, and recently generalized to m classes by Byers (1998) who also devised a computationally efficient method for implementing the energy assignments.

The Descombes et al. MRF requires a fifth-order neighborhood, which for each site i consists of 24 pixels in a 5×5 block centered on i but of course excluding i itself. This defines m^9 cliques C as 3×3 blocks of sites. Thus, for $m = 2$ there are 512 cliques, and for six classes there are 10,077,696 cliques. Imposing translation, rotation, and class invariance reduces the number of distinct configurations. For binary classifications there are 51 of these configurations shown in Figure 2.6. In examining the configurations in Figure 2.6 it is apparent that some represent spatial continuity, some convex and concave object boundaries, some linear features, and others noisy patterns.

Figure 2.7 shows the nine low energy clique configurations, up to translation, rotation, and class invariance, implemented for this research. Type (1) corresponds to the homogeneity within a region. Types (2-6) model region boundaries between any two classes. Types (7-9) represent possible "junctions" between the edges of any three regions. All other possible clique configurations are assumed to represent high energy states having low probability. Note that linear features, especially configuration (26) in Figure 2.6, are not to be preserved in applying this model to a forest structure classification. However, if "roads" were chosen to be a class, one would want to assign configuration (26) associated with that class a lower energy. Thus, constructing an

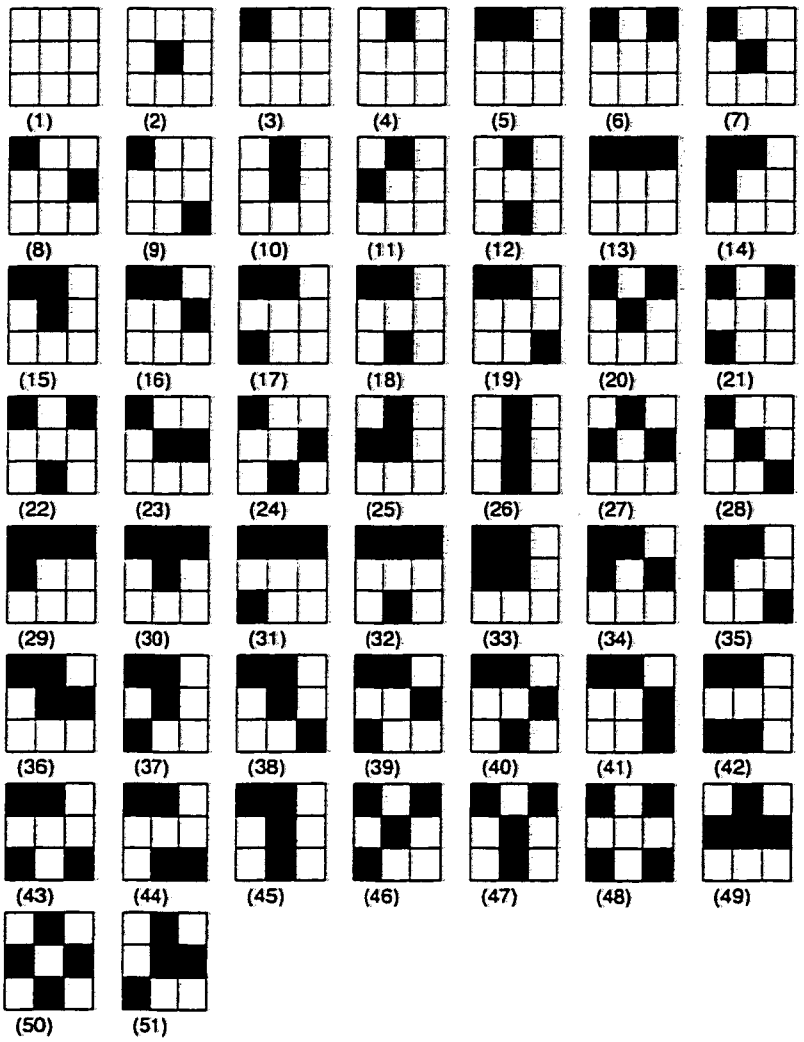


Figure 2.6: Clique configurations for the Descombes et al. prior up to rotation and class invariance.

MRF of desired characteristics amounts to interpreting each of the parameters in an appropriate fashion.

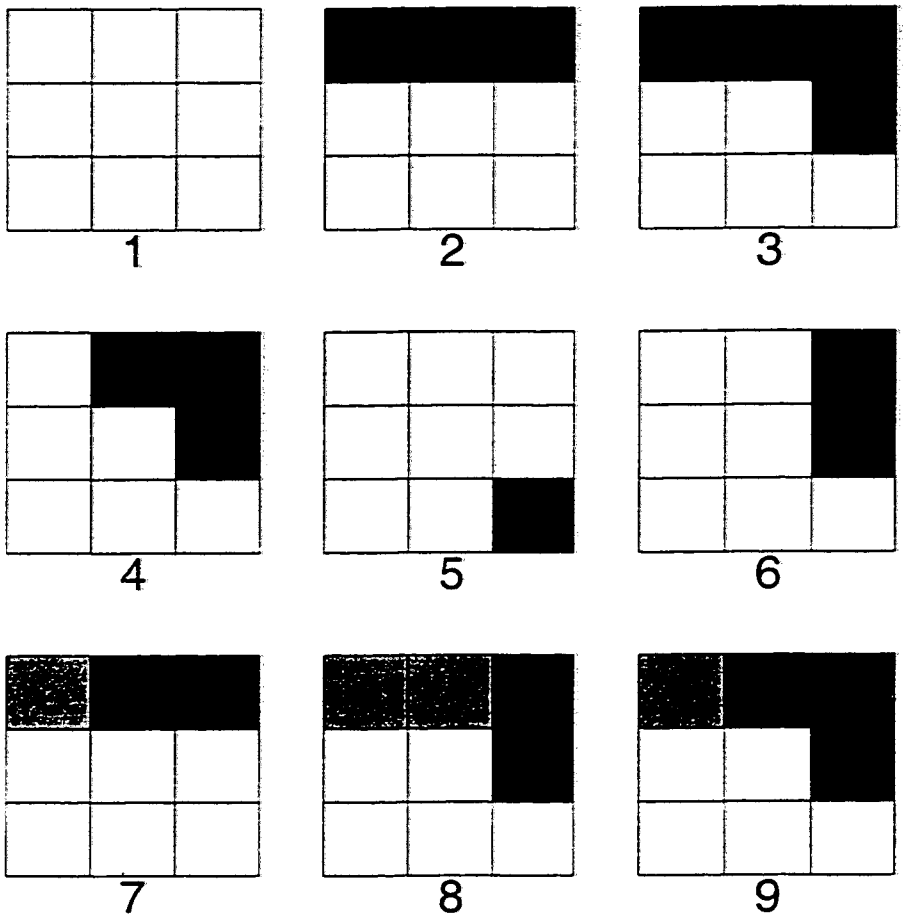


Figure 2.7: Low energy clique configurations for the Byers version of the Descombes et al. prior up to translation, rotation, and class invariance.

Descombes et al. (1995) specified the clique potential functions, V_C, in terms of three interaction parameters in order to regulate the suppression of noise, and preserve edges and lines. Similarly, Byers (1998) specified potential functions in terms of interaction parameters for noise (E_N), edges between two regions (E_E), and junctions between the edges of three regions (E_J). These potential functions are more suited

to classifying forest areas without concern for linear features. The following linear equations give the energies for the nine clique configurations shown in Figure 2.7:

$$
\begin{aligned}
E_1 &= 0 & E_6 &= \tfrac{13}{20}\left(\tfrac{\sqrt{5}}{2}E_E - \tfrac{\sqrt{2}}{4}E_E\right) \\
E_2 &= \tfrac{1}{2}E_E & E_7 &= \tfrac{1}{4}(3E_E + E_J) \\
E_3 &= \tfrac{7}{20}\left(\tfrac{\sqrt{5}}{2}E_E - \tfrac{\sqrt{2}}{4}E_E\right) & E_8 &= \tfrac{1}{4}(3E_E + E_J) \\
E_4 &= \tfrac{\sqrt{2}}{4}E_E & E_9 &= \tfrac{1}{4}(3E_E + E_J) \\
E_5 &= \tfrac{\sqrt{2}}{4}E_E.
\end{aligned}
$$

The continuous nature of different regions are explicitly considered as all energy assignments are relative to the type (1) cliques shown in Figure 2.7. These configurations are assigned zero energy, assuming larger homogeneous areas of all classes are equally likely to appear in the classification. All cliques not shown in Figure 2.7 are assigned an energy equivalent to the constant for noise. By systematically parameterizing the model in this manner the spatial configuration of the MRF is controlled by regulating the three interaction parameters.

Simultaneously modelling noisy configurations, boundaries, and homogeneous interiors enables the Descombes et al. framework to reduce noise without producing undesirable long-range order. Figure 2.8 shows a realization from a simulation of the Byers version of the model. Clearly, the global properties of this model are much more closely related to a desired forest classification than the global properties of the Potts or Ising models. Indeed, Descombes et al. (1995) demonstrated improvements over the Ising model for restoring lines and objects in a binary image. A preliminary version of this MRF showed an improvement over the Potts prior for classifying forest structure (Moffett et al. 1997).

2.3 Hierarchical models

Different processes, such as atmospheric scattering and topographic variation, degrade image data. The Bayesian approach allows a succession of sub-models for each source

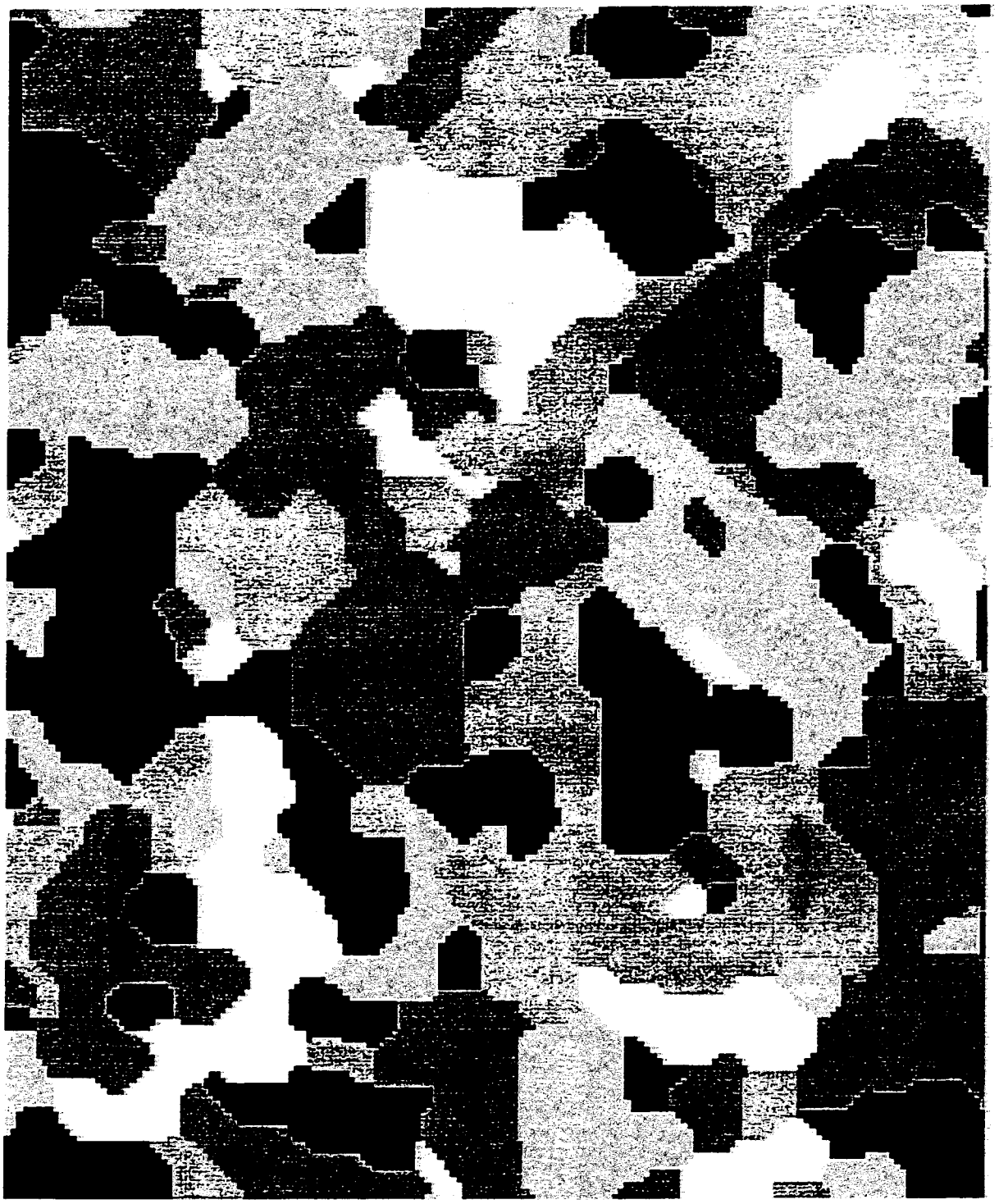


Figure 2.8: A realization from the Byers version of the Descombes et al. prior distribution for $E_N = 2.5, E_E = 1.5$, and $E_J = 2.0$.

of variation within a *hierarchical* framework. Rather than preprocessing the image data to correct for, or "remove" unwanted variation, this approach links the raw image data with the eventual classification through a series of models for each stage of the overall remote sensing process. Given that image data is degraded by unique and different process, a distinct model is required to physically represent the variation induced by each process. Further, Teillet (1986) argued that all radiometric corrections must be combined in a single algorithm in order to preserve proper radiance records, as the application of more than one linear transformation is not a commutative process. Further, deterministic approaches to image correction frequently lead to problems having more variables than equations.

The term "hierarchical" has been interpreted in many ways within the Bayesian image analysis literature. The simple model given by Equation (2.1) is hierarchical in that it links a likelihood function for the spectral information with a prior distribution for the spatial information. Derin and Elliott (1987) and Kelly et al. (1988) are examples two-level hierarchical models of this sort.

Physically-based models are more justifiable and intuitively appealing than statistical classifiers merely based on correlation. The likelihood adopted by Kelly et al. (1988) represents the physical characteristics of the radar speckle process. Furthermore, in principle physically-based models should provide a means of developing reliable and readily extendible remote sensing technology.

Additional models for each form of degradation can be linked through successive conditioning (Smith and Roberts 1993) of the form

$$\pi(x|z_1), \pi(z_1|z_2), \ldots, \pi(z_k|y). \tag{2.7}$$

Each source of variation is modelled by a conditional density as a stage within a hierarchical framework whereby each stage is treated as an additional prior distribution (Besag 1989). The resulting posterior represents the combined effects of the degradation processes and the spatial information, thus fidelity to the data is main-

tained.

Each stage of the hierarchy may represent an unobserved attribute process, for example the line process modelled by Geman and Geman (1984), or additional data sources (Solberg and Taxt 1994; Peytavin et al. 1995). Note that the classification of multiple data sources is not necessarily a physically-based approach. Wilson et al. (1994) linked models of the sensor point spread function and banding degradation with a joint distribution for the classes to classify univariate data. This model assumes unobserved processes in the form of a blurred image and a "true reflectances" image.

The hierarchy presented in this study links models of the system point spread function, bi-directional reflectance, and a higher-order model of spatial interaction. The first two stages represent physically-based models of remote sensing processes. The bi-directional reflectance model assumes an unobserved image and incorporates additional terrain elevation data. Some rather preliminary results were presented in Moffett et al. (1997). Chapters 3 and 4 describe the remote sensing models in detail.

2.4 Summary

The random vectors X and Y have been defined on corresponding rectangular finite lattices representing the classification and the image data, respectively. The components of X are referred to as sites, the components of Y as pixels. Bayes theorem can be used to hierarchically integrate probabilistic models of the spatial patterns of forest classes and physical models of remote sensing processes. By adopting a conditional probability perspective spatial interaction can be modelled as a locally dependent Markov random field. Pairwise MRFs have been proven successful in imaging applications due to consistent local characteristics and global constraints induced by the data. Higher-order MRFs exhibit local and global consistency, thereby providing a means of simulating more realistic forest classifications.

Chapter 3

BIDIRECTIONAL REFLECTANCE

The appropriate use of satellite image data for any purpose requires a knowledge of the data. The first step in developing a reliable classification model is understanding *what* is being recorded by the satellite sensor and *how* it is recorded. This chapter begins with a review of the physical processes underlying surface irradiance and directional reflectance. Field measurements of forest canopy reflectances are reviewed in the second section. The following section introduces a likelihood function that models forest reflectance as the product of a normalized bidirectional reflectance factor and the mean normalized spectral reflectance of each forest class. The fourth section reviews canopy reflectance models, which can be used to estimate the reflectance factors. Section 3.5 outlines the application of a radiosity model for estimating the reflectance factors. The following chapter reviews the measurement of reflected radiance by a remote sensor and presents a model of that process.

3.1 Review of radiometric concepts

The earth reflects *solar radiant energy*. This energy, measured in joules, is transferred by photons in the form of electromagnetic waves. *Radiant flux* (Φ), measured in watts, is the rate at which radiant energy is transferred from one point to another with respect to time. *Irradiance* (E) measures the flux passing through an area on a plane perpendicular to the direction of incidence in watts per square meter. The rate of change in Φ with respect to direction is measured as *radiant intensity* (I) in watts per steradian (unit solid angle).

Radiance (L) provides both spatial and directional information about the transfer of radiant energy. Schott (1997) describes radiance as the most complex, useful, and ubiquitous radiometric term. Radiance is defined as radiant flux per unit projected area per unit solid angle in units of watts per square meter per steradian. Radiance can represent flux emerging from, passing through, or incident upon a surface and in general is expressed as

$$L(A;\omega) = \frac{\partial^2 \Phi(A;\omega)}{\partial A \, \cos\theta \, \partial\omega} \quad [\mathrm{Wm^{-2}sr^{-1}}], \tag{3.1}$$

where A is a surface area, $\omega \equiv (\theta, \phi)$ defines a direction relative to the surface plane normal in terms of a zenith and azimuth angle, respectively, and $\cos\theta\partial\omega$ is an element of projected solid angle about that direction. Passive remote sensors measure and record a portion of the solar radiant energy that is reflected by the earth into a particular direction as radiance emerging from the surface of the earth. Thus, a satellite image contains information about the spectral and directional reflectance properties of the surface being sensed.

3.1.1 Radiance reflected into the direction of the sensor

The following equation describes the components of the upwelling reflected spectral radiance from a given surface and into a specific direction,

$$\begin{aligned} dL_r(\omega_z, \omega_r; A, B; \lambda) \;=\; & L_z(\omega_z;\lambda)\, \cos\theta_z \, d\omega_z \, e^{-\tau^{\downarrow}(\lambda)\, \sec\theta_z} \, f_r(\omega_z, \omega_r; A; \lambda) \\ & + \int_0^{2\pi} L_d(\omega_d;\lambda)\, \cos\theta_d \, f_r(\omega_d, \omega_r; A; \lambda)\, d\omega_d \\ & + \int_B L_b(\omega_b;\lambda)\, \cos\theta_b \, f_r(\omega_b, \omega_r; A; \lambda)\, d\omega_b \, dB \\ & [\mathrm{Wm^{-2}sr^{-1}}], \end{aligned} \tag{3.2}$$

where

$$f_r(\omega_i, \omega_r; A; \lambda) = \int_A F_r(\omega_i, \omega_r; (x,y); \lambda)\, dA \quad \text{for } i = z, d, b, \tag{3.3}$$

and

dL_r = radiance of the reflected radiant flux (reflected radiance) originating from the incident radiance within a solid angle in the direction defined by ω_z and contained within an infinitesimal element of solid angle around the direction defined by ω_r [$\mathrm{Wm^{-2}sr^{-1}}$],

ω_i = a given direction defined relative to the surface plane normal by the zenith and azimuth angles (θ_i, ϕ_i), respectively, for the subscripted directions z (direct solar radiance), r (reflected radiance), d (diffuse radiance), and b (multiply scattered radiance) [dimensionless],

A = an area of surface larger than a point, but small enough to be considered topographically invariant and uniformly reflective [$\mathrm{m^2}$],

λ = spectral wavelength, [μm],

L_z = direct solar radiance from the direction defined by ω_z [$\mathrm{Wm^{-2}sr^{-1}}$],

$d\omega_i$ = $\sin\theta_i d\theta_i d\phi_i$, an infinitesimal element of solid angle around the direction defined by ω_i [sr],

$\cos\theta_i \, d\omega_i$ = the horizontal projection of an element of solid angle, or an element of projected solid angle [sr],

$\tau^{\downarrow}$ = atmospheric extinction coefficient for the downwelling radiance [dimensionless],

θ_z = solar zenith angle, degrees,

f_r = the bidirectional reflectance distribution function (BRDF), the ratio of dL_r to the irradiance of the incident radiance for a given surface [$\mathrm{sr^{-1}}$],

dL_d = diffuse radiance (radiance scattered by the atmosphere) originating anisotropically from the hemisphere above the surface A [$\mathrm{Wm^{-2}sr^{-1}}$],

B = a surface including A, the points of which may reflect radiance into the direction of $A[\mathrm{m}^2]$.

dL_b = radiance reflected between different points of a surface(s) $[\mathrm{Wm}^{-2}\mathrm{sr}^{-1}]$.

F_r = a proportionality function defined as the ratio of reflected radiance to incident radiant flux for a point (x, y) $[\mathrm{m}^{-2}\mathrm{sr}^{-1}]$.

In other words, the total reflected radiance is the sum of radiance reflected from the direct solar irradiance, diffuse irradiance, and multiply scattered irradiance (see Figure 3.1). Equation (3.2) integrates concepts presented by several authors, each emphasizing a different piece of the radiance puzzle. In general the remote sensing literature adopts a broad view emphasizing the range of factors that combine to form the total radiance incident upon a sensor (Slater 1980; Duggin 1985; Schott 1997). Nicodemus (1970), Nicodemus et al. (1977), and Kasten and Raschke (1974) specifically address the issues of bidirectional reflectance. Borel et al. (1991) and Li (1997) concentrate on multiple scattering. Equation (3.2) does not include terms for thermal emissions from either the atmosphere or surface, nor does it account for radiance transmitted by surface elements. Emitted energy was left out as this research focuses on the visible to shortwave infrared (SWIR) portion of the electromagnetic (EM) spectrum. Modelling transmitted radiance is beyond the scope of this research.

Equation (3.2) shows just how many parameters determine the radiance leaving a surface area. Image acquisition introduces additional parameters. Given that radiance is what is actually being measured and classified, radiometric equations such as (3.2) are not only useful, but necessary in order to construct physically-based models of remote sensing processes for classification or other purposes. While the purpose of this research is to model the sensing process in sufficient detail, it may be of greater value in drawing attention to the complexity of the problem in a spirit similar to Duggin and Robinove (1990). Ignoring the complexity will not make it go away.

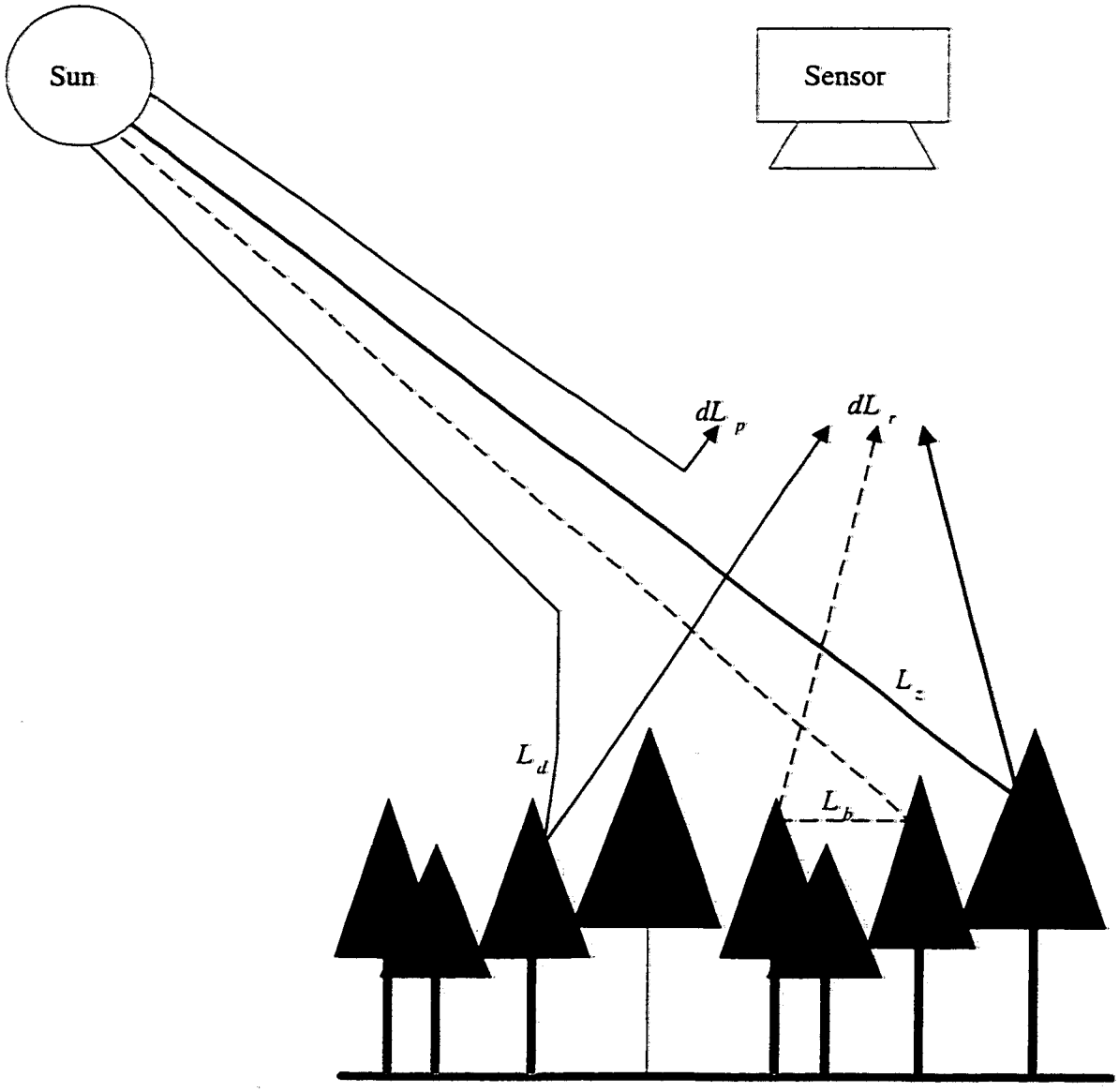


Figure 3.1: The components of reflected radiance and path radiance in the direction of the sensor.

3.1.2 Direct solar irradiance

In Equation (3.2) the downwelling radiance of the direct solar flux is written as L_z. As it strikes the surface the radiance of the flux is projected onto a larger area by a factor of $\cos\theta_z\, d\omega_z$. The irradiance from direct sunlight onto the surface is expressed as $L_z(\omega_z; \lambda)\ \cos\theta_z\, d\omega_z$ [$\mathrm{Wm}^{-2}sr^{-1}$]. Solar radiant flux varies irregularly up to two percent. In addition, another three percent in seasonal variation can be expected attributable to the earth's non-circular orbit around the Sun (Duggin and Robinove 1990). The quantity L_z also depends on wavelength as shown in Equation (3.2).

It is important to note that in Equation (3.2) all of the directions are relative to the surface plane normal. In this case the incidence angle of the flux is equal to the zenith angle of the flux as written. For actual remote sensing of rugged terrain the direction surface plane normal varies continuously with topography, while the solar zenith and azimuth angles are assumed constant per image. In this case the cosine of the incidence angle (ι) can be written as

$$\cos\iota = \cos\theta_z\ \cos\theta_n + \sin\theta_z\ \sin\theta_n\ \cos(\phi_z - \phi_n), \tag{3.4}$$

where θ_n and ϕ_n are surface zenith and azimuth angles, respectively, and the solar angles are defined relative to the earth's surface tangent plane normal. The projected area effect is significant in remote sensing because $\cos\iota$ varies with topography. In other words, dL_r will vary with topography regardless of changes in surface cover.

If the illuminated surface is perpendicular to the direction of incident flux, the incidence angle is zero and the irradiance on that surface will be $L_z d\omega_z$. Basic topographic correction schemes normalize topographic variation to zero incidence angle by dividing measured radiance by $\cos\iota$ and making significant simplifying assumptions.

A portion of the downwelling solar flux is attenuated as it passes through the earth's atmosphere. This reduction is characterized by the downwelling atmospheric extinction coefficient $\tau^{\downarrow(\lambda)}$. Atmospheric absorption depends on wavelength and is highly variable across space, even within the extent of a single Landsat image. Within

the MSS bandpasses (Duggin 1974) measured from 1.4 to 13.4 percent variation in solar irradiance depending upon atmospheric conditions.

3.1.3 Diffuse radiance

The second term of Equation (3.2) accounts for the diffuse irradiance. Diffuse irradiance, or skylight, is radiance reflected onto the surface by particles in the atmosphere. The same principles of projection described above apply here, only that diffuse radiance may emanate anisotropically from all directions in the hemisphere above the surface. The level of diffuse radiance also varies with atmospheric content.

The total amount of diffuse irradiance on the surface is found by integrating over all possible incidence angles. Determining the contribution of the total to the radiance reflected into the direction of the sensor, dL_r, requires that the BRDF fall within the integral as shown. For an anisotropic BRDF, dL_r will vary with only a change in the *distribution* of hemispherical irradiance (Kriebel 1976).

3.1.4 Multiple scattering

The third term of Equation (3.2) represents the irradiance incident upon the surface emanating from other surface points. Photons may reflect many times among surface elements before being scattered into the hemisphere above the surface. This radiance is referred to as multiple scattering. The contribution of multiple scattering to dL_r is determined in a manner similar to that from diffuse irradiance, although the radiance emanates from other surface points rather than the hemisphere above the surface.

Multiple scattering occurs at many measurement scales. For example, a photon may reflect off of a leaf and strike an adjacent leaf, or a photon may reflect off the leaf of one tree in the direction of a leaf on an adjacent tree. If the reflecting leaf is on steep terrain the reflected photon may even be incident on another hillside. The area B in Equation (3.2) accounts for all points which may scatter radiance onto

the surface A. The degree of multiple scattering is depends on the roughness of the surface and the nature of the surrounding topography (Li 1997).

3.1.5 The bidirectional reflectance distribution function

The BRDF is by definition the ratio of the radiance reflected within an element of solid angle around a given direction, originating from a given direction of incidence, to the irradiance incident upon the surface from that direction of incidence (see Figure 3.2; adapted from Nicodemus et al. (1977)).

By considering only direct sunlight and neglecting atmospheric attenuation, the BRDF shown in Equation (3.2) can be rewritten as

$$f_r(\omega_z, \omega_r; A; \lambda) = \frac{dL_r(\omega_z, \omega_r; A; \lambda)}{L_z(\omega_z; \lambda)\ \cos\theta_z\ d\omega_z} \quad [\mathrm{sr}^{-1}]. \tag{3.5}$$

The BRDF characterizes reflected radiance in terms of all possible sun-surface-sensor geometries. Slater (1980) states that it would be more appropriately termed the bidirectional reflected radiance distribution function, but that nomenclature has not been adopted. As the ratio of two derivative quantities, the values of which can never actually be measured, the BRDF is an intrinsic measure of the directional scattering of a surface. The BRDF applies to all sources of irradiance shown in Equation (3.2) and does not depend on atmospheric conditions, however the magnitude of reflected radiance measured above the atmosphere will vary with atmospheric content. Thus, the simplifications made to derive Equation (3.5) from Equation (3.2) are very reasonable for understanding the BRDF. The relative neglect of the effect of bidirectional reflectance in many remote sensing applications may stem from its complexity. As Slater (1980) states, "unfortunately, BRDF and general reflectance nomenclature involve complicated relationships that are seldom fully understood at first encounter."

In principle, the BRDF is a function of two directions, with values ranging from zero to infinity [sr^{-1}]. An infinite value represents perfectly specular reflection from a perfectly flat reflecting surface (Slater 1980). In contrast, reflectance is the ratio

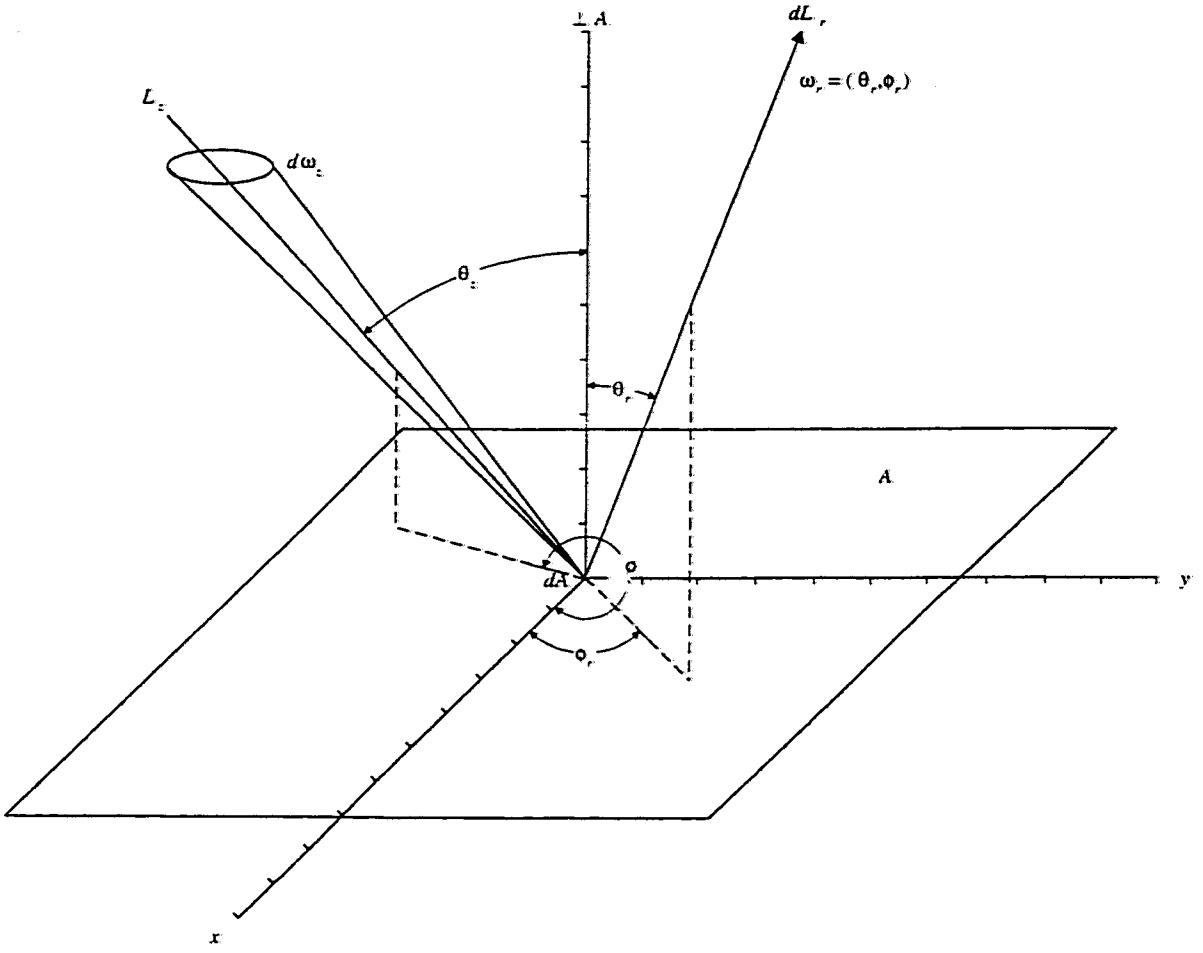


Figure 3.2: Geometry of the bidirectional reflectance distribution function.

of reflected radiant flux to incident radiant flux independent of geometry. Based on the principle of energy conservation reflectance values will range from zero to one, inclusive.

The BRDF is generally considered to be independent of wavelength (Slater 1980). Thus, the shape of the distribution is not expected to vary with wavelength. However, the absolute magnitude of reflectance in a given direction depends upon the spectral reflectance properties of the surface (Slater 1980; Kriebel 1978). For example, photosynthetic vegetation is known to be highly reflective in the near infrared (NIR), but not in the red wavelengths. Therefore, the values of the BRDF contain both directional and bio-physical information about the surface being sensed.

Bidirectional reflectance is often expressed in a unitless form as a bidirectional reflectance factor (BRF), which, unless otherwise defined, is the ratio of the radiance reflected into a particular direction to the radiance that would be reflected into the same direction an ideal Lambertian surface (Nicodemus et al. 1977; Schott 1997). The denominator can be shown to be a constant value of $\frac{1}{\pi\omega_o}[\mathrm{sr}^{-1}]$, where ω_o = a unit of solid angle [sr]. The numerator, dL_r, can never be measured.

The biconical reflectance factor is also a dimensionless reflectance factor that can be measured in principle(Kriebel 1978; Kriebel 1996). This is also the ratio of reflected radiance to that reflected by an ideal Lambertian surface. However, biconical reflectance is based on the use of finite solid angles rather than elements of solid angle, simply derived by integrating over the elements of the solid angles involved (Nicodemus et al. 1977; Kriebel 1996).

3.1.6 Dependence of radiance on measurement intervals

Equation (3.2) expresses reflected radiance, dL_r, as a derivative quantity; the amount of radiance contained within an infinitesimal element of solid angle. Although dL_r can never be measured, as a theoretical concept it provides a valuable understanding of reflected radiance independent of measurement scale.

The reflecting surface, A, by definition of the BRDF, is not an infinitesimal area. For continuously varying natural surfaces, the notion of a uniform area, even a very small one, may not apply. Slater (1980) proposes the bidirectional reflected intensity distribution function as the ratio of reflected intensity, dI_r, to incident intensity, dI_i, with respect to a point surface; but this is rarely used.

In practice, measuring reflected radiance requires that dL_r be integrated over non-zero intervals of area, direction, and wavelength each of which requires a scale of measurement. In addition to the many parameters that determine dL_r the reflected radiance actually measured will depend on the size of these intervals. The spatial resolution depends on the sensor optics, spectral resolution on the width and location of the bandpass(es) (see Chapter 4). The illumination and viewing geometries also depend on the design and orbit of the sensor.

Scales of measurement will be implicit in efforts to model radiance data (see Section 3.3). Simulating bidirectional reflectance requires similar scale specifications as evident in Section 3.4. See Burgess et al. (1997) for a comparison of a BRDF derived from 30 cm airborne data to the BRDF implied by multiple 1.1 km AVHRR data.

Measuring reflected radiance over finite intervals has an averaging effect, reducing the information content of the data. At smaller intervals, two sides of a curved leaf may reflect different levels of radiance into the same direction. The reflected radiance between leaves and a branch will also differ. Yet if reflected radiance is spatially resolved at the scale of the entire tree crown, these different quantities will be averaged into one measurement. Averaging in this way is analogous for spectral resolution. In regard to direction, under certain illumination and viewing geometries, the leaf and branch may reflect the same level of radiance into a particular direction. Thus, knowledge of the surface BRDF can be a valuable tool for modelling radiance data.

3.2 Field measurements of bidirectional reflectance

Several field campaigns have been conducted to measure the radiance from forest canopies in an effort to quantify actual forest reflectance factors. Section 3.1.6 discussed the implicit requirement that radiance be measured at some scale and that the measurements will depend on the scale chosen. At several scales, field measurements confirm that forest canopies scatter irradiance anisotropically in most cases. Anisotropic reflectance is orientation-dependent, while isotropic reflectance is defined to be constant with respect to direction.

Kriebel (1978) used a Cessna aircraft to measure coniferous forest reflectance in Germany. Radiance for seven bandpasses were averaged over a solid angle corresponding to a 10 m linear distance on the ground and directional irradiance was derived from the solution of a radiative transfer equation that accounted for multiple scattering. Thus, Kriebel measured and reported biconical reflectance factors. He then solved a system of linear equations to derive biconical reflectance factors as the ratio of radiance into a given direction to irradiance from all directions for a given set of cones. These results quantified the anisotropy of the conifer canopy. In general, reflectance increased as the zenith angle of reflection (view angle) increased. As the direction of viewing ranges from directly over a horizontal conifer canopy to the horizon, few between-crown shadows are apparent, thus more radiance is reflected in those directions. This effect of between-crown shadows is commonly referred to as mutual shadowing. Kriebel (1978) also showed that other vegetated surfaces, including savannah and bog, exhibited anisotropic reflectance but that the reflectance patterns were different for each surface type.

Kleman (1987) used a helicopter and a field radiometer to measure radiance from spruce and pine stands in eastern Sweden. Irradiance was calculated applying a cosine function to calibrated measurements of a barium sulfate reference panel taken before and after take off. The measurements were recorded at nadir at a scale corresponding

to the TM sensor. Kleman found strong anisotropy in the principal plane associated with changes in solar zenith angle, and more so for spruce than pine. This is evidence of a higher degree of mutual shadowing within the spruce stand, as it exhibited denser biomass and a more pronounced vertical structure. He also found greater anisotropy in the red wavelength than the NIR, attributing that to a greater degree of transmittance in the NIR. Syrén (1994) adopted a similar measurement approach, designed to simulate the TM sensor in spectral and spatial resolution (scale), also for measuring radiance from spruce and pine forests in Sweden. His results demonstrate nadir reflectance increases as solar zenith angle decreases. Both of these studies occurred on fairly level terrain. Thus, as solar zenith angle decreases (towards zero incidence angle) the area of mutual shadows (visible from nadir) decreases. In other words, when the source of illumination is directly over horizontal terrain, trees do not cast shadows on one another.

At a smaller scale, Deering et al. (1994) mounted a field radiometer and the Portable Apparatus for Rapid Acquisition of Bidirectional Observations of Land and Atmosphere (PARABOLA) sensor on a tower extending above a canopy of 70 year old red spruce and eastern Hemlock in Maine. The study site was on level terrain. At the nadir view angle the sensor achieved a spatial resolution of 2.6 m in diameter. Off-nadir resolution increased to 11.0 m in diameter at 60°. PARABOLA has three channels in each of the visible ($0.65 - 0.67\mu$m), NIR ($0.81 - 0.84\mu$m), and SWIR ($1.62 - 1.69\mu$m) regions. Although they do not provide an explicit definition of a bidirectional reflectance factor, they measured the radiance from a reference barium sulfate panel, suggesting a similar approach to the one taken by Kleman (1987). Deering et al. (1994) found that the spruce-hemlock canopy was highly anisotropic in the solar principal plane, but much more isotropic in the plane perpendicular to the solar principal plane. Maximum reflectances occurred in the backscatter directions, with a hot spot of maximum reflectance at the zero phase angle (viewed from same direction as the source of illumination) in all three bandpasses. Nadir reflectance

was found to be slightly less than forward reflectance, but due to the nature of the mechanism, PARABOLA measurements are less reliable from the nadir view angle. PARABOLA data acquired for a mature jack pine stand in Saskatchewan gave similar BRF patterns (North et al. 1996).

Deering et al. (1994) noted a favorable comparison between their reflectance factors to those of Kleman (1987) in terms of the anisotropy. Although Kleman's factors are generally lower in magnitude, the spruce stand he measured was half as dense as the Maine stand. Differences between the two sets of factors may also be due to the difference in the measurement intervals.

In conjunction with the ground-based measurements by Deering et al. (1994), Ranson et al. (1994) derived BRFs for the same site in Maine from measurements acquired by the airborne Advanced Solid-State Array Spectroradiometer (ASAS). The ASAS is a multiangular remote sensing device with 29 visible and NIR bands and a cross-track spatial resolution of 4.25 m. They used an atmospheric radiative transfer model with training sites within the ASAS data set to predict surface reflectance factors. These reflectance factors are consistent with the results cited above, although they noted minimum reflectances in the forward scatter direction of the solar principle plane. The ASAS has also been used to measure anisotropic reflectance from Pacific Northwest forests (Abuelgasim and Strahler 1994). Finally, forest anisotropy in France and Saskatchewan has been observed using the airborne POLDER (POLarazation and Directionality of Earth's Reflectances) sensor (North et al. 1996; Takemata et al. 1996).

At the scale of branches and leaves, Williams (1991) measured the anisotropic reflectance of spruce, pine, and maple needles/leaves and branches for input to radiative transfer models. Walter-Shea et al. (1996) quantified the anisotropy of jack pine shoots.

In general, these studies show that forest canopy reflectance is dominated by strong backscattering. Mutual shadowing reduces reflectance in the forward and nadir scat-

tering directions. Anisotropy is most evident in the solar principle plane. Three dimensional plots of the BRF curves consistently exhibit a bowl shape, with a hot spot maximum reflectance peak within the bowl. The hot spot occurs at the zero phase angle. Although these studies were conducted on areas of little relief, they demonstrate that forest reflectance varies with illumination and viewing orientations. Topographic variation in conjunction with the time of acquisition and latitude determines these orientations.

While these reflectance characteristics appear to be similar for a variety of forest canopies and other vegetated surfaces, these and other studies suggest the BRF curves depend on the type of surface cover. Kriebel (1978) measured different BRF curves for different surface covers. Ranson et al. (1994) found more BRF variability associated with dense spruce than either aspen-birch or bog shrubs, possibly explained by the notion that conifers transmit less radiation and therefore cast darker shadows. Russell et al. (1996), using the ASAS, located the hot spot in one direction for black spruce and aspen, but in another direction for jack pine. Finally, Brown de Colstoun et al. (1996) found bidirectional reflectance to vary with land cover in the Sahel. These studies suggest that rather than complicating the classification process, bidirectional reflectance may yield information for discriminating different forest and land cover types.

3.3 Classification of spectral information

At the scale corresponding to Landsat TM, reflected radiance is largely dependent upon the solar illumination angle, surface topography, the atmosphere, and the spectral and directional reflectance characteristics of the surface cover. Only the variation in the latter contains information useful for classification purposes. For example, the spectral radiance of a young forest on a north facing slope could resemble that of an older forest on a south facing slope, because the larger projected area of the north

facing slope relative to the south facing slope could mimic the effect of larger tree shadows in the older forest for certain geometries. In this case the larger tree shadows are pertinent information, whereas variation in projected area is not.

Several studies have confirmed variation in remotely sensed radiance attributable solely to topography. Holben and Justice (1980) conducted experiments with Landsat MSS data and concluded that a wide range of pixel values can be measured for a single cover type due solely to variations in slope and aspect. Walsh (1987) examined MSS data of forests in the Oregon Cascades and found that the spectral variability attributable to topography was much greater than the spectral variation induced by tree species and canopy size and density. Leprieur et al. (1988) also found that surface geometry has a significant effect on radiance values for TM imagery of mixed forests on mountainous terrain in France.

Normalizing the projected area effect produced by topographic variation, while capturing the bidirectional reflectance information provided by topographic variation is an important task for accurately interpreting remotely sensed data. Developing a reliable, geometrically-invariant classification model requires modelling the sun-surface-sensor geometry in conjunction with the BRDF of the surface at a given scale. Supervised classification itself requires that continuous measurements of spectral radiance be discriminated on the basis of mean values associated with discrete classes.

For predetermined measurement intervals this task can be accomplished by modelling reflected radiance into the direction of the sensor as the product of the mean reflected radiance for each class from a normalized surface orientation and a dimensionless, surface cover-dependent, normalized bidirectional reflectance factor (NBRF). Where the NBRF is, for a given cover type, the ratio of reflected radiance from a given surface orientation to that of the reflected radiance from the normalized surface orientation. In other words, for a given solar position, the reflected radiance into the direction of the sensor can be modelled as the product of a normalized mean value for

each class and a factor that is the ratio of reflected radiance from any given surface orientation to the radiance that would be reflected from the normalized orientation.

The upwelling reflected radiance field over a large area can be modelled as an unobserved image $z = \{z_i : i \in S\}$, where the size of each pixel i corresponds to the TM sampling interval (30 m, see Section 4.1.4). The natural variation in forest structure and reflectance along with the uncertainty arising from averaging reflected radiance at this scale are characterized by assuming the z_i to be Gaussian and conditionally independent, given x. This leads to the following distribution

$$\pi(z|x) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(z_i - T_{x_i}^{(i)}\mu_{x_i})' \Theta^{-1}(z_i - T_{x_i}^{(i)}\mu_{x_i})\right\}, \tag{3.6}$$

where $T_{x_i}^{(i)}$ is a diagonal matrix representing the NBRFs and the mean normalized spectral reflectances are given by μ_{x_i} for class $c = x_i$ for all $c \in \Lambda$. The common covariance matrix is Θ. Thus, this model implies an additive noise term. In principle the covariance will vary depending on the nature of each class, however in the context of a fully Bayesian approach it is difficult to estimate directly given that z is unobserved. The assumption of normality is arbitrary and may present a limitation of this particular model. However, the consensus of many remote sensing studies based on the assumption of independent and normally distributed data has been that the independence assumption is a greater limitation than that of normality. In principle negative radiance values have positive probability under this assumption, although these probabilities are extremely small. Modelling reflected radiance with a gamma distribution should be considered in future research. Recently, Smits and Dellepiane (1997) modelled SAR data as gamma distributed.

Rather than attempting to remove unwanted variation attributable to topography, this approach models the effect of topographic variation on the bidirectional reflectance of the surface covers being sensed. Bidirectional reflectance is one aspect of the entire process of image acquisition and classification. Equation (3.6) is one stage of a hierarchical process and is physically-based to the extent that it is a statistical

model of a physical process. Another stage is required for modelling the measurement of the upwelling reflected radiance field. Thus, z is considered an unobserved image. This approach links the application of the data to the physical processes characterized by the data.

The fact that μ_x is not a radiometric term should also be stressed. The radiometric principles described in Section 3.1 do not address data applications such as classification. As discussed in Section 1.2 forest structure is inherently variable within each class as classification by nature applies discrete categories to continuous stages of forest structure development. Therefore, μ_x represents the mean level of reflected radiance for a range of forest structure and associated radiance values. The following sections discuss the estimation of the NBRFs, while the estimation of μ_x can be found in Chapter 5.

3.4 Review of bidirectional reflectance models

Most of the field measurements presented in Section 3.2 quantified bidirectional reflectance by varying the illumination and/or viewing directions. For the Landsat satellites, the position of the sensor is frequently assumed to be at nadir. For classifying a single image, the solar position can also be assumed static. Thus, the *effective* illumination and viewing angles with vary only with topography. Classifying multiple images of the same scene illuminated from different solar positions adds variation to the effective illumination angles considered.

In order to model topographic and solar position variation, it is necessary to have knowledge of the surface BRDF. However, for a given forest class, the BRDF changes as a function of slope (Hugli and Frei 1983; Schaaf et al. 1994), because trees grow geotropically (see Figure 3.3; adapted from Hugli and Frei (1983)). In other words, a potentially infinite range of BRDFs will be associated with a single cover class over mountainous terrain. It is infeasible to quantify numerous BRF curves on the

basis of field measurements. Therefore, the NBRF values required for Equation (3.6) must be based on modelled rather than empirical estimates. This Section reviews the literature on the subject of modelling bidirectional reflectance for forest remote sensing, and Section 3.5 outlines a particular method chosen for this study.

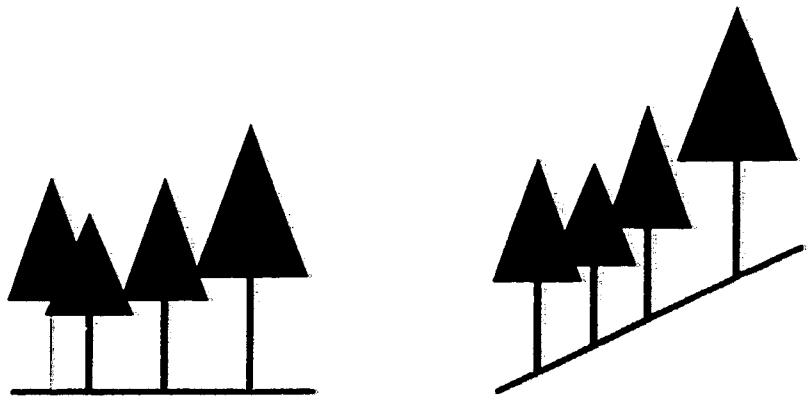


Figure 3.3: Geotropic tree growth on horizontal and sloping terrain.

3.4.1 Simple models

Neglecting forest canopy anisotropy and assuming the canopy to be a perfectly diffuse (Lambertian) reflector is the simplest method for modelling topographic variation. In this case the BRDF is constant and topographic variation can be removed by dividing measured radiance by $\cos \iota$ for a given scale (see Section 3.1.2). This approach further requires neglecting the effects of diffuse radiance and multiple scattering. This technique has been widely adopted, especially for its computational feasibility, but given the number of simplifying assumptions required, this method had not been shown to work well (Smith et al. 1980; Itten and Meyer 1993). Gu and Gillespie (1998) introduce a modification of the Lambertian model for conifer canopies that is also computationally inexpensive. This model represents the mutual shadowing occurring between geotropic conifer trees by dividing $\cos \iota$ by the cosine of the surface slope and the cosine of the solar zenith angle. While their results showed an improvement

over the Lambertian model, neither of these methods is capable of capturing the variation in the BRDF between different canopies; a potential source of information for discriminating classes. Other variations on the Lambertian model have not worked well (Itten and Meyer 1993; Gu and Gillespie 1998).

The Minnaert model was introduced in 1941 as a general anisotropic scattering model for studying the reflectance of the moon. This model also neglects diffuse irradiance and multiple scattering, but in principle could be used to model different cover types. Although the lunar surface is far different from any forest canopy, the Minnaert model has yielded better results than the Lambertian model in forest applications (Smith et al. 1980; Colby 1991; Itten and Meyer 1993). In spite of these studies, a photometric function developed for the moon cannot be considered a physically based approach for modelling forest canopies.

Statistical BRDF models have been developed for the purpose of fitting multiangular data sets (Liang and Strahler 1994; Brown de Colstoun et al. 1996). While these models have fit the data well, they require multiangular data. Applying statistical models to Landsat data acquired under different illumination angles, would require very accurate co-registration and calibration. Statistical correlations may or may not have physical meaning.

3.4.2 Canopy reflectance models

Explicitly modelling canopy reflectance provides a means of capturing the changes to the surface BRDF induced by topographic variation, as well as a method for estimating BRDF sample values from a given sun-surface-sensor geometry. Section 3.1.6 discussed the dependence of radiance on measurement interval. An ideal canopy reflectance model would be able to capture the effects of direct, diffuse, and multiple scattering at all scales. Although the complexity of the models being developed has been increasing to this end, a model of all scales has yet to be developed. Forest reflectance models generally fall into one of three scales: branch/leaf, tree, and land-

scape. Brakke et al. (1996) modelled leaf anisotropy. Several models discussed below have been shown to work well at either the tree scale or branch and tree scales. At the landscape scale, Kimes and Kirchner (1981) modelled the scattering between facing slopes.

While its reasonable to assume that the TM sensor resolves radiance from all three scales mentioned above, the measurements presented in Section 3.2, especially those of Kleman (1987) and Syrén (1994) suggest that mutual shadowing between trees has a predominant effect at the TM scale. Geometric-optical models capture bidirectional anisotropy caused by the single scattering and mutual shadowing of individual tree crowns. All of these models have been found to capture the hot spot and bowl shade of the BRDF explained above; see, for example, Li and Strahler (1992).

The geometric-optical approach models forest canopies as many discrete, three-dimensional trees illuminated and viewed from specific directions. Portions of sunlit and shadowed crown are projected onto a two-dimensional background that is either sunlit or shadowed. Reflectance values are assigned to each of these components. The total reflectance observed from a given direction is the area-weighted sum of the components. The Li-Strahler model is perhaps the most frequently cited geometric-optical model. Li and Strahler (1986) modeled an open conifer canopy as discrete cones of random heights and spacing. Only direct irradiance was considered. Each tree was assumed to be a Lambertian reflector. The model accounted for transmittance, but not multiple scattering. Li and Strahler (1992) improved the treatment of mutual shadowing and adapted the Li-Strahler model for spheroid shaped trees. The geometric-optical approach is limited by the need to project three-dimensional objects onto a planar surface and its inability to accurately model multiple scattering.

Gerard and North (1997) developed a similar geometric-optical model and used a deterministic ray-tracing procedure to find the exact proportion of scene elements projected onto the ground. Both the crown surfaces and the background were assumed to be opaque Lambertian reflectors. They modelled ellipsoid trees of varying shape,

size, density, and distribution to analyze the effects of stand structure variation on bidirectional reflectance. The model predicted a hotspot in the NIR only when there was large variation in crown height. Holding density constant, the BRDF varied with the spatial distribution of individual trees.

Several canopy reflectance models have been designed to capture bidirectional reflectance at both the within and between tree scales. Strahler and Jupp (1990) developed a two-stage model consisting of Li-Strahler type models at each scale. Areal proportions at each stage are determined by the principles of geometric-optics and Boolean models. Li et al. (1995) also extends geometric-optics to the leaf level. North (1996) simulated forest light interaction using a Monte Carlo procedure to model photon transport within and between crowns, stems, and the ground. Photons originate from direct and diffuse sources. This method not only accounts for directional scattering over a range of scales, but also captures the effects of multiple scattering. North found good agreement between his model predictions of BRDF and ASAS measurements for a dense spruce forest in Maine (Ranson et al. 1994). These results demonstrate that crown structure and height variation have a significant effect on the BRDF.

To incorporate the effects of multiple scattering, Li et al. (1995) created a hybrid model by integrating a geometric-optical model of single scattering at the leaf and crown scale with a radiative transfer model for higher orders of scattering. Radiative transfer principles were originally developed as an approach for modelling the atmosphere as a volume scattering medium. Cescatti (1997) simulated radiative transfer in discontinuous canopies. Developing radiative transfer models of forest canopies requires the specification of many physical constants and in this application did not fully capture the directional characteristics of multiple scattering. At the landscape scale, Kimes and Kirchner (1981) developed a model of multiple scattering between facing hillsides to test the effectiveness of simple terrain correction procedures. By assuming the Lambertian surfaces, the model showed that the $\cos \iota$ technique pro-

duced significant errors for slopes greater than 20°. When the correction procedure was modified to account for diffuse irradiance, the results improved.

In addition to the geometric-optical and radiative transfer approaches, radiosity models offer another method of simulating bidirectional reflectance from complex forest canopy surfaces. Radiosity models explicitly represent multiple scattering by predicting the total reflected radiance from a surface as the sum of emitted, transmitted, and single and multiply scattered radiance. Radiosity models also have the advantage of being easily applied to a range of scales and complex surfaces (Li et al. 1996). Borel et al. (1991) simulated the radiosity leaving a canopy of discrete leaves, while Borel and Gerstl (1994) used a radiosity method to simulate the multiple scattering between adjacent hillsides. A radiosity model developed and validated by Wen-Hao Li at the University of Washington was adopted as the technique for estimating the NBRFs required in Equation (3.6) (Li 1997). Section 3.5 describes this procedure.

3.4.3 Inverting canopy reflectance models

Inverting canopy reflectance models in order to retrieve forest parameters and indices from remotely sensed data is a frequently mentioned objective (Strahler and Jupp 1990; Li and Strahler 1992; Rosema et al. 1992; Liang and Strahler 1994; Li et al. 1995; North 1996; Chen and Leblanc 1996; Li 1997). Recently BRDF model inversion has been implemented with statistical and geometric-optical models. Rosema et al. (1992) inverted a geometric-optical model to retrieve crown coverage, LAI, and "yellowness." Comparison with ground data showed fair results, demonstrating the need for further validation. Franklin and Strahler (1988), Wu and Strahler (1994), and Woodcock et al. (1997) have inverted the Li-Strahler model to recover tree size and cover estimates from TM data. In general tree size estimates proved unreliable, while cover estimates compared more favorably to ground data. In addition to the difficulties of the inversion process itself, inverting the model requires estimates of the component reflectances. These are very difficult to measure reliably

and must also be calibrated with the data. Topographic variation further complicates inversion. Woodcock et al. also found it necessary to pre-classify the TM data they used. While BRDF model inversion remains a desirable objective, it has yet to be effectively demonstrated. Multiangular image acquisition planned for the future will create opportunities for statistical BRDF model inversion (Li and Strahler 1996).

3.5 A radiosity method for estimating bidirectional reflectance factors

The application of a radiosity model for simulating the topographic effects inherent in the classification of remotely sensed data is a unique aspect of this research. At the scale of the Landsat TM spatial resolution, the mutual shadowing between individual tree crowns is a primary factor in determining the bidirectional reflectance of a forest canopy (see Section 3.2). Mutual shadowing is influenced by the slope of the terrain as trees grow geotropically (Schaaf et al. 1994) and by the type and structure of the forest cover; see, for example, Ranson et al. (1994). Therefore, to accurately model the upwelling reflected radiance field it is necessary to account for these effects.

Radiosity models predict the total radiant flux reflected into the hemisphere above an enclosure consisting of reflecting surfaces. The interaction of radiant flux between the surface elements is described in terms of an equilibrium radiation energy balance (Borel et al. 1991). Radiosity models are ideally suited to simulate multiple scattering, but unlike radiative transfer methods radiosity models represent the surface elements as discrete objects having distinct shapes, locations and orientations. Given the discrete nature of forest canopies, either at the leaf or the crown scale, radiosity methods represent a more physically correct description of the canopy as opposed to the volume-averaged approach of radiative transfer theory (Borel et al. 1991). Radiosity methods have the disadvantage of being computationally expensive.

3.5.1 Representation of stand structure on sloping terrain

A radiosity model can be easily applied to simulating forest reflectance by representing a forest surface as individual tree crowns and stems. Although radiosity concepts have been developed on the basis of differential surface area elements, modelling radiosity requires the assumption that reflectance remains constant over finite surface elements (Borel et al. 1991). For this research, a program originally developed by Li (1997) was modified to model a range of common Pacific Northwest forest structures. The simulated tree surfaces for one horizontal hectare consist of digital "elevation" models each having 100 rows and 100 columns, thus each grid cell represents a square meter. The grid cell values give the heights of the corners of the surface elements; note that a surface element does not correspond to a grid cell. Thus, there are actually 99×99 surface elements represented by a surface of 100×100 grid cells. The sizes and dimensions of the surface model are completely arbitrary allowing the radiosity to be simulated for a variety of surface covers at any scale. For this study, each grid cell was assumed to represent one meter.

The following method was used to develop the tree surface model. Stem density, height distribution, live crown ratio, crown apex angle for conifer trees, and axes for ellipsoid trees are inputs to the model. Tree locations are randomly chosen without replacement so that only one tree can occupy a grid cell, although stems can occupy adjacent grid cells. Tree heights are randomly selected from a predetermined distribution, either normal or uniform. The height of the crown extending over adjacent grid cells is determined from the tree height, live crown ratio, and shape parameters. For conifer trees, the apex angle (36.8°) corresponds to a crown radius equal to one-third the crown height (Mitchell 1975). If two crowns are selected to occupy the same grid cell, the tallest crown height is retained. This replicates natural stand dynamics in which trees develop asymmetrical crowns when spatially aggregated (Oliver and Larson 1990). Finally, the tree surface is enclosed by a wall; required to maintain the

energy balance. For a given tree surface the wall height corresponds to the mean tree height.

Once the surface model is developed, surface reflectivities are assigned to the corresponding surface elements (not grid cells) depending whether or not the element represents a crown, stem, or the ground. The primary limiting assumption of this model is that each surface element is considered to be an opaque Lambertian reflector, as will be shown below. Nevertheless, by aggregating the surface elements to the resolution of the Landsat sensors the model has been shown to represent forest anisotropy attributable to the mutual shadowing effect (Li 1997). Although it is feasible to model transmission through each surface (Li 1997), to reduce the computational complexity, transmission was not considered for this research. A simple trigonometric algorithm distinguishes between crowns and stems on the basis of the slope of the plane perpendicular to the normal vector of an element. Elements having the steepest slopes are considered stems, otherwise sloping elements are assigned the crown reflectance. Horizontal elements are considered ground elements. Originally this algorithm was developed for trees having the same height, but was modified for this research to accommodate tree height variability.

Figures 3.4 and 3.5 illustrate the tree surfaces developed for the advanced initiation and understory reinitiation structures, respectively. After the crown, stem, and ground elements have been identified, another program transforms a given forest stand on a horizontal surface to the corresponding stand on a sloped surface. Figure 3.6 shows the trees of Figure 3.5 growing on a hillside. Although the individual trees are exactly the same, the mutual shadowing occurring between the trees of Figure 3.6 will differ from the shadowing patterns of the trees in Figure 3.5 because trees grow geotropically.

Although a variety of canopy reflectance models have been developed in recent years only Schaaf et al. (1994) and Li (1997) consider the influence of sloping terrain on tree crown alignment, but only for a single forest type. Gerard and North (1997)

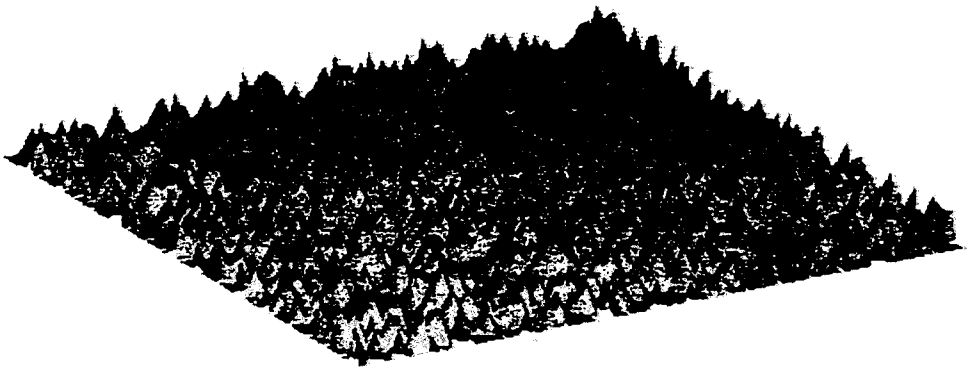


Figure 3.4: Tree surface model of advanced initiation structure on a horizontal surface.

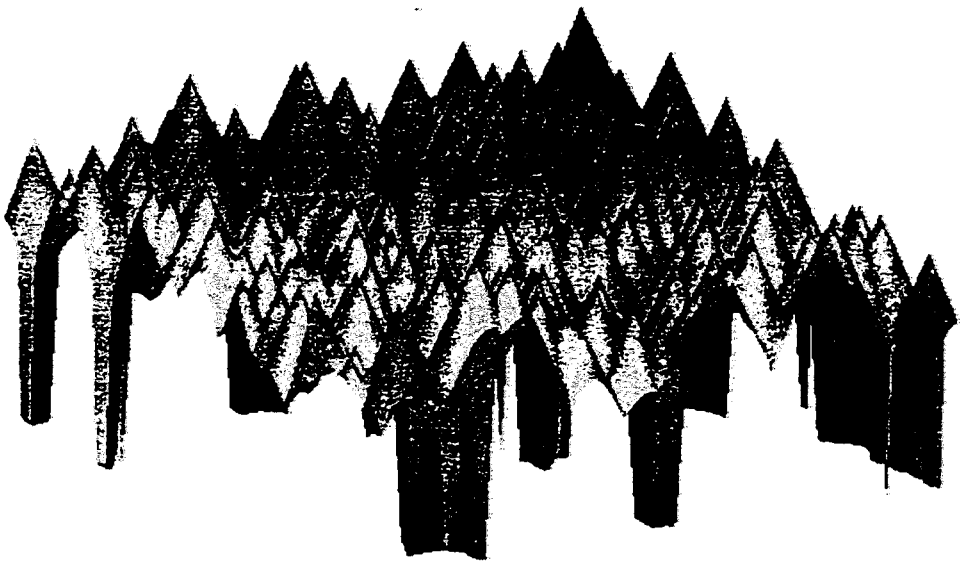


Figure 3.5: Tree surface model of understory reinitiation structure on a horizontal surface.

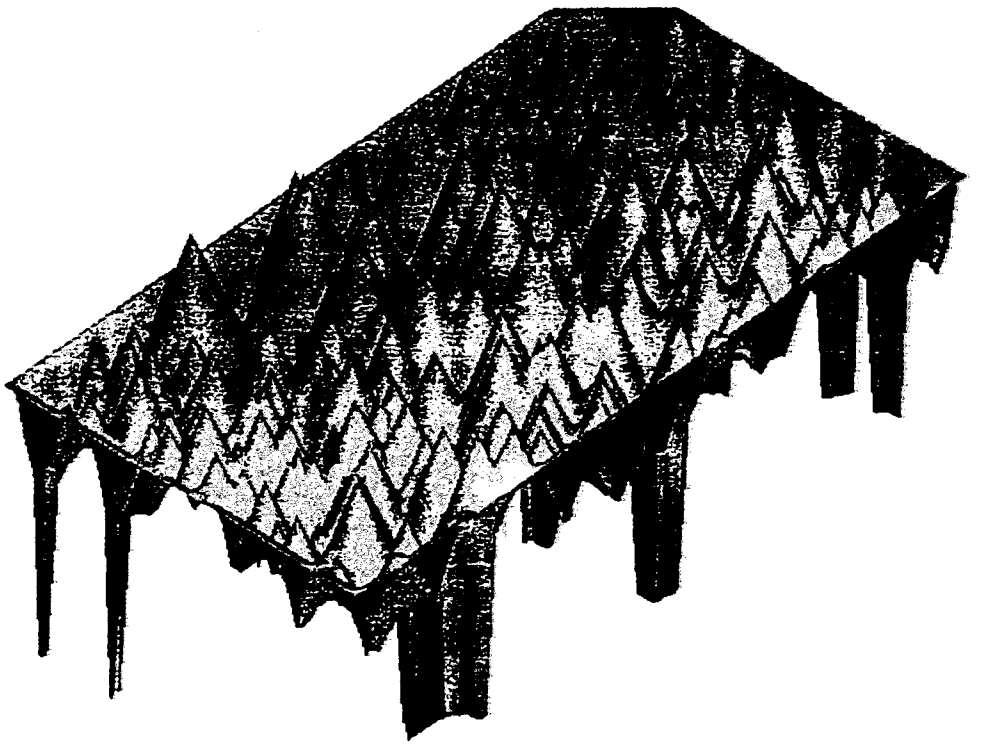


Figure 3.6: Tree surface model of understory reinitiation structure on a 30° slope.

modelled the effects of stand structure variation on canopy BRDFs, but only for stands on a horizontal surface. In order to accurately classify various stand structure types in rugged terrain, both topographic and stand structure influence on reflected radiance must be taken into account; a primary focus of this research.

3.5.2 Form and visibility factor calculations

Once a surface of trees has been developed the next task in applying the radiosity model is to estimate the view or form factors for each surface element. Form factors are defined as the fraction of radiant flux leaving one surface element and reaching another. A surface having l elements will have $\frac{l^2}{2}$ form factors in principle; the factor of one-half accounts for the form factor reciprocity between any pair of elements. The actual number of form factors calculated is reduced by the number of surface elements not visible to one another, and may be further reduced by specifying a maximum distance before extinction. The maximum number of form factors also includes factors for radiant flux incident upon the element it leaves, for each element. These factors are non-zero only for concave elements(Borel et al. 1991).

Calculating the form factors is the central problem of implementing radiosity models. The radiosity model used for this project is a hybrid model in that for each element it selects one of three methods for calculating the form factor in order to balance precision with computing time (Li 1997). The three methods are the double-contour integral, half-contour integral, and the element-to-element method in order of decreasing precision and time. The choice of method depends on a set of threshold values associated with the facing angles, projected areas, and the distance between elements.

Estimating the radiosity visible from a specific view direction requires determining a set of binary detector-view factors for that direction. The number of detector factors is equal to the number of surface elements and specifies whether or not each surface element is visible from a given view angle. For sensors at nadir, such as the

Landsat sensors, all surface elements are considered visible. All of the radiosity values calculated for this research are based on a nadir view direction. For other applications, this approach can be used to model the anisotropic reflectance characteristics, such as the hot spot, by altering the viewing and illumination directions in addition to the surface orientation (Borel et al. 1991; Li 1997).

3.5.3 Simulating reflected radiance

By assuming each of l surface elements to be an opaque Lambertian reflector the radiosity leaving element w', denoted as $M_{w'}$, can be written as

$$\begin{aligned} M_{w'}(\omega_z; A_{w'}, B; \lambda) &= [L_z(\omega_z; \lambda)\ \cos \iota_{w'}\ d\omega_{\iota_{w'}} \\ &\quad + \sum_{w=1}^{l} M_w F F_{w \rightarrow w'} \\ &\quad + E_d(\lambda)]\ \rho_{w'}(\lambda) \quad [\mathrm{Wm}^{-2}], \end{aligned} \tag{3.7}$$

where

$$E_d(\lambda) = \int_0^{2\pi} L_d(\omega_d; \lambda)\ \cos \theta_d\ d\omega_d, \tag{3.8}$$

$A_{w'}$ = the surface area of element w',

B = the total area of surface elements that reflect radiance in the direction of v,

$\iota_{w'}$ = the incidence angle between the surface plane normal and the direction of direct illumination for each surface element w',

$\rho_{w'}$ = surface reflectance of element w',

M_w = the radiosity leaving element w [Wm^{-2}],

$FF_{w \rightarrow w'}$ = the form factor for the radiosity that leaves element w and reaches element w',

E_d = the total diffuse irradiance incident upon element w'.

and all other terms are as defined for Equation (3.2). The incident radiance is projected onto each surface element as a function of the illumination direction and the orientation of the element. Surface elements not directly illuminated are considered shadows, representing the mutual shadowing between crowns. At near zero phase angles, few, if any, shadow pixels are visible giving rise to the hot spot. Likewise, in the forward scatter direction reflectance would be reduced by viewing a higher percentage of shadow elements.

Computing the l unknown radiosity values requires solving a system of l linear equations. For 99^2 surface elements direct inversion methods are not practical. However, Gauss-Seidel iteration provides an approximate solution for a surface of this size (Li 1997).

Having computed the radiosity leaving a surface element, the reflected radiance into the direction of the sensor can be found on the basis of the Lambertian assumption. By definition a Lambertian surface is perfectly diffuse, reflecting equally into all directions in the hemisphere above the surface. Therefore, the radiance reflected into a particular direction can determined by dividing the radiosity leaving the surface by $\pi\omega_o$. The estimate of total reflected radiance leaving the surface, $\tilde{L}_r$, is the sum of radiance leaving the individual elements, written as

$$\tilde{L}_r(\omega_{z'}, \omega_{r'}; A, B; \lambda) = \sum_{w'=1}^{l} \frac{M_{w'}}{\pi\omega_o} = \frac{\sum_{w'=1}^{l} M_{w'}}{\pi\omega_o}, \tag{3.9}$$

where $\omega_{z'}$ and $\omega_{r'}$ define the directions of solar illumination and reflection relative to the normal to the plane of the entire tree surface A. In practice, a large border of surface elements are not included in the summation in order to minimize the effect of the artificial wall enclosing the surface. As mentioned above, for this research the direction of reflection is towards a sensor at nadir, and all radiosity estimates have been calculated for that view angle. In addition, the solar position corresponds to the solar position of the image being classified. Figure 3.7 shows a graphic representation

of the application of this model. For this research it is not necessary to estimate full BRDFs and this has not been done. Rather the objective of modelling bidirectional reflectance in this way is simulate specific directional estimates corresponding to the forest structure and surface orientation variation actually sensed. By estimating $\tilde{L}_r$ as the reflected radiance from a surface of many trees, the radiosity model characterizes forest anisotropy at a scale corresponding to the Landsat sensors.

3.5.4 Applying the radiosity model

Equation (3.6) requires NBRFs for every forest structure-wavelength-slope-azimuth-solar position combination. As shown in the previous section, estimating the reflected radiance for each of these combinations requires inverting a very large matrix. To reduce the computational intensity of this task, radiosity estimates were calculated only for discrete slope, azimuth, and wavelength classes. A total of 130 surface slope-azimuth combinations were considered. Slopes ranged in 10° increments from 60° facing the source of illumination to 60° in the opposite direction. Slope is defined as the angle between the surface plane and plane tangent to the earth's surface. Ten azimuth angles ranged in 10° increments from the solar principal plane to the direction perpendicular to the solar principal place. Symmetry about the solar principal plane was assumed.

The degree of multiple scattering is a function of wavelength (Li 1997). In order to simulate the effect of multiple scattering at a scale corresponding to the spectral resolution of the TM, it is necessary to estimate the radiosity values over a range of reflectances associated with the surface elements, $\rho_{w'}$. To limit the number of crown-stem-ground reflectance combinations, stem and ground reflectances were assumed constant and crown reflectances ranged from 5 to 55 percent in 10 percent increments.

The level of diffuse sky irradiance also varies with wavelength. Not including the term for diffuse irradiance shown in Equation (3.7) further limits the number of radiosity estimates required. Instead, a term for reflected diffuse irradiance can

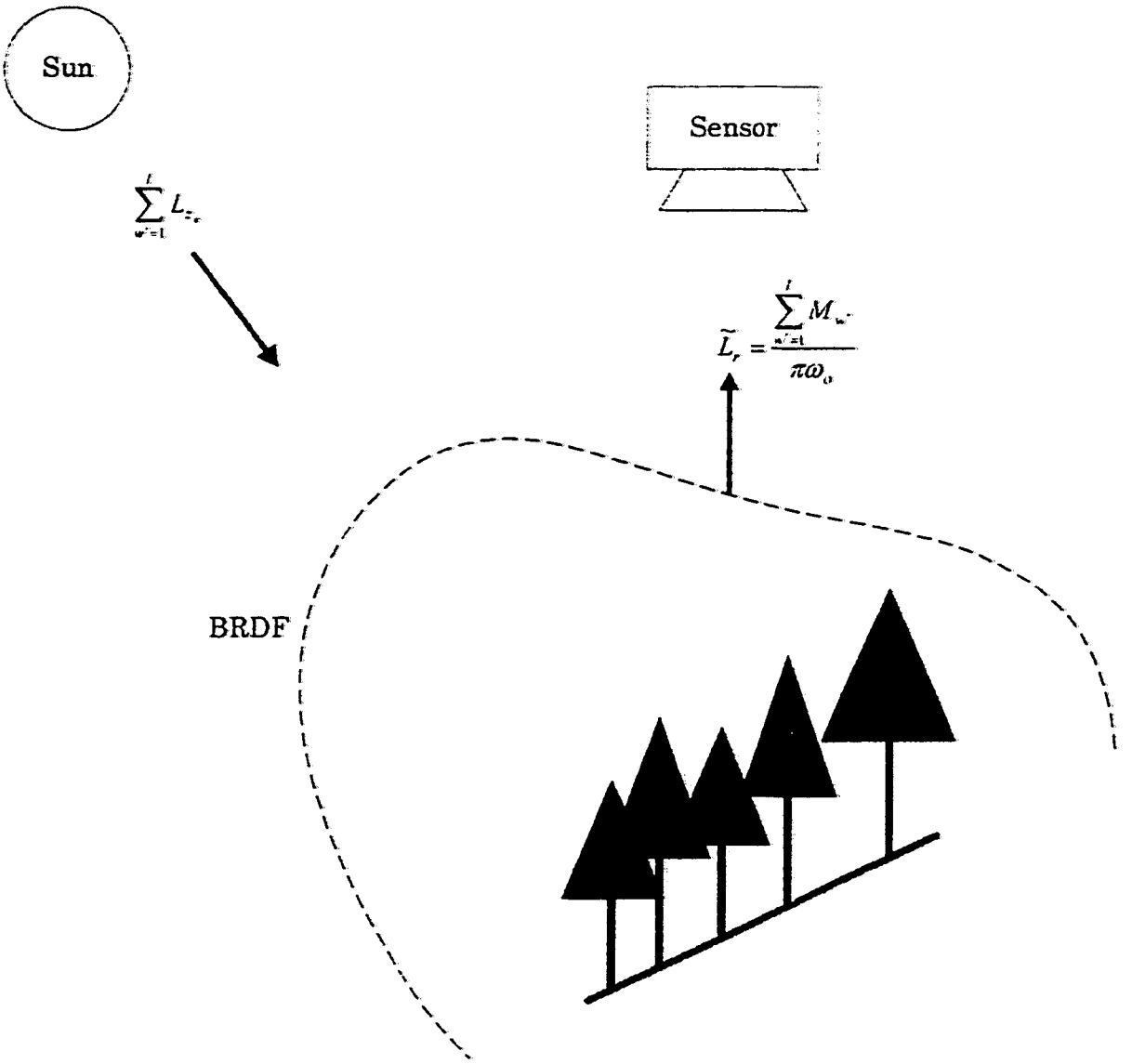


Figure 3.7: Nadir view of radiosity leaving a tree surface.

be added to the radiosity estimates as a function of diffuse irradiance (as a percentage) expected for each wavelength and the class-crown reflectance specified for that wavelength. Calculated in this way, the reflected radiance from a tree surface can be rewritten as

$$\begin{aligned}\tilde{L}_r(\omega_{z'}, \omega_{r'}; A, B; \lambda) = \frac{1}{\pi\omega_o}\Bigg\{ & \sum_{w'=1}^{l} \rho_{w'}(\lambda)\left(L_z(\omega_z;\lambda)\cos\iota_{w'}\, d\omega_{\iota_{w'}}\right. \\ & \left. + \sum_{w=1}^{l} M_w F F_{w\to w'}\right) \\ & + l\left[L_z(\omega_z;\lambda)\cos\theta_{z'} d\omega_{z'}\right] DI(\lambda)\,\frac{\rho_{crown}(\lambda)}{\pi}\Bigg\}, \quad (3.10)\end{aligned}$$

where $DI(\lambda)$ a wavelength-dependent percentage representing the level of diffuse irradiance relative to direct irradiance and ρ_{crown} gives the crown reflectance for the corresponding wavelength. Given that only total downwelling diffuse irradiance was considered in Equation (3.8) and that crown reflectances comprise most of the surface element reflectances, the radiance estimates calculated using Equation (3.10) should not differ significantly from those that would have been calculated using Equation (3.9).

In spite of these limitations, 780 radiosity estimates are still needed for simulating the reflected radiance leaving each forest structure-solar position combination. A final reduction occurs when not all reflectances are required for modelling a particular forest structure. Calculating all of the radiosity estimates needed for this project kept 2 DEC Alpha 3000-276 MHz workstations running almost continuously for about a month. However, for a given solar position the radiosity estimates are as spatially extendible as the forest classes they represent.

3.5.5 Normalization

The direct application of Equation (3.10) would require, at the very least, measurements of solar incident radiance at the time of image acquisition along with the very difficult task of calibrating the image data to radiance. However, direct application

is not necessary for the purpose of classification as discussed in Section 3.3. In Equation (3.6) each forest class is characterized by a unique spectral mean associated with a normalized surface orientation. For this research, class means correspond to horizontal surfaces. The NBRFs, as the ratios of radiance from a sloping surface to radiance from a horizontal surface, transform the mean value to the value associated with a given surface orientation and forest class. The NBRFs are relative measures of the topographic effects.

$$NBRF_{x_i}(\lambda) = \frac{\tilde{L}_r(\omega_{z'}, \omega_{r'}; A', B'; \lambda)}{\tilde{L}_r(\omega_z, \omega_r; A, B; \lambda)} \tag{3.11}$$

As the ratio of simulated values, the specification of L_z for estimating the NBRFs can be arbitrary. The accuracy of the multiple scattering effect depends on the difference between the specified and actual surface reflectances; and actual surface reflectance is not characterized by discrete uniform elements. The accuracy of the representation of reflected diffuse irradiance depends on the proportion (DI) specified for each wavelength as well as the averaging effect evident in Equation (3.8). While the validity of the radiosity simulations will depend on each of the assumptions, primarily the specification of Lambertian surface elements. Yet, at the scale of mutual shadowing, this may be a legitimate assumption for capturing the the relative effects of topographic and forest structure variation.

The final step in applying the radiosity model involves associating the NBRFs calculated for discrete slope and azimuth classes with the slope and azimuths of the terrain being imaged. Estimates of the latter can be derived from Digital Elevation Models (DEMs). Actual terrain slopes can take on values from 0 to 90° and azimuths from 0 to 360°. NBRFs for slopes and azimuths not directly estimated using the radiosity model need to be interpolated from the NBRFs estimated for the 130 slope-azimuth combinations. Following Li (1997), a quartic polynomial surface was fit to the 130 points. Thus, a polynomial surface was estimated for every forest structure-wavelength-sun position combination, allowing the corresponding NBRFs

to be computed for a continuous range of terrain slopes and azimuths.

3.6 Summary

This chapter reviewed the basic concepts of radiometry and the primary factors contributing the the upwelling reflected radiance upwelling leaving the earth's surface. Measurements of reflected radiance are scale dependent. At several scales, field measurements of reflected radiance leaving forest canopies show strong anisotropy attributable to the mutual shadowing between tree crowns. Topography varies the projected area of the solar irradiance and induces different different BRDFs for a given forest structure class. Canopy reflectance models are required to represent these interacting effects. In order to classify forest structure on the basis of a normalized terrain orientation a model of the upwelling reflected radiance field has been proposed that considers mean class reflectance from a normalized surface orientation relative to actual surface orientations. A radiosity method was employed to estimate the change in forest reflectance between different structures for all possible terrain orientations. The results of this procedure are presented in Section 7.4. Chapter 4 reviews the measurement of the upwelling reflected radiance field from an orbiting satellite.

Chapter 4

RADIANCE DATA ACQUISITION

Satellite remote sensors record discrete samples of the continuous upwelling reflected radiance field. The reflected radiance reaching the sensor is modulated first by the atmosphere and again by the detection processes. The spatial frequency response of the sensor determines the fidelity of the recorded image to the actual radiance field. The total effective modulation transfer determines the effective scale of measurement and has implications for statistical image classification techniques. This chapter first reviews image acquisition primarily in terms of the design and operation of the Landsat TM sensor, although some comparisons to the MSS sensor are presented. The second section gives a detailed description of the system modulation transfer and point spread functions, and the effective spatial resolution of image data. The third section reviews previous research to evaluate and model dependent image data. Section 4.4 presents a likelihood function for an observed image given the upwelling reflected radiance field by modelling the system point spread function (PSF). Finally, the chapter reviews the techniques and results of previous research to quantify effective spatial resolution.

4.1 Review of data acquisition process

After leaving the earth's surface into the direction of the sensor, the upwelling radiance is attenuated and scattered by the atmosphere. The reflected radiance incident upon the sensor is projected as irradiance which is resolved spatially, spectrally, and radiometrically. The resulting image is recorded transmitted to Earth as an array of

discrete samples. For each pixel the sampled signal is recorded as a dimensionless digital number (DN). The following equation describes this process

$$\begin{aligned} \mathrm{DN}_i(\lambda) \leftarrow\ & g(\delta(i,\lambda)) \int_{\underline{\lambda}}^{\bar{\lambda}} R(\delta(i,\lambda)) \left\{ \left[\left(\int_{A_{\delta(i,\lambda)}} \int_{\omega_r} dL_r(\omega_z, \omega_r; A, B; \lambda) \right.\right.\right. \\ & \left. \cos\theta_r d\omega_r\, dA_{\delta(i,\lambda)}\, e^{-\tau^{\uparrow}(\lambda)\sec\theta_r} \right) \\ & \left.\left.\left. + \int_{A_{\delta(i,\lambda)}} \int_{\omega_r} dL_p(\omega_z, \omega_r; \lambda) d\omega_r\, dA_{\delta(i,\lambda)} \right] \otimes h(\lambda) \right\} d\lambda \\ & + o(\delta(i,\lambda)) \quad \text{[dimensionless]}, \end{aligned} \tag{4.1}$$

where

DN_i = the digital number for pixel i and bandpass having mean wavelength λ,

g = the per volt gain factor for the sample from detector δ, $[\mathrm{V}^{-1}]$,

$\delta(i,\lambda))$ = the detector sensing pixel i and bandpass having mean wavelength λ,

$\underline{\lambda}, \bar{\lambda}$ = zero power cutoffs for the bandpass having mean wavelength λ,

R = a detector response function,$[\mathrm{VW}^{-1}]$,

$\int_{\omega_r} d\omega_r$ = integral over a given solid angle in the direction of ω_r [sr],

$\int_{A_{\delta(i,\lambda)}} dA_{\delta(i,\lambda)}$ = integral over the area of detector $\delta(i,\lambda)$ $[\mathrm{m}^2]$,

$\tau^{\uparrow}$ = atmospheric extinction coefficient for the reflected radiance, dimensionless,

θ_r = declination angle from the surface plane normal to the sensor, degrees,

dL_p = path radiance, solar irradiance reflected by the atmosphere into a given direction, $[\mathrm{Wm}^{-2}\mathrm{sr}^{-1}]$,

$h(\lambda)$ = the system point spread function, or impulse response function,

$\otimes$ = the convolution operation,

o = the offset for the sample from detector δ [digital number].

4.1.1 Irradiance at the sensor

The radiance reflected into the direction of the sensor that is contained within a finite solid angle, $\int_{\omega_r} d\omega_r$, is projected onto the focal plane of the sensor by a factor of $\cos\theta_r d\omega_r$. The projection of reflected radiance onto the focal plane becomes the irradiance incident upon the sensor in [Wm^{-2}]. For a sensor viewing from nadir, $\theta_r = 0°$.

Section 3.1.2 described the attenuation of direct solar radiance by the atmosphere. Reflected radiance is also attenuated as it passes through the atmosphere, but not necessarily to the same extent due to the variability of the atmosphere. The coefficient $\tau^{\uparrow}$ in Equation (4.1) distinguishes the reduction in the upwelling reflected radiance field from that of downwelling solar radiance in Equation (3.2). The temporal and spatial variation in the level of atmospheric absorption preclude comparison of absolute radiance measurements acquired at different times and locations.

4.1.2 Path radiance

A portion of the radiant flux is scattered directly into the direction of the sensor by the atmosphere. This portion of the radiance incident upon the sensor L_p is commonly called path radiance. Note that the bidirectional scattering of solar irradiance by the atmosphere is implied, but not explicitly shown, by Equation (4.1).

Rayleigh scattering is the dominant atmospheric effect at shorter wavelengths ($< 0.7\mu$m). The sky appears blue because the air molecules in the atmosphere are highly reflective in the visible blue range of the EM spectrum. Larger particles, such as smoke and haze, scatter irradiance across a broader spectrum, referred to a Mie scattering (Richards 1986). At longer wavelengths ($> 1.0\mu$m) atmospheric absorption

becomes the dominant effect. As a result path radiance is most apparent in the visible spectrum.

Because these photons never reach the earth, path radiance contains no information about the surface being sensed. Unlike the complex interactions between the atmosphere and direct solar and reflected radiance, path radiance is additive. As with other atmospheric effects, the degree of path radiance will vary from image to image and within images, but once estimated, the effect of path radiance can be removed from the recorded radiance by subtracting its proportion.

4.1.3 Nominal spatial resolution

The optical system of a passive remote sensor resembles a camera with a telescopic lens. Detectors mounted on the focal plane receive radiance through a telescope with a given aperture and focal length. At any instant of time the angular field of view apparent to a detector is given by

$$\text{IFOV} = \frac{D}{f} \quad [\text{rad}], \tag{4.2}$$

where IFOV refers to the instantaneous field of view in radians, D is the side dimension of a square detector, and f gives the focal length of the telescope. The corresponding surface area visible at any instant is referred to as the linear, or ground-projected, instantaneous field of view, written as

$$\text{IFOV} = \frac{H_t D}{f} \quad [\text{m}], \tag{4.3}$$

where H_t is the altitude of the lens. Note that angular and linear IFOVs are distinguished by their units of measurement. The linear IFOV is the length of a side of a square area on the surface being sensed. The linear IFOV determines the nominal spatial resolution of the sensor and is frequently cited as the TM spatial resolution. The TM primary focal plane detectors measure 0.01036 cm and the nominal telescope focal length measures 243.8 cm (Markham 1985). Thus, a nominal spacecraft altitude

of 705.3 km gives TM a nominal resolution of 30 m. The dimensions of the primary focal plane projections of the cooled focal plane detectors (bands 5, 6, and 7) are slightly larger, though they are sampled at the same rate as the prime focal plane detectors in the in-track direction (Markham 1985). See Section 4.1.4 for a description of the focal plane layouts.

While f and D are likely to remain constant over the life of the sensor, the nominal resolution varies with the altitude of the spacecraft. For the first three MSSs flown, up to four percent variation in the nominal resolution as attributable to changes in spacecraft altitude attributable to the non-circularity of the orbital path and the non-sphericity of the earth (Slater 1979). Thus, the nominal TM altitude of 705.3 km corresponds to 712.5 km at 40 degrees north latitude, for which the TM scanning parameters are optimized (Engel and Weinstein 1983). The actual value of H_t will vary with topography for any given IFOV sampled.

4.1.4 Sampling patterns

The TM and MSS sensors carried on the Landsat satellites are mechanical scanning devices that acquire data by scanning the surface in strips perpendicular to the motion of the satellite. An oscillating mirror scans in the across-track direction thereby reflecting irradiance into the telescope. For the TM, the telescope directs the radiant flux to a scan line corrector which transfers the flux onto the primary focal plane. Relay optics further direct the flux onto a secondary, cooled focal plane. The scan corrector is a rotating two mirror system that produces scans perpendicular to the satellite velocity vector, allowing the TM to scan in both directions unlike the MSS. Scanning in both directions increases the effective dwell time of the TM, allowing for increased spectral and spatial resolution(Schott 1997). The dwell time is further increased by simultaneously scanning 16 lines per bandpass. In contrast the MSS detectors (6 per bandpass) are not located on the focal plane. Instead fiber optic bundles conduct light to photo-multiplier tubes and photo-diodes (Wrigley et al. 1984).

The TM uses photon detectors for each of the reflective bands. The photons from the incident radiant flux [W] interact with the detector material to produce electrical signals, measured in volts [V]. The response function, R, expresses this interaction in $[VW^{-1}]$. A channel refers to the signal from an individual detector. The photo-voltaic detectors for bands 1-4 are made of Silicon and are located on the primary focal plane. Bands 5 and 7 are sensed by indium antimonide photo-conductive detectors located on the cooled focal plane.

A single TM image contains as many as 200 million radiance samples sensed by 100 detectors (including the 4 thermal detectors) and takes over 22 seconds to produce. Gordon (1983) provides an excellent description of scanning device sampling patterns highlighted here. For each reflective band there is a staggered array of 16 detectors each sensing one line per scan. The detectors are staggered to accommodate their size and their amplifiers size, which are in turn determined by resolution and signal-to-noise ratio specifications. Thus, for each scan, 16 lines are sensed for each band, although adjacent lines are not sensed coincidentally. There is a 2 to 3 IFOV locational translation between odd and even lines. The combined effect of scan motion and focal plane layout results in a 2 IFOV translation in the forward scan and 3 in the reverse scan. Likewise, given there is a separate array for each band the bands are not inherently co-registered. In fact at any given time the band 1 detectors will be viewing a surface area 4380 m from the area viewed by the band 5 detectors, solely attributable to the focal plane layout. Further the reverse scan is 180° out of phase from the forward scan.

Once downloaded to ground receiving stations the raw image data must be aligned or "geometrically corrected" to account for these displacements. The reverse scan data must be effectively flipped and all other displacements be corrected based on knowledge of the pertinent rates, and the focal play layout. However, unanticipated variation may preclude the same IFOVs from being viewed by each band. Note that geometric correction is not the same process as registration to the coordinate system

of an earth surface projection.

The rate of sampling determines the linear sampling interval (pixel size) in the across-track direction. The TM samples at a rate such that the ground-projected pixel area corresponds to the dimensions of the ground-projected IFOV. This is described as a 1:1 aspect ratio in the direction of the scan. The MSS is designed to have a 1:1.4 aspect ratio. This oversampling is intended to reduce *aliasing*, a form of sampling induced degradation described in Section 4.2.4. Note that with a 1:1.4 aspect ratio, the size of an MSS sample does not correspond to the MSS IFOV (57 and 76 m, respectively, in the across track direction).

Several factors may induce unintentional oversampling or undersampling. Variations in the mirror scan velocity and spacecraft vibrations may cause displacements of up to 8 TM IFOV in the across-track direction. Variation in spacecraft altitude and vibrations may also induce up to one IFOV oversampling between scans in the in-track direction (Fischel 1984). The net effect of oversampling is that the adjacent pixels are not unique (Duggin and Robinove 1990).

Evaluation of the first TM on Landsat-4 (TM-4) data found that the band-to-band registrations between bands of a common focal plane were all less than the 0.2 pixel specification (Anuta et al. 1984; Wrigley et al. 1984). For band-to-band registration between focal planes the specification is 0.3 pixel, but Anuta et al. (1984) found misregistrations up to 0.6 pixel and Wrigley et al. (1984) found a consistent mean misregistration of 0.5 pixel in the across-track direction and up to 0.3 in the in-track direction. Malaret et al. (1985) used a slightly different evaluation technique than Anuta et al. (1984) and concluded that between focal plane registration met with specifications, but that band 4 was misregistered with band 1 and 3 by 0.38 pixel. Bryant et al. (1985) found between focal plane misregistration ranged from 0.75 to 1.25 pixels in one direction, but was within specifications for bands of a common focal plane.

Evaluation of TM-5 band-to-band registration has been determined to be better

than for TM-4 images and well within specifications for within and between the bands of each focal plane (Malaret et al. 1985). Bryant et al. (1985) confirmed within focal plane intraband registration to be very good, but again noted a consistent misregistration in one direction between bands 1 and 5 usually less than 0.35 pixel, evidence of a systematic misregistration between those bands. Bryant et al. (1985) also determined that individual bands were properly aligned between forward and reverse scans.

Although the Landsat sensors nominally view from nadir, the view angle does not remain constant over the 185 km swath width. As a result the irradiance incident upon the sensor will be reduced by the projection factor $\cos\theta_r d\omega_r$. Although it is difficult to quantify this effect due to scene and atmospheric variability over a 185 km distance. Anuta et al. (1984) suggest it may be as great as 5 DN.

4.1.5 Spectral resolution

The spectral resolution of a remote sensor is defined by the number of wavelengths, or bands, it senses and the width of each bandpass. The TM nominally records reflected radiance in three visible and three infrared bands, shown in the first two columns of Table 4.1. Slater (1980) provides the rationale for the nominal bandpass selections. Band 1 spans a region between significant atmospheric attenuation and the upper limit of blue chlorophyll absorption, thought to provide better discrimination of conifers. Band 2 covers the length of green reflectance of healthy vegetation. Band 3 exhibits the lowest atmospheric haze of the visible spectrum. This spans the region of red chlorophyll absorption and is critical for separating vegetation from soil, rock, and other non-vegetated surface covers. Band 4 in the near infrared corresponds to the peak reflectance of vegetation. Ratios of bands 2 and 4 vary with the amount of green biomass and vegetation moisture. In band 5, vegetation reflectance varies with moisture content, important for crop drought detection. Band 7 was primarily included for geological purposes. Finally, band 6 records thermal radiation at a much

larger interval.

Table 4.1: Nominal and effective band edges of the Landsat-5 TM in nanometers.

Band	Nominal	50 % NPR	Moments Method
1	450-520	452.4-517.8	451.3-521.4
2	520-600	528.0-609.3	526.2-615.0
3	630-690	626.4-693.2	622.6-698.8
4	760-900	776.4-904.5	771.0-905.3
5	1550-1750	1567.5-1784.2	1564.0-1790.0
7	2080-2350	2097.2-2349.0	2083.0-2351.0

The nominal spectral resolution is defined by the zero power wavelengths, $\underline{\lambda}, \bar{\lambda}$, of the optical filters bounding the primary response region for each bandpass. The TM uses one filter for all 16 detectors in each band, whereas there is one filter per detector for the MSS. Several normalization techniques have been used for calculating effective bandwidths (Palmer 1984). Markham and Barker (1985) considered band edges at 50 percent of the normalized peak response. Palmer (1984) developed a more elaborate technique based on a moments analysis of the spectral responsivity curve. These calculations are also presented in Table 4.1.

The response function, $R(\lambda)$, varies with wavelength for each detector within the primary response region. Thus, $R(\lambda)$ is integrated over the bandpass with respect to wavelength (see Equation (4.1). As a result two different spectral reflectance patterns may yield the same values over a given bandpass. In fact all of the primary factors determining the amount of reflected radiance reaching the sensor are wavelength dependent. For this reason spectral resolution is inversely related to bandwidth. Table 4.1 shows that the effective bandpasses generally resemble the nominal specifications. However, the upper edge of band 5 significantly deviates into the atmospheric

water absorption region, increasing the sensitivity of the band to atmospheric water content variability (Markham and Barker 1985). The band 1 and 3 filters have also been found to transmit small amounts of flux in the NIR (Markham and Barker 1985).

Spectral resolution is further reduced by detector-to-detector (striping) and scan-to-scan (banding) calibration errors. The detectors are designed to have linear response functions. Ideally, each detector would have the same response function such that the same electrical signal was output in response to a given flux input for each detector. In practice the slope and offsets may vary slightly from detector to detector resulting in an image striping effect. For the TM flown on Landsat-5, Markham and Barker (1985) reported the only significant striping problems were associated with two of the band 4 channels. Malaret et al. (1985) reported striping in bands 4, 7, and most evident in one band 5 channel having a standard deviation of 0.8 DN. Slater (1979) demonstrates how different response curves between two detectors of the same band can yield different data values for the same surface reflectance pattern. Variation in detector response precludes direct comparison of the output from the two TM sensors on Landsats-4 and -5 (Metzler and Malila 1985). Finally, there is the possibility for the response characteristics of sensor detectors to degrade over time.

Banding occurs between successive scans. All of the detectors for each band are calibrated against on-board references as the start of each scan. Bias is introduced as the detectors receive flux across the scan. Since the TM scans in both directions the direction of bias drift will alternate directions with each of the 16 lines in each scan. Malaret et al. (1985) reported variation of up to one DN attributable to banding for band 4 of the Landsat-5 TM. For Landsat-5 most of the striping and banding errors are within tolerable limits and system specifications (Markham and Barker 1985).

A portion of the sampled voltage is attributable to coherent, or periodic, noise. Several studies of the TM-4 have reported noise frequencies of 0.055, 0.07, and 0.31 cycles/sample for the primary focal plane bands (Anuta et al. 1984; Wrigley et al. 1984). The term *cycle* is described in Section 4.2.1. The 0.31 frequency is the strongest

and may be derived from electronic coupling with the Payload Correction Data as their frequencies correspond (Wrigley et al. 1984). Analysis of Landsat-5 TM data does not show evidence of the 0.31 frequency, but otherwise shows noise frequencies corresponding to Landsat-4 TM frequencies or harmonics of them, suggesting they may be from the same source (Malaret et al. 1985; Wrigley et al. 1985). The source of this noise was not identified, but electronic and mechanical oscillations were suspected. Although the periodic noise identified represents a small portion of the total noise, it is significant enough to degrade the quality of low contrast areas (Wrigley et al. 1985), reducing class separability within those areas.

IN summary, variability between detector response functions, filter transmission, and the averaging of spectral reflectance over the width of a bandpass gives rise to variation in the data samples of reflected radiance recorded for even a single forest structure on a normalized surface.

4.1.6 Radiometric resolution

The detectors transform radiant flux into a continuously varying electrical signal. These low level electric signals are amplified by current mode pre-amplifiers and frequency boosting post-amplifiers (Engel and Weinstein 1983). After amplification the signals are low-pass filtered to reduce aliasing (see Section 4.2.4). In order to store and transmit data as digital images the amplified analogue signal must be sampled in time to produce discrete records across the scan line. Periodic samples of the output voltage are recorded as DNs. This is known as analogue-to-digital (A/D) conversion. The signal from each set of odd/even detectors for each band on the TM focal plane is A/D converted by an individual submultiplexer. The submultiplexer outputs are then encoded by the main multiplexer (Gordon 1983) to form an 85 M bit/second data stream (Engel and Weinstein 1983).

The dynamic range of the data is given by the number of bits used for quantization. This determines the radiometric resolution of the sensor. The dynamic range of the

TM multiplexers is 8 bits, allowing for DNs ranging from zero to 255. In contrast the dynamic range of the MSS is 6 bits giving DN values from 0 to only 63. The TM offers higher radiometric precision than the MSS.

4.1.7 Resolution trade-offs

Occasionally, remote sensing studies conclude that classification accuracy can be improved by increasing the spatial resolution of the data. This is not necessarily true and the perception may stem from the advent of digital data which are easily manipulated by those with no photogrammetric training unaware of other factors determining image quality (Forshaw et al. 1983). For example, Croft and Kessler (1996) clearly imply that data quality improves with greater spatial resolution. Several studies have shown that increasing spatial resolution cannot be expected to improve results for a given objective. In theory, there is no unique spatial resolution for discriminating all classes in a complex forest environment (Marceau et al. 1994) and the optimal resolution for each class may differ within a given classification scheme (Marceau et al. 1994). Indeed, Christensen et al. (1988) used 5.6 m airborne MSS data to map South Carolina wetlands but still encountered considerable classification problems. Gjertsen (1993) found that SPOT's 20 m resolution was not significantly better than the 30 m TM resolution for mapping tree size and density in Norway. For classifying forests in Alberta, Franklin (1994) achieved the same level of accuracy using both a 2.5 m aerial sensor (CASI) and 30 m TM data. As spatial resolution increases, classification problems at one scale are exchanged for classification problems at another scale!

At issue are the trade-offs between spatial, spectral, and radiometric resolution. Increased spatial resolution alone, without corresponding improvements in spectral and radiometric resolution cannot be expected to improve classification results. As an extreme example, a single bandpass image having one cm spatial resolution but only one bit radiometric resolution will only allow 2 classes to be extracted. If the two classes are forest and water, this will not be a problem. However, one cm spatial

resolution is rather unnecessary for mapping forest and water. If the objective is to separate various tree species, then one bit resolution is clearly insufficient. On the other hand there is a danger in having resolution that is too high. Many sources of useful and unwanted image variation have been described here. While high resolution data may capture subtle differences in the spectral reflectance properties of two tree species, the same data will be more sensitive to error induced variation such as gain differences between detectors. Again, the one bit data necessary for two classes will not likely resolve gain errors, but will not resolve more than two classes either.

Increasing sensor resolution is expensive and tends to reduce the signal-to-noise ratio (SNR). The desired detector output (signal) corresponding to the flux incident upon the detector is further subject to random variations attributable to thermal and thermal exchange variation, random photon events, and other random events. These variations about the mean signal level are referred to as noise, also measured in volts (Schott 1997). The SNR indicates the level of noise relative to the level of signal. This incoherent noise is also characterized as noise equivalent reflectance, NER(λ), the percent change in reflectance equivalent to the level of noise. Table 4.2 gives the TM specified NER(λ) for each band, and the measured and specified SNRs for each band (Engel and Weinstein 1983).

Equation (4.3) shows that the linear IFOV can be narrowed by reducing the altitude and/or detector size or by increasing the focal length. The reduced TM IFOV relative to the MSS IFOV is partially attributable to the lower orbiting altitude of Landsat-4 and -5. However, this is feasible only up to a point. Decreasing detector size and increasing focal length will each lower the SNR. This can be offset by increasing bandwidth, but that will reduce spectral resolution. Improving the modulation transfer characteristics of the system components will increase the signal without sacrificing resolution but increases costs. See Slater (1980) for a more detailed discussion of SNR trade-offs.

Ultimately the number and definition of classes should be determined by the map-

Table 4.2: TM NER(λ) and SNR specifications and SNR measurements.

Band	NER(λ)	Specified SNR		Measured SNR	
		Min	Max	Min	Max
1	0.8	32	85	52	143
2	0.5	35	170	60	279
3	0.5	26	143	48	248
4	0.5	32	240	35	342
5	1.0	13	75	40	194
7	2.4	5	45	21	164

ping objective and the necessary sensor resolutions specified accordingly. In practice there is often little opportunity to specify sensor resolutions and the accuracy of Landsat image classifications depends upon the appropriateness of the designated classes relative to the sensor resolutions. Thus, higher accuracies are usually obtained by reducing the number of classes (Felix and Binney 1989; Itten and Meyer 1993). Unfortunately there is no exact formula for determining the classes resolvable by a given sensor.

4.2 Effective spatial resolution

For several reasons the actual reflected radiance sampled at any time originates from an area greater than the ground-projected IFOV. This larger area becomes the effective spatial resolution of the remote sensor. Whereas the IFOV is derived only from the geometry of the optical system, the effective ground-projected area resolved is greater due to factors such as atmospheric scattering and the motion of the satellite. The effective spatial resolution of an imaging system is best characterized by the modulation transfer function (MTF) of the system in the spatial frequency domain

and by its two-dimensional Fourier inverse, the system point spread function (PSF) in the spatial domain. A line spread function (LSF) is a one-dimensional cross section of a PSF. Likewise modulation transfer can be characterized in one-dimension.

4.2.1 Modulation transfer

An MTF can be used to describe how a system modifies, or modulates, the amplitude of an input frequency by assuming a linear and shift-invariant system (Markham 1985). Phase shifts, described by the phase transfer function, are frequently ignored (Forshaw et al. 1983). For a sinusoidal distribution of a certain frequency of a light wave the wavelength is the distance between the wave crests, representing one complete cycle (Slater 1980). When the sinusoidal wave is plotted as power against distance, the wave modulation is defined as the difference between the maximum and minimum powers divided by the sum of the maximum and minimum powers. The MTF is then the spatial frequency response curve for an optical system defined as the ratio of the modulation of the output wave to the modulation of its input wave over a range of spatial frequencies in units of cycles per unit distance or angle. In other words, the MTF of a component or the whole system describes how well the sinusoidal distribution of radiance input to the system is reproduced. High MTF values represent high fidelity. A perfectly reproduced image would have an MTF of unity at all spatial frequencies (Schott 1997). In general the curve falls off at higher frequencies as shown in Figure 4.1, because high frequencies represent high spatial variation in the distribution of radiance which is harder to reproduce than low frequencies.

Alternatively, the spatial response of an imaging system can be characterized in the spatial domain by the PSF, which also assumes a linear, shift-invariant system. The PSF defines the radiance recorded on focal plane for a point source at infinity. On the basis of the linearity assumption the system is reversible, thus the zero power boundaries of the rear-projected PSF define the region that produces a point image on the focal plane (Duggin and Robinove 1990). The MTF is the Fourier transform

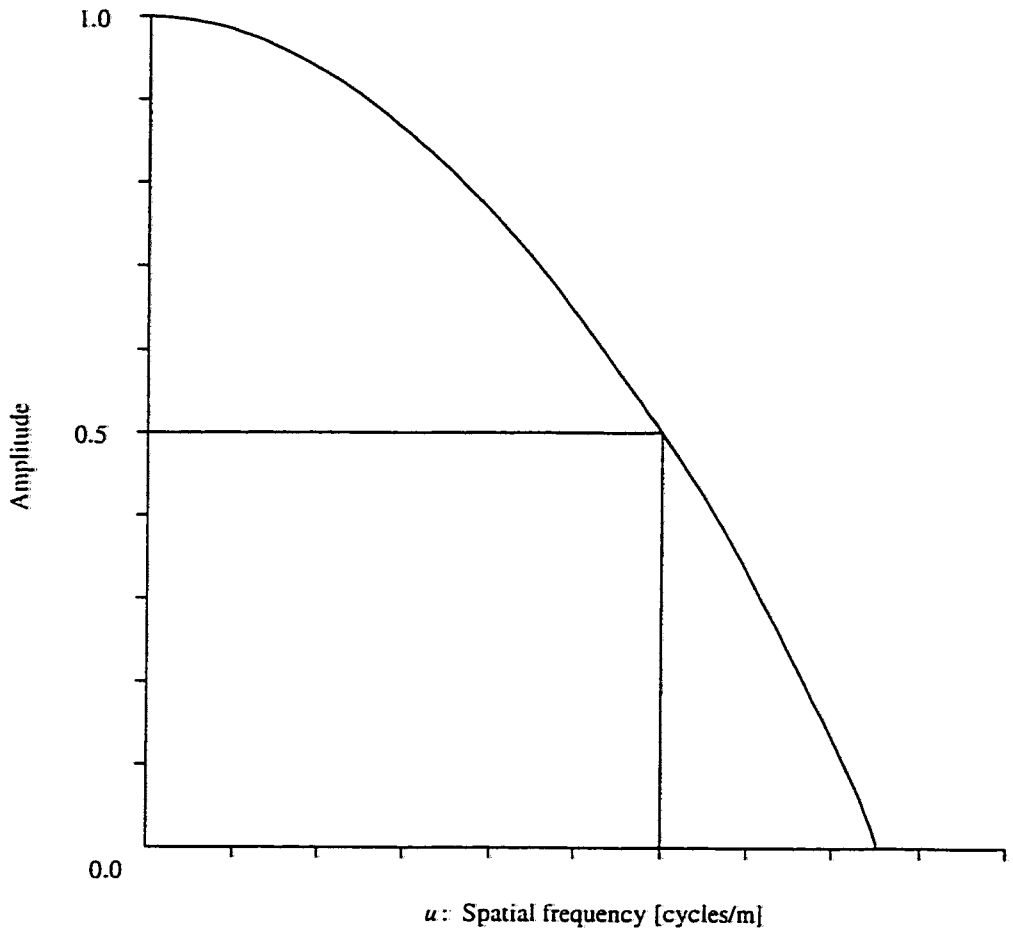


Figure 4.1: A hypothetical modulation transfer function showing the spatial frequency corresponding to $EIFOV_{0.5}$.

of the PSF. A recorded image is produced by a two-dimensional convolution of the upwelling radiance field with the rear-projected PSF, written as

$$y(a,b) = o(a,b) \otimes \mathrm{PSF}(a,b), \tag{4.4}$$

where y is the image recorded by the scanning system, o is the object of interest, in this case spatially varying patterns of radiance, and (a,b) are orthogonal coordinate axes.

The convolution operation within the spatial domain transforms to multiplication in the spatial frequency domain. Thus, the MTFs of the system components can be multiplied, or cascaded, together to give the MTF for the whole system. The following equation describes the system PSF, PSF_{sys}, in terms of the MTFs for each of the primary attenuating factors in an optical remote sensing system

$$\mathrm{PSF_{sys}} = \mathcal{F}^{-1}(\mathrm{MTF_{atm}} \cdot \mathrm{MTF_o} \cdot \mathrm{MTF_d} \cdot \mathrm{MTF_e} \cdot \mathrm{MTF_{samp}} \cdot \mathrm{MTF_{resamp}} \cdot \mathrm{MTF_{reg}}), \tag{4.5}$$

where the subscripts atm, o, d, e, samp, resamp, and reg denote MTFs corresponding to atmospheric transmission, the optical system, the detectors, the electronics, the sampling pattern, geometric resampling, and registration, respectively. The system MTF is the product of these individual MTFs which are discussed below, following a presentation of response metrics.

4.2.2 Spatial response metrics

On basis of these response functions either of two comparable metrics within the spatial and frequency domains are often used to quantify the effective spatial resolution of a recorded image. The effective IFOV is defined as the field of view corresponding the the spatial frequency at which the output radiance distribution is one-half the input distribution, in other words the spatial resolution corresponding to one-half the

MTF (see Figure 4.1). This can be written as

$$\text{EIFOV} = \frac{H_t}{2uf} \quad [\text{m}], \tag{4.6}$$

where u is the spatial frequency at which the MTF equals one-half. Occasionally, this is written as $\text{EIFOV}_{0.5}$ to distinguish of other possible definitions of EIFOV.

Effective resolution in the spatial domain is measured as the ground projected half amplitude width ($W_{\frac{1}{2}}$), or full width at half maximum (half-width), of the PSF. In other words, the ground-projected region for which the magnitude of the PSF is greater than half its maximum value. Figure 4.2 illustrates the effective resolution given by $W_{\frac{1}{2}}$ for a cross section of a PSF. Half-width estimates for TM data are presented in Section 4.5. Other metrics have been proposed which characterize effective resolution for the purpose of target detection (see, for example, (Forshaw et al. 1983)), but these are less meaningful for image classification.

4.2.3 Atmospheric scattering

The atmosphere both attenuates and scatters the upwelling reflected radiance field. The scattering may cause the radiance leaving a given ground-projected IFOV to be reflected into another direction and thus be recorded as though leaving from another surface location. Kaufman (1985) describes this as the adjacency effect, which is primarily attributable to atmospheric turbidity (Schott 1997). Turbidity is the ratio of the total normal optical thickness of the atmosphere to the scattering optical thickness of the atmospheric gases for a given wavelength (Fraser et al. 1977).

The adjacency effect attenuates the amplitude of the radiance reaching the sensor thereby further reducing the system MTF beyond the combined effect of the other components. In contrast to the MTFs associated with the optics and detectors the atmospheric MTF is highly variable. By reducing the system MTF, atmospheric scattering reduces the SNR in a way that is not easily overcome. Atmospheric scattering also reduces radiometric resolution by decreasing the range of radiance values incident

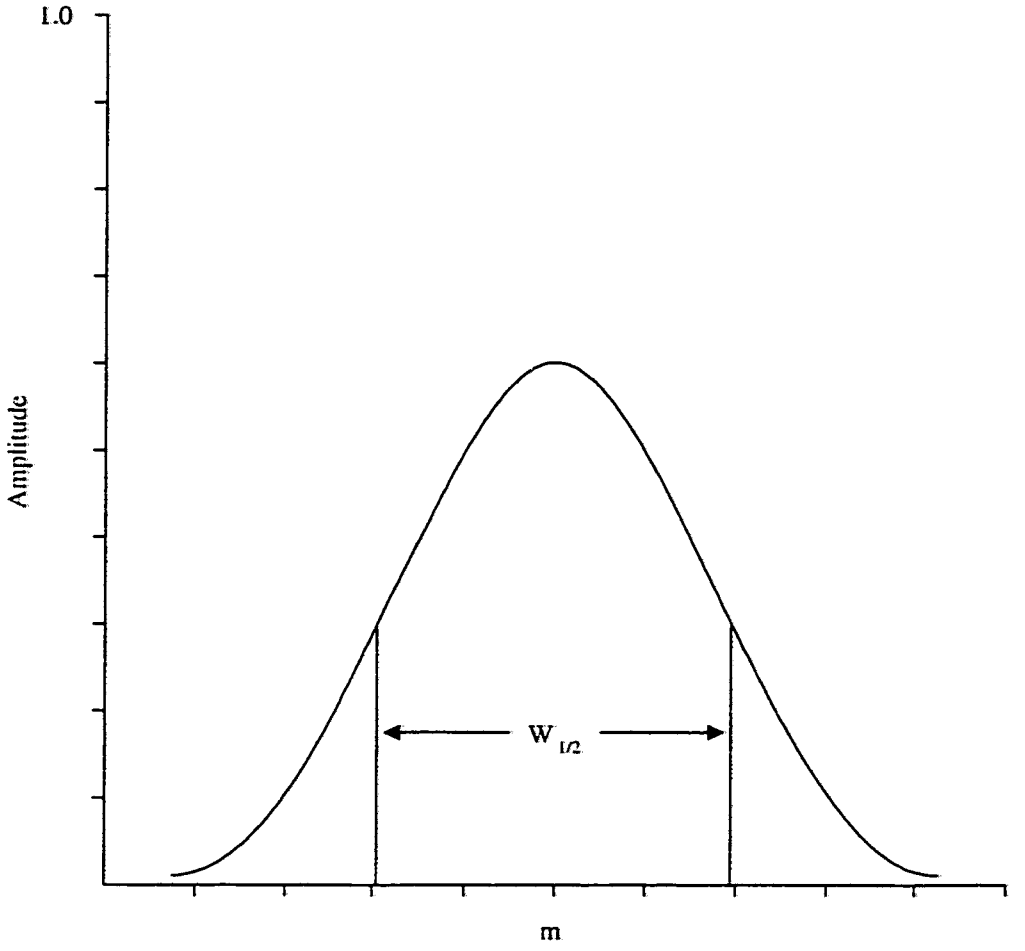


Figure 4.2: Point spread function showing the half amplitude width.

upon the sensor (Kaufman 1985).

4.2.4 Sensor component and sampling MTFs

The TM telescope and scan mirror, forming the optical system, each have associated MTFs. Few data have been collected on these MTFs, although optical scan and track LSFs were derived from combined optical and detector LSF measurements made by the TM manufacturer, Hughes Aircraft (Markham 1985). The TM was nominally designed so that the optical system PSF would be less than 10 percent of the detector size. Thus the combined sensor components MTF is dominated by the detector size (Slater 1980). The MTFs of the detector and sampling interval can be calculated from their design measurements using an ideal reconstruction filter, the sinc function (Slater 1980; Markham 1985). The MTF for frequency u is given by the modulus of the sinc function, written in one-dimension as

$$\mathrm{MTF}(u) = \left|\frac{\sin \pi D u}{\pi D u}\right| = \left|\mathrm{sinc}\, \pi D u\right| . \tag{4.7}$$

Given the 1:1 aspect ratio, the MTF will be the same for both detector and sampling interval. The sinc function is the Fourier transform of a rectangular (rect) function in the spatial domain (Schott 1997). Also note that the EIFOV of a square detector will be greater by a factor of $\sqrt{2}$ at 45° to the across- and in-track directions (Slater 1980).

Detailed frequency response measurements for the detectors and electronics assemblies were conducted by Hughes Aircraft and summarized by Markham (1985). The electronics response has a negligible effect in the in-track direction as the effective in-track scanning velocity is much slower than the mirror scanning velocity. The temporal frequency modulation of the electronics readout translates to a spatial frequency modulation (Schott 1997). For the TM, pre- and post-launch presample measurements have shown the EIFOV to be 2 to 6 m larger than the IFOV (Markham 1985).

Sampling by the TM submultiplexers requires further modulation transfer. According to signal processing sampling theory the highest spatial frequency that can be reproduced by an ideal sampling system is defined by the Nyquist limit of the sampling process (Richards 1986; Schott 1997). The Nyquist frequency is defined as one-half the sampling rate. Sampling higher frequencies causes the segments of the sampled power spectrum to overlap. This form of degradation is known as aliasing. Ideally the sampling interval would correspond to a small fraction of the IFOV, but is usually limited in order to restrict the data rate (Forshaw et al. 1983). Edges and boundaries in terrestrial scenes tend to produce frequencies greater than the Nyquist limit. For the TM 1:1 aspect ratio corresponds to a limit of $\frac{1}{2}$ cycles/pixel or $\frac{1}{60}$ cycles/m. Thus, to reduce aliasing the signal is low-pass filtered, which in turn decreases spatial resolution. As mentioned in Section 4.1.4 the MSS was designed with a 1:1.4 aspect ratio (oversample), in addition to low-pass filtering, to decrease aliasing (Markham 1985). Since the sampling occurs in the scanning direction more aliasing will occur in that direction. The movement of the sensor during the integration time of sampling will also attenuate the signal. If the ground speed and integration time are known this MTF can be derived from a sinc function (Slater 1980).

4.2.5 Resampling

Once transmitted to the ground, the image data is resampled to correct for the staggered array effects, spectral band offsets, offsets between scan directions, and nominally corrected for the rotation of the earth during the exposure time (Schott 1997). The image data may need to be further resampled to an earth surface projection system by the end-user to allow multi-image or multi-data comparisons. Cubic convolution and nearest neighbor resampling methods are most common. The former provides more spatial accuracy, but reduces the radiometric and spatial resolution of the data by averaging the DN values. Wrigley et al. (1984) derived an MTF curve for the cubic convolution resampling technique used for geometrically correcting TM

data showing the fall-off attributable to this technique. The nearest neighbor method preserves the recorded data values at the expense of positional accuracy up to a factor of $\frac{\sqrt{2}}{2}$.

4.2.6 Spectral mixture implications

Within the area of the TM IFOV one or more forest types being classified may occur. In the context of probabilistic classification such samples will be classified with less certainty. Problems arise when the radiance from a *mixed* pixel resembles the "pure" radiance from yet another cover type being classified. Although contextual information may reduce this problem, mixed pixels become more likely with an increasing EIFOV. As an alternative to probabilistically classifying a site as one class or another, Kent and Mardia (1988) developed an MRF model to classify the proportion of each class in each site.

It would be even more likely to find a mixture of a single forest class and another surface cover not described by any class such as a gravel bar along a stream. Even for an MTF of unity, the reflected radiance from such an IFOV would represent and average of the surface covers sensed. This is the basis for spectral mixture analysis (Adams et al. 1989). Each pixel is assumed represent a mixture of spectrally unique surface materials such as soil and photosynthetic vegetation.

The mixed pixel effect resembles that of atmospheric scattering, thus the effective resolution can depend upon the size of the areas being classified (Kaufman 1984). Campbell (1987) points out that decreased resolution may increase the proportion of mixed pixels. This is more likely to occur when the scale corresponding to the sensor resolution is not appropriate for the scale of classes. For example, for a very small IFOV mixtures of leaves and branches may complicate the classification of forest and pasture.

Another effect of the system PSF is that the signal recorded will be more sensitive to the central portion of the IFOV (Duggin 1985). Thus, the average signal for a

mixture of a road and surrounding forest will vary with the location of the road within the IFOV. Reduced effective spatial resolution may be more of a problem for target detection than classification. Wu and Schowengerdt (1993) consider the effect of the system PSF on linear mixture models.

4.3 Statistical implications of the effective spatial resolution

Landsat images are systematic samples of the upwelling reflected radiance field. Because the effective spatial resolution is greater than the sample size, the samples "overlap." In other words, each pixel represents a sample of radiance from an area greater than the area of the ground-projected pixel size. Therefore, the samples are not unique. In statistical terms the DNs are spatially correlated (Tubbs and Coberly 1978; Kittler 1983) and not independent. Further, they are not even independent given knowledge of the underlying surface cover. Spatial correlation is also attributable to the fact that areas having similar reflectance patterns are much larger than the ground-projected pixel areas.

4.3.1 Conditionally dependent image data

The ML classifier requires that image DNs be independent, conditioned on knowing the classes. This is the assumption of class-conditional independence. While Landsat image data clearly violates this assumption, ML and other image processing techniques are frequently used in the absence of better methods; see, for example, Harding and Scott (1978) and Green et al. (1994).

From the outset of the Landsat program, some attention has been given to the issue of spatially correlated image data, but a widely accepted alternative to the assumption of independence has yet to be developed. More than 20 years ago Basu and Odell (1974) demonstrated that using spatially correlated *training samples* (samples from a known land cover class) for ML classification would yield biased estimates

of class means and covariance matrices and increase the rate of misclassification. On the basis of this research several models of Landsat MSS data correlation were proposed (Basu and Odell 1974; Craig 1979; Tubbs 1980). Such low-order autoregressive and autoregressive moving-average models fit MSS data reasonably well (Kittler 1983). All of the early research on Landsat spatial correlation pertains to MSS data. The 1:1.4 scan line aspect ratio introduced 23 m of intentional oversampling.

Campbell (1981) directly evaluated the correlation of MSS images of forested areas in Virginia finding positive autocorrelation in all cases and also observed that the degree of continuity varied from date-to-date and band-to-band for a given site. Several authors have suggested using non-adjacent, randomly selected pixels as training data in order to overcome the effects of spatial correlation (Tubbs and Coberly 1978; Craig 1979; Campbell 1981). Labovitz (1986) also found that contiguous training samples would produce biased estimates due to autocorrelation in MSS data and also proposed using widely spaced training pixels as a solution to the problem. Bouman and Shapiro (1994) used a random selection of widely spaced clusters of training pixels. Chapter 5 examines parameter estimation in more detail.

While increased sample spacing can reduce the bias of estimated mean and covariance matrices, the conventional ML function is nevertheless based on the assumption of conditionally independent image data. Sensitivity analysis of the independence assumption implicit in the classifier developed by Haslett showed that high spatial autocorrelation reduced the relative improvement of a contextual classifier over a non-contextual classifier (Haslett 1985). Zhang et al. (1990) found their classifier performed better using simulated (independent) data than for the classification of actual MSS data. Likewise, the advantages of the SMAP estimator were diminished when classifying real SPOT image data as opposed to synthetic images (Bouman and Shapiro 1994).

4.3.2 Contextual classifiers modelling spatial autocorrelation

These results suggest the improvement of contextual classifiers over non-contextual classifiers is reduced by spatially autocorrelated image data. While specifically modelling the spatial continuity of the classification with an MRF approach, each of the above classifiers used a conditional density function based on the assumption of class conditional independence. Several studies have attempted to avoid this assumption by employing conditionally dependent likelihood functions in the context of Bayesian image analysis. See Switzer (1980) and Switzer (1983) for early and relevant proposals for modelling dependent data and mixed pixels. More recently, Schultz and Stevenson (1995) modelled multispectral image data as an MRF having both spatial and spectral cliques for the purpose of image restoration, but not classification.

Hjort and Mohn (1984) and Hjort et al. (1985) proposed modelling the image data as the sum of a normally distributed class-dependent signal and a contaminating, independent autocorrelated noise process. Mohn et al. (1987) presents results of this likelihood function in conjunction with several non-Markov contextual information models and in comparison to a non-contextual model. Using simulated data for several spatial patterns these results show that in general misclassification increased with increasing dependence in the noise term. However, the classifier explicitly modelling the dependent noise process showed fewer errors than all other methods for the spatial pattern resembling an actual land cover classification. Several methods of estimating the class means and covariances were employed and variances were modelled as class dependent. Khazenie and Crawford (1990) and Smreker and Khazenie (1990) extend this approach to classify multiple image sets correlated in time. The latter compared the non-Markov spatial classifier to an MRF approach and found similar results using simulated data, although the MRF model was more computationally expensive.

Markov random field models offer another method of modelling conditionally dependent image data. Kittler and Föglein (1984) used a local auto-normal scheme (Be-

sag 1974) to represent the data and contextual prior distribution for the class labels, this is also referred to as a Gaussian MRF. This model approximately corresponds to the Hjort et al. (1985) parameterized such that the data is modelled as the sum of the mean signal and the autocorrelated noise (Hjort et al. 1985). This technique was applied to MSS data, but only different estimation algorithms were compared. Kittler and Pairman (1985) and Kittler and Pairman (1989) modelled AVHRR data as a similar auto-normal noise process and also included a contextual prior for class labels. Parameter estimation for the prior was a particular problem however as they modelled specific cloud configurations, making it difficult to understand the effect of the auto model for the data. Schröder et al. (1998) used an auto-binomial MRF to model SAR and optical data.

Kiiveri and Campbell (1992) employed a Gaussian MRF as a conditional autoregressive model for the image data along with the Potts prior and evaluated it using simulated data. They made an effort to ensure unbiased estimates of class means and covariances. Their findings show there is little to be gained from modelling spatially correlated data and conclude that it is redundant to model both the correlation of the data and the spatial pattern of the classes. Owen (1984) suggests that he tried to model conditional dependence, but the accuracy was not improved. His published results assume conditional independence and he did not describe his method of modelling dependence. Wilson et al. (1994) proposed an elaborate model for the joint distribution of the classes and the "true" undegraded and unobserved reflectances to capture their spatial correlation. Finally, Conradsen and Nielsen (1987) conclude direct modelling of dependence would lead to a very complicated model with little practical value. Therefore, it remains undetermined as to whether or not explicitly modelling the dependence will improve the performance of contextual classifiers.

4.4 *Modelling the effective spatial resolution of image data*

A hierarchical method provides a method for explicitly modelling the image data as a convolution of the upwelling reflected radiance field. Consider each data sample, y_i, to be normal with a mean that is a weighted average of the z_j for pixels j in the proximity of pixel i. If the y_i are assumed conditionally independent, given z, then

$$\pi(y|z) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i - \sum_j H_{ij}z_j)'\,\Psi^{-1}(y_i - \sum_j H_{ij}z_j)\right\}, \tag{4.8}$$

where H is an array estimated PSF half-widths in terms of pixels j in the proximity of pixel i, and including i itself. Note that there is no requirement for the distribution of the j pixels to be symmetric about i. This research extends the model to multispectral images by specifying a wavelength dependent half-width. In contrast to the system response itself $W_{\frac{1}{2}}$ is finite, although its edges may not correspond to pixel edges. Further, the response is not uniform within the area of $W_{\frac{1}{2}}$. Thus, for each bandpass the elements of H will vary according to the proportion of the $W_{\frac{1}{2}}$ for pixel i that they represent. This amounts to assigning a greater weight to z_i than the other z_j (see Section 7.6). The common covariance matrix is Ψ implying additive white noise. Given that the effective resolution is the same for all forest classes, the assumption of a common covariance matrix here is reasonable. Again the assumption of normality is made for convenience, but not necessary (see Section 3.3).

By nature the process of classifying forest structure from Landsat data is one of inferring dependent random variables from dependent data samples. Modelling the PSF as a stage of a hierarchical model avoids the assumption that the y_i are conditionally independent given x and provides a convenient method for characterizing the physical process by which the data is acquired. Although the y_i are dependent upon a set of z_j this method does not model the image data as conditionally dependent (Besag 1986). Wilson et al. (1994) used this model for univariate data.

4.5 *Estimating the effective system spatial resolution*

While the PSFs associated with the optics, detectors, electronics, and sampling can be expected to remain relatively stable over time, atmospheric and resampling variability require that a unique system PSF be estimated for each image from the image data itself. While the presample EIFOVs calculated by Markham (1985) range up to 36 m for the TM, a hazy atmosphere can increase the EIFOV to over 100 m (Kaufman 1984).

In practice, estimates of the system PSF and MTF are based on measurements in one-dimension. In the spatial domain the LSF determines the system response to an infinitesimally narrow line separating different levels of reflected radiance. As the integral of the PSF in the direction of the line it is more feasible to measure than the variation about a point source (Markham 1985). The one-dimensional Fourier transform of a LSF gives the corresponding one-dimensional MTF. Alternatively, an MTF in one dimension can be estimated directly as the ratio of the modulation of a spatial frequency response to a known wave modulation (radiance distribution) with respect to the spatial frequency in the direction perpendicular to a linear discontinuity, such a road or a bridge over water.

Consider a spatial coordinate system defined by two orthogonal axes (a, b). If the two-dimensional PSF is assumed separable with respect to a and b, then it can be derived from the product of two functions of independent variables, $\mathrm{PSF}(a, b) = \mathrm{LSF}_a(a) \cdot \mathrm{LSF}_b(b)$. Likewise, $\mathrm{MTF}(u, v) = \mathrm{MTF}(u) \cdot \mathrm{MTF}(v)$, for spatial frequencies (u, v). Even if inseparable, the one-dimensional response functions still represent cross sections of the two-dimensional functions. For calculating one-dimensional responses from the system design and measurements it is convenient to assume that the (a, b) axes and (u, v) frequencies correspond to the across-track and in-track directions, respectively. The separability assumption is generally considered a good approximation for most passive remote sensing applications (Schott 1997), although for the reasons

given in Sections 4.1 and 4.2 it is unlikely the responses will be equal in the scan and track directions. Due to the nature of the targets used for image derived response functions, the directions of these one-dimensional response functions may not correspond to either the scan or track directions.

4.5.1 One-dimensional MTF method

Wrigley et al. (1984) and Schowengerdt et al. (1985) presented several approaches for quantifying the system MTF from image data. Each of these methods involves calculating the spatial frequency response of the system as the modulus of the fast Fourier transform (FFT) of an image subset containing discontinuities. This is equivalent to the square root of the image power spectrum.

Wrigley et al. (1984) tested for modulation transfer differences between the forward and reverse scans of the Landsat-4 TM by taking square root of the ratio of the forward and reverse scan power spectrums. The image data they used contained high contrast along the shores of Chesapeake Bay. Wrigley et al. (1984) also confirmed the modulation transfer effect of cubic convolution resampling by taking the square root of the ratio of the power spectrum of geometrically corrected data to that of uncorrected data and comparing the result to the cubic convolution MTF. By evaluating the ratio of power spectrum square roots, this approach avoids the need for a target having specific dimensions, but does not provide a means of estimating a system EIFOV.

Schowengerdt et al. (1985) calculated a system MTF in one dimension as the ratio of recorded frequency response to the actual radiance distribution of a one-dimensional target. For this he chose the San Mateo Bridge over San Francisco Bay. The actual wave modulation of the bridge over a uniform high contrast background is modelled as the modulus of a sinc function. This assumes that the radiance distribution of the bridge in the direction perpendicular to the bridge is a perfect rect function. The frequency response given by the data is again calculated as the modulus of the FFT

of the image data. Thus, the one-dimensional system MTF is given by

$$\mathrm{MTF}(u') = \frac{\sqrt{\mathrm{PS}}(u')}{|\mathrm{sinc}(W_b\, u')|}, \tag{4.9}$$

where PS is the image power spectrum, W_b is the width of the bridge and u' is the spatial frequency in the direction perpendicular to the bridge, 58.9° off the scan direction. Although this oblique angle changes the spatial phase between pixel samples from line to line, it was assumed that this phase information would be averaged out as the image response was averaged over many lines. Finally , the ratio function in Equation 4.9 was smoothed using a least squares polynomial fit. Using this method Schowengerdt et al. (1985) reported EIFOVs of 40.8 m for band 4 of a December TM-4 image and 48.6 m as an average of all reflective bands in an August TM-4 image.

Schowengerdt et al. (1985) also demonstrated a method for estimating a two-dimensional system MTF that does not require known high-contrast targets. Instead this method requires an airborne scanned image to be acquired at the same time, view angle, and have the same spectral resolution as the satellite image. The scene being imaged should contain a variety of high spatial frequency features such as edges, but no particular targets are required. The two images are assumed to be an object-image pair, such that the TM image results from the convolution of the airborne image by the TM system PSF. In the spatial frequency domain this is represented as

$$|\mathcal{F}[y_{\mathrm{TM}}(u,v)]| = |\mathcal{F}[y_{\mathrm{air}}(u,v)]| \cdot \mathrm{MTF}_{\mathrm{TM\ sys}}(u,v), \tag{4.10}$$

where $|\mathcal{F}[y_{\mathrm{TM}}(u,v)]|$ is the modulus of the Fourier transform of the TM image, y_{air} is the aircraft image, $\mathrm{MTF}_{\mathrm{TM\ sys}}$ is the TM system MTF, and (u,v) are the spatial frequencies corresponding to the scan and track directions, respectively. Thus, the TM system MTF is given as the following ratio

$$\mathrm{MTF}_{\mathrm{TM\ sys}}(u,v) = \frac{|\mathcal{F}[y_{\mathrm{TM}}(u,v)]|}{|\mathcal{F}[y_{\mathrm{air}}(u,v)]|}. \tag{4.11}$$

This method further requires that the two images be co-registered and calibrated. The resampling necessary to register the aircraft image to the TM image reduces contrast. In order to account for this reduction the denominator of Equation (4.11) was multiplied by an MTF estimated for the resampling process. The band average and smoothed MTFs calculated using this technique yielded EIFOVs of 40.4 and 42.4 m in the scan and track directions for a subset of San Jose and 38.7 and 46.0 m in the scan and track directions for a subset of Stockton. These EIFOVs were calculated from the same August TM-4 image used to estimate the 48.6 m EIFOV using the bridge target.

Although it is somewhat reassuring that the results of these two methods roughly correspond, in the absence of high contrast linear features of known dimension (as is often the case for mountainous forested terrain) and simultaneous aircraft underflight, other methods must be considered.

4.5.2 One-dimensional LSF method

By describing effective spatial resolution in terms of $W_{\frac{1}{2}}$ and assuming a separable PSF, the PSF half-width can be constructed from two orthogonal LSF half-widths. This is the approach adopted for this research. Several methods have been described in the remote sensing literature for estimating LSFs from image data. All of these methods are based on the spatial domain relationship given by Equation (4.4) applied to one-dimensional targets.

Malaret et al. (1985) modelled the radiance distribution across a narrow road as a rect function in the spatial domain in a similar fashion as Schowengerdt et al. (1985) modelled the bridge in the frequency domain. For a 6 m wide road and a half-width greater than or equal to one pixel, it can be shown that the system LSF is equivalent to a smoothed representation of the image response, where the image response is taken from a subpixel sampling procedure. Using a Landsat-5 TM image, trigonometric smoothing produced a band average half-width of 39.14 m compared

to 36.15 m for spline smoothing.

McGillem et al. (1985) represents a LSF as a finite set of non-overlapping rectangular basis functions. The basis function coefficients are found by solving a system of linear equations. The half-width can then be derived from the resulting function. This method gives a direct solution to an approximate form of the LSF. An alternative method proposed by Anuta et al. (1984) was shown to yield similar results (McGillem et al. 1985). For a linear edge between two homogeneous and spectrally different areas, the image can be modelled as a step superimposed on a constant background

$$o(a) = B_k + A_m\, u(a), \tag{4.12}$$

where o is the varying pattern of reflected radiance in the direction perpendicular to the edge, B_k is the background response, and A_m is the amplitude change across the edge. The recorded image is given by substituting Equation (4.12) into Equation (4.4) in one dimension. By taking the derivative of both sides of (4.4), the LSF is found as

$$\mathrm{LSF}(a) = \frac{y'(a)}{A_m}. \tag{4.13}$$

This gives the LSF as the derivative of the measured step function. For an edge spanning many rows or columns it is best to estimate the LSF from the average response. Either trigonometric polynomials or cubic splines can be used to smooth the data. Anuta et al. (1984) found that spline smoothing tended to underestimate the half-width by ten percent, and trigonometric smoothing overestimated the half-width by 15 percent. McGillem et al. (1985) estimated an LSF of 38.8 m from spline smoothed TM-4 band 4 data. After correcting for under- and overestimation Anuta et al. (1984) estimated a band 4 LSF of 39.3 m using spline smoothing and 43.2 m using trigonometric smoothing. For band 5 the results were 43.8 m and 44.7, respectively.

As system responses these estimates include the effects of atmospheric scattering and resampling which will vary from image to image. Thus, LSFs must be estimated

for each image classified. Historical harvest patterns in managed forests provide the necessary patterns for LSF estimation using the step function derivative.

4.6 Summary

A satellite image is composed of discrete samples of the continuously upwelling reflected radiance field. The recording and sampling mechanisms introduce unwanted spectral and spatial variation. The atmosphere further complicates the process by absorbing and scattering radiance, and by directly scattering solar irradiance into the direction of the sensor. The radiometric sensitivity to these variations is determined by the radiometric resolution. According to the effective spatial resolution the individual data samples are not conditionally independent given the classes. This research characterizes the system PSF by modelling the data samples as spatial averages of discrete upwelling radiance. In the context of a hierarchical model the data are assumed conditionally independent given the reflected radiance field. Several methods exist for estimating the effective spatial resolution of a remote sensing system.

Chapter 5

HYPERPARAMETER ESTIMATION

The models given by Equations (3.6) and (4.8) along with a prior distribution for x contain several unknown *hyperparameters*. The estimation of these parameters is a complicated and difficult problem. Although a number of estimation techniques have been developed, none of them are completely satisfactory (Besag 1989; Demoment 1989; Wilson et al. 1994). As a general rule in developing image models the number of adjustable parameters should be kept to a minimum. Sections 3.5 and 4.5 presented elaborate techniques for estimating the NBRFs and the PSF half-widths, respectively. This study treats those estimates as fixed quantities.

The conventional supervised ML approach has been to estimate the class means and precision matrices for each class from training data. Of the literature reviewed in Section 4.3.1, several papers presented suggestions for improving precision estimates when the independence assumption is violated. The Bayesian approach provides a method for probabilistically estimating parameter values. This chapter considers the specification of prior distributions for the hyperparameters μ_c, Θ^{-1}, and Ψ^{-1}. Additionally, this provides a method for estimating the marginal posterior distributions of these parameters (see Chapter 6).

Not all hyperparameters can be estimated in this manner. Section 2.2.2 mentioned the problems associated with specifying a prior distribution for the Potts model parameter β. For this study the hyperparameters governing the potential functions are arbitrarily chosen, reflecting their dependence on the subjective nature of the desired classification.

5.1 Class means

Supervised ML classification segments images into specifically defined classes, whereby the class means represent the spectral signatures to be extended. Representative training sites are then identified on the basis of the class definitions. In principle, the image data corresponding to the training sites can be used to estimate the spectral signatures. In Equation (3.6) the class means, μ_c, are the mean level of unblurred reflected radiance recorded for each class from a normalized surface. The actual radiance recorded is reflected by a topographic surface and attenuated by the modulation transfer of the sensor system. Thus, the class means cannot be estimated directly from the image data.

5.1.1 Point estimation

The simplest approach is to topographically normalize the training samples as a preprocessing step. Then randomly select a set of training samples to approximately correct for the effects of the system PSF (see Section 4.3.1) . While a random selection does not guarantee adequate spacing, if the selection represents a small proportion most of the samples will at least not be adjacent. Further, dependence will only bias the covariances, not the means. By assuming a distribution, ML estimates of the class means can be obtained from these preprocessed samples.

5.1.2 A single prior distribution for each class

As mentioned above, the uncertainty of estimating μ_c can be modelled probabilistically by specifying prior distributions. One alternative assigns the same *conjugate* normal distribution to each class mean. A prior distribution is considered conjugate if it combines with a conditional density to produce a posterior having the same distribution as the prior. In this case, a normal prior leads to a normal full conditional

for the class means. The prior is written as follows

$$\mu_c \sim N_r(\eta, \Gamma), \tag{5.1}$$

for all $c \in \Lambda$ and where η and Γ are a mean vector and covariance matrix, respectively. While this prior captures the uncertainty of estimating μ_c, in and of itself it does not contribute to a supervised classification. By considering the resulting full conditional distributions for μ_c (see Chapter 6) it can be seen that this prior would produce an unsupervised classification except for the uniqueness of the NBRFs for each class. An unsupervised classification segments an image into classes without predetermined definitions. However, in practice there does not appear to be enough information contained in the NBRFs to supervise a classification. Wilson et al. (1994) developed an unsupervised classifier by adopting an exchangeable uniform prior for univariate means. Hartt et al. (1989) adopted a more conventional approach for unsupervised classification within the context of Bayesian image analysis, although most unsupervised classifiers are not Bayesian.

Moffett et al. (1997) used the prior distribution given by Equation (5.1) for a supervised classification by weighting the full conditional distributions for the means. The weighting was accomplished by fixing a certain number of x_i in each class corresponding to training sites for each class. Further, the components of Γ^{-1} were assumed zero to represent vague prior information. This approach requires that training areas be located within the extent of the image or image subset being classified.

While this approach has achieved reasonable results, other research conducted for this study raises questions about the reliability of this supervision technique. In the first place, there is no reason to believe the prior is correct. A conjugate was chosen for convenience. While Landsat data is generally regarded as normally distributed, this assigns positive probability to negative values. Although weighting distinguishes the classes in a meaningful way, the prior implies all the classes are the same. This is true whether or not Γ^{-1} is set to zero. More importantly, it is possible for the

training pixels, weighting the full conditionals, to be outweighed themselves. This can occur as the class mean full conditionals together consider all image pixels [see Equations (6.32) and (6.33)] and, site for site, the fixed training sites do not carry greater weight than the unfixed sites being classified. Indeed, within this framework, estimates of the class means are taken from the full conditionals, not the priors! For reasons related to cost and extensibility, the number of training sites required should be kept to a minimum. To the extent that there are few fixed training sites relative to the total number of sites being classified, it is possible for a large number of incorrectly classified sites to outweigh the training sites for that class and shape the full conditional accordingly. There is an increased tendency for this to occur when classes are spectrally similar to begin with and for models having limited mobility due to high bonding strengths.

5.1.3 Unique prior distributions for each class

As another alternative, consider the specification of a conjugate normal distribution for each class mean, written as

$$\mu_c \sim N_r(\eta_c, \Gamma_c), \tag{5.2}$$

for all $c \in \Lambda$ and where η_c and Γ_c are the mean vectors and covariance matrices for each class, respectively. This approach opens up more options for supervision. Point estimates of η_c and Γ_c can be obtained from preprocessed data as described in Section 5.1.1. This approach requires some faith in the ability to estimate parameters from the training data, but in this case the classifier is trained by an estimated distribution for each class mean, as opposed to point estimates, incorporating prior information about the spectral variability within each class. Estimates of η_c and Γ_c do not necessarily need to come from the image data being classified and therefore the training sites themselves are not required. While this may represent the holy grail for remote sensing, the calibration required to produce relevant estimates of η_c and

Γ_c would not be easy in this case.

Incidentally, Green et al. (1994) indirectly adopted this approach. They first assumed the same conjugate normal for all the classes as described in Section 5.1.2. Then they derived conditional densities for the means and precisions from the prior distribution and an assumed normal density for each class, parameterized by estimates taken from a random selection of training data. Finally, they set $\Gamma^{-1} = 0$, arriving at a unique conditional density for each class mean. Unfortunately, this formulation gave disappointing results. Although many boundaries were identified and the classification was spatially homogeneous, large areas were misclassified.

While the authors contend that the errors were attributable to a misspecified likelihood function, other reasons may exist. Indeed, as with the weighting scheme, the full conditional distributions incorporate the information content of the parameter priors with that of the pixels being classified. This is reflected in the values sampled. To the extent that the training site information is outweighed, more misclassifications are likely, further distorting the full conditionals. Given the high degree of spectral variability to be expected, there is a danger in allowing too much uncertainty in estimating the class means. There should not be large discrepancies between sampled means and the point estimates taken from training data, unless large areas being classified are not spectrally well represented by the classes chosen. The tendency to distort the full conditionals is exacerbated when the posterior probability of x accepts less likely classifications at the site level in favor of preserving spatial contiguity. As a possible remedy for this problem, additional weight can be placed on the training site data by adopting the the weighting scheme in conjunction with a prior for each class.

5.1.4 Extensibility

Because of the variability of atmospheric absorption, correlations drawn between samples of reflected radiance and surface cover cannot be accurately extrapolated over

space and time unless the interaction between radiant flux and the atmosphere can be estimated accurately (Mather 1989). Thus, DNs must be calibrated to reflectance, either explicitly or implicitly, before multi-temporal or multi-image data sets can be compared. Even calibration to radiance will not allow comparison of images acquired under different illumination angles. Accurate calibration methods are time consuming, costly, and difficult to validate. Teillet (1986) and Duggin and Robinove (1990) review several of the many proposed calibration methods. However, the atmospheric model adopted by default has been to ignore the presence of the atmosphere (Teillet 1986). Arguably, this conclusion still applies.

For the supervised classification of multiple images this means that unless the data have been calibrated, parameter estimates from training sites must be calculated for each image. For monitoring the same area over time, this is not an unreasonable burden and this is the approach adopted for this research. The need for calibration does limit the ability to extend training data from one region to another. This is not a new problem. Fraser et al. (1977) found significant classification errors resulting from atmospheric variability and concluded that training samples should be remeasured for each scene being classified.

5.2 Precision matrices

In Equation (3.6) the z_i are assumed to be distributed with common covariance Θ, and in Equation (4.8) the observed y_i are assumed to have a common covariance Ψ.

5.2.1 Point estimation

Maximum likelihood can also be used to estimate the precision matrices as described for the class means in Section 5.1.1, although the training data need not be normalized. For estimating common precisions, the training data for all the classes can be grouped. However, unbiased precision estimates further require the training data to

be independently distributed. Although a random selection of the training samples may improve the estimates, given that z is unobserved and the data is dependently distributed, the specification of priors for these parameters is a logical alternative to direct estimation.

5.2.2 Wishart prior distributions

Following the convention of Gelfand et al. (1990) and Green et al. (1994) the corresponding precision matrices, Θ^{-1} and Ψ^{-1}, are assigned the same conjugate Wishart distributions (Anderson 1984) as follows

$$\Theta^{-1} \sim W_r([\omega R]^{-1}, \omega) \tag{5.3}$$

and

$$\Psi^{-1} \sim W_r([\omega R]^{-1}, \omega) \tag{5.4}$$

where ω is the degrees of freedom and $[\omega R]^{-1}$ is the covariance. Thus,

$$\pi(\Theta^{-1}) \propto |\Theta^{-1}|^{\frac{\omega-r-1}{2}} \exp\{-\frac{1}{2}\mathrm{tr}([\omega R]\Theta^{-1})\} \tag{5.5}$$

and likewise for $\pi(\Psi^{-1})$, where tr is the trace function. Note that although a precision matrix is the inverse of the covariance matrix, these matrices are assumed Wishart distributed, not inverse Wishart distributed. For this research, vague precision priors are established by setting $\omega = 0$, indicating a lack of prior information regarding these parameters.

5.2.3 Unique precisions for each class

Green et al. (1994) adopted the same approach for the precision matrices as they did for the class means (see Section 5.1.3), which essentially produced a unique prior for each class. There are many reasons to expect each class the exhibit a unique covariance. However, preliminary research for the model presented here suggests that

the specification of Θ_c^{-1} for all $c \in \Lambda$, in the context of weighting the full conditionals, introduces too much uncertainty into the model as z must be updated each cycle based on the current estimate of Θ^{-1}. If a Θ_c^{-1} is specified for each class, then there will be the tendency for the covariances of a few classes to disperse dramatically increasing the tendency for pixels to be misclassified, further dispersing the covariances. This effect will also lead to more varied updates of z.

These preliminary results should be considered exploratory not conclusive. Aside from the issue of specifying single or unique precisions, there is the issue of adequately estimating the parameters in the prior(s). The specification of conjugate distributions is convenient, but not necessary.

5.3 Summary

The estimation of hyperparameters in any model is a difficult problem. When direct methods are not available, the Bayesian framework allows prior distributions to be specified as an alternative. Determining the correct prior is also difficult, but allows for uncertainty. The class means effectively supervise the classification. The two precision matrices are most suited for prior distributions. Hyperparameters in the prior for x are left to the users discretion.

Chapter 6

IMAGE ESTIMATION

The primary objective of forest classification within the context of Bayesian image analysis is to estimate x^* along with an assessment of uncertainty. Section 2.1.2 described how inferences about x^* are based on a conditional posterior distribution for x given data y, formed using Bayes theorem. The posterior distribution for the hierarchical model developed for this study is presented in Section 6.1. Section 6.2 reviews the most frequently used image point estimates and a spatially explicit method of quantifying the uncertainty of one particular estimate. Given the high dimensionality of x for typical image sizes, the optimization required to find an exact solution is infeasible. Section 6.3 describes the application of Markov chain Monte Carlo (MCMC) methods, whereby inferences about x^* are based on samples drawn from a Markov chain having the desired posterior distribution as its limit distribution. Finally, common algorithms for simulating Markov chains are presented in Section 6.4.

6.1 The joint posterior distribution

The full joint posterior distribution for the hierarchical model is formed using Bayes theorem to combine the sub-models given by Equations (2.3), (3.6), and (4.8), along with the prior distributions for the parameters given by Equations (5.1), (5.3), and (5.4). Hence the full posterior is written as

$$\pi(x, z, \mu_x, \Psi, \Theta|y) \propto \pi(y|z, \Psi)\, \pi(z|x, \mu_x, \Theta)\, \pi(x)\, \pi(\mu_x)\, \pi(\Psi)\, \pi(\Theta). \qquad (6.1)$$

Inferences about x^* are based on the marginal posterior distribution for the classification given the data $\pi(x|y)$. This marginal distribution arises from integrating out

all of the other variables from full posterior such that

$$\pi(x|y) = \int\int\int\int \pi(x, z, \mu_x, \Psi, \Theta|y)\, dz\, d\mu_x\, d\Psi\, d\Theta. \tag{6.2}$$

In practice the very high dimensional integration required to find $\pi(x|y)$ is intractable. Even if it were possible, $\pi(x|y)$ itself would remain n-dimensional. Further marginalization over x_{-i} gives the marginal posterior probability $\pi(x_i = c|y)$ that site i falls in any particular class c given the observed data. In practice all such marginalizations are carried out automatically (see Section 6.3).

6.2 *Image estimates*

Within this framework, the final classification is a point estimate of x^*. The two most common methods are the maximum *a posteriori* (MAP) and the marginal posterior modes (MPM) estimates, each implying an expectation of a particular classification. These probabilistic statements require the estimation of the marginal distribution $\pi(x|y)$ which is carried out indirectly using the MCMC methods described in Section 6.3.

6.2.1 *Maximum* a posteriori *estimate (MAP)*

The MAP estimate of x^* seeks the x having maximum probability, written as

$$\hat{x} = \arg\max \pi(x|y). \tag{6.3}$$

The MAP estimate identifies the mode of the posterior distribution and can be thought of as the global maximum. However, MAP does not guarantee that each component of x will be classified with maximum probability. In the context of decision theory, $\hat{x}$ represents a zero-one loss function on whole classifications, minimizing the probability of any pixel being misclassified (Besag 1986). In other words, MAP assigns a unit loss to a classification whether a single or all sites are incorrect.

For moderately strong interactions, second-order MRFs converge to single class realizations over significant distances (see Section 2.2.2). The MAP estimate represents a balance between the information represented by the data and the spatial continuity modelled by the prior at the global level. Thus, seeking the mode of the posterior distribution exacerbates this tendency for long-range order, which may be manifest in $\hat{x}$. Marroquin et al. (1987) demonstrate this limitation using the Ising prior.

In principle the MAP estimate can be found by simulating enough whole images from $\pi(x|y)$ and identifying the most frequently occurring image. However, for an array of n sites and m classes there are m^n possible images, so producing enough realizations becomes impossible. Instead, the MAP estimate can be found via simulated annealing (Kirkpatrick et al. 1983). Geman and Geman (1984) implemented a *simulated annealing* algorithm using the Gibbs sampler (see Section 6.4.1). As an example, consider the simple posterior given by Equation (2.1). Rather than generating realizations from $\pi(x \mid y)$, samples are collected from

$$\pi_T(x|y) \propto \{l(y \mid x)\pi(x)\}^{\frac{1}{T}}, \tag{6.4}$$

where T is a global parameter controlling the "temperature" of the system. As $T \to \infty$ $\pi_T(x|y)$ converges to its maximum. Again there is an analogy to thermal equilibrium whereby low temperatures correspond to stable states having high probability. In practice, T must be reduced sufficiently slowly and the actual cooling schedule required may be prohibitively slow. In this case, simulated annealing provides an approximation to the MAP estimate. Application of simulated annealing, or similar algorithms, for obtaining the MAP estimate can be found in the remote sensing literature; see, for example, Derin and Elliott (1987), Kelly et al. (1988), Zhang et al. (1990), Kato et al. (1992), and Peytavin et al. (1995).

The iterated conditional modes (ICM) algorithm provides an alternative for MAP approximation provided the log of the posterior is a strictly concave differentiable

function of x, in which case the posterior has a single maximum (Besag 1986; Besag 1989). In contrast to the Gibbs sampler, ICM is a *gradient ascent* algorithm that maximizes the posterior by always choosing the mode of the full conditional distributions during each iteration. In other words, ICM corresponds to simulated annealing at zero temperature. ICM converges rapidly and is therefore computationally inexpensive. Kittler and Föglein (1984), Kittler and Pairman (1985), Solberg and Taxt (1994), and Yamazaki and Gingras (1995) applied such an algorithm to satellite data. Kiiveri and Campbell (1992) apply it to synthetic images. While the influence of the data in this context prevents classifications from having only one class, there is still a tendency for excessive homogeneity when seeking the MAP estimate. Further, there is no measure of uncertainty associated with MAP estimates.

6.2.2 Marginal posterior modes estimate (MPM)

The MPM estimate minimizes the expected number of misclassifications under the model by choosing, for each pixel, the class that has maximum posterior probability given the data, and is written as follows

$$\hat{x}_i = \arg\max \pi(x_i|y), \tag{6.5}$$

where the MPM estimate of x^* has the components $\hat{x}_i$. Computing the MPM estimate with the Gibbs sampler corresponds to simulated annealing at a constant temperature $T = 1$. The the most frequently sampled class for site i is the estimate of x_i^*. Note that the MPM estimate requires the marginalization described at the end of Section 6.1.

Essentially the MPM estimate maximizes the marginal probability with which each site is classified and corresponds to a zero-one loss function on each pixel. In contrast to MAP, MPM will not guarantee that the image as a whole is classified with maximum probability. Unlike annealing where the primary convergence issue is to identify the appropriate cooling schedule, the MPM estimate requires determining the appropriate number of realizations to be sampled from the posterior distribution.

In addition to a point estimate, this procedure estimates a marginal posterior probability distribution for each pixel. In this way MPM provides spatially explicit information about the certainty of the point estimate classification. Although spatial placement of error is not considered in distinguishing between estimates (Bouman and Shapiro 1994). Estimation of probability surfaces in this manner is a highly desirable feature of MPM estimation and not viable when seeking the MAP estimate. The MPM estimate has been adopted in recent years as a classification method (Besag et al. 1991; Wilson et al. 1994; Green et al. 1994; Moffett and Besag 1996; Moffett et al. 1997). Other probabilistic estimates can be used to characterize a point estimate including the posterior median; see Besag (1989) and Tjelmeland and Besag (1998) for more details and an example, respectively.

6.3 Markov chain Monte Carlo

In principle the MAP and MPM estimates require computing the probabilities for each of the m^n realizations (for minimal sample space Ω) of x under $\pi(x|y)$ as given by Equation (6.2). However, the integrals involved for exact optimization cannot be handled by standard analytical or deterministic numerical approximation techniques due to the high dimensionality of the problem. Standard Monte Carlo simulation methods estimate probabilities of specific events by directly sampling from a given distribution or density. However, direct simulation of $\pi(x, z, \mu_x, \Psi, \Theta|y)$ is not possible due to the dimensionality of the integrals involved. Besag et al. (1995) provides a detailed explanation of this limitation with examples.

Rather than drawing samples directly from $\pi(x, z, \mu_x, \Psi, \Theta|y)$, MCMC methods seek to simulate the full posterior indirectly by sampling from a very long realization of a *discrete-time Markov chain* having $\pi(x, z, \mu_x, \Psi, \Theta|y)$ as its *equilibrium distribution* and minimal sample space Ω. Image estimates can be derived from these samples. Information about any subset of the components is extracted from the output, simply

by ignoring the remaining components. This automatically produces estimates of any marginal distribution that may be of interest. For example, $\pi(x_i = c|y)$ is estimated by the relative frequency of class c on pixel i and hence the MPM estimate $\hat{x}$ is obtained by noting the most frequently occurring class for each site.

MCMC methods have become standard for Bayesian computation in recent years, see Smith and Roberts (1993), Besag and Green (1993), Besag et al. (1995), (Higdon 1994) and the references therein for a more detailed discussion. These papers are summarized here as they pertain to this application. MCMC provides a last resort for simulating complex distributions typically characterized by non-standard form, dependent components, and highly multivariate.

6.3.1 Markov chains

A Markov chain is defined as a sequence of random variables $X^0, X^1, X^2, \ldots$ whereby each X^{t+1} depends only on the previous random variable, X^t, written as

$$\Pr(X^{t+1} = x^{t+1} | X^1 = x^1, \ldots, X^t = x^t) = \Pr(X^{t+1} = x^{t+1} | X^t = x^t), \tag{6.6}$$

for all time steps $t = 1, 2, \ldots$, $x \in \Omega$, and where Pr denotes a probability. *Time homogeneous Markov chains* are defined as

$$\Pr(X^{t+1} = x' | X^t = x) = \Pr(X^1 = x' | X^0 = x), \tag{6.7}$$

for all t and $x, x' \in \Omega$. A discrete-time Markov chain is characterized by discrete steps (time) and does not inherently correspond to a discrete-state MRF in which each X_i takes on only discrete values.

Simulating a Markov chain having Equation (6.1) as its limit distribution generates a sequence of s stochastically dependent classifications $x^1, x^2, \ldots, x^s$. This is accomplished by constructing a discrete-time Markov chain *transition probability matrix* P with the desired *limit distribution* and state space. For a given seed x^0, each new realization depends only on the previous one. The probability associated with a

move from x to x' is written as $P(x \rightarrow x')$. Also note that P is an $m^n \times m^n$ matrix. For a time-homogeneous Markov chain, P does not vary with time.

6.3.2 Ergodicity

More generally, it is necessary for P to be an *ergodic* transition probability matrix that satisfies *general balance*. For a given distribution $\pi(x)$ estimates of its various properties can be expressed as expectations of a corresponding function f written as

$$\mathbf{E}_\pi f = \sum_{x \in \Omega} f(x)\pi(x), \tag{6.8}$$

where $\pi(x)$ is any distribution of interest including the posterior distribution of interest for this study, also an MRF. When f is an indicator function, $\mathbf{E}_\pi f$ provides probability statements of specific events distributed as π. For any subset $G \in \Omega$ the probability $\Pr(X \in G)$ is given by

$$\Pr(X \in G) = \sum_{x \in G} \pi(x). \tag{6.9}$$

The ability to directly approximate such probabilities from complex distributions is a primary feature of MCMC.

The ergodic theorem for Markov chains implies that for any seed x^0 the sequence of random samples given by

$$\bar{f}_s = \frac{1}{s} \sum_{t=1}^{s} f(x^t) \tag{6.10}$$

converges almost surely to $\mathbf{E}_\pi f$ as $s \rightarrow \infty$, where s is the number of samples. Thus, if P is ergodic, then the empirical average $\bar{f}_s$, computed from long sequence of realizations $x^1, x^2, \ldots, x^s$ of the chain, can be used to approximate $\mathbf{E}_\pi f$. In other words, given any seed and enough samples, any corresponding sequence of random variables or vectors can be used to compute properties of the limit distribution. For ergodic transition matrices the equilibrium distribution is also the limit distribution.

6.3.3 General and detailed balance

General balance ensures that P maintains π, in which case

$$\pi P = \pi, \tag{6.11}$$

ensuring that π is the marginal distribution for each subsequent realization x^t of the chain. If P satisfies general balance, then π is the *stationary distribution* for P, the limit distribution of the chain. In practice verification of general balance is problematic due to the complicated summation required, shown by rewriting Equation (6.11) as

$$\sum_{x \in \Omega} \pi(x) P(x \to x') = \pi(x'), \tag{6.12}$$

where x is the current state and x' is an alternative state for all $x, x' \in \Omega$. This difficulty is overcome by requiring P to satisfy *detailed balance*

$$\pi(x) P(x \to x') = \pi(x') P(x' \to x), \tag{6.13}$$

for all $x, x' \in \Omega$. Detailed balance implies general balance and only needs to be checked for $x \neq x'$. The restriction to detailed balance also ensures that P is *time reversible*, in which case it is impossible to distinguish between forward and reverse simulations of the chain.

Ergodic transition matrices are *irreducible* and *aperiodic*. An irreducible transition matrix allows any realization x' to be reached from any other realization x in a finite number of transitions for $x, x' \in \Omega$. In other words, it is possible to achieve any realization from any other realization. Aperiodic Markov chains have a period of one, where the period of the chain is the least number of steps required to return to a given state.

In practice P is specified as the product of individually reducible but reversible P_is written as either

$$P = P_1 P_2 \dots P_n \tag{6.14}$$

or

$$P = \frac{1}{n}(P_1 + P_2 + \ldots + P_n) \tag{6.15}$$

in which $\pi(x)$ is the stationary distribution of each P_i. If each P_i maintains $\pi(x)$ so does P, whose elements are therefore zero whenever x and x' differ on more than one site. Provided a transition from x to x' is possible for at least one P_i, irreducibility is ensured for P. For P formed as in Equation (6.14), if the P_is are arranged in a fixed permutation at each time step, the reversibility of P_i does not guarantee the reversibility of P. However, randomizing the permutation at each step overcomes this limitation.

6.3.4 Initialization and convergence

Returning to the ergodic theorem, several issues must be considered in order to properly estimate expectations using MCMC. The first concerns the choice of the seed x^0. In principle this should be sampled from the limit distribution itself, but if this were possible there would be no need for MCMC in the first place. Thus, the choice of x^0 becomes arbitrary. Any bias created by this choice is erased in the limit as $\bar{f}_s$ converges to $\mathbf{E}_\pi f$ as $s \to \infty$. However, in practice s is a finite number and it is necessary to consider the possible bias introduced by the initialization of the algorithm.

Generating and discarding successive realizations x^t until the sequence has converged to its equilibrium distribution and then initiating sampling from that point on is a practical approach for avoiding initialization bias. The predetermined number of discarded samples is referred to as the length of the "burn-in" period. Although some formal methods of diagnosing convergence exist, this is an area of active research; see Besag et al. (1995) for references.

More often informal functions of chain output are plotted against iteration number with the idea that the rate of convergence of these functions mimic that of the Markov chain being simulated. In remote sensing applications, Wilson et al. (1994) considered

several functions of image components and Green et al. (1994) looked for stability in the number of sites in each class after each iteration. Moffett et al. (1997) assessed convergence by monitoring the stability of the number of pixels in each class and their mean spectral reflectances, as well as the number of pixels that changed on each iteration. Another approach is to evaluate the convergence of several simulations each beginning in very different locations of the state space (Wilson et al. 1994; Tjelmeland and Besag 1998; Moffett et al. 1997). Although these informal methods offer no guarantee that the correct limiting distribution has been reached, using them together should increase confidence.

Finally, the size of s must also be predetermined. Again there are no specific requirements. Greater numbers of samples will lead to better estimates, thus storage space and run times become the limiting factors. Note that disk space and speed will depend on the size of the image being classified, how its modelled, and the choice of MCMC algorithm. The latter is discussed in the following sections.

6.4 Hastings algorithms

Hastings algorithms (Hastings 1970) provide a general methodology for constructing ergodic Markov transition probability matrices that satisfy detailed balance for a desired limit distribution.

Suppose that $R(x \to x')$ is an arbitrary Markov transition probability matrix with state space Ω, and that

$$R(x \to x') > 0 \Leftrightarrow R(x' \to x) > 0, \tag{6.16}$$

for all $x, x' \in \Omega$. Given R, $P(x \to x')$ can be defined as

$$P(x \to x') = R(x \to x')A(x \to x'), \tag{6.17}$$

for all $x' \neq x \in \Omega$, and where

$$A(x \to x') = \min\left\{1, \frac{\pi(x')R(x' \to x)}{\pi(x)R(x \to x')}\right\}. \tag{6.18}$$

Thus,

$$P(x \to x') = \begin{cases} R(x \to x')A(x \to x'), & x' \neq x \\ 1 - \sum_{x' \neq x} P(x \to x'), & x' = x \end{cases} \tag{6.19}$$

For $x \neq x'$ detailed balance is easily verified for P having stationary distribution π, even for arbitrary R! In this case, a transition from x to x' is *proposed* with probability $R(x \to x')$, and *accepted* with probability $A(x \to x')$, else $x = x'$, retaining x as the subsequent state of the chain. Note that $\pi(x)$ need only be known up to scale for Equation (6.18) since the normalizing constant cancels.

6.4.1 Componentwise Hastings algorithms

Ergodicity is ensured upon implementation. Constructing P as in Equation (6.14) and cycling through the P_is in a fixed or random permutation establishes irreducibility. An R_i and an A_i are then required for each P_i. For site i the current state is denoted as x_i and a transition to state x_i' is proposed with probability $R_i(x_i \to x_i' : x_{-i})$ given $x'_{-i} = x_{-i}$. The acceptance probability for the proposed transition is given by

$$A_i(x_i \to x_i' : x_{-i}) = \min\left\{1, \frac{\pi(x')R_i(x_i' \to x_i; x_{-i})}{\pi(x)R_i(x_i \to x_i'; x_{-i})}\right\}. \tag{6.20}$$

If not accepted, $x_i = x_i'$.

By considering only a single site at a time, A_i can be specified in terms the corresponding full conditional distribution $\pi(x_i|x_{-i})$. This follows from the general result that for any $x \in \Omega$ and $N \subset \mathcal{N}$,

$$\pi(x_N|x_{-N}) \propto \pi(x), \tag{6.21}$$

where $\pi(x_N|x_{-N})$ is the full conditional distribution of x_N. Note that each component of x is represented on the left hand side of Equation (6.21). This, perhaps the most important result in componentwise MCMC, is easily proven by

$$\pi(x_N|x_{-N}) = \frac{\pi(x)}{\pi(x_{-N})} \propto \pi(x). \tag{6.22}$$

Also if $x, x' \in \Omega$ and $x_{-i} = x'_{-i}$,

$$\frac{\pi(x'_i|x'_{-i})}{\pi(x_i|x_{-i})} = \frac{\pi(x')}{\pi(x)}. \tag{6.23}$$

Thus, Equation (6.20) can be rewritten as

$$A_i(x_i \to x'_i; x_{-i}) = \min\left\{1, \frac{\pi(x'_i|x_{-i})R_i(x'_i \to x_i; x_{-i})}{\pi(x_i|x_{-i})R_i(x_i \to x'_i; x_{-i})}\right\}. \tag{6.24}$$

Ultimately, Equation (6.21) significantly reduces the computational expense of implementing such MCMC algorithms. For any $\pi(x)$ written in product form, common for locally dependent MRFs, only the terms involving the components of x_i are relevant. Further, in Equation (6.24) the normalizing constant as a function of x_{-i} requires only a one-dimensional summation over x_i, but cancels out regardless. Thus, Equation (6.24) requires the n full conditional only up to scale!

Componentwise Hastings algorithms define an ergodic P that maintains $\pi(x)$ by visiting each site during a single iteration or cycle. At each site x_i is updated while leaving x_{-i} unchanged. Cycle by cycle this generates a sequence of stochastically dependent realizations $x^1, x^2, \ldots$ used for estimating $\mathbf{E}_\pi f$. Notice that when $\pi(x)$ is a posterior distribution, each x^t is a classified image given some image data.

As a final step the proposal distributions R_i must be selected. There is a great deal of flexibility in the choice of R_i and the distribution need not be the same for each site, depending on the application. In general, proposals should be produced rapidly and acceptance probabilities easy to calculate. These objectives are more easily for univariate sites and corresponding full conditionals. For multivariate random variables at each site, univariate implementation is possible by considering each component individually. The Gibbs sampler and Metropolis algorithms correspond to specific choices of R_i and have been widely adopted in the literature.

6.4.2 The Gibbs sampler

The term *Gibbs sampler* first appeared in Geman and Geman (1984) referring to sampling from Gibbs distributions in statistical physics, although the algorithm has

earlier origins; see Besag et al. (1995) for references. Gibbs distributions are equivalent to the full conditional distributions of an MRF (Besag 1974). Thus, the Gibbs sampler simulates the full joint posterior distribution in Equation (6.1) by sampling from its full conditional distributions.

At each site x_i is replaced by x_i' drawn randomly from its full conditional, in which case the proposal distribution is given by the full conditional,

$$R_i(x_i \to x_i'; x_{-i}) = \pi(x_i'|x_{-i}), \tag{6.25}$$

and proposals are always accepted. In other words, the Gibbs sampler is a Hastings algorithm with zero probability of rejection as $A_i = 1$. Further, the Gibbs sampler is easily extended to continuous distributions of common form.

More directly, it is easily proven that updating x_i with x_i' drawn from $\pi(x_i|x_{-i})$ maintains $\pi(x)$. The corresponding transition probability matrix is defined as

$$P_i(x_i \to x_i') = \pi(x_i'|x_{-i})I[x_{-i} = x'_{-i}], \tag{6.26}$$

where I is an indicator function. Positivity ensures irreducibility since $P(x \to x') > 0$ for $x \in \Omega$.

6.4.3 Metropolis algorithms

Metropolis algorithms were originally developed in the field of statistical physics for simulating the Ising model (Metropolis et al. 1953). This method considers R_i to be a symmetric matrix, thus proposals are randomly sampled from an arbitrary symmetric distribution, in which case

$$R_i(x_i \to x_i'; x_{-i}) = R_i(x_i' \to x_i; x_{-i}). \tag{6.27}$$

The probability of accepting the proposal is then given by

$$A_i(x_i \to x_i'; x_{-i}) = \min\left\{1, \frac{\pi(x_i'|x_{-i})}{\pi(x_i|x_{-i})}\right\}, \tag{6.28}$$

else x_i is retained. Note that proposals having higher probability are always accepted.

The symmetric proposal distribution allows the R_i to cancel out. When simulating the Ising model the opposite value is always proposed. For random variables taking on more values, it is usually convenient to select x_i' uniformly from the state space of x_i excluding the current value of x_i. For continuous random variables, x_i' can be sampled from any symmetric distribution, such as uniform or normal, centered on x_i. Experimentation may be required to determine the variance giving a desired rate of acceptance.

6.4.4 Choosing between algorithms

The best choice of algorithms really depends upon the application and its objectives. Computational expense and statistical efficiency should each be considered. The Gibbs sampler offers the intuitive appeal of sampling from the full conditionals, provided they can be easily sampled. Since all proposals are accepted, A_i does not need to be computed. Green et al. (1994) and Moffett and Besag (1996) demonstrate the use of the Gibbs sampler for simulating relatively simple posterior distributions for image classification.

For updating discrete random variables, Metropolis requires the full conditional probability for only the current state and the proposed state. In comparison, Gibbs requires the full conditional for each value in the state space of x_i. Thus, Metropolis has a clear advantage computationally as the size of the state space increases.

The Hastings algorithms other than Gibbs tend to increase the mobility around the state space by making use of the current state of x_i in proposing x_i'. Although uniform Metropolis proposals are not necessarily more efficient, since unlikely states are proposed with the same probability as more likely states. This is more of a drawback for larger state spaces, and is offset by the fact that Metropolis requires dramatically less cpu than Gibbs for larger state spaces. In the case of the Ising model Metropolis converges faster than Gibbs for strong interactions.

Statistical efficiency can increased by adopting an asymmetric proposal distribution. Early research for this study implemented a Hastings algorithm whereby states other than the current were proposed with probabilities weighted by their full conditionals and likewise accepted. While this approach is more efficient, it requires computing all of the full conditionals and the acceptance probability. The same result may be achieved more quickly by a longer run of a faster algorithm, depending on the rate of convergence.

While most of these algorithms are rather computationally intensive, improved computing technology continues to reduce the burden of MCMC. In any case, they are all relatively easy to program.

6.4.5 Updating schedules

In order to ensure that componentwise updating algorithms define irreducible transition probability matrices each site must be visited during each cycle (see Section 6.4.1). There is no requirement that R_i be the same for each site, just that all sites be updated. There is some flexibility in the order the components are updated.

Scanning the array in a fixed order given by Equation (6.14) is easy to program and does not require additional computation. Fixed scans of this nature are not irreducible and may introduce undesirable directional effects when the sites are visited according to their spatial arrangement. Reversing the scan order for alternate scans is one approach for preventing such problems.

Another alternative determines the scan order by a unique random permutation for each iteration. A random permutation is required as simply selecting sites at random does not guarantee that all sites are updated. A random permutation generates an irreducible Markov chain with no potential for drift, but at the expense of sampling the permutation at each iteration.

Finally, coding sets (Besag 1974) allow mutually independent sets of sites to be

updated simultaneously. Visiting the sets in random order during each cycle ensures irreducibility. Further, coding sets are naturally extendible to parallel processing.

6.4.6 Full conditional distributions

During each cycle all of the components of x, z, μ_x, Ψ^{-1} and Θ^{-1} are updated in accordance with their full conditional distributions arising from the full posterior distribution in Equation (6.1). These distributions are given in this section. For multivariate sites, such as z_i, each component is cycled through individually and updated according to its full conditional. All of the components, except for those of x, have common distributions and are updated using Gibbs steps.

The components of x are updated via the Metropolis algorithm. Given the current class x_i, a proposal x'_i, chosen uniformly from the remaining classes, which is then accepted with probability

$$\min\left\{1, \frac{\pi(x'_i|\cdots)}{\pi(x_i|\cdots)}\right\}, \tag{6.29}$$

else x_i is retained. Alternatively, Gibbs steps could easily be implemented here as well, but at an increased computational expense.

The full conditional for z_i^p, the component of z at pixel i, band p, is Gaussian with mean $\mu_{z_i^p}$ and variance $\sigma^2_{z_i^p}$ defined as follows:

$$\begin{aligned}\frac{\mu_{z_i^p}}{\sigma^2_{z_i^p}} &= \Theta^{-1}_{p,p}\, t^{i,p}_{x_i}\, \mu^p_{x_i} + \sum_{q\neq p} \Theta^{-1}_{p,q}\,(t^{i,q}_{x_i}\,\mu^q_{x_i} - z_i^q) \\ &\quad + \Psi^{-1}_{p,p}\left\{\sum_{k\in(\nu_i\cup i)} h^p_{ki}\, y^p_k - h_{ii}\sum_{j\in\nu_i} h^p_{ij}\, z^p_j - \sum_{k\in\nu_i} h^p_{ki}\sum_{j\in\xi^i_k} h^p_{kj}\, z^p_j\right\} \\ &\quad + \sum_{q\neq p}\Psi^{-1}_{p,q}\left\{\sum_{k\in(\nu_i\cup i)} h^q_{ki} y^q_k - z^q_i\sum_{k\in(\nu_i\cup i)}(h^q_{ki})^2 - h^q_{ii}\sum_{j\in\nu_i} h^q_{ij}\, z^q_j - \sum_{k\in\nu_i} h^q_{ki}\sum_{j\in\xi^i_k} h^q_{kj}\, z^q_j\right\},\end{aligned} \tag{6.30}$$

where ν_i represents the pixels in the proximity of i as given by the PSF, excluding pixel i, and $\xi^i_k = (\nu_k \cup k)\backslash i$ indicates all of the pixels in the neighborhood of k,

including pixel k but excluding pixel i:

$$\sigma^2_{z_i^p} = \frac{1}{\Theta^{-1}_{p,p} + \Psi^{-1}_{p,p} \sum_{k \in (\nu_i \cup i)} (h^p_{ki})^2}. \tag{6.31}$$

The full conditional for μ_c^p, the mean DN of class c, band p, is also a normal distribution with mean given by,

$$\begin{aligned} \frac{\mu_c^p}{\sigma^2_{\mu_c^p}} &= \Theta^{-1}_{p,p} \sum_{i=1}^{n_c} t^{i,p}_{c_i} z_i^p + \sum_{q \neq p}^{r} \Theta^{-1}_{p,q} \sum_{i=1}^{n_c} (t^{i,p}_{c_i} z_i^q - t^{i,p}_{c_i} t^{i,q}_{c_i} \mu_c^q) \\ &\quad + \Gamma^{-1}_{p,p} \eta_p + \sum_{q \neq p}^{r} \Gamma^{-1}_{p,q} (\eta_q - \mu_c^q) \end{aligned} \tag{6.32}$$

where n_c is the number of pixels in class c and variance given by,

$$\sigma^2_{\mu_c^p} = \frac{1}{\Theta^{-1}_{p,p} \sum_{i=1}^{n_c} (t^{i,p}_{c_i})^2 + \Gamma^{-1}_{p,p}}. \tag{6.33}$$

The full conditional for the precision matrix Θ^{-1} is a Wishart distribution, written as

$$\pi(\Theta^{-1} | \ldots) \sim W_r \left(\left[(\omega R) + \sum_{i=1}^{n} (z_i - T^{(i)}_{x_i} \mu_{x_i})(z_i - T^{(i)}_{x_i} \mu_{x_i})' \right]^{-1}, n + \omega \right), \tag{6.34}$$

where "..." denotes "all other variables." Similarly, the full conditional for Ψ^{-1} is given by

$$\pi(\Psi^{-1} \mid \cdots) \sim W_r \left(\left[(\omega R) + \sum_{i=1}^{n} (y_i - \sum_j H_{ij} z_j)(y_i - \sum_j H_{ij} z_j)' \right]^{-1}, n + \alpha \right). \tag{6.35}$$

Finally, with the Potts MRF, the full conditional for $\pi(x_i = c | \ldots)$ is given by

$$\begin{aligned} \pi(x_i = c | \ldots) \quad \propto \quad & \exp \left\{ \sum_{p=1}^{r} \Theta^{-1}_{p,p} (z_i^p \, t^{i,p}_{c_i} \, \mu_c^p - \frac{1}{2} (t^{i,p}_{c_i})^2 \, (\mu_c^p)^2) \right. \\ & \left. + \sum_{p<q} \Theta^{-1}_{p,q} \, (z_i^p \, t^{i,q}_{c_i} \, \mu_c^q + z_i^q \, t^{i,p}_{c_i} \, \mu_c^p - t^{i,p}_{c_i} \, t^{i,q}_{c_i} \, \mu_c^p \, \mu_c^q) + \sum_{j \in \partial i} \beta_{ij} 1_{[x_i = x_j]} \right\}, \end{aligned} \tag{6.36}$$

with an adjustment to the final term for the Descombes et al. prior distribution.

6.5 Summary

Bayes theorem can be used to create a posterior probability distribution from various likelihood densities and prior distributions. In the present context this allows physically-based models of remote sensing processes to be integrated with a model of the spatial configuration of forest classes. Point estimates of the desired classification are obtained from the posterior distribution. Given the high dimensionality of the integrals involved exact optimization is impossible. Indirect estimates can be derived using stochastic simulation techniques. MCMC generates a sequence of dependent classifications that form a Markov chain. Estimates of image attributes are based on these samples, formed by sampling from the full conditional distribution of each component of the posterior. Componentwise MCMC dramatically reduces the computational expense and complexity.

Chapter 7

RESULTS

An operational forest structure classification model must yield consistent and accurate solutions. Even before a formal accuracy assessment is carried out the stability of its results should be evaluated with regard to topography, illumination geometry, modulation transfer characteristics, and training area selection. Furthermore, the model solutions may depend on specific formulations and parameter estimates. This classification model requires the specification of particular BRDF and MRF models. Hyperparameters are required to fully define the MRF. Several techniques are available for supervising the classification. As discussed in the previous chapters, all of these choices involve numerous assumptions and trade-offs.

The first seven sections of this chapter outline the necessary steps to set up, parameterize, and run the model. First, Section 7.1 describes the remotely sensed image data used. Section 7.2 reviews several common forest classification schemes in the Pacific Northwest and the choice of classes for this evaluation. Training stand and GIS information is discussed in Section 7.3. The assumptions required for estimating the NBRFs from radiosity simulations are presented in Section 7.4. With these estimates, maximum likelihood was used to produce point estimates of normalized class means and precisions. These estimates can be used directly or indirectly, as prior parameters, to supervise the classification. Section 7.6 describes the construction of PSF half-widths from LSF estimates. The implementation of MCMC is outlined in Section 7.7.

Repeated classifications under different circumstances and model formulations is a logical and useful means of evaluating the stability of the solutions (Demoment

1989). Section 7.8 considers four methods of training the classifier. The choice and parameterization of the spatial prior distribution is considered in Section 7.9. The effective topographic correction is examined in Section 7.10. Section 7.11 evaluates the sensitivity of the model to variation in the illumination geometry. The stability of the solutions with regard to training area selection is assessed in Section 7.12. A cursory test of spatial extensibility is presented in Section 7.13. Chapter 8 discusses these results.

7.1 The remotely sensed data

Several Landsat TM data sets were made available for the North and South Cascades in Washington State and in northern California. Although some preliminary analysis was conducted with the California data, the distance from Seattle precluded necessary field surveys. High quality field data accompanied the North Cascade imagery, but almost the entire study area was on a south facing slope. In order to evaluate the model's ability to account for anisotropic reflectance a variety of forest types are needed on a full range of slopes and aspects. Therefore, a study area was selected at the north end of the Gifford Pinchot National Forest in the Cispus Ranger District. The area includes both managed and unmanaged forests covering mountainous terrain in the South Cascades.

At least two images are necessary for evaluating variation in illumination geometry and atmospheric scattering. The first image was acquired by the TM sensor on board the Landsat-5 satellite on July 7, 1991. The solar elevation and azimuth angles were 56° and 125°, respectively, at the time of acquisition. The data was originally resampled to the Universal Transverse Mercator (UTM) projection system using the cubic convolution algorithm and resulting in a pixel size of 25 m. The second image was also acquired by the TM-5 sensor on August 10, 1992. The solar elevation and azimuth angles were 51° and 132°, respectively. The data was originally resampled

to the Space Oblique Mercator (SOM) projection system using the cubic convolution algorithm and resulting in a pixel size of 28.5 m.

Several steps were required to prepare these images for classification and analysis. The band 5 data did not conform well with the other bands. Thus, band 5 was shifted vertically one pixel and horizontally 2 pixels to conform with the other bands. The August image was co-registered to the July image in UTM coordinates using a forward affine transformation with 11 registration points. The transformation RMS error was 20.7 and 16.3 m along the horizontal and vertical directions, respectively. At just over half a pixel size this should be considered quite good. The UTM projection was chosen for two reasons. The ancillary GIS and DEM information was already available in UTM coordinates and previous research suggests that the SOM products do not conform as well to surface geometry as UTM products (Bryant et al. 1985). Nevertheless, visual inspection revealed registration discrepancies between the co-registered images and the digital map of stand boundaries. All bands were shifted vertically one pixel and horizontally 3 pixels to conform to the ancillary data.

A subset of 250 rows by 200 columns was clipped from the original July image. Then both images were resampled to 30 m to corresponds to the TM sampling interval (see Section 4.1.4) and to the DEM resolution. The nearest neighbor algorithm was used for resampling. This produced subsets of 208 rows by 167 columns.

7.1.1 Path radiance offsets

Path radiance offsets for each image were qualitatively assessed by considering the DN values for large reservoirs near the subset areas which should have near zero reflectance. These assessments were adjusted slightly after inspection of DN values for dense vegetation in the subset regions. Vegetation is more reflective in the green band than either the red or blue bands. By assuming uniform path radiance over the extent of the subset region the offsets can be subtracted from image DNs prior to classification as was done for this study. Table 7.1 shows these offsets.

7.1.2 Percentages of diffuse irradiance

Without detailed measurements at the time of image acquisition it is very difficult to estimate the proportion of diffuse irradiance. The percentages $DI(\lambda)$ are required for Equation 3.10 and influence the estimation of the NBRFs. Deering et al. (1994) reported rough estimates and in general the trend will resemble that of path radiance as scattering decreases with increasing wavelength. On this basis these percentages were arbitrarily determined and are also given in Table 7.1. Given such limited information the same proportions are used for each image and assumed uniform over the subset region.

Table 7.1: Diffuse irradiance proportions and path radiance offsets.

Band	$DI(\lambda)$	DN offsets	
		July	August
1	0.18	48	36
2	0.12	7	4
3	0.10	5	2
4	0.08	2	1
5	0.07	0	0
7	0.06	0	0

7.2 The classification scheme

The classification of continuous forest growth patterns into discrete stages facilitates an understanding of stand dynamics as well as mapping their distribution. Several classification schemes have been developed to describe Pacific Northwest forest structures and stand dynamics. From the forest research perspective, the systems proposed

by Brown (1985) and Oliver and Larson (1990) have become widely adopted. Brown (1985) describes six stand conditions that follow clearcutting and broadcast burning:

- Grass-forb
- Shrub
- Open sapling-pole
- Closed sapling-pole
- Large sawtimber
- Old growth

Oliver and Larson (1990) outline four stages of stand development:

- Stand initiation
- Stem exclusion
- Understory reinitiation
- Old growth

Resource planners and field foresters have used these general schemes to guide the development of specific classification systems suited to particular objectives. Hitchcock (1996) adapted eight classes from Brown (1985) for designing a landscape management plan for 15,000 acres of forest managed by the Washington Department of Natural Resources (DNR). These classes describe eight stand structures:

- Clearcut
- Grass-forb-shrub

- Open sapling-pole
- Closed sapling-pole
- Owl dispersal habitat
- Large sawtimber
- Old growth
- Hardwood dominant

At a larger scale the DNR has developed a digital database for Washington State's 20 million acres of forest land (Green et al. 1993). One layer of this database considers the following four cover classes:

- Cleared forest
- Early seral stage
- Mid-seral stage
- Late seral stage

Although these classification schemes are similar, the degree of similarity depends on the criteria used to define each class. Oliver and Larson's scheme describes processes such as initiation and exclusion, characterizing the continuous nature of forest development. Various structures may be associated with a single process. For example, Figures 7.1 and 7.2 show two structures associated with the stand initiation process. In contrast the seral stages adopted by the DNR are defined on the basis of percent crown closure and tree size. For example, the late seral stage has a total crown closure greater than 70 percent, of which 10 percent is in trees greater than 21 inches in diameter, and less than 75 percent of the canopy hardwood.

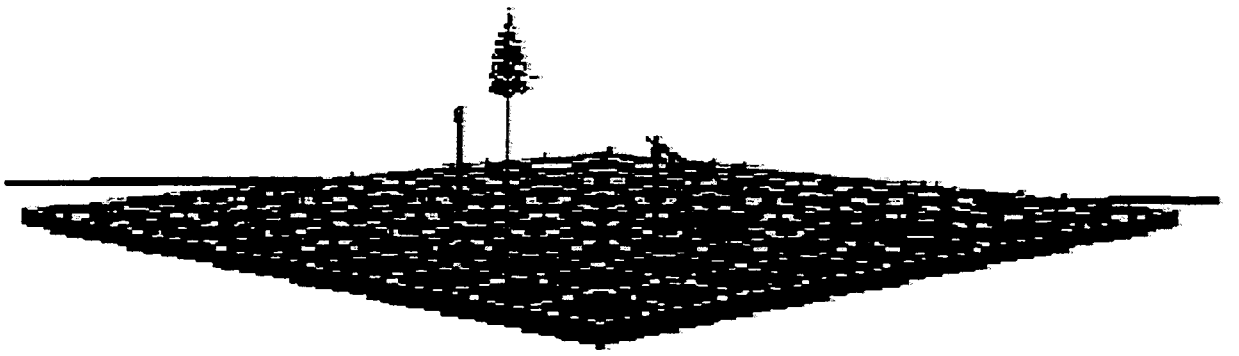


Figure 7.1: Clearcut forest characteristic of the stand initiation process.

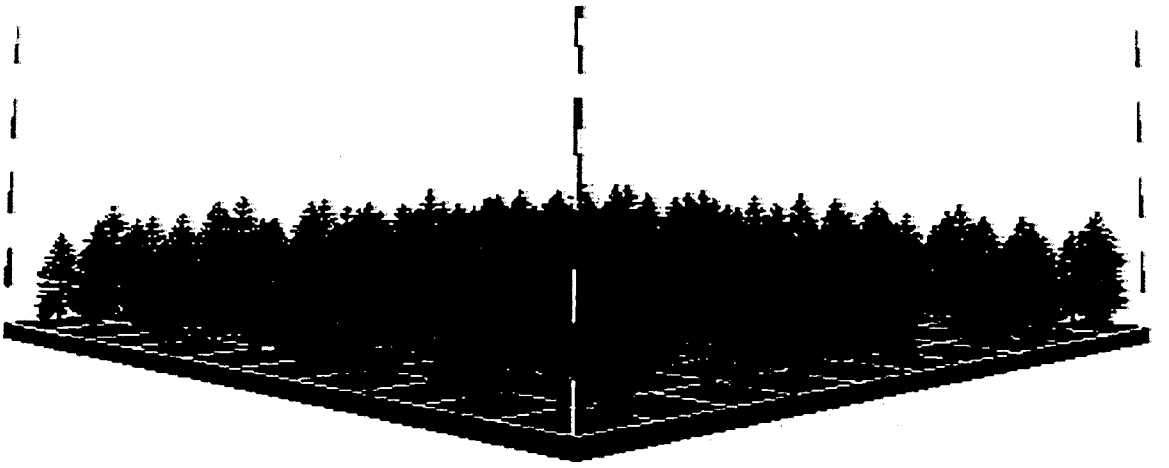


Figure 7.2: Advanced initiation structure characteristic of the stand initiation process.

In general these classification systems do not distinguish between species, although in the Northwest they have been applied to conifer dominated forests. The system of Green et al. (1993) considers all forests having greater than 75 percent hardwood crown closure as cleared forest. Hitchcock (1996) designates a separate class for hardwood dominated stands and discriminates dispersal habitat from large saw timber on the basis of the hardwood component of a stand. Hardwoods such as Red Alder predominantly grow in riparian zones.

Individual remotely sensed images can, at best, be used to infer forest structures, but not processes. The following six classes have been defined for evaluating the model developed for this study:

- Clearcut (CC)
- Advanced initiation (AI)
- Early stem exclusion (EE)
- Late stem exclusion (LE)
- Understory reinitiation-old growth (UR)
- Hardwood dominant (HW)

These classes are intended to represent the forest structures commonly found in the Northwest and integrate the concepts presented by the most widely used classification schemes. While mixed hardwood-conifer stands are not represented in this system, in the context of probabilistic classification mixed stands may be classified as having moderate probability of being either hardwood or conifer provided each of those classes are represented and the mixture does not represent another class in the scheme. As indicated by Figures 7.1 and 7.2 the stand initiation stage has been broken down into two distinct structures, recent clearcut and advanced initiation. Also note that the

model shown in Figure 3.4 is intended to represent the canopy of the stand shown in Figure 7.2. Stands may remain in the stem exclusion process for decades. Thus, the stem exclusion process is separated into early and late structures. Finally, the understory reinitiation and old growth structures have been combined.

7.3 Training stand and GIS information

Estimating parameters from image data corresponding to training stands represents one approach for supervising the classifier (see Chapter 5). Measurements of training stands are also needed for parameterizing the tree surface model as described in Section 3.5.1. To accomplish this task several field surveys were conducted to sample plots within stands that epitomized each of the six structures described above. In principle training stands should be as small as possible as the need for training limits the utility of remote sensing in the first place. Summaries of the training plot measurements are presented in Table 7.2 as they are required for the tree surface model. Mean heights are reported in meters, density is in units of trees per hectare (TPH). Live crown ratios (LCR) are also reported. The clearcut parameters were arbitrarily designated. The legend colors describe the classes in the color figures below.

Table 7.2: Class definitions.

Class	Shape	Height Distribution	LCR	TPH	Legend
CC	cone	$N(1.5, 0.25)$	0.95	3000	cyan
HW	circle	$N(11.0, 8.4)$	0.125	2285	yellow
AI	cone	$N(5.9, 1.64)$	0.9	988	brown
EE	cone	$N(14.6, 20.0)$	0.55	1605	red
LE	cone	$N(21.3, 41.0)$	0.36	1350	green
UR	cone	$N(50.6, 17.5)$	0.4	238	blue

The present stand structures reflect past management practices. The relatively low density of the AI class is indicative of precommercial thinning, while the higher density stem exclusion classes pre-date this era of intensive management. Knowledge of stand areas and past management practices guided training stand selection. Stand boundary, vegetation, and other resource inventory information was available from the Forest Service's GIS database. For the purpose of comparison to the image classifications, the information contained in the GIS was also classified into one of the six classes on a stand by stand basis for each stand in the subset area. Figure 7.3 shows the classification of the GIS data. Much of this data is based on photo interpretation, and to that extent Figure 7.3 represents a classification of an interpretation. However, some of the Forest Service information includes records of previous silvicultural treatments, particularly useful for classifying younger structures. In neither case does Figure 7.3 represent the "true" classification that would be observed at a scale corresponding the TM data. Rather the GIS classification provides a rough gauge for qualitatively assessing the MPM estimates. A gray scale representation of the corresponding DEM data is shown in Figure 7.4, with lighter shades representing higher elevations. This is useful for identifying patterns in the classifications that may relate to topography. Elevations in this subset range from 396 to 1513 meters.

Equation 4.8 is based on the assumption that the image data is normally distributed given z. However, as z is an unobserved image it is not possible to directly evaluate this assumption. For reference, histograms of the data for each training stand and image are presented in Appendix A. The ML classifier assumes that y is normally distributed given x. Many of the histograms vary from image to image; see, for example, CC bands 1 and 2, AI band 4, and UR band 5. Many histograms are skewed; see, for example, HW band 5 in July, and EE and LE band 4 in August. The data captures the projected area effects of topography and illumination geometry which z models. Thus, it is possible that if z could be observed, then y would not be skewed.

Figure 7.3: Classification of ancillary GIS information for Cispus subset.

Figure 7.4: USGS Digital Elevation Model. Brighter shades correspond with higher elevations.

Also in Appendix A, bands 4 and 5 exhibit more variation than the other bands. Given the greater reflectivity of photosynthetic vegetation in these bands, they are more sensitive to variation in mutual shadowing. Clearcuts are more variable in all bands as each data sample represents a mixture of bare ground and shrubby vegetation each having rather distinct reflectance patterns. This is also the case for the AI training stand in which pre-commercial thinning has left some bare ground exposed. The large proportion of path radiance in band 1 is also evident. Since the two images have not been calibrated to one another, the absolute DN values between images cannot be compared. The August values are lower because of the lower sun elevation.

7.4 NBRF estimation via radiosity simulation

7.4.1 Modelling mutual shadowing

Section 3.5.1 described how tree surface models can be generated from the parameters given in Table 7.2. Examples are shown in Figures 3.4, 3.5, and 3.6. For six classes on 13 slopes there are 78 tree surface models needed for this study. The programs generating these surfaces have been written in C.

Section 3.2 described the significance of mutual shadowing between crowns in a forest canopy. Crown shadowing patterns are a major factor in determining anisotropic reflectance at the canopy scale. The radiosity model explicitly captures this effect as illustrated in Figure 7.5. For each of the two solar elevation angles corresponding to the images being classified these graphs show the proportion of the tree surface in shadow as viewed from nadir for slopes perpendicular to the solar principal plane. Negative slopes face away from the source of illumination.

As expected Figure 7.5 shows the percentage of mutual shadowing increases for decreasing solar elevation. At $-60°$ the entire tree surface is shadowed by the hill slope with the exception of a few exposed tree tops. Among the conifer classes, as

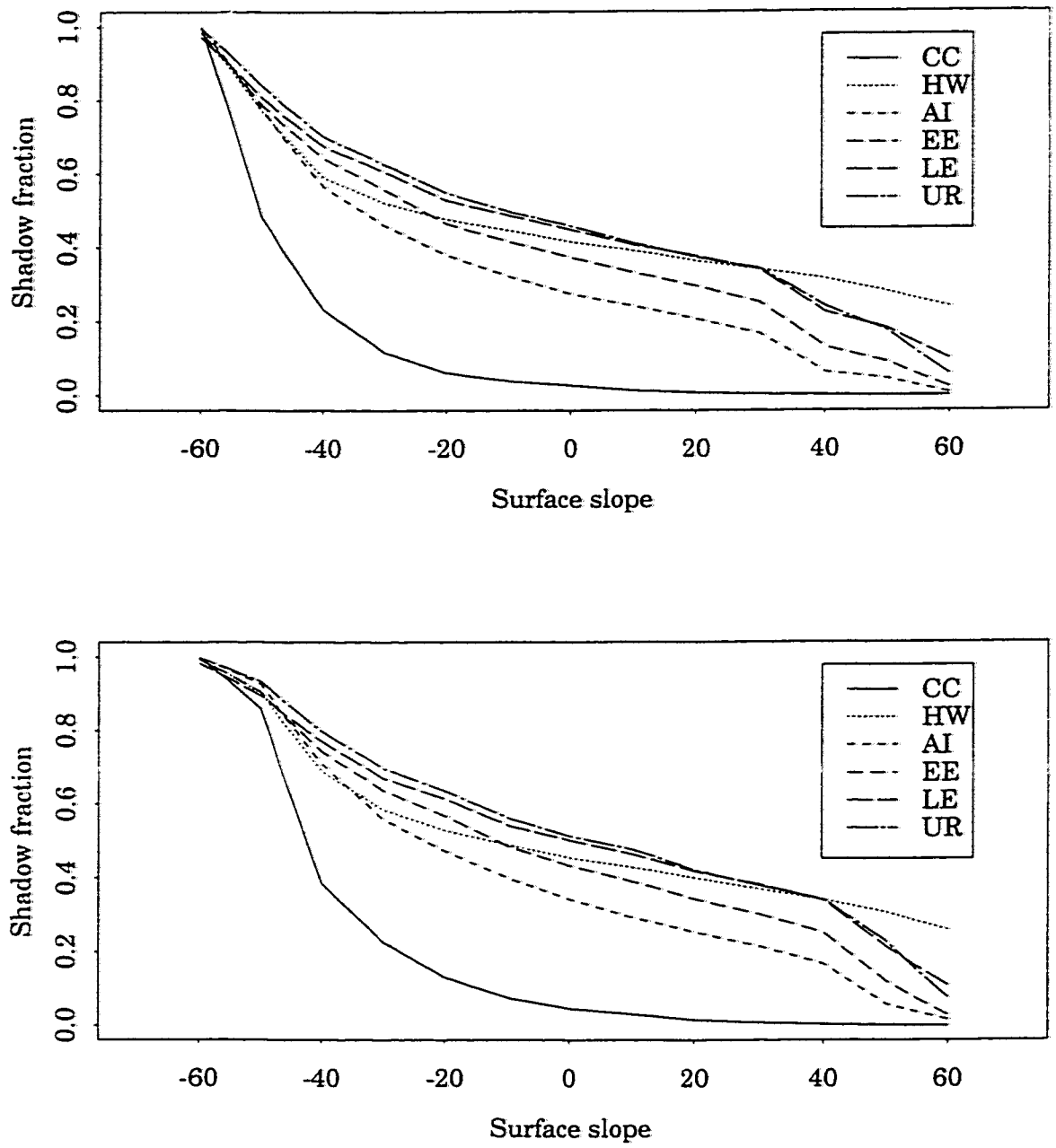


Figure 7.5: Principal plane proportions of between-crown shadows as viewed from nadir as a function of surface slope, class, and solar elevations 56° (top) and 51° (bottom).

tree sizes increase and density decreases, shadowing increases. The denser spherically modelled hardwood canopy is less sensitive to slope variation than the conifer canopies. Clearcuts exhibit much less shadowing than the tree surfaces. These encouraging results support the notion that the radiosity approach is a physically-based method for simulating anisotropic forest reflectance.

7.4.2 Canopy reflectance

In order to simulate multiple scattering, reflectance values must be assigned to the tree surfaces for each class. The procedure for distinguishing between crowns, stems, and the ground was presented in Section 3.5.1. Reference spectra obtained from the remote sensing literature were used to guide these assumptions (Elachi 1987; Williams 1991; Lillesand and Kieffer 1994). Although measuring the training stands with a field spectrometer may have been more appropriate, given the opaque assumption, such a field campaign was not undertaken. Table 7.3 lists the tree crown reflectances assumed for the radiosity simulations.

Table 7.3: Class reflectances specified for the radiosity model.

Band	CC	HW	AI	EE	LE	UR
1	0.15	0.05	0.05	0.05	0.05	0.05
2	0.15	0.05	0.05	0.05	0.05	0.05
3	0.15	0.05	0.05	0.05	0.05	0.05
4	0.25	0.55	0.45	0.35	0.35	0.25
5	0.35	0.35	0.35	0.15	0.15	0.15
7	0.15	0.05	0.05	0.05	0.05	0.05

Tree stem and ground reflectances were assumed not to vary with wavelength in order to moderate the already significant computational burden. Only a small

percentage of visible surface elements were either stems or bare ground, thus this assumption should not significantly influence the results. Conifer stems were assigned 0.05 reflectance, hardwood stems 0.1, and clearcut shrubs 0.15. The bare ground under the conifer and hardwood canopies was designated as 0.2 reflective and clearcut soil 0.25 reflective.

7.4.3 Radiosity simulation and normalization

The general procedures for applying the radiosity model have been outlined in Sections 3.5.4 and 3.5.5. All C programs needed to implement the model were optimized at the -O4 level. Several checks showed that the results were not sensitive to the level of optimization. As a total of 9,360 radiosity simulations were needed, a UNIX script program was written to automate the process.

At this point radiosity estimates are available for each of the 130 slope-aspect combinations. An S program was written to add the the corresponding component of diffuse radiance to each of these estimates, normalize to a horizontal surface, and estimate the quartic polynomial. The output of the S program is a coefficient matrix for each class-wavelength-solar elevation combination.

For each image being classified the NBRFs for each class and bandpass are computed for each pixel depending on its slope and aspect. Slopes and aspects are in turn calculated from a DEM co-registered to the image data. Although the DEM arrived co-registered to the other Forest Service GIS data, a further check was conducted to assure the DEM was accurately registered to the image data. No discrepancies were noted in displaying two bands of image data along with a raster of $\cos \iota$ as RGB.

The quality of the DEM will influence the results. In the United States many have been manually digitized from mylar contour maps. Producing accurate contour maps of forested terrain is a difficult task. Raster DEMs were interpolated from the digital contour maps. Further, when calculating slope and aspect values for DEM grid cells as a function of adjacent cells, the resulting values correspond to an area greater than

the original cell resolution (Hodgson 1995).

7.5 Normalized class mean and precision estimation

Maximum likelihood estimates of μ_c, η_c, and Γ_c can be obtained from preprocessed image data as described in Section 5.1.1. Each training stand consists of approximately 130 pixels. Histograms for each class and image are presented in Appendix B. In general these histograms are less skewed than those presented in Appendix A; see, for example, HW band 5 in July, and EE and LE band 4 in August. This suggests that topography may be contributing to the shape of the histograms in Appendix A. In a few instances, notably LE band 5 for July and CC band 7 for August, normalization increased skew. Further, the symmetry of these histograms increases the confidence with which ML estimates of the class means can be obtained.

From each training stand 50 pixels were randomly selected and normalized using the corresponding reflectance factors. Mean vectors and covariance matrices were then calculated from these samples. Arc/Info macro language and S script programs were written to implement each of these steps. These estimates of the class means are presented in Tables 7.4 and 7.5 for the July and August images, respectively.

Table 7.4: Class means estimated from normalized low slope data for July.

Band	CC	HW	AI	EE	LE	UR
1	33.7	19.0	22.0	15.2	16.1	12.5
2	27.4	24.5	25.2	18.0	17.7	16.4
3	37.8	20.0	25.6	15.4	15.5	15.5
4	50.0	148.6	96.9	94.9	76.1	49.6
5	109.9	77.6	84.0	42.3	41.9	33.5
7	68.1	21.7	31.7	10.3	11.8	11.0

Table 7.5: Class means estimated from normalized low slope data for August.

Band	CC	HW	AI	EE	LE	UR
1	21.5	13.0	14.3	11.3	12.2	9.2
2	20.0	15.9	16.5	12.8	13.0	12.3
3	22.8	12.8	15.2	10.1	10.8	10.4
4	51.1	103.7	71.2	70.8	57.7	39.2
5	79.9	54.0	52.0	29.8	31.1	24.4
7	37.7	12.6	16.4	6.3	7.7	6.7

These DN values are consistent with the references cited in Section 7.4.2. The larger proportion of non-photosynthetic vegetation in the clearcuts results in that class having a higher reflectance in the visible and SWIR bandpasses. The broad leaves and high density of the HW class produces very high levels of reflected radiance in the NIR. As the canopies of the older conifer classes open up, gaps between trees leave shadows that reduces the reflectance of those stands.

7.6 System PSF half-width estimation

At the outset of this study a high degree of class conditional dependence was confirmed by regressing the mean of a pixel's neighbors against its value for all pixels in a given training area. All regressions showed strong positive correlation. Estimates of the system PSF half-widths required to define the model given by Equation 4.8 were obtained by assuming separability and calculating orthogonal LSFs using the method of Anuta et al. (1984) described in Section 4.5.2. An S program was written to implement this method.

The study area includes managed stands and several harvest unit boundaries run horizontally and vertically across the image. Clearcut-forest boundaries provide the

step edge contrast necessary, except for band 4. The NIR bandpass required young forest-old forest boundaries to provide the necessary contrast. Thus, a total of four boundaries were identified and sampled for five pixels to either side of the edge along a length of ten pixels. As with path radiance and diffuse irradiance, the atmosphere and sensor effect have been assumed constant over the subset region. Boundaries characterized by taller trees casting shadows onto smaller trees were avoided. Figure 7.6 shows the pixels sampled to estimate the horizontal LSFs for the July image.

A cubic B-spline was fitted to the data assuming 6 degrees of freedom. In Figure 7.7 the fitted splines are shown relative to the mean sample values. The dashed line represents the step edge between the mean values of each homogeneous region. Finally, the LSF is found by taking the derivative of the spline and dividing by the amplitude change. The LSFs for each of the directions and both images are presented in Appendix C. These graphs also indicate $W_{\frac{1}{2}}$ spans and values. The half-widths range from 1.73 to 2.54 sampling intervals, implying effective resolutions from 52 to 76 m. In general the half-widths are wider for the IR than the visible bands and also wider for the August than July image.

The nature of the model requires that the system PSF be approximated by discrete groups of pixels. All estimated half-widths are longer than the TM linear IFOV, but less than three IFOV. Thus, they can be projected onto nine pixel regions, whereas the zero power boundary may encompass a much larger area. A linear approximation was used to model non-uniform responses over the region. Figures 7.8 and 7.9 display the estimated arrays for each image and bandpass. Note the values shown have been rounded for presentation.

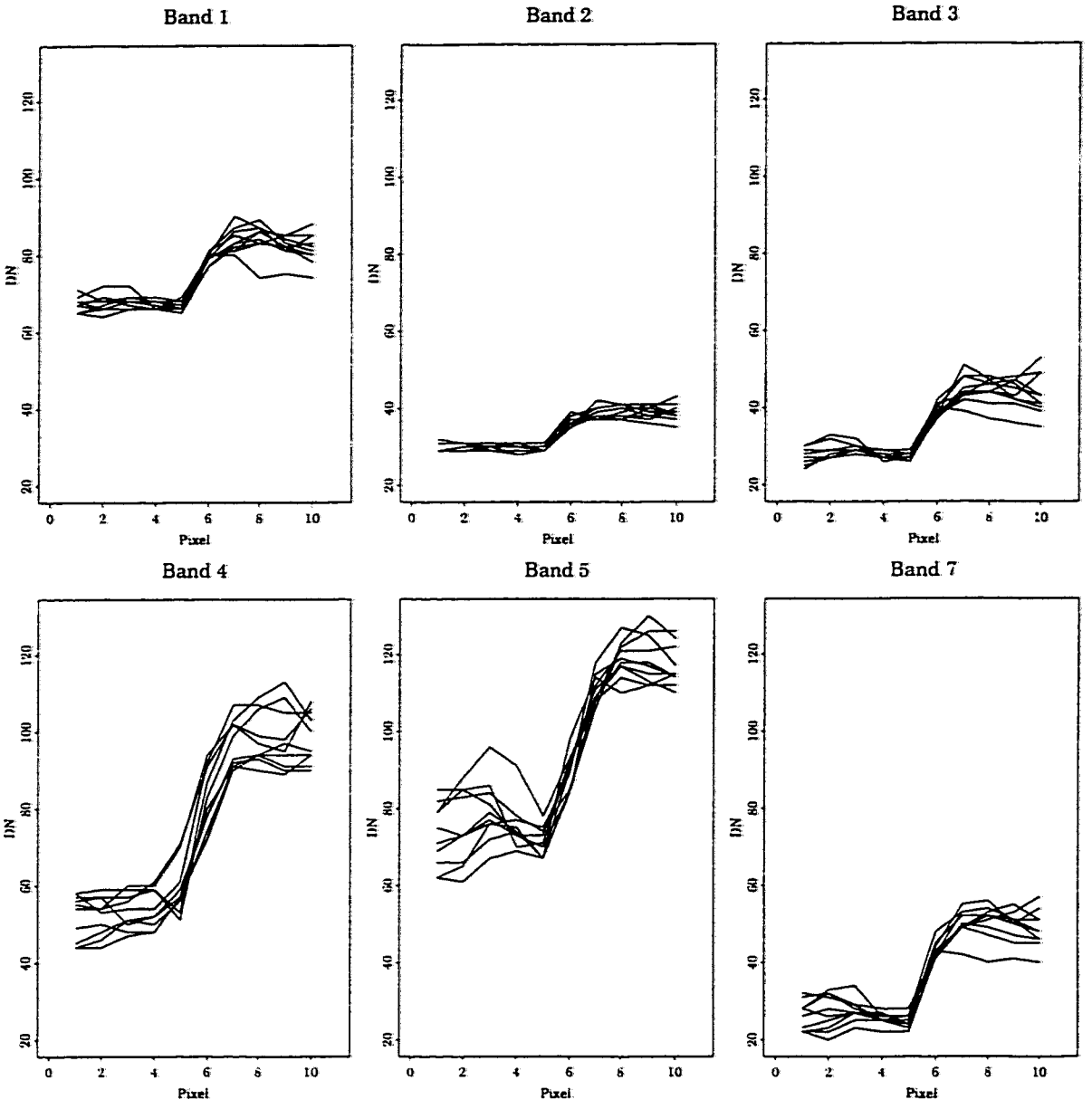


Figure 7.6: July image data sampled across horizontal contrast boundaries.

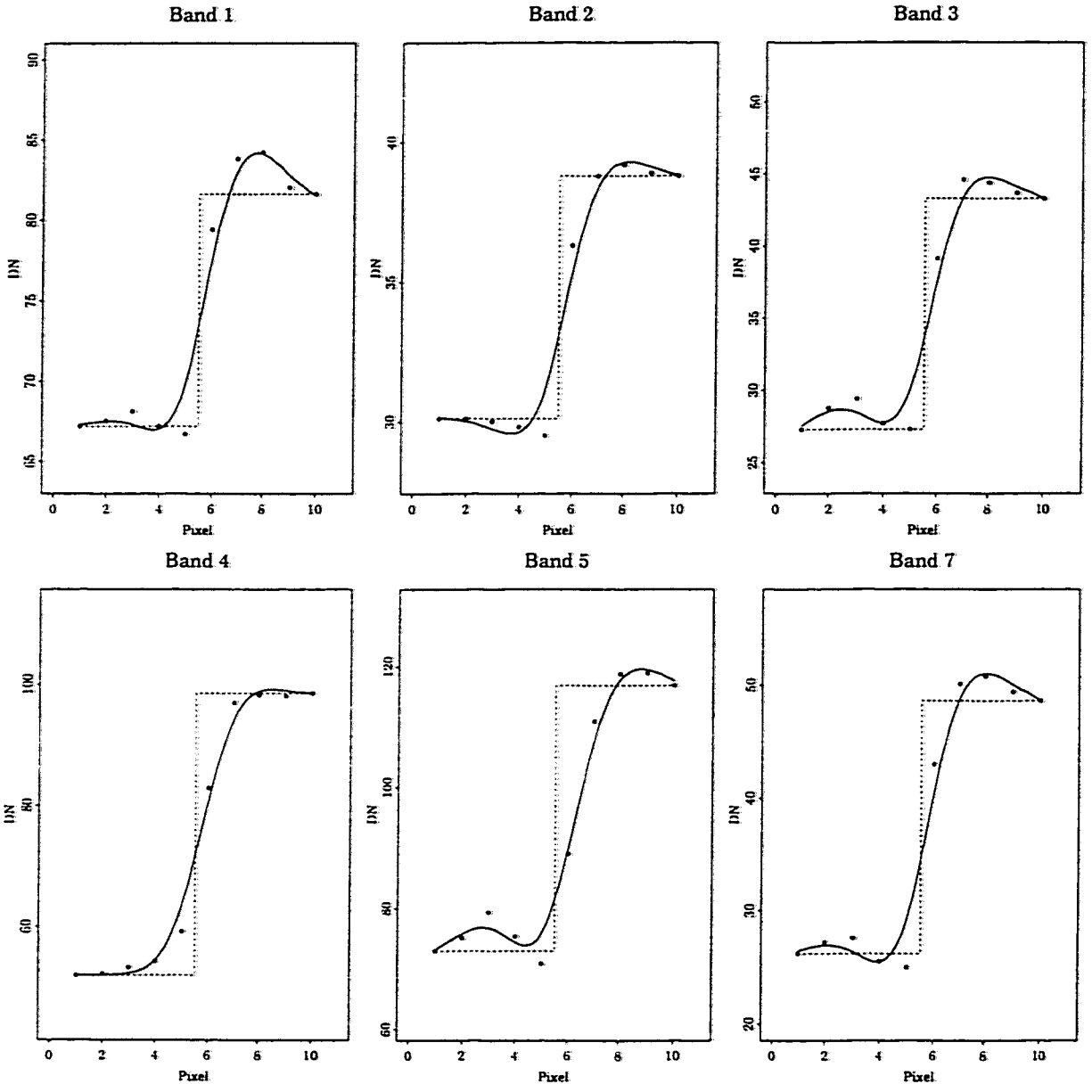


Figure 7.7: Cubic B-splines fitted to July image horizontal contrast boundaries.

Band 1

.048	.102	.048
.134	.336	.134
.048	.102	.048

Band 2

.047	.093	.047
.144	.340	.144
.047	.093	.047

Band 3

.049	.100	.049
.137	.332	.137
.049	.100	.049

Band 4

.075	.115	.075
.123	.223	.123
.075	.115	.075

Band 5

.074	.114	.074
.124	.229	.124
.074	.114	.074

Band 7

.061	.123	.061
.116	.280	.116
.061	.123	.061

Figure 7.8: System PSF arrays estimated for Equation 4.8 from July data.

Band 1

.066	.109	.066
.132	.256	.132
.066	.109	.066

Band 2

.069	.110	.069
.130	.246	.130
.069	.110	.069

Band 3

.067	.118	.067
.122	.253	.122
.067	.118	.067

Band 4

.076	.111	.076
.127	.221	.127
.076	.111	.076

Band 5

.073	.125	.073
.114	.230	.114
.073	.125	.073

Band 7

.074	.121	.074
.117	.226	.117
.074	.121	.074

Figure 7.9: System PSF arrays estimated for Equation 4.8 from August data.

7.7 Simulation

7.7.1 Implementation and optimization

Markov chain simulation and sampling was implemented in the C programming language on various UNIX platforms. The program has over 3,100 lines of code. As this method is computationally expensive, many efforts were made to decrease run times. On Sun workstations the programs were optimized at the `-xO5` level and at the `-O4` level on DEC workstations. Several simulations were run without optimization to assure that the optimization did not influence the results. On average the program ran at 700 iterations per hour. Profiles showed that updating the z array accounts for up to 70 percent of the run time as each of its almost 200,000 components requires numerous multiplications and a random sample from a normal distribution for every iteration. Where possible repetitive operations were calculated once and saved at the expense of increased memory requirements. For simulations with the fifth-order MRF a look-up table was developed by Byers (1998) to store and retrieve the energy values assigned to each of the 10,077,696 cliques. While the look-up table reduces the number of operations per iteration, it requires 80 Mbit of memory. All variables and arrays were specified as double precision. Preliminary experiments indicated slight differences between single and double precision. Although the program is memory intensive, given the relatively small subset size, double precision was chosen to minimize precision related errors.

7.7.2 Updating the full conditionals

For every iteration the parameters and array components are updated in the following order, Θ^{-1}, Ψ^{-1}, z, x, and μ_c. Updating Θ^{-1} and Ψ^{-1} first obviates the need for their initialization. The data y are used to initialize z and various arbitrary classifications are used to initialize x. The updating order of each component of z and x is determined by a unique random permutation for each iteration (see Sec-

tion 6.4.5). For simulations updating μ_c, the maximum likelihood estimates are taken as initial values. All random number generation is carried out using the `drand48()` function. Except for the precision parameters updating the full conditionals presented in Section 6.4.6 is trivial. Updating the precisions via Gibbs steps requires randomly sampling from Wishart distributions. This is achieved using the method of Odell and Feiveson (1966). For $W \sim W_r(\Sigma, df)$,

$$W^* = \frac{1}{df} L A^* L^T \sim W_r(\Sigma, df), \tag{7.1}$$

where $LL^T = \Sigma$ is the Cholesky decomposition of Σ and A^* is a matrix assembled from randomly sampled normal and chi-square variates.

7.7.3 Convergence diagnostics

Several methods for diagnosing convergence were employed (see Section 6.3.4). Primary consideration was given to stability in the number of sites in each class along with initializing x to random and single class configurations. Although the number of accepted updates per iteration and two image functionals were also computed.

$$f_1(z, x) = \sum_{i=1}^{n} (z_i - T_{x_i}^{(i)} \mu_{x_i})' (z_i - T_{x_i}^{(i)} \mu_{x_i}), \tag{7.2}$$

$$f_2(y, z) = \sum_{i=1}^{n} (y_i - \sum_j H_{ij} z_j)' (y_i - \sum_j H_{ij} z_j). \tag{7.3}$$

Although each of these informal methods produced consistent results, the convergence functions tended to stabilize before the number of sites per class. Different starting points highlighted problems associated with specifying a prior distribution(s) for the class means (see Sections 5.1.2, 5.1.3, and 7.8). The length of the burn-in period was determined for each simulation on an individual basis. Simulations from single class initializations required a much longer burn-in period than random initializations.

7.7.4 *Sample sizes*

Sample sizes were evaluated on the basis of simulations supervised by point estimates of the class means and using the fifth-order MRF defined by $E_N = 2.5/1.35$, $E_E = 1.5/1.35$, and $E_J = 2.0/1.35$. The MPM estimates shown in Figures 7.10 and 7.11 are classifications of the August image from a random and an AI initialization, respectively. Each estimate is based on 5,000 samples per site. They appear very similar having 94.9 percent of the sites assigned to the same class. Figure 7.10 required a burn-in of 7,000 iterations, while Figure 7.11 required a burn-in of 17,000 iterations.

Like classifications of the July image from different initializations and sampled for 5,000 iterations yielded very similar results with 94.9 percent of the sites comparably classified. Increasing the sample size to 10,000 brings the two MPM estimates slightly closer together with 95.7 percent of the sites comparably classified. Figures 7.12 and 7.13 show the two MPM estimates based on 10,000 samples from random and AI initializations, respectively.

Returning to the August image, if the sample size is increased to 20,000, the percentage of sites having the same class increases to 95.9 percent. These estimates are shown in Figures 7.14 and 7.15, again from random and AI initializations, respectively. The sample size was increased by continuing the simulations that produced Figures 7.10 and 7.11, thus their respective burn-in periods were 12,000 and 22,000 iterations.

In each of these cases increasing the sample size brings the MPM estimates from two very arbitrary and different points in the state space closer together. While not entirely conclusive, this suggests that these samples have been obtained from the target distribution and that in the limit these estimates would converge to the actual MPM as given by the ergodic theorem (see Section 6.3.2). Given that sample sizes greater than 5,000 produced only marginal improvements that was taken as the sample size for most of the simulations, although some estimates are based on 10,000

Figure 7.10: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35$, and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = random; 5,000 samples; radiosity estimated NBRFs; low slope training stands.

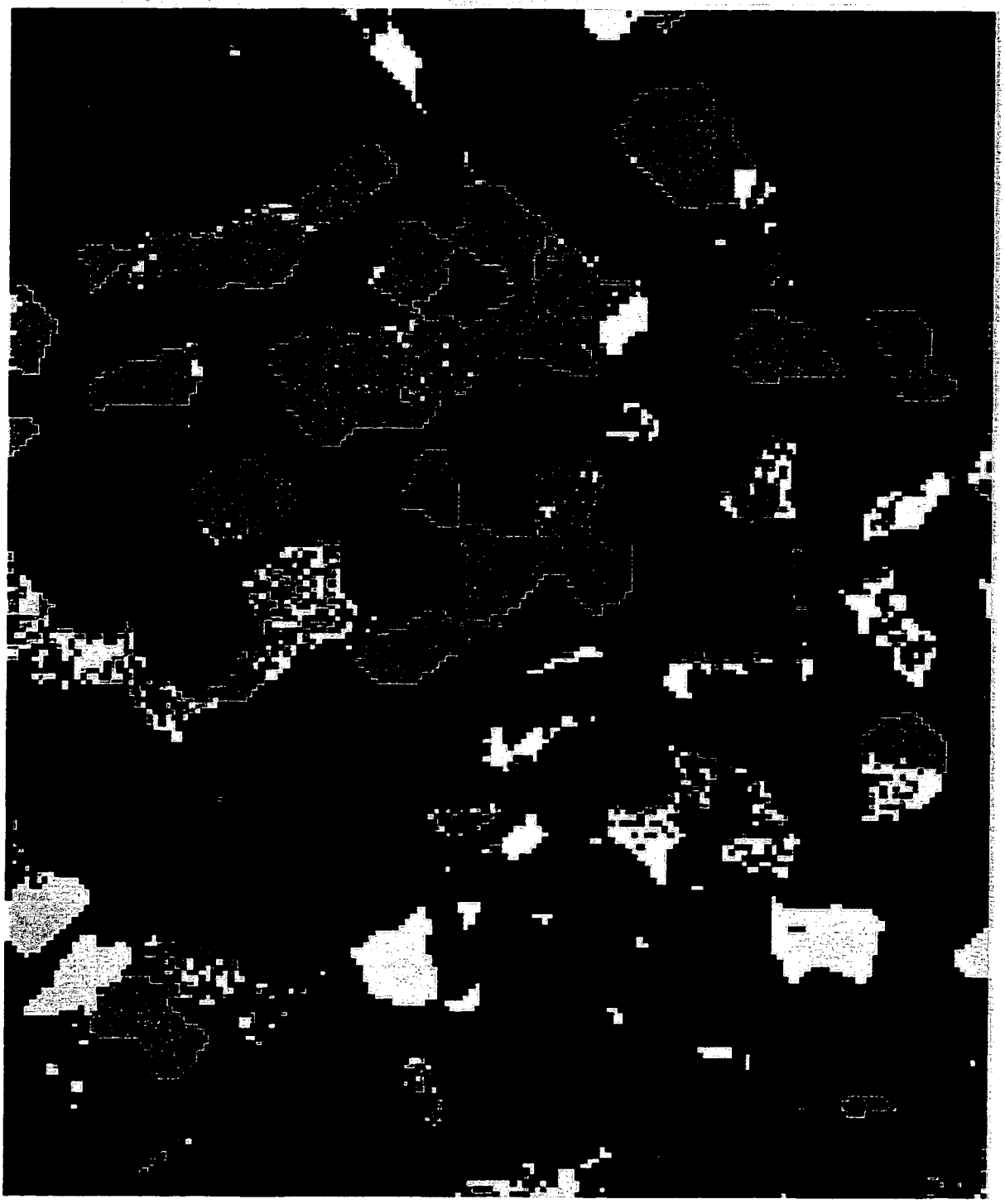


Figure 7.11: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35$, and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = AI class; 5,000 samples; radiosity estimated NBRFs; low slope training stands.

Figure 7.12: MPM: July; $E_N = 2.5/1.35, E_E = 1.5/1.35,$ and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = random; 10,000 samples; radiosity estimated NBRFs; low slope training stands.

Figure 7.13: MPM: July; $E_N = 2.5/1.35, E_E = 1.5/1.35$, and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = AI class; 10,000 samples; radiosity estimated NBRFs; low slope training stands.

Figure 7.14: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35$, and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = random; 20,000 samples; radiosity estimated NBRFs; low slope training stands.

Figure 7.15: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35$, and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = AI class; 20,000 samples; radiosity estimated NBRFs; low slope training stands.

samples for added confidence.

7.7.5 Border sites and pixels

Problems arise at the image border for updating those components of z and x. Given the PSF the full conditional of z_i considers pixel values two sites away in each direction. Likewise, the potential functions for the fifth-order random field require site values two pixels away. Thus, a two pixel wide border is set at the initial values and not updated. In the case of z the border pixels correspond to the data values. For x the border pixels are set to the classes obtained from maximum likelihood classification, regardless of the initialization of the interior sites.

7.8 Supervision technique

Chapter 5 presented four techniques for supervising the classification: ML estimates; a single prior distribution for the class means weighted by the training sites; unique priors for each class parameterized by the ML estimates; and unique priors parameterized by the ML estimates and weighted by the training sites. Each approach was evaluated by inspecting the MPM estimates corresponding to random and single class initializations.

7.8.1 Point estimation

Results obtained by treating μ_c, for all $c \in \Lambda$, as fixed were presented in Section 7.7.4. In Figures 7.14 and 7.15 approximately 95 percent of the training stand pixels are correctly classified and the same solution is reached from two starting points. With the exception of some areas of LE or UR structures, the results compare favorably with the GIS classification.

7.8.2 A single prior distribution

This method of supervision requires that the training stand sites remain fixed so as to uniquely influence the full conditional distributions for μ_c (see Section 5.1.2). Figures 7.16 and 7.17 are MPM estimates derived from random and CC initializations, respectively. As in Section 7.7.4 the fifth-order MRF defined by $E_N = 2.5/1.35$, $E_E = 1.5/1.35$, and $E_J = 2.0/1.35$ was used and 5,000 samples were collected. The burn-in period for Figure 7.16 was 20,000 iterations, and 105,000 iterations for Figure 7.17. With a prior distribution for the means it takes much longer for a single class initialization to burn-in as μ_c is allowed to vary and the prior for x favors spatial continuity.

The discrepancies between Figures 7.16 and 7.17 brings into question the reliability of this approach to supervision. Only 26.3 percent of the sites are consistently classified. While not unreasonable, Figure 7.16 is rather different from any of the fixed mean classifications and from the GIS classification. However, Figure 7.17 is troublesome. Large areas of mature forest structures are classified as clearcut, most of the clearcuts are AI, while the AI is primarily classified as either EE or HW. This is not surprising given the class means being sampled during equilibrium. Table 7.6 gives the means sampled during the final iteration. In equilibrium the standard deviation of the class mean full conditionals is 0.1 on average, thus the sampled means range vary little from iteration to iteration.

As expected the CC class mean resembles the ML estimate of the UR class mean given in Table 7.5, while the sampled UR class mean falls between the LE and UR means in Table 7.5. Similarly, the sampled AI mean compares with the ML estimate of the CC class. Of the 7,940 sites classified as CC in Figure 7.17 only 346 are classified as CC by the GIS data, and only 126 of those are fixed training sites. Thus, by allowing so much variability in the class means the possibility exists for the training information to be outweighed.

The class means represent spectral information, yet each site is also classified on

Figure 7.16: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35,$ and $E_J = 2.0/1.35$; single prior for all μ_c; x initialization = random; 5,000 samples; radiosity estimated NBRFs; low slope training stands.

Figure 7.17: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35,$ and $E_J = 2.0/1.35$; single prior for all μ_c; x initialization = CC; 5,000 samples; radiosity estimated NBRFs; low slope training stands.

Table 7.6: Class means corresponding to Figure 7.17 sampled during the final iteration.

Band	CC	HW	AI	EE	LE	UR
1	9.0	13.4	26.4	11.9	10.4	9.9
2	11.6	16.9	23.6	14.4	12.1	13.3
3	10.0	14.5	28.5	12.1	9.9	11.0
4	43.0	100.6	50.6	80.4	58.1	46.4
5	24.9	69.6	90.3	41.4	29.9	25.0
7	6.8	18.3	42.6	10.4	7.2	6.4

the basis of spatial information. The prior distribution for x in this model induces spatial homogeneity. Figure 7.17 was initialized as all CC and 4,109 of the CC sites in MPM estimate never had proposals accepted. In other words, most those sites have been incorrectly classified with 100 percent probability. While part of that is explained by the varied means, the lack of mobility is also attributable to the strength of the prior: approximately 1,000 sites per iteration have proposals accepted. This notion can be tested by down weighting the spatial information. Figures 7.18 and 7.19 compare to Figures 7.16 and 7.17, respectively, except that the MRF is defined by $E_N = 1.25, E_E = 0.75,$ and $E_J = 1.00$.

Reducing the strength of the prior increased the number of similarly classified sites from 26.3 to 66.5 percent at the expense of spatial homogeneity. As for the CC initialization, the increased mobility (roughly 5,500 proposals accepted per iteration) allows the simulation to move away from preserving the large homogeneous areas of incorrectly classified CC sites found in Figure 7.17. The sampled class means adjust accordingly. Table 7.7 shows the class means sampled during the last iteration of the simulation corresponding to Figure 7.19. With the exception of the AI class, these values are much more consistent with those of Table 7.5. Indeed, in Figure 7.19 many

Figure 7.18: MPM: August; $E_N = 1.25$, $E_E = 0.75$, and $E_J = 1.00$; single prior for all μ_c; x initialization = random; 10,000 samples; radiosity estimated NBRFs; low slope training stands.

Figure 7.19: MPM: August; $E_N = 1.25, E_E = 0.75$, and $E_J = 1.00$; single prior for all μ_c; x initialization = CC; 10,000 samples; radiosity estimated NBRFs; low slope training stands.

sites are classified as AI that correspond to older structures and this is evident in the class mean sampled for AI. Further Table 7.7 emphasizes the contrast presented by the sampled means associated with a stronger prior, shown in Table 7.6.

Table 7.7: Class means corresponding to Figure 7.19 sampled during the final iteration.

Band	CC	HW	AI	EE	LE	UR
1	27.7	12.2	8.7	10.9	10.1	8.4
2	24.5	14.7	11.4	11.8	13.9	11.7
3	30.0	12.3	10.3	9.6	10.7	9.8
4	51.9	100.6	44.1	60.5	46.8	35.9
5	97.8	56.1	27.4	25.4	23.7	21.5
7	46.2	13.4	7.6	5.6	5.6	5.7

In considering the random initialization, much fewer sites are classified as AI in Figure 7.18 and consequently the final mean sampled for AI band 4 is 70.4, almost identical to the corresponding value of 70.8 found in Table 7.5.

7.8.3 Unique prior distributions

The classifier may also be supervised by using ML estimates of the class means and covariances to parameterize a unique prior distribution for each class mean (see Section 5.1.3). This does not require the training sites to be fixed, nor must the hyperparameter estimates be taken from the image data being classified. Figures 7.20 and 7.21 show MPM estimates based on this supervision technique from random and CC initializations, respectively. Each estimate is obtained from 10,000 samples and uses the MRF defined by $E_N = 2.5/1.35, E_E = 1.5/1.35,$ and $E_J = 2.0/1.35$. Figure 7.20 required a burn-in period of 30,000 iterations, while Figure 7.21 required 185,000

iterations to reach equilibrium.

Between these two estimates 45.0 percent of the sites are similarly classified. As in Figures 7.16 and 7.17 a large area in the lower left portion of the image is inconsistently classified as either EE or LE, two spectrally similar classes. In Figure 7.20 the last sampled mean of class EE band 4 was 62.1, indicative of the larger area of LE and/or UR forest structures classified as EE. Furthermore, 89 percent of the sites in Figures 7.20 and 7.16 are similarly classified; they were initialized with the same random classification. This suggests that the initial classification may be influencing the sampled means as much, if not more, than the training data. However, as evident in Figure 7.19, the initialization has no influence on the results when the prior is down weighted.

In Figure 7.21 several areas are incorrectly classified as CC, but not to the same extent as in Figure 7.17. Only 3.7 percent of the CC sites in Figure 7.21 are classified as CC in Figure 7.3. Most of the clearcuts are classified as AI and many of the AI stands are classified as either EE or HW. As a result only 34 percent of the training stands are correctly classified, compared to 75 percent in Figure 7.20, primarily the LE and UR stands. All of these patterns are reflected in the sampled class means presented in Table 7.8.

Reducing the weight of the spatial information to $E_N = 1.25$, $E_E = 0.75$, and $E_J = 1.00$ increases the number of similarly classified sites between the two initializations from 45.0 to 55.4 percent. The resulting MPM estimates are remarkably similar to those in Figures 7.18 and 7.19 and are therefore not shown. With the unique mean priors there is less of an increase in mobility (4,000 to 5,000 proposals accepted per iteration) when compared to the single prior for the means. For the CC initialization the number of correctly classified training sites increased from 34 to 59 percent, but decreased from 75 to 52 percent for the random initialization.

Figure 7.20: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35,$ and $E_J = 2.0/1.35$; unique priors for each μ_c; x initialization = random; 10,000 samples; radiosity estimated NBRFs; low slope training stands.

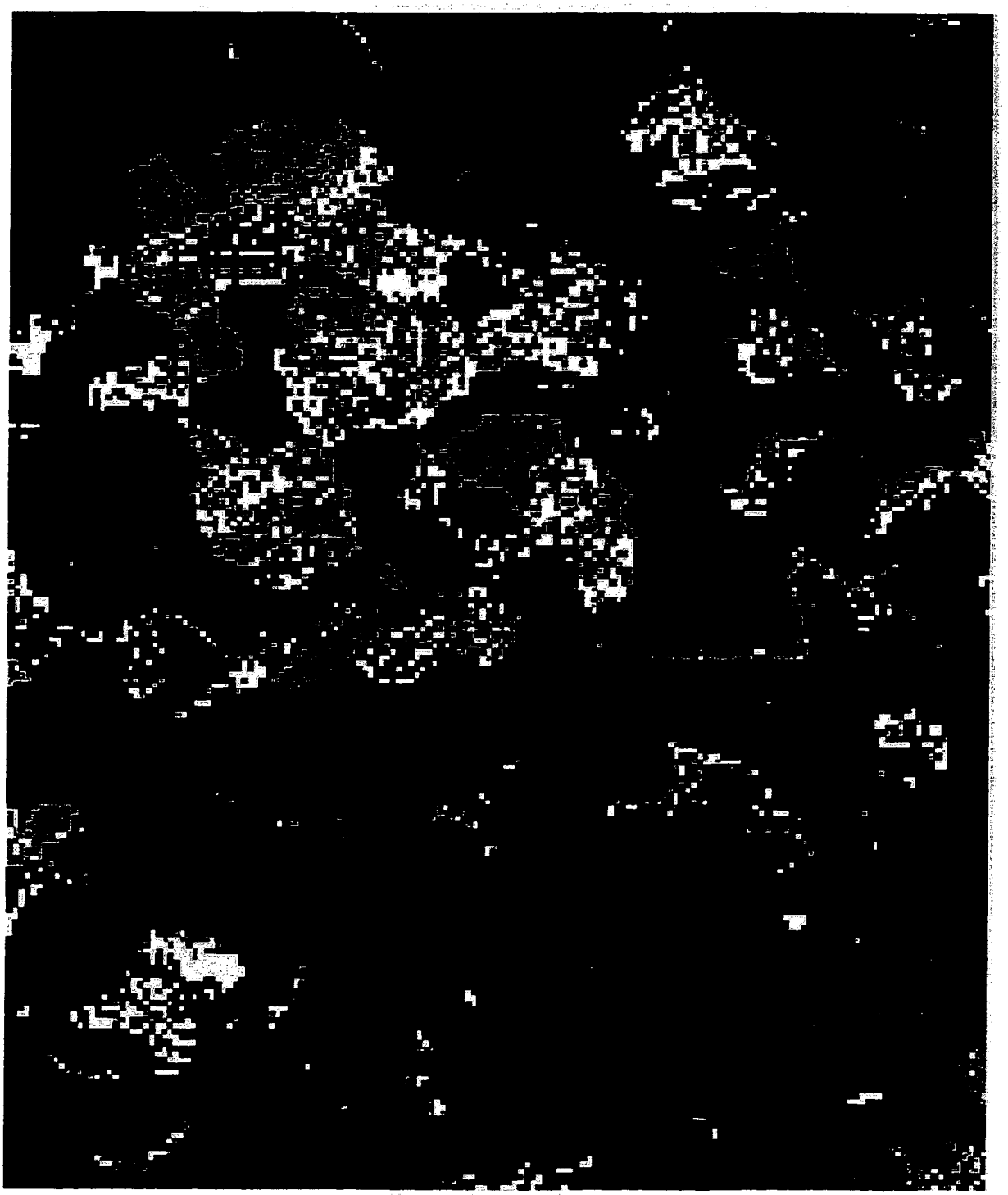


Figure 7.21: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35,$ and $E_J = 2.0/1.35$; unique priors for each μ_c; x initialization = CC; 10,000 samples; radiosity estimated NBRFs; low slope training stands.

Table 7.8: Class means corresponding to Figure 7.21 sampled during the final iteration.

Band	CC	HW	AI	EE	LE	UR
1	12.4	15.8	28.6	11.6	10.3	8.9
2	17.1	18.5	24.5	14.2	12.1	11.7
3	14.1	18.0	30.7	11.7	9.8	9.8
4	56.5	79.7	47.6	89.7	58.6	43.5
5	28.6	79.1	96.7	44.9	29.4	23.6
7	7.3	24.9	47.9	10.5	7.0	6.1

7.8.4 Weighted unique prior distributions

Increased weight can be assigned to the training data by combining the previous two supervision techniques. As a result the sites corresponding to the training stands are fixed and ML estimates are used to parameterize unique prior distributions for each of the class means. Figures 7.22 and 7.23 show MPM estimates based on this supervision technique from random and CC initializations, respectively. Each estimate is obtained from 5,000 samples and uses the MRF defined by $E_N = 2.5/1.35, E_E = 1.5/1.35$, and $E_J = 2.0/1.35$. Figure 7.22 required a burn-in period of 35,000 iterations, while Figure 7.23 required 130,000 iterations to converge.

These two estimates have 56.0 percent of their sites similarly classified. This suggests that better estimates are achieved when more weight is given to the training data supervising the classification. As above, the lower left portion of the image is inconsistently classified as either EE or LE forest. Figure 7.22, based on the same random initialization, has 89 percent of its sites in common with Figures 7.16 and 7.20. However, in Figure 7.23 older forest structures are no longer classified as clearcuts. In fact, Figure 7.23 resembles Figures 7.14 and 7.15 in which the class means were

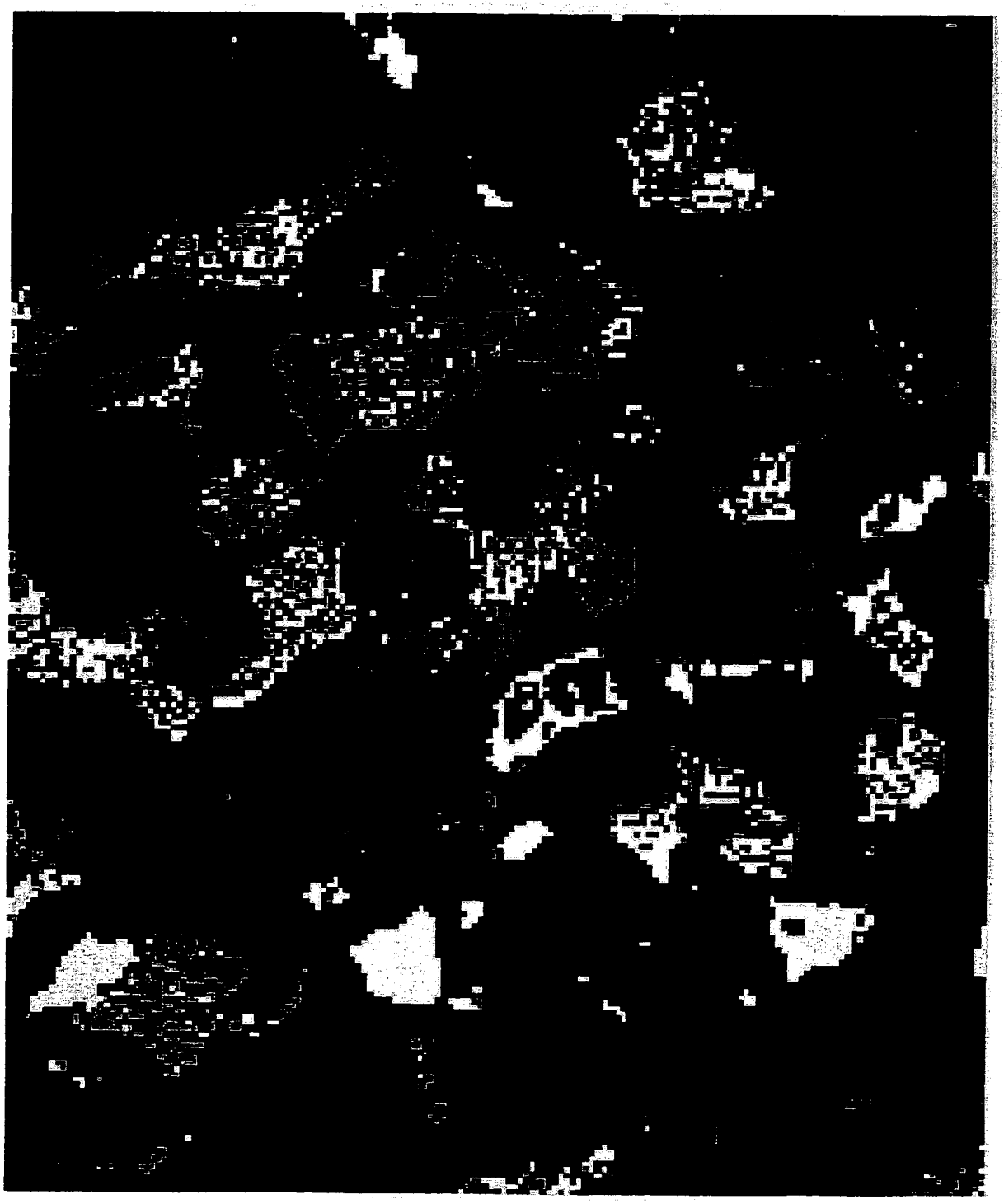


Figure 7.22: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35,$ and $E_J = 2.0/1.35$; weighted unique priors for each μ_c; x initialization = random; 5,000 samples; radiosity estimated NBRFs; low slope training stands.

Figure 7.23: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35$, and $E_J = 2.0/1.35$; weighted unique priors for each μ_c; x initialization = CC; 5,000 samples; radiosity estimated NBRFs; low slope training stands.

treated as fixed.

Down weighting the spatial prior as above increases the number of similarly classified sites between the two initializations from 56.0 to 64.0 percent, with estimates comparable to those shown in Figures 7.18 and 7.19. The increase in mobility compares to that of single mean prior scheme with approximately 6,500 proposals accepted per iteration.

In order to estimate spatially homogeneous and consistent solutions this analysis indicates that it is necessary to supervise the classifications by fixing the class means to the ML point estimates. While specifying prior distributions is an intuitively appealing approach for capturing the uncertainty with which the normalized class means have been estimated, the variation in the mean values sampled from the full conditional distributions precludes the assurance of convergence to the target distribution.

7.9 Markov random fields

7.9.1 Descombes et al. prior distribution

MPM estimates

The GIS classification shown in Figure 7.3 typifies a desirable spatial configuration. The version of the Descombes et al. prior used for this study (see Section 2.2.3) induces desirable patterns at the local and global levels, as demonstrated by the realization shown in Figure 2.8. Without the influence of data this realization is void of noisy configurations, while still maintaining class areas of varying shape and size. This prior will tend to suppress noisy patterns and induce homogeneity to a degree determined by the values assigned to $E_N, E_E,$ and E_J. The combination of $E_N = 2.5, E_E = 1.5,$ and $E_J = 2.0$ proved to yield the most satisfying realizations from the prior (see Figure 2.8).

In applying the fifth-order prior to image data it was necessary to down weight

the strength of the prior. Section 7.8 presented MPM estimates associated with $E_N = 2.5/1.35, E_E = 1.5/1.35$, and $E_J = 2.0/1.35$. These estimates exhibit large homogeneous areas of a single structure along with clusters of noisy patterns representing areas not easily classified as a particular structure. However, the performance of the prior is diminished for $E_N = 1.25, E_E = 0.75$, and $E_J = 1.00$ as shown in Figures 7.18 and 7.19. Although a few homogeneous areas are evident, these estimates are primarily characterized by noisy patterns.

At each site, proposals are accepted or rejected based on the spectral information contained in the data and the spatial information represented by the neighboring sites. In effect the three interaction parameters determine the weight placed on the spatial information. For larger values more weight is placed on the spatial information inducing homogeneity. To the extent that less weight is given to the spatial information and the information content of the data is low, there will be a tendency for more proposals to be accepted. If there is less information with which to classify a site, then the site will be more likely to fall into a larger number of classes, hence more proposals are accepted. In other words, less information implies more uncertainty. This explains the greater mobility and noisy MPM estimates associated with the $E_N = 1.25, E_E = 0.75$, and $E_J = 1.00$ simulations described in Section 7.8. Also note that the information content of the data is partially determined by the system PSF.

Consider Figures 7.24 and 7.25, each MPM estimates from the August image. For Figure 7.24 the three interaction constants were set at $E_N = 2.5/1.2, E_E = 1.5/1.2$, and $E_J = 2.0/1.2$, while lower values of $E_N = 2.5/1.5, E_E = 1.5/1.5$, and $E_J = 2.0/1.5$ were used for Figure 7.25. By comparison Figure 7.24 is almost void of noisy configurations, while noise characterizes the classification of many stands in Figure 7.25. Many of those are younger conifer stands either exhibiting greater spectral variability due to precommercial thinning or more easily confused with hardwoods due to their higher density. In any event, for the prior specified in Figure 7.25 there

is not enough information to homogeneously classify them into a particular structure. This is an example of maintaining fidelity the data. While even a weak prior will induce homogeneity, the data prevents spectrally ambiguous areas from being homogeneously classified.

For example, consider stand 727, highlighted in Figure 7.24. According to Forest Service records this stand was clearcut in 1966 and precommercially thinned in 1982, likely to 300 trees per acre. The image was acquired in 1992. Field plots taken in 1997 found 350 trees per acre, primarily Douglas and Pacific Silver fir. Also in 1997 tree heights ranged from eight to 44 feet and diameters from 1.6 to 8.4 inches. Given its orientation, this stand has a very low incidence angle. While not knowing the exact structural characteristics in 1992, it is reasonable to assume that it would have been making the transition from AI to EE at that time. Thus, while the increased spatial information used to estimate Figure 7.24 favors a homogeneous AI classification of stand 727, the MPM estimate in Figure 7.25 shows a heterogeneous classification of AI, EE, HW, and even some CC and UR. The latter two classes may be indicative of exposed and shaded dirt road beds.

Posterior probability estimates

The posterior probability estimates vary with the prior parameter values. By assigning more weight to the spatial information in Figure 7.24 a higher degree of homogeneity is achieved, but at the expense of mobility. On average only about 300 proposals were accepted per iteration. In contrast about 2000 proposals per iteration were accepted during the simulation for Figure 7.25. Greater mobility means that more sites are changing classes during each iteration, thus having some probability of belonging to more than one class. Figure 7.26 shows the posterior probability that sites belong to the AI class for the Figure 7.24 MPM estimate. Brighter shades represent higher probabilities. Considering stand 727 again, in Figure 7.26 most of the sites that comprise the stand belong to the AI class with a probability of one. An exception

Figure 7.24: MPM: August; $E_N = 2.5/1.2, E_E = 1.5/1.2,$ and $E_J = 2.0/1.2$; fixed ML estimates for all μ_c; x initialization = random; 5,000 samples; radiosity estimated NBRFs; low slope training stands.

Figure 7.25: MPM: August; $E_N = 2.5/1.5, E_E = 1.5/1.5,$ and $E_J = 2.0/1.5$; fixed ML estimates for all μ_c; x initialization = random; 5,000 samples; radiosity estimated NBRFs; low slope training stands.

occurs in the lower left portion of the stand where the sites have about a 90 percent chance of belonging to the EE class and ten percent to the AI class. In contrast Figures 7.27 and 7.28 show that with the weaker prior stand 727 has roughly a 30 percent chance of belonging to each of the AI and EE classes, with 40 percent of the probability divided amount the remaining four classes. Also note that even with the weaker prior there are still areas classified with near 100 percent probability as influenced by the data.

Choosing values for the three interaction parameters depends upon the objective and in turn will influence the posterior probability estimates. While the spatial homogeneity in Figure 7.24 is desirable, the associated posterior probabilities indicate very little uncertainty. Although formal accuracy assessment of the posterior probabilities is beyond the scope of this research, based on field visits there are several stands, such as 727, that should have some probability of belonging to several classes. The key is to identify interaction parameters that preserve a balance between a homogeneous MPM estimate and realistic posterior probability estimates. For this reason, the values $E_N = 2.5/1.35, E_E = 1.5/1.35,$ and $E_J = 2.0/1.35$, as shown in Section 7.7.4, were chosen to best meet these needs. Figures 7.29 and 7.30 show the posterior probability images associated with Figure 7.14 for the AI and EE classes, respectively. During the simulation corresponding to Figure 7.14 approximately 1,000 proposals were accepted per iteration.

The sensitivity of the posterior probability estimates can be seen by comparing Figures 7.26, 7.29, and 7.27. In Figure 7.26 most of the sites classified as AI are done so with very high probability. In Figure 7.29 several stands are classified as AI with high probability, yet several others have only moderate probability of belonging to the AI class. Finally, in Figure 7.27 few stands are classified AI with high probability, including the AI training stand, while a greater number of areas are found to have moderate probability of belonging to the AI class. Consequently, those areas exhibit little homogeneity in the MPM estimate.

Figure 7.26: The estimated AI posterior probabilities corresponding to Figure 7.24. Brighter shades represent higher probabilities.

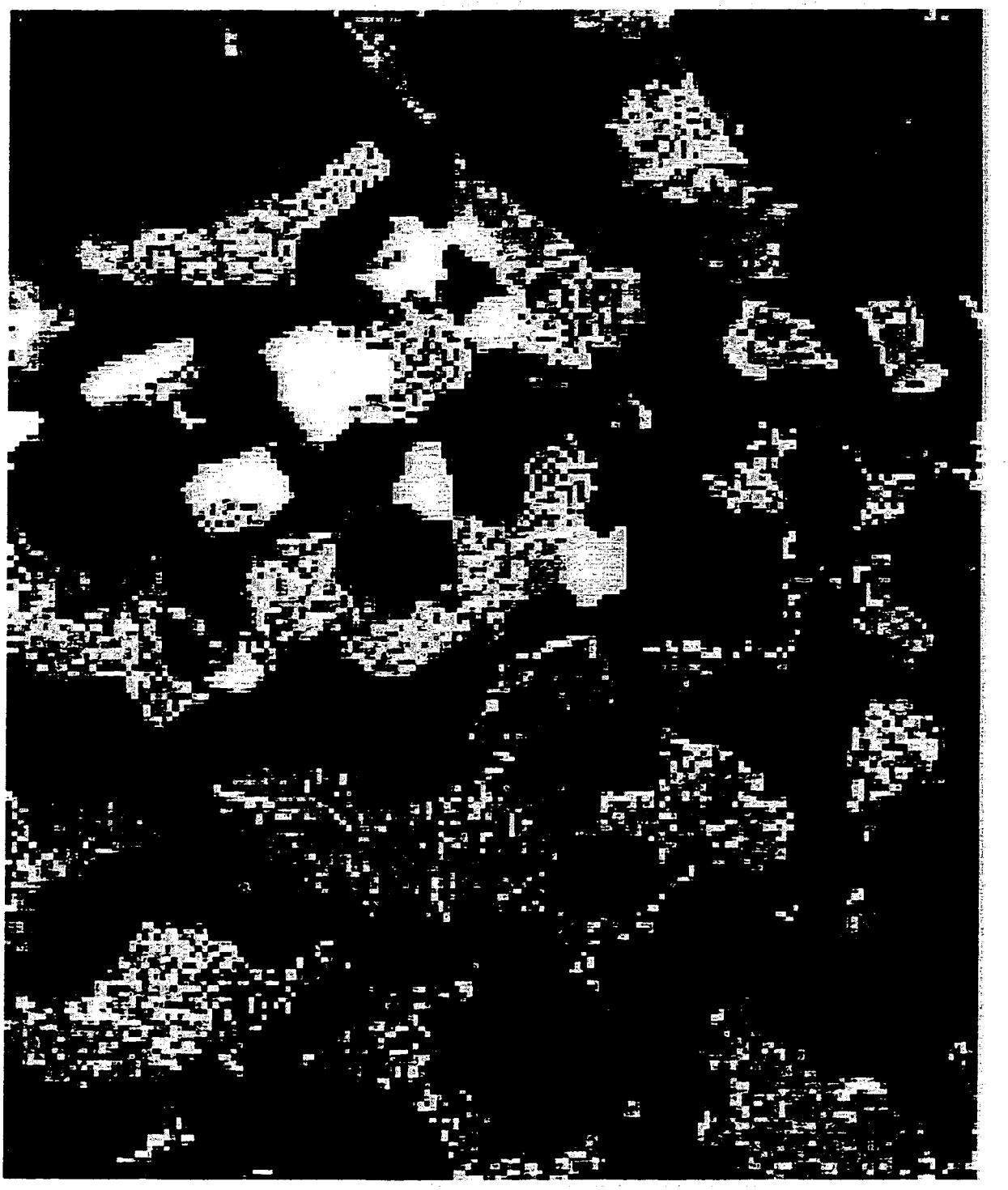


Figure 7.27: The estimated AI posterior probabilities corresponding to Figure 7.25. Brighter shades represent higher probabilities.

Figure 7.28: The estimated EE posterior probabilities corresponding to Figure 7.25. Brighter shades represent higher probabilities.

Figure 7.29: The estimated AI posterior probabilities corresponding to Figure 7.14. Brighter shades represent higher probabilities.

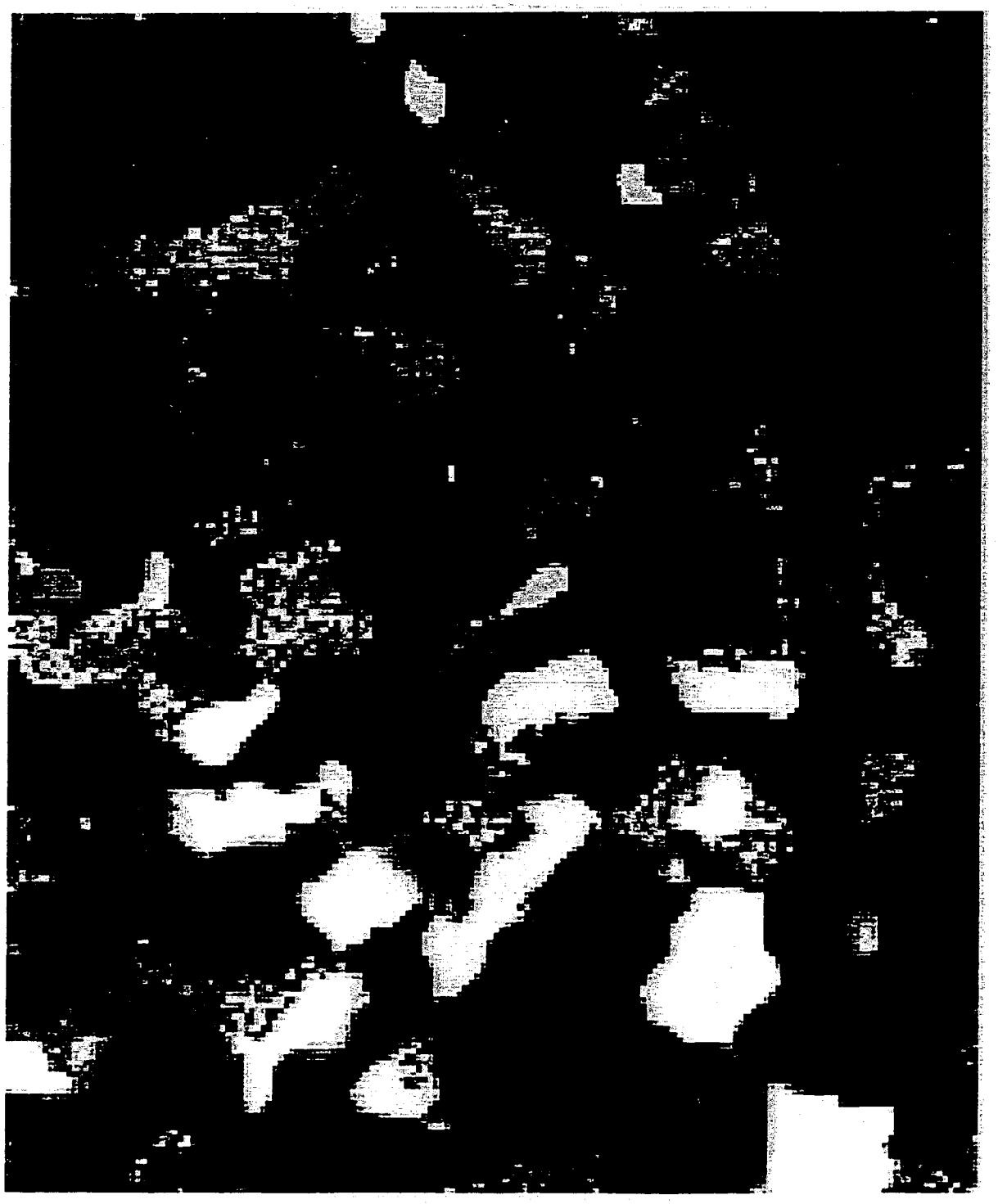


Figure 7.30: The estimated EE posterior probabilities corresponding to Figure 7.14. Brighter shades represent higher probabilities.

7.9.2 Potts prior distribution

MPM estimates

Pairwise interaction random fields induce spatial continuity at the local level, whereas the global properties must be supported by the data (see Section 2.2.2). They are also easy to program and computationally less demanding than higher-order MRFs. Classifying the July image, Figures 7.31 and 7.32 are MPM estimates based on the Potts MRF (see Equation 2.6) with β set at 0.5 and 1.0, respectively. In each case diagonal neighbors are down weighted by a factor of $\sqrt{2}$.

Clearly higher values of β yield more homogeneity. The characteristic of random fields to maintain fidelity to the data is evident by the number of sites classified as CC or AI that actually correspond to forest roads amidst large tracts for forest. If β were large enough to eliminate such "individuals", small stands would begin to disappear as well. Thus, the Potts model is not able to eliminate noisy patterns without producing undesirable homogeneity. Lower values of β merely allow for more heterogeneity. For comparison Figure 7.33 shows the corresponding ML classification, which does not take into account spatial information. Even with $\beta = 0.5$ the effect of the prior in Figure 7.31 is apparent. Many of the heterogeneous areas actually consist of small blocks of sites (local order).

Posterior probability estimates

The limitations of the pairwise random field is more evident upon considering the posterior probability estimates. To achieve the homogeneity in Figure 7.32 so much weight is given to the spatial information that only 500 proposals per iteration are accepted on average. Thus, almost all the sites are classified with very high probability, perhaps a desirable, but not reasonable solution.

By reducing β to 0.5 the number of accepted proposal increases to roughly 2400 per iteration. Figure 7.34 shows the posterior probability of belonging to the EE

Figure 7.31: MPM: July; $\beta = 0.5$; fixed ML estimates for all μ_c; x initialization = random; 5,000 samples; radiosity estimated NBRFs; low slope training stands.

Figure 7.32: MPM: July; $\beta = 1.0$; fixed ML estimates for all μ_c; x initialization = random; 5,000 samples; radiosity estimated NBRFs; low slope training stands.

Figure 7.33: ML estimate: July image; low slope training stands.

class. Many of the stands classified as EE in the MPM estimate (Figure 7.31) have a high probability of being EE, and many other sites have at least some probability of belonging to the EE class. However, the spatial coherence of Figure 7.34 is not desirable. Whole areas homogeneously classified as UR have some probability of being EE. Most of the sites within stand 727 are either AI or EE in the MPM image and the probabilities of these classes within the area of stand 727 range from 0.0 to 0.8.

In contrast the global characteristics maintained by the Descombes et al. prior are perhaps more evident in the resulting posterior probabilities. Figure 7.35 provides a comparison. Here all of the sites having moderate probability of belonging to EE are found clustered in areas corresponding to distinct stands. Within these stands the probabilities are much more consistent than is the case in Figure 7.34, even if the corresponding stand in the MPM is fairly noisy. Other sites either have very low or high probability of belonging to EE and again they are homogeneously arranged (see also Figure 7.30 for a comparable estimate using the August data).

7.10 Topographic variation

Topography determines the bidirectional reflectance in the direction of the sensor and the angle of the direct solar irradiance. By modelling the sun-surface-sensor geometry this classifier seeks to produce geometrically invariant MPM estimates (see Section 3.3). Ignoring the illumination and surface orientation angles yields topographically dependent classifications. Figure 7.36 shows an ML classification of the original August image. Many of the sun facing slopes with lower incidence angles have been classified into the younger conifer or hardwood classes characterized by higher reflectances. High incidence angle slopes tend to be classified as UR.

Assessing the geometric dependence of a classification is a difficult task. Posterior probability estimates may reflect topographically related uncertainties. In principle the GIS classification could be used for comparison, however to the extent that the

Figure 7.34: The estimated EE posterior probabilities corresponding to Figure 7.31. Brighter shades represent higher probabilities.

Figure 7.35: The estimated EE posterior probabilities corresponding to Figure 7.12. Brighter shades represent higher probabilities.

Figure 7.36: ML estimate: August image; low slope training stands.

stand polygons were identified via air photo interpretation, the GIS classification may incorporate topographic variation. Aerial photography is sensitive to bidirectional reflectance as well.

Another method for assessing geometric dependence considers informal plots of original and preprocessed DN values against $\cos\iota$. This approach evaluates the ability of the bidirectional reflectance model to correct the data for topographic dependence and assumes it will influence the MPM estimates in similar fashion. In addition to evaluating the radiosity technique described in Section 3.5, the SCS correction proposed by Gu and Gillespie (1998) was also evaluated (see Section 3.4.1). The latter model was chosen for its simplicity and because exploratory analysis confirmed its improvement over the Lambertian model. Figure 7.37 plots the observed DN values for band 4 of each image for the pixels corresponding to the UR class areas in the GIS classification. Only those pixels were used as the radiosity reflectances are class dependent and the UR class areas provided the greatest range of variability in surface orientation. These plots show a positive correlation between the DN values and $\cos\iota$. As expected, reflected radiance recorded increases as incidence angle decreases. Also note the degree of variation even for a particular incidence angle. Figures 7.38 and 7.39 show plots of the same pixels having been preprocessed with using the radiosity and SCS techniques, respectively. Both of these methods remove the positive correlation, although the SCS introduces a slight negative correlation, or over correction, especially at high incidence angles. Over correction was even more apparent for the Lambertian model.

7.11 Illumination angle variation

Comparing the MPM and posterior probability estimates from images acquired under different illumination angles provides another means for evaluating the simulated bidirectional reflectances. This is also a straightforward and simple approach for

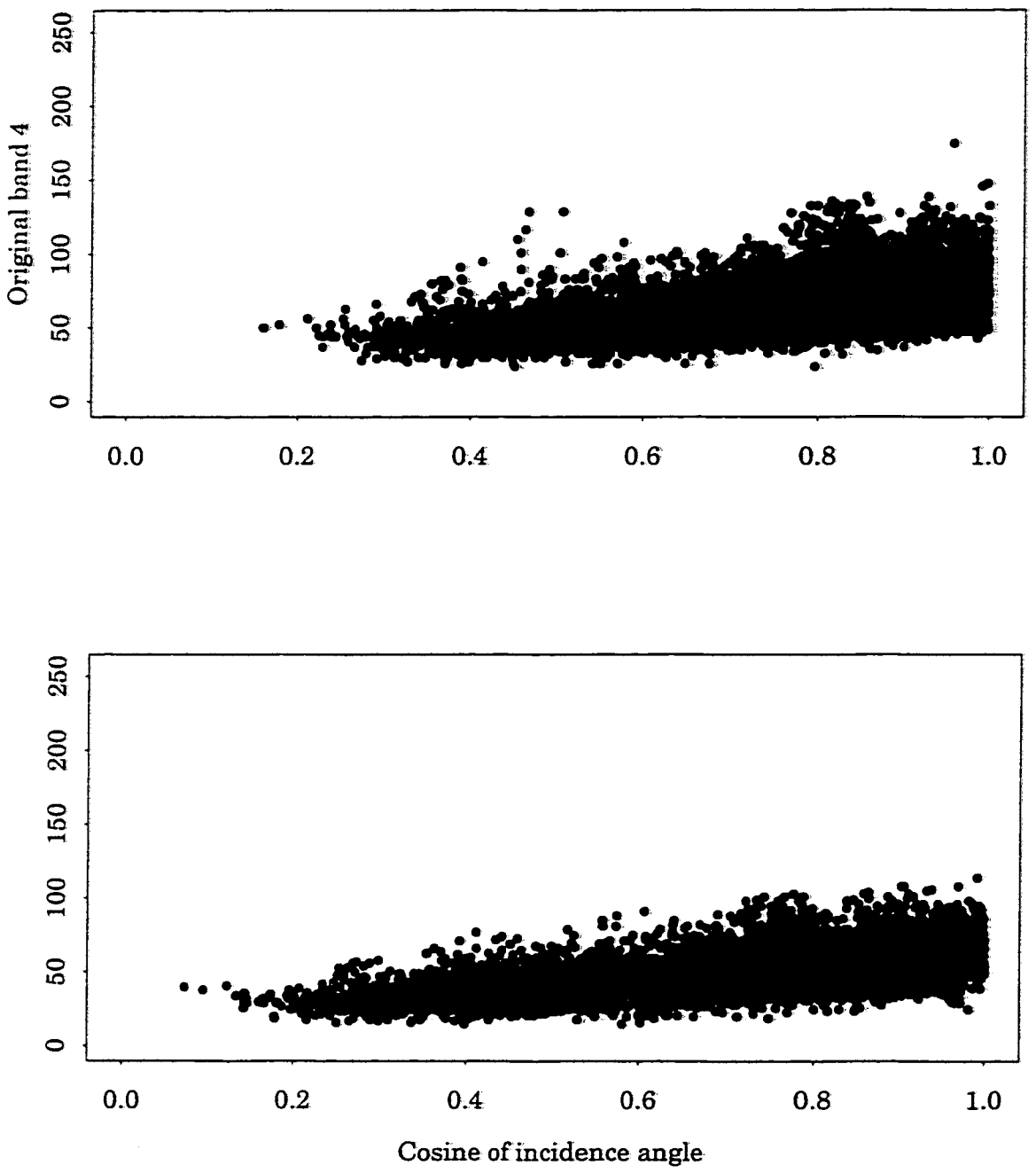


Figure 7.37: Original band 4 data plotted against $\cos \iota$ for UR class: July image (top) and August image (bottom).

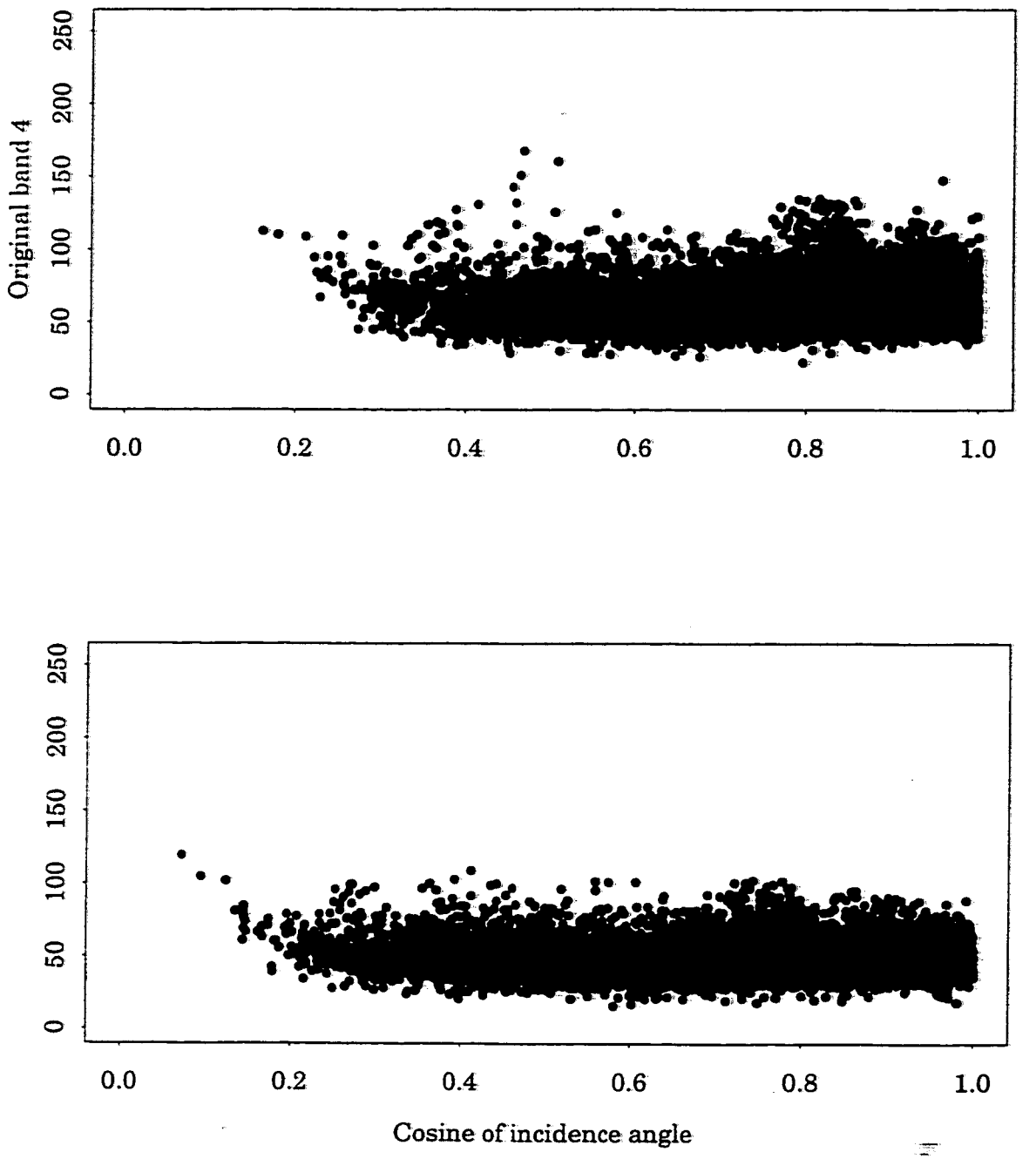


Figure 7.38: Normalized band 4 data (radiosity method) plotted against $\cos \iota$ for UR class: July image (top) and August image (bottom).

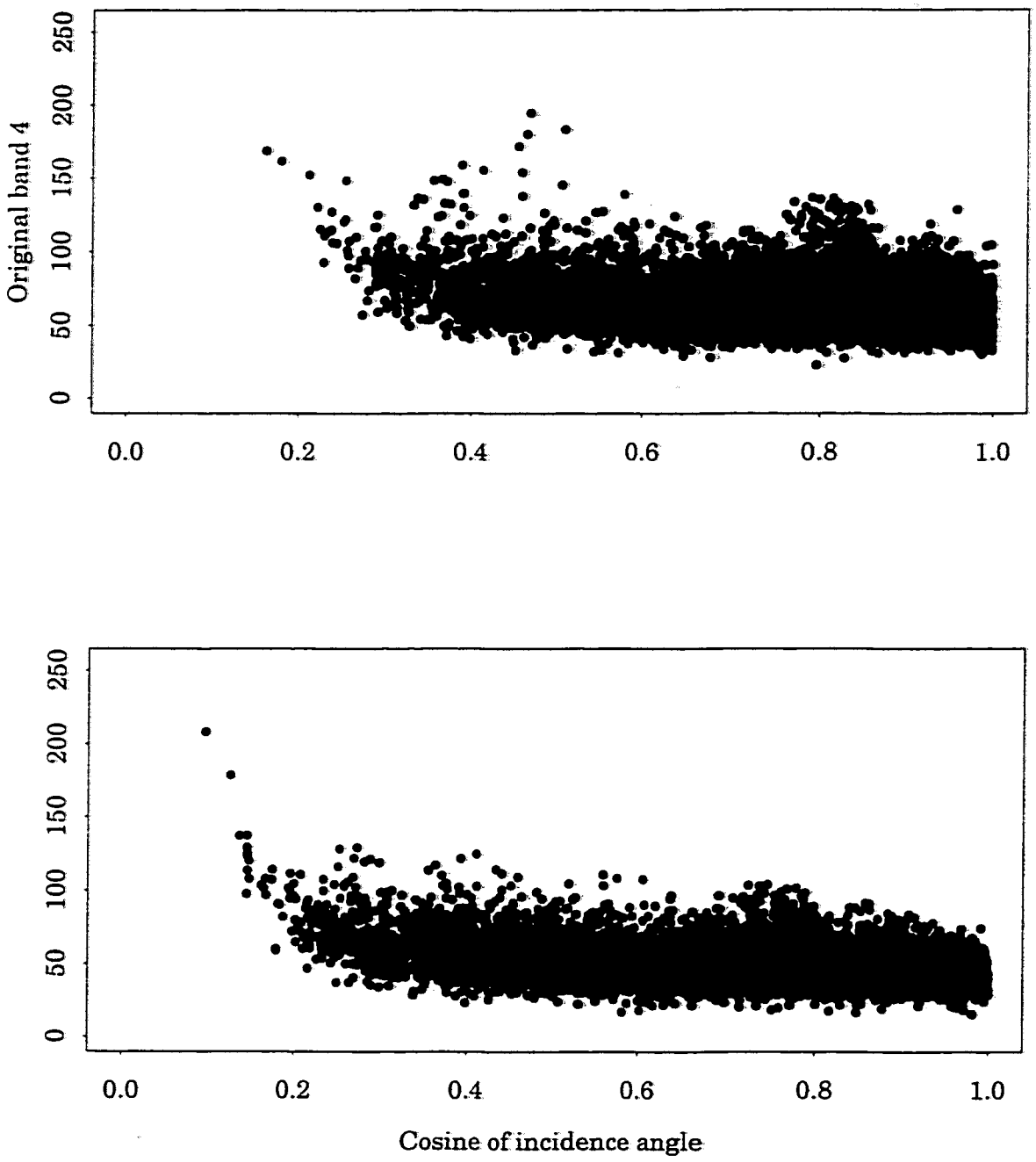


Figure 7.39: Normalized band 4 data (SCS model) plotted against $\cos\iota$ for UR class: July image (top) and August image (bottom).

assessing the sensitivity of image classifications to variation in the illumination geometry. The MPM estimates shown in Figures 7.12 and 7.14 are based on the same prior distribution. As mentioned in Section 7.1 there is only a five degree difference in solar elevation and seven degree difference in solar azimuth between the two images, which is taken into account by the NBRFs. The PSF half-widths model the unique atmospheric conditions and resampling methods associated with each image.

Although the two images were acquired one year apart, it is reasonable to expect that an ongoing forest monitoring program might be working with images on an annual basis acquired under various solar positions. Furthermore, trees grow slowly, thus it can be assumed that the actual change in structure due to forest growth during the one year interval is negligent. No management activities appear to have occurred during this year either.

7.11.1 Classifications using the radiosity estimated NBRFs

Figures 7.12 and 7.14 compare favorably with 76 percent of the sites having the same classification. Neither of these estimates exhibit the topographic dependencies evident in Figures 7.33 and 7.36. Figure 7.40 highlights the remaining 24 percent of the sites. With the exception of the north facing hillside in the lower right portion of the image, there does not appear to be a topographically related pattern to the distribution of the differently classified sites. However, these inconsistencies are reflected in the posterior probability estimates. The mean MPM of the mismatched sites is 77 and 74 percent for the July and August images, respectively. In contrast the mean MPM of the similarly classified sites is 95 and 94 percent, respectively.

Examination the clearcut areas shows a trend in seasonal variation of that class. Clearcuts are rather spectrally distinct (see Tables 7.4 and 7.5). These areas are very homogeneous in Figure 7.14, but mixed with many clusters of AI in Figure 7.12. The mean reflectance of AI is more characteristic of a mix of photosynthetic vegetation and bare ground. As a likely explanation for this difference, a variety of herbacious

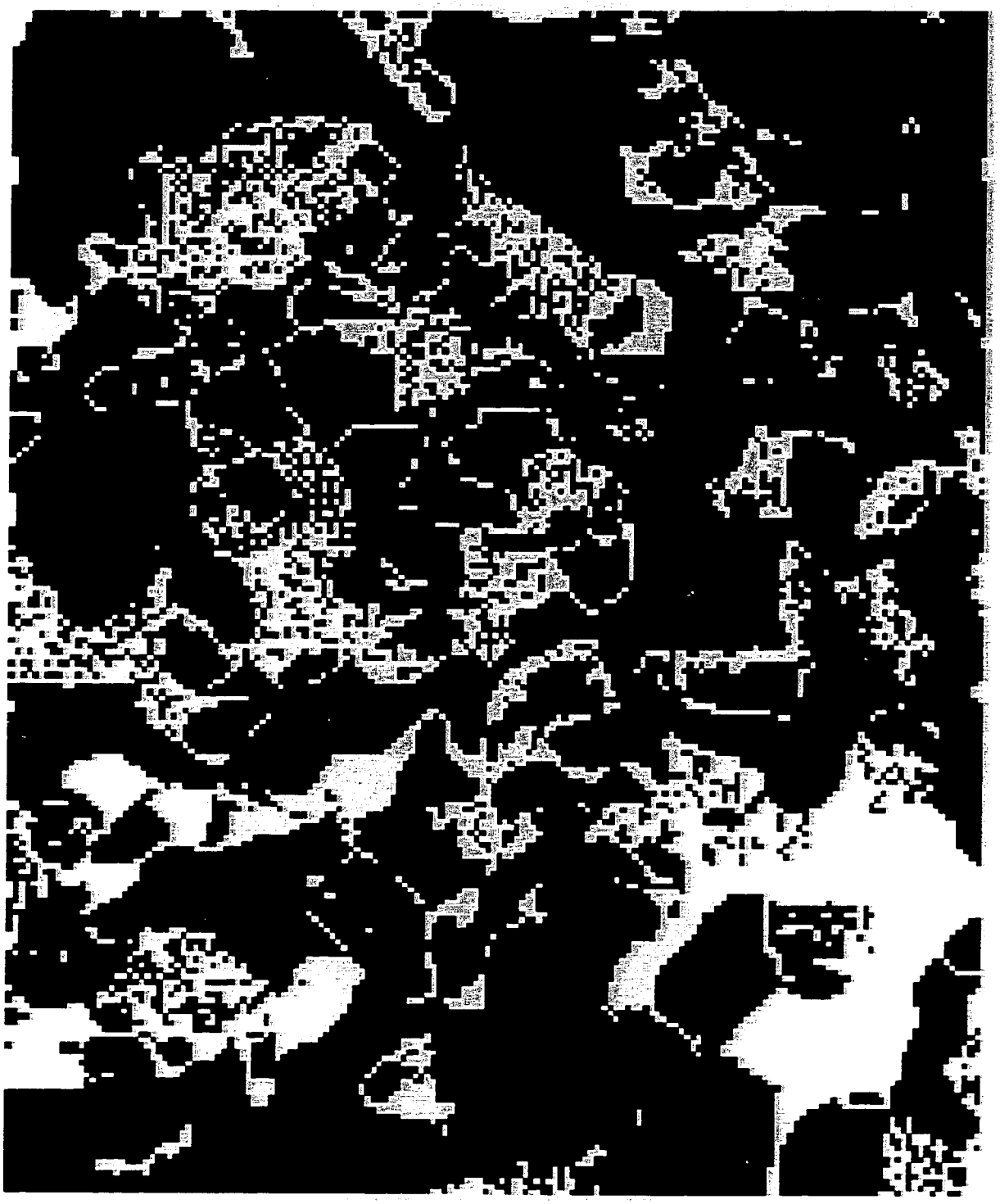


Figure 7.40: Comparison of Figures 7.12 and 7.14 showing inconsistent site classifications.

plants, such as fireweed, commonly grow in clearcuts. In early July, the reflectance of this vegetation would contrast the reflectance of the bare soil and woody debris also found in clearcuts. However, by August many of these plants would have died as a result of the dry Northwest summers, leaving only non-photosynthetic surface cover. Such seasonal variation is not accounted for by this model. When the GIS classified CC sites are masked out, the percent of similarly classified sites increases from 76 to 78 percent.

7.11.2 Classifications using the SCS estimated NBRFs

For the MPM estimates shown in Figures 7.41 and 7.42 the SCS model was used to estimate normalized reflectance factors. Indeed, these estimates resemble those using the radiosity estimated NBRFs. The estimates in Figures 7.41 and 7.42 are also 76 percent similar. The seasonal trend within the clearcuts is also apparent.

The mean MPM of the mismatched sites is 78 and 76 percent for the July and August images, respectively. In contrast the mean MPM of the similarly classified sites is 96 and 95 percent, respectively. Indeed, mobility is slightly reduced when using the SCS model. The class means estimated using the SCS model span a greater range than the means presented in Tables 7.4 and 7.5. For example, with the SCS model the minimum band 4 value is 49.3 (CC class) and the maximum value is 150.7 (HW class). The increased HW band 4 mean DN is likely explained by the model's tendency to over correct. Closer mean values reduce the certainty of classification, hence increasing mobility.

7.11.3 Maximum likelihood classifications using the SCS model

Preprocessing image data to correct for topography combined with maximum likelihood provides a simple method for classification that is readily implemented using commercial GIS and remote sensing software. Figures 7.43 and 7.44 present such classifications of the July and August images, respectively. Fewer sites, only 62 percent,

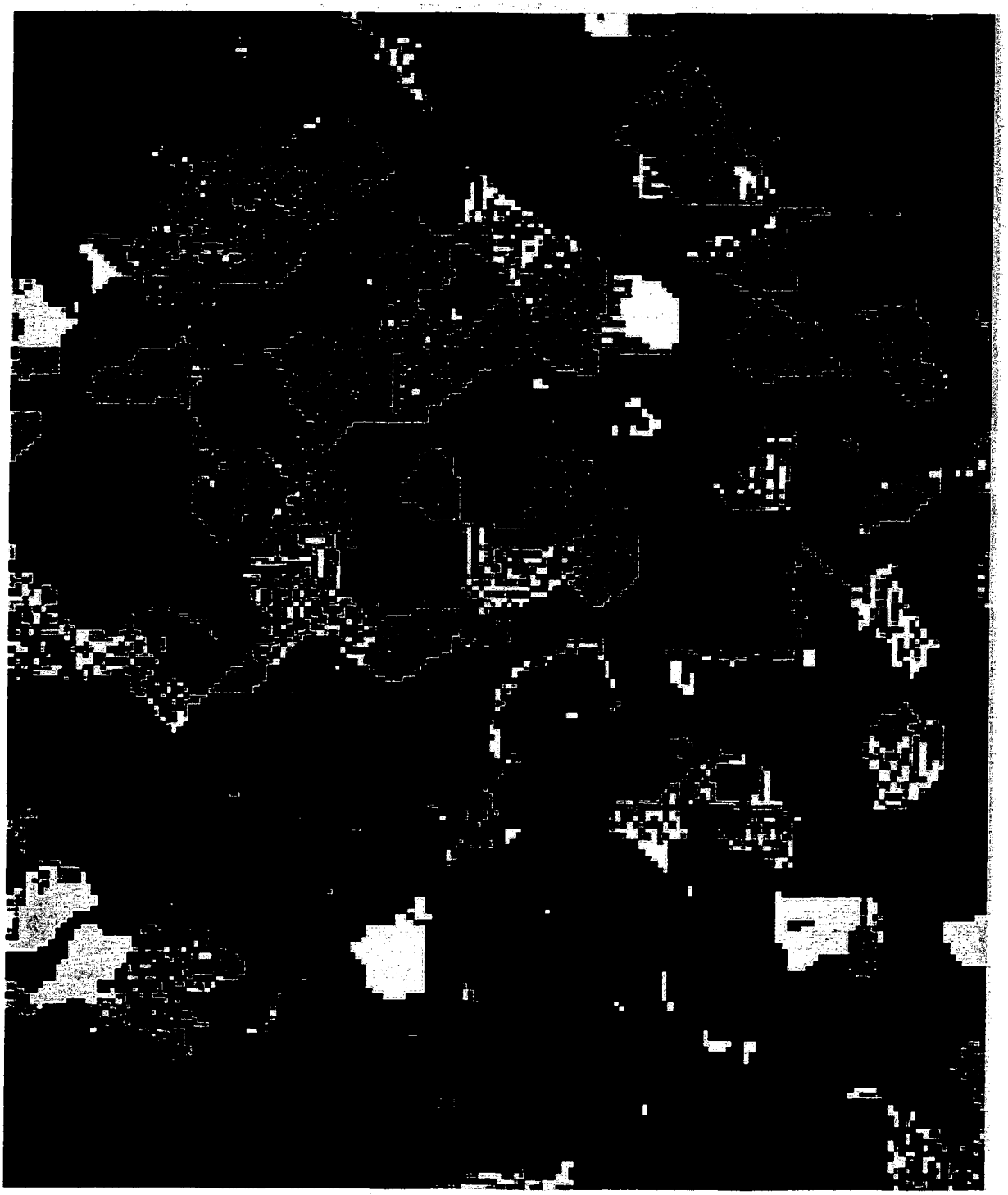


Figure 7.41: MPM: July; $E_N = 2.5/1.35$, $E_E = 1.5/1.35$, and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = random; 5,000 samples; SCS estimated NBRFs; low slope training stands.

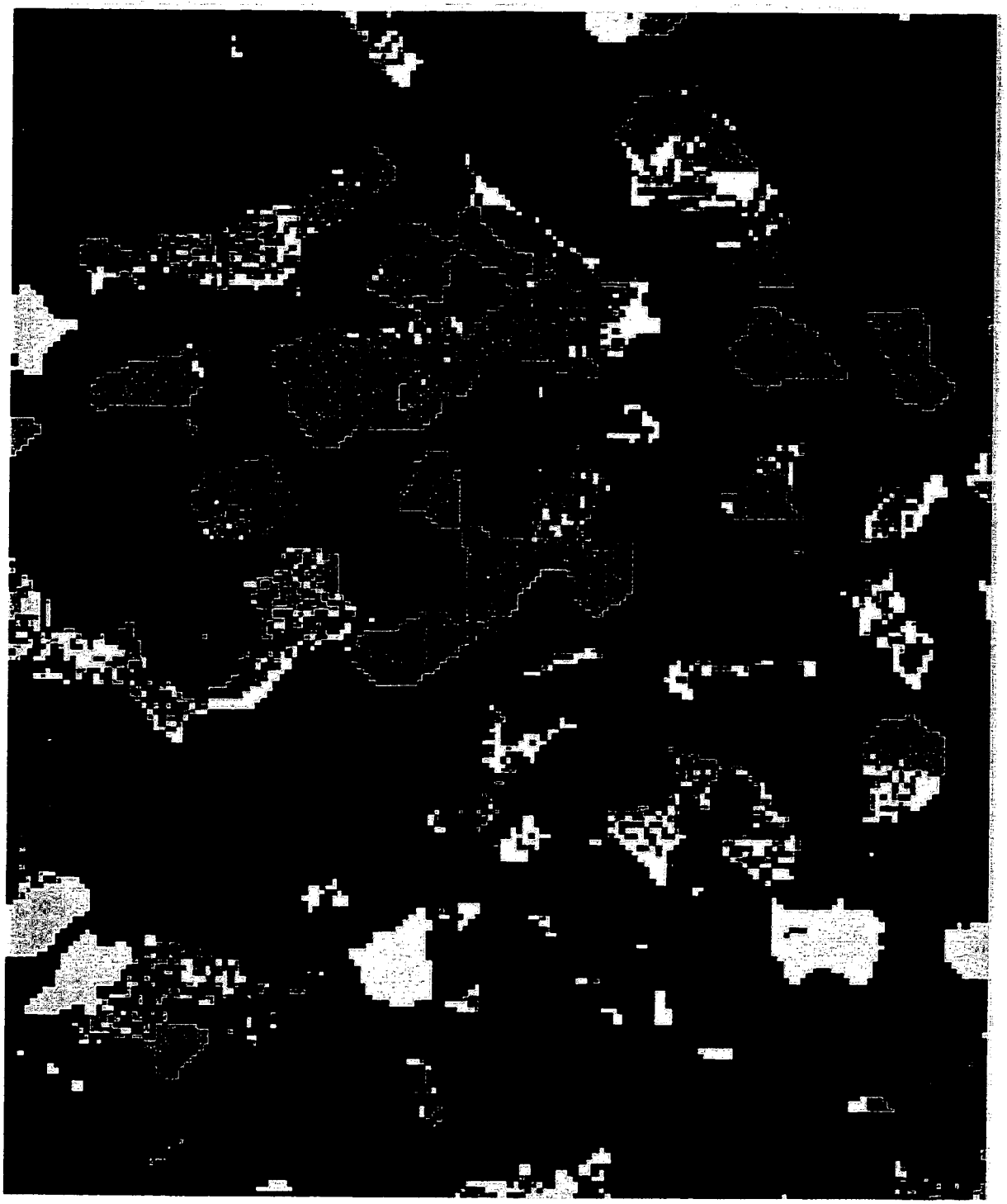


Figure 7.42: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35$, and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = random; 5,000 samples; SCS estimated NBRFs; low slope training stands.

are consistently classified from month to month.

The propensity of the SCS model to over compensate is especially evident in Figure 7.44 where the north facing slopes in the lower right and upper left portions of the image are incorrectly classified as the more reflective HW class. This also suggests that the over compensation is dampened when the SCS model is used in conjunction with the PSF and MRF models in a Bayesian context (see Figure 7.42).

The seasonal variation within the clearcuts is not quite as apparent either. Given that this variation is evident in Figures 7.41 and 7.42, and that the MRF seeks to suppress such patterns, the possibility of AI sites occurring within the clearcuts reflects the uncertainty induced by PSF model.

7.11.4 Variation in effective spatial resolution

Comparing the classifications of two images using the same prior distribution provides a means of assessing the impact of varying atmospheric conditions on the stability of the solutions. The half-width estimates presented in Section 7.6 show a broader system PSF for the August image compared to the July image. As a result less information is contained in the August data and the classifications of that data are less certain. On averages 800 proposals are accepted per iteration for the MPM estimate in Figure 7.12 as compared to 1,000 per iteration for the MPM estimate in Figure 7.14.

7.12 Training stand selection

The training areas indicated in Figure 7.3 fall on relatively level terrain. In principle, to the extent that bidirectional reflectance is correctly modelled, the surface orientation of the training stands should not matter. A further method of assessing the stability of the MPM solutions and their reliability is manifest by varying the orientation of the training stands.

Figure 7.43: ML estimate: July image; data preprocessed for topographic correction using SCS model; low slope training stands.

Figure 7.44: ML estimate: August image; data preprocessed for topographic correction using SCS model; low slope training stands.

7.12.1 Classifications using the radiosity estimated NBRFs

To carry this out class means for the EE, LE, and UR classes were obtained from training areas on steeper terrain as indicated by the polygons in Figure 7.45. The EE and LE areas face northwest, giving the associated pixels high incidence angles. In contrast the UR training area faces south, thus those pixels have low incidence angles. The resulting class means are presented in Tables 7.9 and 7.10. Note that the means for CC, HW, and AI have not been changed. In principle, the EE, LE, and UR means in Tables 7.9 and 7.10 should resemble those shown in Tables 7.4 and 7.5. As there is no way to know if the comparable stands share a "true" mean reflectance from a normalized surface, the fact that the mean value for the EE class has changed little is promising. However, in the case of the LE class, shifting to a northwest facing slope yields lower normalized means. Likewise, shifting the UR training area to a south facing slope increases the estimated mean. Thus, either the two LE and UR stands are inherently different or the radiosity technique is not modelling bidirectional reflectance properly. In actuality, both explanations are likely.

Table 7.9: Class means estimated from normalized high slope data for July.

Band	CC	HW	AI	EE	LE	UR
1	33.7	19.0	22.0	16.0	14.9	13.6
2	27.4	24.5	25.2	20.3	18.4	15.8
3	37.8	20.0	25.6	17.0	15.8	14.2
4	50.0	148.6	96.9	96.1	72.8	58.5
5	109.9	77.6	84.0	44.0	37.2	34.3
7	68.1	21.7	31.7	12.2	11.3	10.6

Figures 7.45 and 7.46 show MPM estimates classifying the July and August images, respectively. These estimates demonstrate just how sensitive the solutions are

Table 7.10: Class means estimated from normalized high slope data for August.

Band	CC	HW	AI	EE	LE	UR
1	21.5	13.0	14.3	12.1	11.0	9.5
2	20.0	15.9	16.5	15.6	14.3	11.4
3	22.8	12.8	15.2	12.3	11.4	9.6
4	51.1	103.7	71.2	78.0	53.3	46.7
5	79.9	54.0	52.0	34.3	26.4	25.0
7	37.7	12.6	16.4	8.0	6.3	6.3

to the selection of training areas. The proportion of similarly classified sites falls to 70 percent. Again those sites are classified with more certainty on average than the differing sites; 95 percent compared to 81 percent for Figure 7.45, and 94 percent compared to 76 percent for Figure 7.46.

The most notable effect of changing the training areas is seen by comparing Figure 7.45 with Figure 7.12 and Figure 7.46 with Figure 7.14. In both cases it is clear that topography has influenced the MPM estimates in Figures 7.45 and 7.46. For the classes having the same or similar means, there is little effect. However, of the two most mature forest structures the areas classified as LE correspond to north facing slopes, while much of the south facing terrain previously classified as LE is now estimated to be UR. This is especially evident in Figure 7.46 in which north facing slopes receive that much less irradiance due to the lower sun elevation. Incidentally, the mobility was increased for the August simulation, but not the July simulation.

7.12.2 Classifications using the SCS estimated NBRFs

When SCS model is used in conjunction with the steeper training sites the results compare with those using the NBRFs, with the exception of the EE class. Figures 7.47

Figure 7.45: MPM: July; $E_N = 2.5/1.35, E_E = 1.5/1.35$, and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = random; 10,000 samples; radiosity estimated NBRFs; high slope training stands.

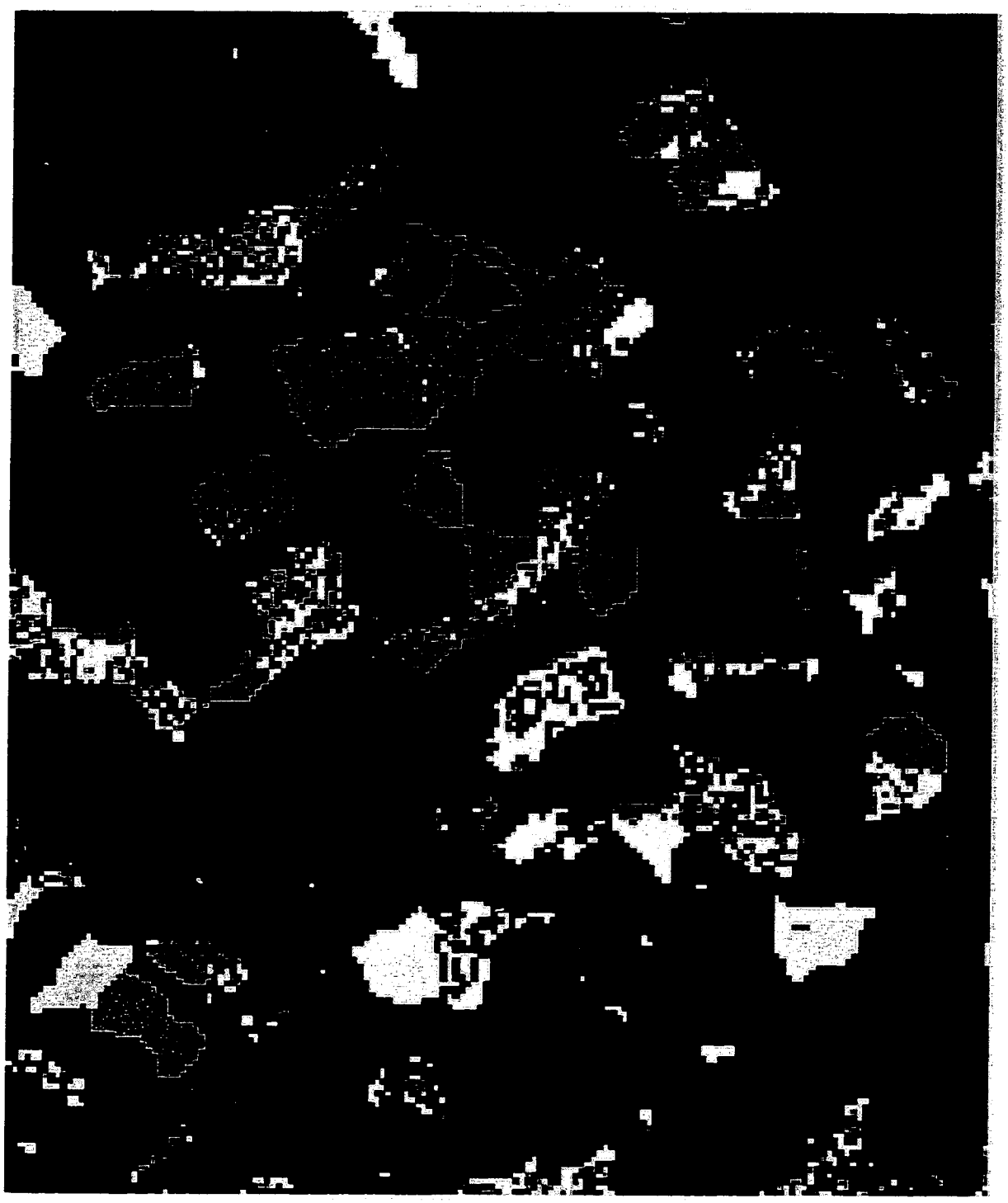


Figure 7.46: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35,$ and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = random; 5,000 samples; radiosity estimated NBRFs; high slope training stands.

and 7.48 show these results for each image. While the radiosity model accounted for the change in the orientation of the EE training data, the SCS model has over corrected the data associated with high incidence angles. As a result the several EE stands are virtually eliminated from the MPM estimates, with those sites classified as either HW or UR. While the mean vector for the EE class in Table 7.9 is very similar to that of Table 7.4, the EE mean vector estimated using the SCS model with steeper terrain is: [18.7, 23.6, 19.6, 109.8, 50.5, 14.0]. Finally, the higher incidence training data leads to increased mobility in the context of the SCS model; from approximately 700 to 800, and 900 to 1,000 proposals accepted per iteration for the July and August images, respectively.

7.12.3 Maximum likelihood classifications using the SCS model

Varying the orientation of the training areas reduces the performance of the ML classifier. In Figures 7.49 and 7.50 topographic variation is readily apparent. The classification of mature forests correlates with the orientation of the LE and UR training areas. In Figure 7.50 many of the EE stands have been incorrectly classified as the more reflective HW, including many of the sites within the former EE training stand. Between these two estimates, 65 percent of the sites are similarly classified.

A comparison of Figures 7.49 and 7.47 shows that the addition of the PSF model and spatial information caused most of the forest area in the lower left portion of the image to be classified as UR, whereas many of those sites were classified as LE in Figure 7.49. Thus, in this instance the Bayesian model exacerbates the inability of the SCS model to represent the bidirectional reflectance of the UR forest structure.

All of the results presented in this Section suggest that, regardless of modelling strategy, forest classifications are very sensitive to the selection of training stands.

Figure 7.47: MPM: July; $E_N = 2.5/1.35$, $E_E = 1.5/1.35$, and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = random; 5,000 samples; SCS estimated NBRFs; high slope training stands.

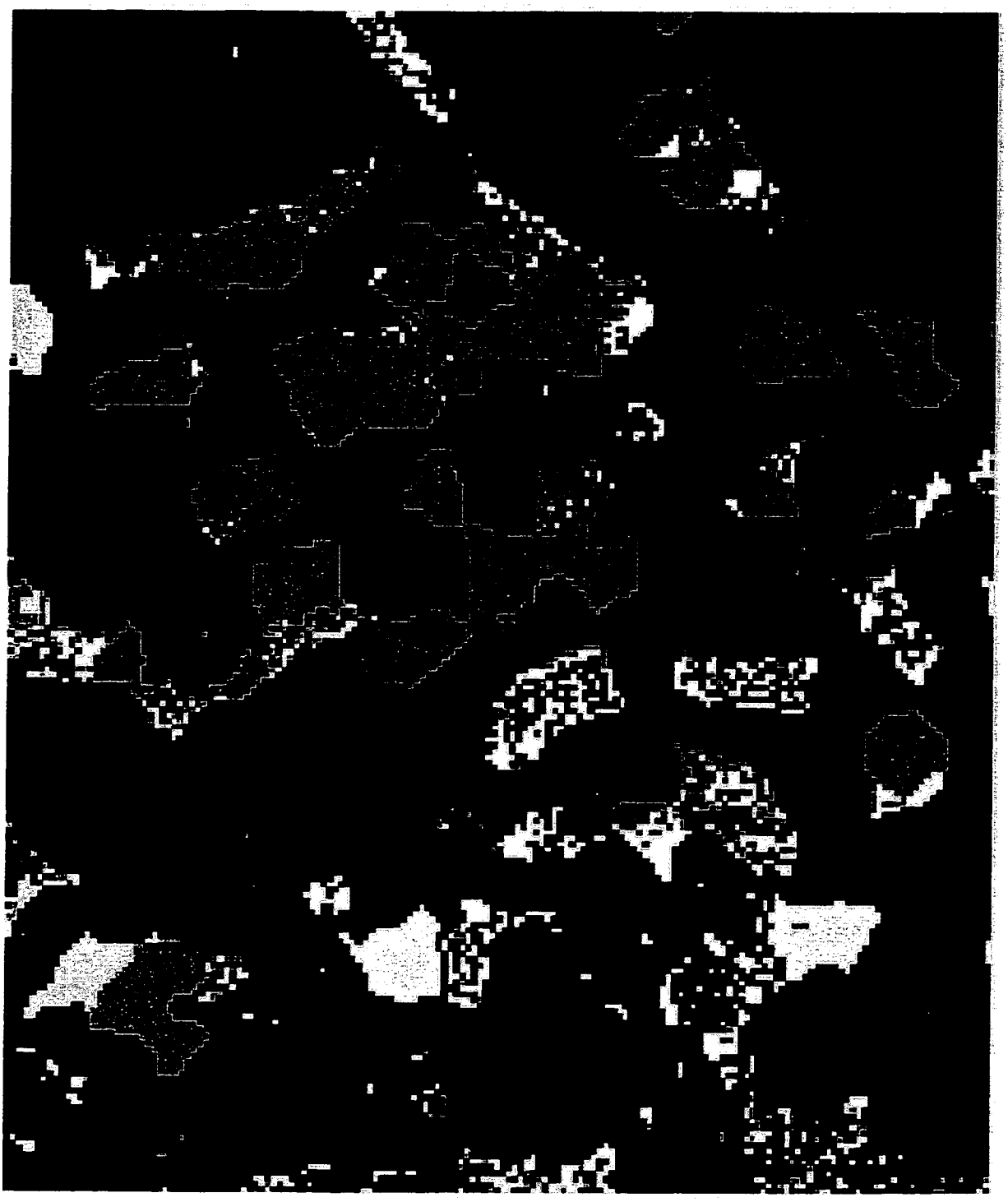


Figure 7.48: MPM: August; $E_N = 2.5/1.35, E_E = 1.5/1.35,$ and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c; x initialization = random; 5,000 samples; SCS estimated NBRFs; high slope training stands.

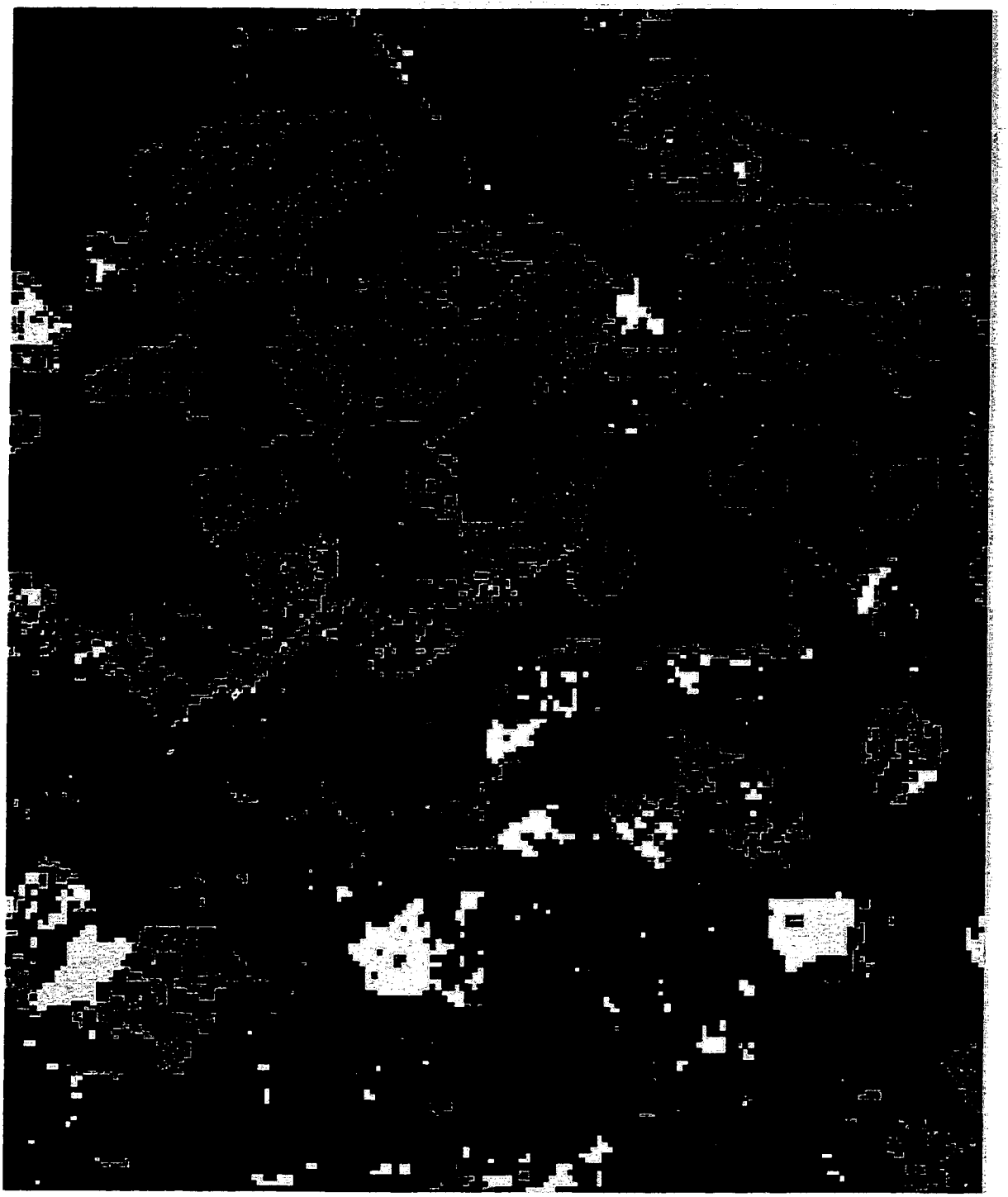


Figure 7.49: ML estimate: July image; data preprocessed for topographic correction using SCS model; high slope training stands.

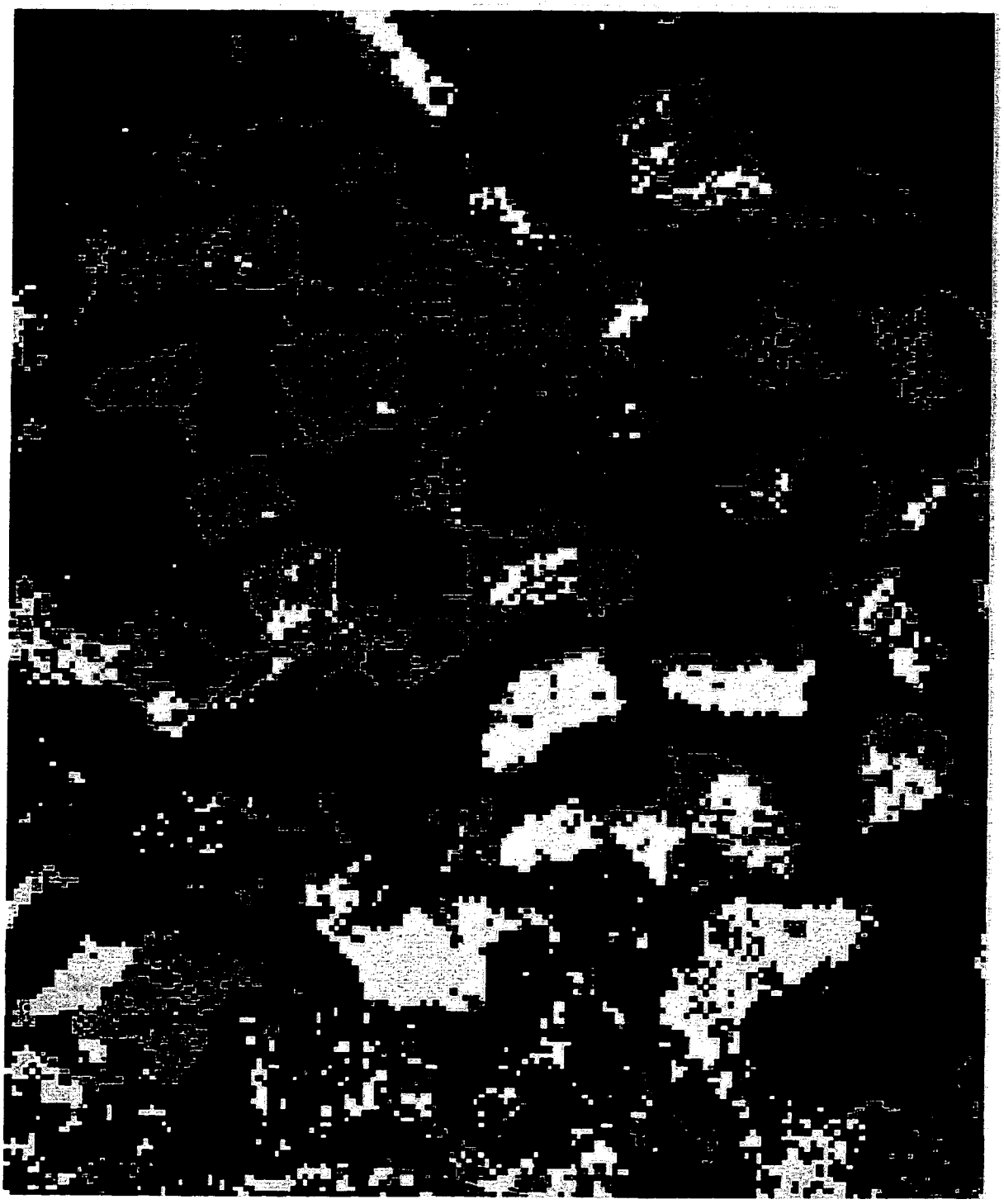


Figure 7.50: ML estimate: August image; data preprocessed for topographic correction using SCS model; high slope training stands.

7.13 Temporal and spatial extensibility

The extensibility of a remote sensing model is another important measure of its performance. So far these results have demonstrated that this approach is extensible in time provided identifiable training areas exist within the subset being classified. In the absence of calibration, this allowed unique sets of class means to be estimated for each image (see Sections 4.1.1 and 5.1.4). The unique half-widths estimated for each image required the presence particular features in the subset area. The reflectance factors must be recalculated for varying solar positions.

Spatial extensibility is a more challenging task. As a very modest test a different subset taken from the full July image was classified using the estimates class means shown in Table 7.4. This subset, near the Wind River, lies approximately 40 miles to the south of the Cispus subset. Although the atmosphere may be different at that distance, to evaluate the applicability of using this model to classify an entire image, the half-widths were not re-estimated. The average DEM value was almost 300 m less for the Wind River subset.

Figures 7.51 and 7.52 show the classified GIS data and an MPM estimate from the image data, respectively. The black areas in Figure 7.51 indicate areas of rock outcrops. Many of these areas are classified with a high degree of uncertainty as they should be. Several of the stands visited or sampled in this area have been reasonably classified in the MPM estimate and topographic variation does not appear to have influenced this estimate. However, there are significant inconsistencies between the MPM estimate and the GIS classification for EE and LE on the left side of the image and for LE and UR in lower right portion of the image. It is somewhat encouraging that these are similar classes, but unfortunately in the MPM estimate these areas are classified with a high degree of certainty. Their spatial homogeneity suggests that the spatial prior may have outweighed possible distinctions contained within the data. A more thorough evaluation of extensibility would require a field survey of this area.

Figure 7.51: Classification of ancillary GIS information for Wind River subset.

Figure 7.52: MPM: July, Wind River; $E_N = 2.5/1.35, E_E = 1.5/1.35$, and $E_J = 2.0/1.35$; fixed ML estimates for all μ_c corresponding to Table 7.4; x initialization = random; 10,000 samples; radiosity estimated NBRFs.

Calibration would be required to apply the model beyond the extent of a single image, to the Olympic Peninsula for example. However, without some prior knowledge of the land cover there is no way to parameterize the tree surface for the radiosity model. Likewise, to apply the Cispus training data to an area 40 miles to the east would possible yield poor results given the different characteristics of forests in eastern Washington. In a drier climate stand density decreases. Thus, more mature forests would tend to be classified as AI. For this reason, the classifier developed in this study is more appropriate for the ongoing monitoring of a given mountainous area.

Chapter 8

DISCUSSION AND CONCLUSIONS

Consistent solutions are an important prerequisite for an operational remote sensing system. The results of this study show that for a single image classification, the MPM estimate will depend upon the particular model formulation and parameter specifications. For a given formulation and set of parameters, the MPM estimate will vary with time. However, the hierarchical model yields more consistent results than the spectral-only classification of preprocessed image data. Although some estimates will be more accurate than others, until the variation in the sun-surface-sensor geometry, atmospheric conditions, and other, perhaps as yet unknown, sources of variation can be adequately modelled or normalized there is little use for time consuming and expensive accuracy assessments. High accuracy is of less value if the results are not reliable and extensible. Acceptable levels of accuracy, as well as consistency, depend upon the task at hand. However, accurate and appropriate field data remain useful for technology development. In the process of sampling the training stands in the field, most of the stands in the subset area were personally observed. Many of the conclusions drawn in Chapter 7 have been based on these observations.

8.1 The classes

This research has focused on improving the classification of a predetermined number of classes. Future research should be aimed at determining the sensitivity of solutions to the defined set of classes. In considering the overall variability of image data for the purpose of classification there is both desirable and undesirable variation. As the

number of classes increases, their spectral contrast (desirable variation) decreases, while the level of unwanted variation remains the same. Figure 8.1 shows a hypothetical graph of these relationships. The proportion of overall variability attributable to the spectral distinctions between the classes decreases as the number of classes is increased.

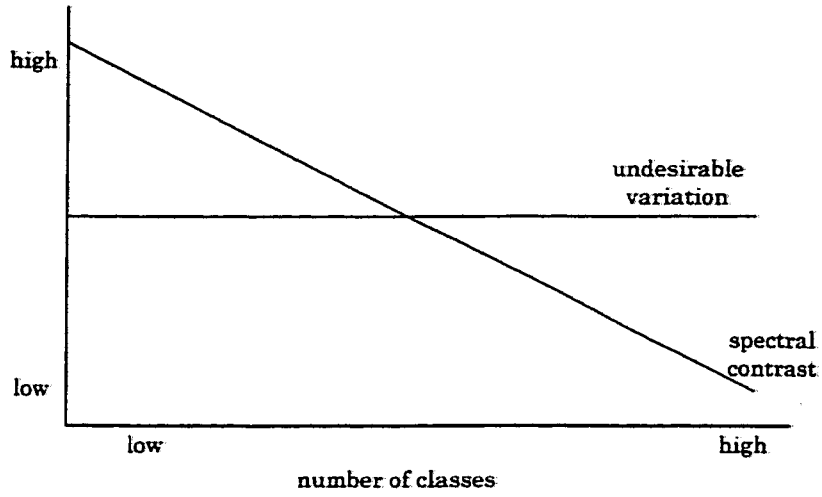
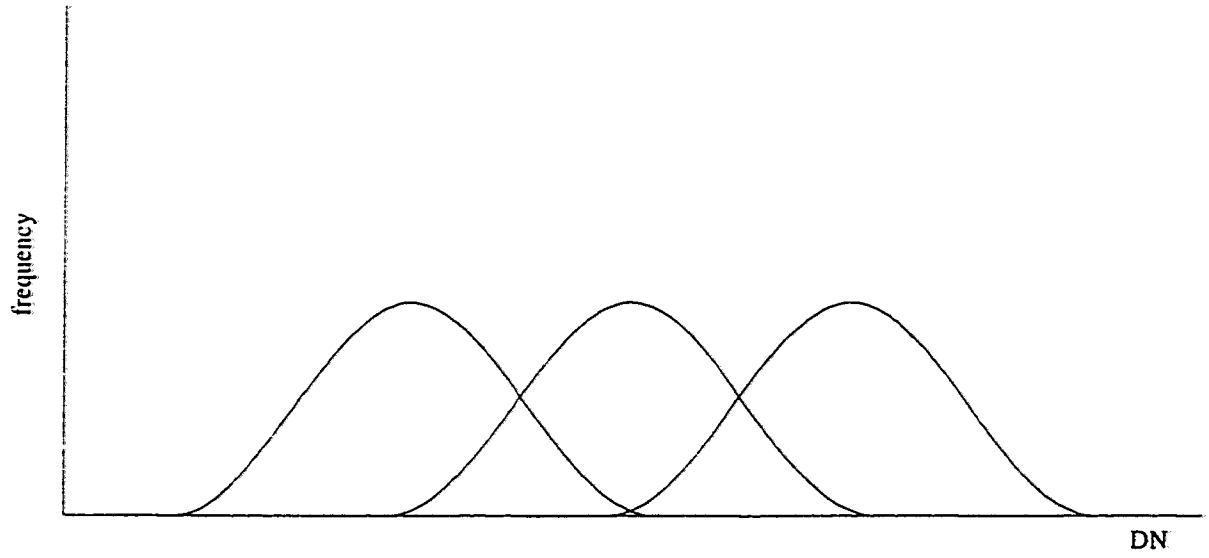
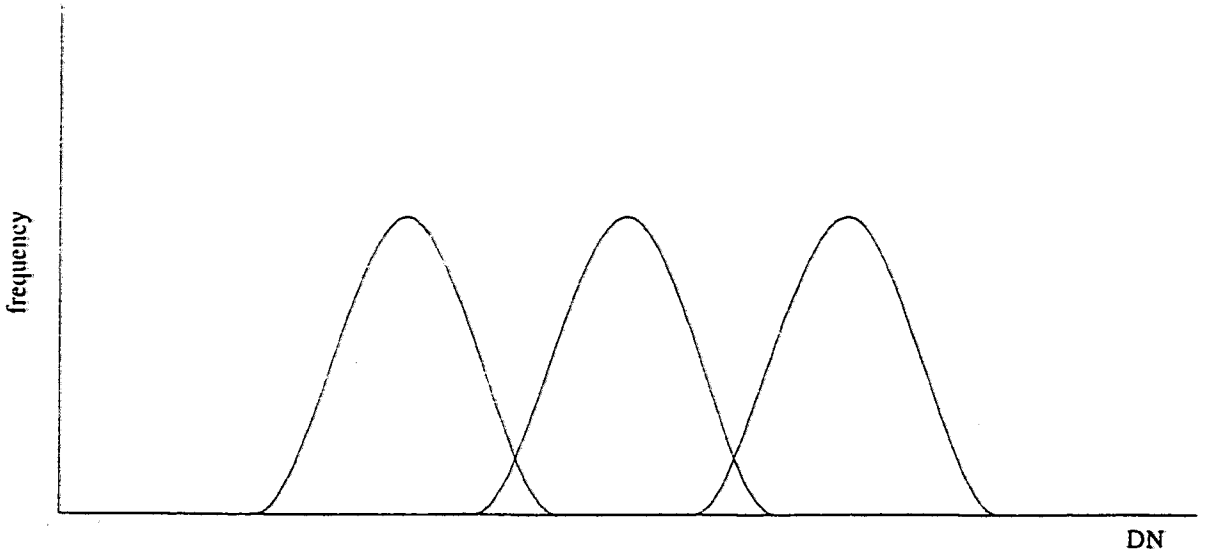


Figure 8.1: Hypothetical relationship between spectral contrast and undesirable variation relative to the number of classes.

The underlying philosophy of this research is that for a given number of classes the separability of the classes can be enhanced in part by normalizing the unwanted variation. In Figure 8.1, lowering the level of unwanted variation effectively increases spectral separability. Figure 8.2 illustrates this principle in terms of spectral distributions in one dimension. Modelling unwanted variation effectively reduces the within class variation, leaving each class more distinct. Although topographic variation and modulation transfer are significant sources of unwanted variation, there are other sources that have not been modelled. Even if all sources of unwanted variation are effectively normalized, spectral "overlap" from inherent class variability may still occur to an extent determined by the selection and number of classes.



(a) spectral contrast reduced by unwanted variation



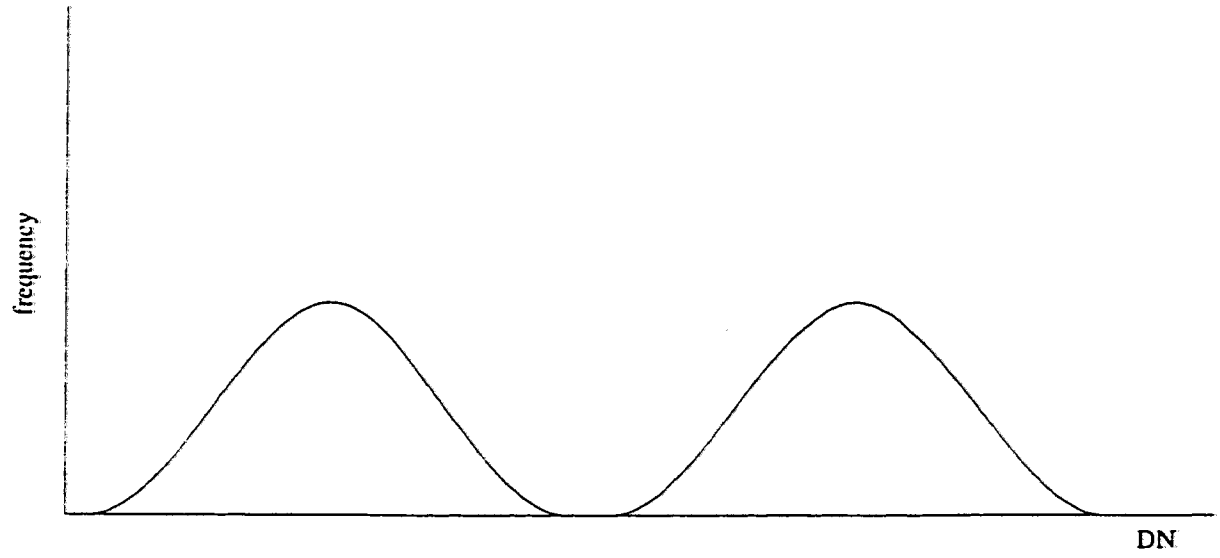
(b) spectral contrast enhanced by normalizing unwanted variation, note some pixels remain ambiguous due to inherent within class variation and/or unwanted variation not modelled

Figure 8.2: Principle of enhancing spectral contrast by normalizing unwanted variation.

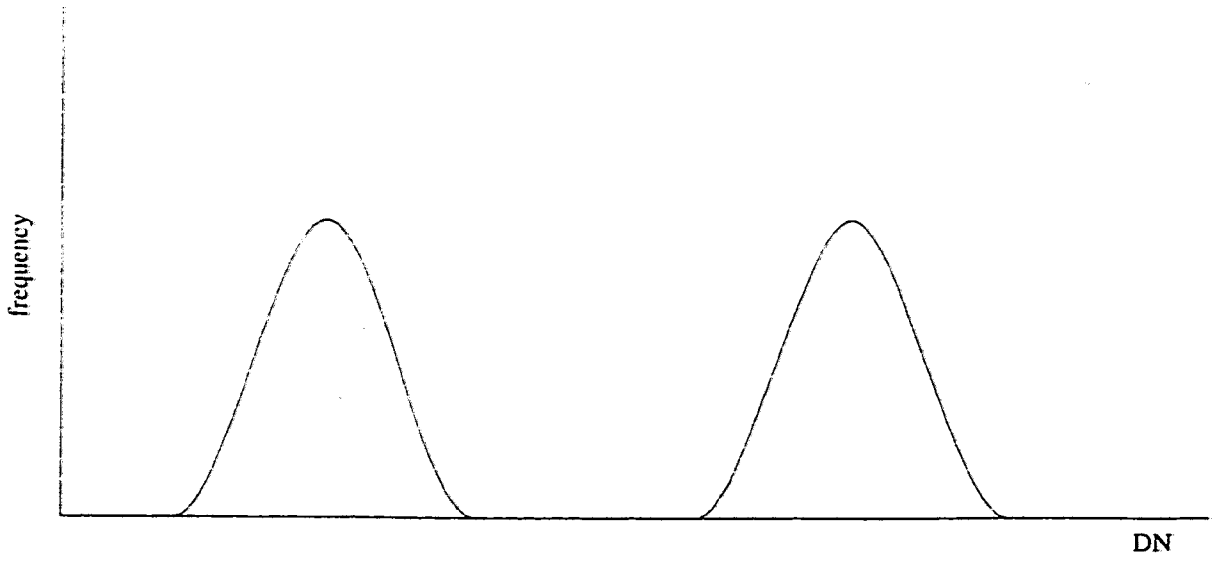
Supervised classification of reflected radiance samples assumes forests having similar structures have common reflectance characteristics. Although increasing the number of classes will tend to reduce the within class variance up to a point, the contrast between classes relative to the overall variability also determines the success of a classification. For example, in separating forests from water, the forest class will exhibit a high degree of within class variability and yet be discriminated on the basis of the high contrast between the two classes. In this case little would be gained from detailed modelling of bidirectional reflectance as the contrast between the classes would outweigh the undesirable variation. Figure 8.3 shows the diminished return from normalizing two spectrally distinct classes. Even for a given number of classes, the certainty of the classification will decrease as the variability in external factors increases. Mapping continuous forest stand parameters in lieu of classes (see Section 8.6) may reduce the problem of within class variation, however the spectral variation attributable to such parameters is also likely to be overwhelmed by the unwanted variation.

As a further experiment, the degree of spectral contrast relative to the undesirable variation could be assessed for varying numbers of classes by considering the seasonal variation within the clearcut areas as undesirable. Successively reducing the number of classes with which each image is classified until the clearcut areas are consistently classified would identify the number of classes at which spectral contrast outweighs the unwanted variation. However, one may find that this point varies not only with the number of classes, but with the spectral character of the classes defined. For example, in Figure 8.3 the two classes would exhibit high contrast having unique distributions and not because there are only two classes. As the number of classes increases, the opportunity for such unique distributions decreases. Thus, it may be more insightful to conduct this experiment with unsupervised classifications.

Spectral mixing occurs regardless of the number of classes. However, as the number of classes is increased there is a greater opportunity for the spectral mixture of two classes to resemble yet another class. Rather than finding that a given pixel has a



(a) high spectral contrast in the presence of unwanted variation
achieved with few spectrally distinct classes



(b) spectral contrast enhanced by normalizing unwanted
variation, little is gained with few spectrally distinct classes

Figure 8.3: The return from normalizing unwanted variation is diminished when spectral contrast is high.

moderate probability of belonging to the two classes being mixed, it may be assigned to the third class with high probability. Spatial information will reduce this tendency, but only with diminishing returns as the number of classes is increased.

A set of classes is spatially extensible provided the surface cover being classified is represented. Forest structure classes cannot be used to classify agricultural fields. This notion suggests that spatial extensibility can be enhanced by increasing the number of classes. However, as the number of classes is increased, the spectral contrast between them tends to diminish. The scale of the objective must be considered. A scheme consisting agricultural and forest classes would extend across Washington State. However, at that scale a large number of forest or agricultural classes may not be necessary or even desirable. This notion suggests that spatial extensibility is improved by reducing the number of classes. The solution lies in the definition of the classes. As the area being classified increases, the definition of the classes must change, not the number of classes. Class names are extremely scale dependent (Adams and Smith 1995). As the classes change, the success of the classifications will vary with their spectral contrast.

The success of a classification ultimately depends on whether or not it meets the intended objective. For a given number of classes, an unsupervised classifications will ensure maximum spectral contrast. Yet, this will be of little value if the resulting classes cannot be adequately interpreted. At the outset of this thesis a lack of clear objectives was cited as a limiting factor for the development of operational remote sensing system. However, even for a well defined objective this research suggests that it may not be possible to overcome the physical limitations of the data. For those cases it may be necessary to reconsider the objectives.

8.2 *Prior distributions and posterior probabilities*

Single value and error matrix accuracy statements (Story and Congalton 1986; Congalton 1991; Fitzgerald and Lees 1994) do not provide spatially explicit information. As Walsh (1989) points out, "thematic overlays in a GIS can be produced in a manner that meets technical accuracy standards but still fails to transmit the needed information to the user." Green and Strawderman (1994) incorporated the personal opinions of remote sensing experts via a prior distribution in a Bayesian context to estimate accuracy.

In seeking the MAP estimate, Mohn et al. (1987) handled uncertainty by specifying a doubt class. Although this assures that the remaining classes are assigned with a high degree of probability, it fails to estimate the class probabilities for those areas. For this hierarchical model, finding the MAP estimate would require fixed estimates for the components of z and the mean and precision parameters. Otherwise x will tend to a single class at their expense. However, estimating those parameters essentially requires simulating the posterior distribution as per seeking the MPM estimate. For the terrain being classified in Figure 7.52, a doubt class would be reasonable given the areas of rock outcrops. Specifying rock as a class would have been better, of course, but that underscores the problem of extending classification schemes.

By sampling from the posterior distribution, MCMC provides a method for estimating spatially explicit and probabilistic classifications. The validity of the posterior probability estimates depends on the prior distribution for x. Tjelmeland and Besag (1998) demonstrated the impact of using an Ising prior to reconstruct a scene generated from a fourth-order binary MRF model. Although visually the MPM compared rather well to the actual scene, the posterior probability distributions for each of three image functionals were incorrect. With extremely high probability, the Ising prior overestimated the total number of objects and underestimated the number objects larger than a certain size. Considering total edge length, the mean and median

Ising prior estimates were well beyond the bounds of the 90 percent credible intervals. In forestry, estimating the total edge length between two classes provides a method of monitoring adjacency constraints remotely. Unfortunately in actual forestry applications the correct prior will not be known.

Clearly, the Ising and Potts prior distributions are incorrect distributions for "true" scenes having desirable global properties. Although the Descombes et al. MRF induces such spatial coherence, it is most likely incorrect as well; Besag (1983) distinguishes the application of spatial models from modelling for this reason. However, because of its global properties, it is a better alternative than the Potts prior for this classification. Section 7.9 shows that the Descombes et al. prior is able to suppress noise without inducing excessive homogeneity. The result being that the noisy configurations are clustered at the global level corresponding to stands about which the classifier is uncertain. In this way the Descombes et al. prior confines the mobility within the sample space providing the posterior probability estimates with desirable global characteristics. Consider Figure 8.4 which shows a sample from the posterior distribution corresponding to Figure 7.14. Indeed the sample closely resembles the MPM estimate, indicating that the mobility is concentrated among the noisy areas in the MPM estimate. In his reply to the discussion, Besag (1986) explains that there is no reason to expect an MPM estimate to resemble a typical classification. Indeed, the MPM corresponds to a loss function aimed at maximizing the percent correctly classified. However, in this case there is little variability in the posterior at the global level and the sample resembles the MPM.

Further, the influence of the prior is determined by its parameterization. Section 7.9 showed that the estimated posterior probabilities are sensitive to the values of the three interaction parameters. Only the Byers (1998) version of the Descombes et al. prior was evaluated. Other clique configurations than those shown in Figure 2.7 and/or other energy assignments may yield improved results. Again, without the actual prior it is impossible to know.

Figure 8.4: Sample x taken from the posterior distribution corresponding to Figure 7.14.

Kiiveri and Campbell (1992) suggest that little is gained by modelling spatially correlated image data within a contextual classification framework. Although an MRF will induce homogeneity regardless of the data model, the validity of the posterior probability estimates will depend on the legitimacy of the data model. As Kiiveri and Campbell pursued MAP estimation via ICM, they would not have been concerned with posterior probabilities. In earlier phases of this research a number of simulations were carried out using arbitrarily narrower half-widths. Although this does not accurately model the data, it nevertheless gives greater weight to the data. Thus, in order to achieve a desired level of homogeneity more weight must be given to the spatial information, reducing mobility and yielding unsatisfactory posterior probability estimates. Estimating unique LSFs for each image captures the atmospheric and resampling variation between images. It would be worthwhile to apply this model to an image acquired under hazy sky conditions, but the availability of such undesirable images is limited. Future research should also consider the sensitivity of the posterior probability estimates to the supervising mechanism by arbitrarily varying the values of the class means.

8.3 Supervision

An accurate estimate of the mean level of reflected radiance leaving a normalized surface is required for each class in order to supervise the classification procedure. The results presented in Section 7.12 suggest even detailed modelling of bidirectional reflectance does not fully normalize the training data. More importantly, Section 7.12 demonstrated the dependency of the MPM estimates on the training data and the need for choosing appropriate training stands. More consistent solutions may have been obtained had training data been acquired from stands on horizontal terrain, rather than from stands on low slopes as was the case for this research. However, in mountainous terrain, horizontal training stands may not occur and the training

data must be normalized. The class means do not need to be obtained from the image data being classified. The use of high resolution image or field radiometer data may lead to improved estimates. However, the use of data other than that being classified requires calibration; likely eliminating the gains achieved by using alternative training data. Further, the reflectance from a particular reference stand, even if perfectly normalized, may not be indicative of the class as a whole. Although the reflected radiance from a training stand for a class should resemble the mean of all of the stands that should fall into that class, identifying such "mean" training stands would require more intensive ground sampling. However, ground sampling is time consuming, expensive, and defeats the purpose of remote sensing.

By specifying a prior distribution(s) for the class means, the Bayesian framework provides an intuitive method of modelling the uncertainty associated with parameter estimation. Section 7.8 evaluated three model formulations for this approach to supervision. However, with each of these methods convergence could not be guaranteed. The prior distribution(s) allowed for too much uncertainty and the sampled means appeared unduly influenced by the initialization and the spatial information. The resulting MPM estimates minimized the expected number of misclassifications, in part, by adjusting the class means. In this way the model behaved as a hybrid supervised-unsupervised classifier. The probability of belonging to ambiguously defined classes is not meaningful; the CC class in Figure 7.17 for example. These results support the contentions of Section 5.1.3 and parallel those of Green et al. (1994) who unfortunately did not report either the sampled or marginal posterior means.

In principle, further information for supervising the classifier may be gained by specifying a unique covariance matrix for each class, as per Green et al. (1994). This was considered during the model building phase of this project. In practice, some classes exhibit greater variability than others and z is updated accordingly. This increases the variability of the most variable classes to the point that just about any pixel could belong to that class. As this occurs the prior for x induces spatial homo-

geneity and the result is an MPM featuring one class, usually CC in this situation. Green et al. (1994) modelled only y given x and did not experience this problem. Finally, the choice of conjugate priors for the means and precisions, although somewhat reasonable, was made for convenience. As this is not a necessary requirement, future research should consider multivariate gamma distributions, having all positive values, or empirical distributions.

8.4 Temporal extensibility

Forest reflectance varies systematically in time with changing illumination angles and spatially with variation in the topographically dependent projected area. Forest canopies reflect radiation anisotropically (see Section 3.2). In addition, topography influences the SNR of the data. For example, consider TM band 3. On north facing slopes having high incidence angles, the increased projected area reduces the reflected radiance (and consequently the SNR). Thus, an AI stand on a north slope may resemble a HW stand on a south slope. The reflected radiance of the AI stand is reduced by the projected area, while HW stand inherently reflects very little in band 3. In order to consistently discriminate forest classes on the basis of their biophysical characteristics, the geometrically induced sources of variation must be sufficiently normalized. The climate on the west side of the Cascade Mountains is such that forest structure is rather independent of topography. In drier climates, surface orientation may have a distinct impact on the nature of vegetation, and therefore on the spatial distribution of the classes themselves.

Many methods have been developed for simulating forest bidirectional reflectance (see Section 3.4) and the radiosity technique applied in this study is more elaborate than most. The results in Section 7.10 suggest this method improves upon the simpler SCS model. While greater improvement would have been observed relative to the Lambertian model, the SCS model was no more difficult to implement and provides

a higher standard for comparison. In the context of a Bayesian classifier that also considers the sensor PSF and spatial context, the radiosity approach did not improve upon the SCS model for consistency between the two MPM estimates. The performance was more evident for the estimated posterior probabilities. It is important to note that the sites inconsistently classified between the two images were done so with less certainty than the consistently classified sites. Further, the hierarchical model offered much more consistent results between the two images than the supervised ML classification, which reduced the performance of the SCS model. The radiosity technique could not be evaluated with ML, because the data could not be preprocessed in a class dependent manner.

Neither the radiosity technique nor the SCS model performed as well when training data was gathered from steeper terrain, although the radiosity technique held up better. Thus, when training data must be collected on steep terrain, the increased complexity of the radiosity technique may be justified. In both cases, the orientation of the estimated classes corresponded to that of the training stand orientation. This suggests that topographic variation was having a greater influence on the results than the variation between stands of the same class.

Bidirectional reflectance is a scale dependent phenomenon that varies continuously across space with forest structure, over time with forest growth, and along the EM spectrum depending on physical properties of the materials being sensed. This project effectively simulated 780 unique BRDFs, however only in discrete increments. Structural continuity was approximated via the polynomial surface, but the spectral increments were not. The spectral resolution of the TM sensor is relatively coarse. Although the spectral separability of the classes is limited by the resolution, even coarse measurements may be influenced by variation at a continuous level that has not been accounted for by this formulation. Future research priorities should include within-crown scattering, transmittance, and compare simulated BRDFs with empirical BRDFs constructed from many TM scenes. The sensitivity of the estimates to the

accuracy of the DEM can be investigated by arbitrarily varying the elevation values. The uncertainty of estimating bidirectional reflectance may also be handled by specifying prior distributions for those parameters. This was considered during an earlier phase of this project. An independent gamma prior was assigned to each component of $T_{x_i}^{(i)}$ having the NBRF estimate as its mean and arbitrary variance. However, this was not pursued as the initial results did not justify the increased computational burden.

8.5 Modelling assumptions

Mathematical and statistical models are necessarily abstractions of real world phenomena. As a modelling process, the statistical classification of image data requires a number of assumptions. Many analysts are often unaware of the implicit assumptions they are making (Duggin and Robinove 1990). Correlations can be misleading and/or unreliable. For example, Horler and Ahern (1986) transformed TM data and labeled the third principal component "swirness," because it appeared to represent a contrast between the first four TM bands and the two short wave infrared bands. Likewise, the models developed for this project assumed normally distributed radiance, whereas in principle reflected radiance is a positive quantity.

The reviews of spatial statistics, MCMC, bidirectional reflectance, and multispectral scanners earlier in this thesis were intended to outline the complexity of the problems involved. Although some general principles of spectral radiance are well understood many unknowns remain, such as the relative contribution of scattered diffuse irradiance to that of multiple scattered radiance. Often research is limited by the expense and/or difficulty of measuring complex physical phenomena.

By way of detailed modelling this project substitutes many explicit assumptions in place of gross assumptions. For example, many more assumptions were required to model forest anisotropy in place of assuming isotropy, with the aim of developing

a more physically representative model. However, what is appropriate at one scale may not be so at another. While the radiosity simulations captured the effects of mutual shadowing between crowns at one scale, clearly the opacity assumption is inappropriate at the scale of the individual crown. The use of a DEM for estimating reflectance factors and the comparison of multiple image classifications requires exact co-registration. Co-registration errors can reduce the effectiveness of the normalization and lead to inconsistent comparisons. Even with perfect registration of the digital products, it is highly unlikely that the IFOVs line up between any two images, creating another source of unwanted variation.

As for detailed physical models, the question remains: how much is enough? This research indicates that the radiosity method produced better estimates of bidirectional reflectance factors than the SCS model. Yet, in the context of the hierarchical model, the radiosity estimates did not improve temporal extensibility, although the uncertainty estimates were affected. This suggests that, all else being equal, undesirable variation outweighed the subtle improvements of the detailed reflectance model. However, the addition of spatial information did improve the consistency of multiple image estimates by more than ten percent. Instead of attempting to account for undesirable spectral variation by explicitly modelling it, Smith et al. (1994) presented a method of separating desired variation from background variation. This approach accounts for all unwanted variation, both known and unknown, without having to model it. Further improvements may be gained by adopting this philosophy in conjunction with contextual models. Indeed, Bayesian image analysis provides a flexible framework. The model developed in this thesis is merely one example.

Finally, the implementation itself may be flawed. During the course of this research the model has been run hundreds, if not thousands, of times. Occasionally, bugs in the software were identified and results discarded. At one stage, simulated images corresponding to the GIS classification were sampled from the model and subsequently classified. This served as a further check on the legitimacy of the results. While it

would be nearly impossible to guarantee that the software is bug-free, this is very likely the case.

8.6 Accuracy assessment

As remote sensing models capable of producing consistent solutions are developed accuracy assessment will present a difficult challenge. Classification accuracy depends on both the number of classes and the accuracy of the classification of the reference data. Moreover, the accuracy of MPM estimates depend on the accuracy of the n posterior probability distributions for which there may be no reference or "ground truth". Reference data is required for computing the most common accuracy measures; see, for example Story and Congalton (1986), Congalton (1991), Green et al. (1993). and Fitzgerald and Lees (1994). However, there are as many ways to classify surface cover as there are ways to classify images. There is not a method for directly and objectively measuring thematic classes. If the abstract cartographic features being mapped do not exist in the real world, then there can be no accurate standard against which to assess data layer accuracy (Veregin 1989). For "ground truth" information to be relevant for classifying image data and assessing accuracy it must represent what should be discriminated for a given set of classes.

The literature reviewed in Chapters 3 and 4 indicates the complexity of forest classification and identifies many potential sources of error in classified satellite images. Likewise, there are many potential sources of error for corresponding GIS classifications (Goodchild and Gopal 1989; Thapa and Bossler 1992; Fenstermaker 1994). Figures 7.3 and 7.51 are examples of thematic maps produced from the classification of field survey data, management records, photo interpretation and stratification, and/or other image classification. Compare the GIS classification in Figure 7.3 with the ML classification in Figure 7.43. The classification of the area slightly up and to the left of stand 727 is remarkably similar, with a narrow zone of UR surrounded by

LE forest. This area was one of the primary discrepancies between the GIS and MPM classifications. In fact much of the area classified as LE by the GIS had been selectively harvested in the recent past although no records for this area were available. Although the average diameter of a plot measured in this area was less than that of other UR stands, the diameters ranged up to 31 inches suggesting that UR may be a reasonable classification; especially if the GIS information was derived from unknown image interpretation.

Digitizing photo stratifications is another source of error, and to use photo interpretation as "ground truth" is a misnomer "that has clearly had a negative impact on the results of satellite image classification accuracy (Bolstad and Smith 1992)." Biging et al. (1994) compared the work of six photo interpreters, each having at least ten years experience, finding a surprising degree of variation between each individual. In another experiment, Chrisman and Lester (1991) found 85 percent consistency between the land cover maps produced from photos by two interpreters. Computer classification methods are actually computer-aided analysis procedures and the accuracy and the results are dependent upon the analyst's ability (Hoffer 1986). A further problem arises in the use of ground and photo data for accuracy assessment in that the issue of scale is implicitly ignored.

Classification accuracy also depends upon the number of classes, with accuracy increasing as the number of classes are reduced (Chrisman 1982; Felix and Binney 1989; Congalton 1991). However, this may lead to situations in which the final set of classes is inappropriate for the initial objective. In principle the classification objective will define the classes and the level of acceptable accuracy, thus determining the resolution of the data required. For this study six classes were considered, comparable to actual regional forest classifications in the Northwest. More than six classes would have significantly increased the already significant computational burden. Nevertheless, some future research should be directed at assessing the sensitivity of the MPM and posterior probability estimates to the number of classes (see Section 8.1). Given

the rate of computational advancement, such implementations should be feasible in the future.

A more formal accuracy evaluation could be conducted on the basis of measurable stand parameters, such as trees per acres, rather than on the basis of the structure classes themselves. For a set of reference stands in a given class, stand parameters could be plotted against the probability of the stand belonging to the class as defined by specific parameter values. Finally, as an alternative to classifying discrete structure classes, a model could be developed to map continuous "intensities" of parameters such as trees per acres.

8.7 Conclusion

This study identifies some limitations of remote sensing, while simultaneously attempting to improve the technology. The approach contrasts studies that have evaluated *what* can be discriminated from remote sensing data given a specific technology. Contemporary image correction and classification routines do not account for all of the variability necessary to extend results over time and space. The complexity of such an algorithm will need to match the the data and information desired (Landgrebe 1976). The level of modelling detail undertaken here was to this end.

Forest structure classification can be improved by evaluating and applying theoretical developments in statistics and engineering. The original aspects of this dissertation include the development of a hierarchical framework for classifying remotely sensed data that overcomes the assumption of class conditional independence. The Bayesian paradigm provides a flexible method of incorporating both spectral and spatial information in the classification of image data. Spatial continuity is enhanced by considering contextual information. The application of a fifth-order random field to forest remote sensing is an original aspect of this research.

Simulating multiple slope and class dependent bidirectional reflectance distribu-

tion functions for the purpose of normalizing projected area variation in rugged terrain and between multiple images is another important and original feature of this research. Temporal extensibility is improved by hierarchical modelling of anisotropic reflectance, effective spatial resolution, and spatial interaction. However, in the context of hierarchical modelling, the solutions are more sensitive to the selection of training data and the parameterization of the MRF than to the estimates of bidirectional reflectance.

This research has provided insights over answers. Depending on the specific mapping objective and the corresponding set of classes, accuracy assessment and operational feasibility may be premature issues. Probabilistic classifications estimate the level of certainty with which each pixel is classified–the logical result of statistical inference.

BIBLIOGRAPHY

Abuelgasim, A. A. and A. H. Strahler (1994). Modeling bidirectional radiance measurements collected by the advanced solid-state array spectroradiometer (ASAS) over Oregon transect conifer forests. *Remote Sensing of Environment 41*, 261–275.

Adams, J. and M. O. Smith (1995). Changes in vegetated landscapes: A process-model approach to multispectral remote sensing. In D. Peter (Ed.), *Remote Sensing for Land Degradation and Desertification Monitoring in the Mediterranean Basin*. Volume DG XII-D2. Commission of the European Communities.

Adams, J., M. O. Smith, and A. R. Gillespie (1989). Simple models for complex natural surfaces: A strategy for the hyperspectral era of remote sensing. In *Remote Sensing: an economic tool for the nineties / 1989 International Geoscience and Remote Sensing Symposium*, pp. 16–21. IEEE, CRSS, URSI.

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer (1976). A land use and land cover classification system for use with remote sensor data. Geological Survey professional paper 964, United States Department of the Interior, Geological Survey.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley.

Anuta, P. E., L. A. Bartolucci, M. E. Dean, D. F. Lozano, E. Malaret, C. D. McGillem, J. Valdés, and C. R. Valenzuela (1984). Landsat-4 MSS and Thematic Mapper data quality and information content analysis. *IEEE Transactions on Geoscience and Remote Sensing 22*, 222–236.

Barber, D. G. and E. F. LeDrew (1991). SAR sea ice discrimination using texture statistics: A multivariate approach. *Photogrammetric Engineering and Remote Sensing 57*, 385–395.

Bartholome, E. (1991). Remote sensing and agricultural production monitoring in Sahelian countries. In A. S. Belward and C. R. Valenzuela (Eds.), *Remote Sensing and Geographical Information Systems for Resource Management in Developing Countries*, pp. 189–214. Kluwer Academic Publishers.

Basu, J. P. and P. L. Odell (1974). Effects of intraclass correlation among training samples on the misclassification probabilities of Bayes' procedure. *Pattern Recognition 6*, 13–16.

Bernknopf, R. L., R. H. Campbell, D. S. Brookshire, and C. D. Shapiro (1988). A probabilistic approach to landslide hazard mapping in Cincinnati, Ohio, with applications for economic evaluation. *Bulletin of the Association of Engineering Geologists 25*, 39–56.

Berry, J. K. (1994). Toward a humane and honest GIS: Traditional mapping in triage. In *Decision Making with GIS: The Fourth Dimension/Proceedings of the Eighth Annual Symposium on Geographic Information Systems in Forestry, Environmental and Natural Resources Management*, Volume 1, pp. 35–39.

Besag, J. (1983). Discussion of paper by P. Switzer. In *Bulletin of the International Statistical Institute: Proceedings of the 44th Session*, Volume L, Book 3, pp. 422–425.

Besag, J. (1989). Towards bayesian image analysis. *Journal of Applied Statistics 16*, 395–407.

Besag, J. and P. J. Green (1993). Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society, Series B 55*, 25–37.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B 36*(2), 192–236.

Besag, J. E. (1986). Statistical analysis of dirty pictures (with Discussion). *Journal of the Royal Statistical Society, Series B 48*, 259–302.

Besag, J. E., P. J. Green, D. M. Higdon, and K. Mengersen (1995). Bayesian computation and stochastic systems (with Discussion). *Statistical Science 10*, 3–66.

Besag, J. E., J. York, and A. Mollié (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics 43*, 1–59.

Best, R. G. (1982). Handbook of remote sensing in fish and wildlife management. Technical Report SDSU-RSI-82-05, Remote Sensing Institute, South Dakota State University.

Biging, G. S., R. G. Congalton, and E. C. Murphy (1994). A comparison of photointerpretation and ground measurements of forest structure. In L. K. Fenstermaker (Ed.), *Remote Sensing Thematic Accuracy: A Compendium*, pp. 148–157. American Society for Photogrammetry and Remote Sensing.

Bolstad, P. V. and J. L. Smith (1992). Errors in GIS: Assessing spatial data accuracy. *Journal of Forestry 90*(11), 21–29.

Bondi, H. (1989). Introduction. In S. R. Brooks (Ed.), *Mathematics in Remote Sensing*, Number 21 in The Institute of Mathematics and its Application Conference Series, pp. 1–4. Clarendon Press.

Bonham-Carter, G. F., F. P. Agterberg, and D. F. Wright (1988). Integration of geological datasets for gold exploration in Nova Scotia. *Photogrammetric Engineering and Remote Sensing 54*, 1585–1592.

Borel, C. C. and S. A. W. Gerstl (1994). Radiosity based model for terrain effects on multi-angular view. In *Surface and Atmospheric Remote Sensing: Technologies, Data Analysis, and Interpretation / 1994 International Geoscience and Remote Sensing Symposium*, pp. 1667–1669. IEEE.

Borel, C. C., S. A. W. Gerstl, and B. J. Powers (1991). The radiosity method in optical remote sensing of structured 3-d surfaces. *Remote Sensing of Environment 36*, 13–44.

Bouman, C. and B. Liu (1991). Multiple resolution segmentation of textured images. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-13*, 99–113.

Bouman, C. A. and M. Shapiro (1994). A multiscale random field model for Bayesian image segmentation. *IEEE Transactions on Image Processing 3*, 162–177.

Brakke, T. W., J. Otterman, J. R. Irons, and F. G. Hall (1996). Assessing canopy biomass and vigor by model-inversion of bidirectional reflectances: Problems and prospects. In *Remote Sensing for a Sustainable Future / 1996 International Geoscience and Remote Sensing Symposium*, pp. 1657–1659. IEEE.

Brown, E. R. (1985). Management of wildlife and fish habitats in forests of western oregon and washington. Technical report, USDA Forest Service, PNW.

Brown de Colstoun, E. C., C. L. Walthall, A. T. Ciallla, E. R. Vermote, and R. N. Halthore (1996). Variability of BRDF with land cover type for the west central HAPEX-Sahel super site. In *Remote Sensing for a Sustainable Future / 1996 International Geoscience and Remote Sensing Symposium*, Volume 4, pp. 1904–1907. IEEE.

Bryant, N. A., A. L. Zobrist, R. E. Walker, and B. Gokhman (1985). An analysis of Landsat Thematic Mapper P-product internal geometry and conformity to

earth surface geometry. *Photogrammetric Engineering and Remote Sensing 51*, 1435–1447.

Burger, D. and G. Pierpoint (1990). Trends in forest site and land classification in Canada. *The Forestry Chronicle 66*(4), 91–96.

Burgess, D. W., J. R. Dymond, D. Pairman, and J. D. Sheperd (1997). Scale integration of bi-directional reflectance effects in remotely sensed vegetated surfaces. In *Remote Sensing - A Scientific Vision for Sustainable Development / 1997 International Geoscience and Remote Sensing Symposium*, pp. 1075–1077. IEEE.

Byers, S. D. (1998). *Bayesian Modeling of Highly Structured Systems using Markov Chain Monte Carlo*. Ph. D. thesis, University of Washington, Department of Statistics.

Campbell, J. B. (1981). Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogrammetric Engineering and Remote Sensing 47*, 355–363.

Campbell, J. B. (1987). *Introduction to Remote Sensing*. The Guilford Press.

Cescatti, A. (1997). Modelling the radiative transfer in discontinuous canopies of asymmetric crowns. I. Model structure and algorithms. *Ecological Modelling 101*, 263–274.

Chen, J. M. and S. G. Leblanc (1996). A four-level bidirectional reflectance model based on canopy architecture and its inversion. In *Remote Sensing for a Sustainable Future / 1996 International Geoscience and Remote Sensing Symposium*, pp. 1651–1653. IEEE.

Chrisman, N. (1997). *Exploring Geographic Information Systems*. John Wiley and Sons.

Chrisman, N. R. (1982). Beyond accuracy assessment: Correction of misclassification. In *Environmental Assessment and Resource Management: Proceedings of the Fifth International Symposium on Computer-Assisted Cartography*, pp. 123–132. American Society of Photogrammetry and American Congress on Surveying and Mapping.

Chrisman, N. R. (1987). Design of geographic information systems based on social and cultural goals. *Photogrammetric Engineering and Remote Sensing 53*, 1367–1370.

Chrisman, N. R. and M. Lester (1991). A diagnostic test for error in categorical maps. In *Proceedings of the Tenth International Symposium on Computer-Assisted Cartography/Auto-Carto 10*, pp. 330–348. ASPRS.

Christensen, E., J. Jensen, E. Ramsey, and J. H.E. Mackey (1988). Aircraft MSS data registration and vegetation classification for wetland change detection. *International Journal of Remote Sensing 9*, 23–38.

Civco, D. L. (1989). Topographic normalization of Landsat Thematic Mapper digital imagery. *Photogrammetric Engineering and Remote Sensing 55*, 1303–1309.

Cohen, W. B. (1994). Applications perspective: Current research on remote sensing of forest structure. In V. A. Sample (Ed.), *Remote Sensing and GIS in Ecosystem Management*, pp. 91–107. Island Press.

Cohen, W. B., T. A. Spies, and M. Fiorella (1995). Estimating age and structure of forests in a multi-ownership landscape of western Oregon, U.S.A. *International Journal of Remote Sensing 16*, 721–746.

Colby, J. D. (1991). Topographic normalization in rugged terrain. *Photogrammetric Engineering and Remote Sensing 57*, 531–537.

Colwell, R. N. (1961). Some practical applications of multiband spectral reconnais-

sance. *American Scientist 49*, 9–36.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment 37*, 35–46.

Congalton, R. G., K. Green, and J. Teply (1993). Mapping old growth forests on National Forest and Park lands in the Pacific Northwest from remotely sensed data. *Photogrammetric Engineering and Remote Sensing 59*, 529–535.

Conradsen, K. and B. K. Nielsen (1987). Textural features useful in the classification of digital images. In T. Pukkila and S. Puntanen (Eds.), *Proceedings of the Second International Tampere Conference in Statistics*, pp. 143–159. University of Tampere.

Corbley, K. P. (1994). Canada counts potatoes and assesses crop health with satellites. *Earth Observation Magazine 3*(3), 36–40.

Craig, R. G. (1979). Autocorrelation in Landsat data. In *Proceedings of the Thirteenth International Symposium on Remote Sensing of Environment*, pp. 1517–1524. Environmental Research Institute of Michigan.

Craighead, J. J., F. L. Craighead, and D. J. Craighead (1985). Using satellites to evaluate ecosystems as grizzly bear habitat. In *Proceedings–Grizzly Bear Habitat Symposium*, pp. 101–112. USDA Forest Service, Intermountain Research Station.

Crea, W. (1973). Agriculture, forestry, range resources. In S. C. Fresden and E. P. Mercanti (Eds.), *Symposium on Significant Results Obtained from Earth Resources Technology Satellite-1*, pp. 1–17. NASA, Goddard Space Flight Center.

Croft, F. and B. Kessler (1996). Remote sensing, image processing, and GIS: Trends and forecasts. *Journal of Forestry 94*(6), 31–35.

Cross, G. R. and A. K. Jain (1983). Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-5*, 25–39.

Davis, F. W. and J. Dozier (1990). Information analysis of a spatial database for ecological land classification. *Photogrammetric Engineering and Remote Sensing 56*, 605–613.

Deering, D. W., E. M. Middleton, and T. F. Eck (1994). Reflectance anisotropy for a spruce-hemlock forest canopy. *Remote Sensing of Environment 47*, 242–260.

Demoment, G. (1989). Image reconstruction and restoration: Overview of common estimation structures and problems. *IEEE Transactions on Acoustics, Speech, and Signal Processing 37*, 2024–2036.

Derin, H. and H. Elliott (1987). Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-9*, 39–55.

Descombes, X., J.-F. Mangin, E. Pechersky, and M. Sigelle (1995). Fine structures preserving Markov model for image processing. In *Proceedings of the 9th Scandinavian Conference on Image Analysis*, pp. 349–356.

Duggin, M. J. (1974). On the limitations of target discrimination by means of spectral discrimination techniques. In *Proceedings of the Ninth International Symposium on Remote Sensing of Environment*, Volume 1, pp. 499–516. Environmental Research Institute of Michigan.

Duggin, M. J. (1985). Factors limiting the discrimination and quantification of terrestrial features using remotely sensed radiance. *International Journal of Remote Sensing 6*, 3–27.

Duggin, M. J. (1990). A survey of remote sensing methodology for forest inventory. In *State-of-the-Art Methodology of Forest Inventory: A Symposium Proceedings*,

pp. 267–285. USDA, Forest Service, PNW Research Station, General Technical Report PNW-GTR-263.

Duggin, M. J. and C. J. Robinove (1990). Assumptions implicit in remote sensing data acquisition and analysis. *International Journal of Remote Sensing 11*, 1669–1694.

Eisenbeis, K. M. (1995). *Privatizing Government Information: The Effects of Policy on Access to Landsat Satellite Data*. The Scarecrow Press.

Elachi, C. (1987). *Introduction to the Physics and Techniques of Remote Sensing*. Wiley.

Eller, R. G. and J. J. Ulliman (1974). ERTS-1 data applications to Minnesota forest land use classification. IARSL Research Report 74-3, Remote Sensing Laboratory, Institute of Agriculture, College of Forestry, University of Minnesota.

Engel, J. L. and O. Weinstein (1983). The Thematic Mapper–an overview. *IEEE Transactions on Geoscience and Remote Sensing 21*, 258–265.

Evanisko, F. (1990). New perspectives. In *Protecting Natural Resources with Remote Sensing: Proceedings from the Third Forest Service Remote Sensing Applications Conference*, pp. 403–413. USDA, Forest Service.

Felix, N. A. and D. L. Binney (1989). Accuracy assessment of a Landsat-assisted vegetation map of the coastal plain of the Artic National Wildlife Refuge. *Photogrammetric Engineering and Remote Sensing 55*, 475–478.

Fenstermaker, L. K. (Ed.) (1994). *Remote Sensing Thematic Accuracy: A Compendium*. American Society for Photogrammetry and Remote Sensing.

Fiorella, M. and W. J. Ripple (1993). Determining successional stage of temperate coniferous forests with Landsat satellite data. *Photogrammetric Engineering and Remote Sensing 59*, 239–246.

Fischel, D. (1984). Validation of the Thematic Mapper radiometric and geometric correction algorithms. *IEEE Transactions on Geoscience and Remote Sensing 22*, 237–242.

Fitzgerald, R. W. and B. G. Lees (1994). Assessing the classification accuracy of multisource remote sensing data. *Remote Sensing of Environment 47*, 362–368.

Forshaw, M. R. B., A. Haskell, P. F. Miller, D. J. Stanley, and J. R. G. Townshend (1983). Spatial resolution of remotely sensed imagery: A review paper. *International Journal of Remote Sensing 4*, 497–520.

Franklin, J., T. L. Logan, C. E. Woodcock, and A. H. Strahler (1986). Coniferous forest classification and inventory using Landsat and digital terrain data. *IEEE Transactions on Geoscience and Remote Sensing 24*, 139–149.

Franklin, J. and A. H. Strahler (1988). Invertible canopy reflectance modeling of vegetation structure in semiarid woodland. *IEEE Transactions on Geoscience and Remote Sensing 26*, 809–825.

Franklin, S. E. (1994). Discrimination of subalpine forest species and canopy density using digital CASI, SPOT, PLA, and Landsat TM data. *Photogrammetric Engineering and Remote Sensing 60*, 1233–1241.

Fraser, R. S., O. P. Bahethi, and A. H. Al-Abbas (1977). The effects of atmosphere on the classification of satellite observations to identify surface features. *Remote Sensing of Environment 6*, 229–249.

Gelfand, A. E., S. E. Hills, A. Racine-Poon, and A. F. M. Smith (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association 85*, 972–985.

Geman, D. (1987). Stochastic model for boundary detection. *Image Vision and Computing 5*, 61–65.

Geman, D., S. Geman, C. Graffigne, and P. Dong (1990). Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-12*, 609–628.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6*, 721–741.

Gemmell, F. M. (1995). Effects of forest cover, terrain, and scale on timber volume estimation with Thematic Mapper data in a Rocky Mountain site. *Remote Sensing of Environment 51*, 291–305.

Gerard, F. F. and P. R. J. North (1997). Analyzing the effect of structural variability and canopy gaps on forest BRDF using a geometric-optical model. *Remote Sensing of Environment 62*, 46–62.

Gjertsen, A. (1993). Testing TM and spot data as input to a canopy reflectance model for mapping density and size of trees in forest stands. In *Better Understanding of Earth Environment / 1993 International Geoscience and Remote Sensing Symposium*, pp. 750–752. IEEE.

Goodchild, M. F. and S. Gopal (Eds.) (1989). *Accuracy of Spatial Databases*. Taylor and Francis.

Goodchild, M. F., S. Guoqing, and Y. Shiren (1992). Development and test of an error model for categorical data. *International Journal of Geographical Informations Systems 6*, 87–104.

Gordon, F. (1983). The time-space relationships among data points from multispectral spatial scanners. *International Journal of Remote Sensing 4*, 555–570.

Gorte, B. and A. Stein (1998). Bayesian classification and class area estimation of satellite images using stratification. *IEEE Transactions on Geoscience and*

Remote Sensing 36, 803–812.

Green, E. J., A. F. M. Smith, and W. E. Strawderman (1994). Construction of thematic maps from satellite imagery. In *Bayesian Statistics 5: Proceedings of the 4th Valencia International Meeting*, pp. 75–104.

Green, E. J. and W. E. Strawderman (1994). Determining accuracy of thematic maps. *The Statistician 43*, 77–85.

Green, E. J., W. E. Strawderman, and T. M. Airola (1993). Assessing classification probabilities for thematic maps. *Photogrammetric Engineering and Remote Sensing 59*, 635–639.

Green, K. S., S. Bernath, L. Lackey, M. Brunengo, and S. Smith (1993). Analyzing the cumulative effects of forest practices: Where do we start? *Geo Info Systems 3*(2), 31–41.

Gu, D. and A. Gillespie (1998). Topographic normalization of Landsat TM images of forest based on subpixel sun-canopy-sensor geometry. *Remote Sensing of Environment 64*, 166–175.

Hahn, J. T., R. E. McRoberts, and W. Befort (1992). Annual forest inventory system: Integrating data base techniques, satellite imagery, annual design sampling, and modelling. In G. Wood and B. Turner (Eds.), *Integrating Forest Information over Space and Time*, pp. 314–324. IUFRO.

Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE 67*, 786–804.

Harding, R. A. and R. B. Scott (1978). Forest inventory with Landsat: Phase II Washington forest productivity study. Technical report, State of Washington Department of Natural Resources.

Hartt, K., M. J. Carlotto, and M. W. Brennan (1989). A method for multi-dimensional image segmentation. In *Remote Sensing: An Economic Tool for the Nineties / 1989 International Geoscience and Remote Sensing Symposium*, pp. 509–512. IEEE.

Haslett, J. (1985). Maximum likelihood discriminant analysis on the plane using a Markovian model of spatial context. *Pattern Recognition 18*, 287–296.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97–109.

Higdon, D. (1994). *Incorporating Uncertainty in Bayesian Inference for Spatial Data*. Ph. D. thesis, Department of Statistics, University of Washington.

Hill, G. J. and G. D. Kelly (1987). Habitat mapping by Landsat for aerial census of kangaroos. *Remote Sensing of Environment 21*, 53–60.

Hitchcock, A. J. (1996). Landscape planning on the Harry Osborne forest. Master's thesis, University of Washington.

Hjort, N. J. and E. Mohn (1984). A comparison of some contextual methods in remote sensing classification. In *Proceedings of the Eighteenth International Symposium on Remote Sensing of Environment*, Volume 3, pp. 1693–1702. Environmental Research Institute of Michigan.

Hjort, N. L., E. Mohn, and G. O. Storvik (1985). Contextual classification of remotely sensed data, based on an autocorrelation model. In H. V. Sæbø et al. (Eds.), *Contextual Classification of Remotely Sensed Data: Statistical Methods and Development of a System*. Norwegian Computing Center. Technical Report 768.

Hodgson, M. E. (1995). What cell size does the computed slope/aspect angle represent? *Photogrammetric Engineering and Remote Sensing 61*, 513–517.

Hoffer, R. M. (1986). Digital analysis techniques for forestry applications. *Remote Sensing Reviews 2*, 61–110.

Hoffer, R. M. (1994). Challenges in developing and applying remote sensing to ecosystem management. In V. A. Sample (Ed.), *Remote Sensing and GIS in Ecosystem Management*, pp. 25–40. Island Press.

Holben, B. N. and C. O. Justice (1980). The topographic effect on spectral response from nadir-pointing sensors. *Photogrammetric Engineering and Remote Sensing 46*, 1191–1200.

Holmgren, P. (1995). Geographic information for forestry planning: Forest mapping in the 1820's and 1990's. Is there a difference? Reports in Forest Ecology and Forest Soils 68, Department of Forest Soils, Swedish University of Agricultural Sciences.

Horler, D. N. H. and F. J. Ahern (1986). Forestry information content of Thematic Mapper data. *International Journal of Remote Sensing 7*, 405–428.

Hugli, H. and W. Frei (1983). Understanding anisotropic reflectance in mountainous terrain. *Photogrammetric Engineering and Remote Sensing 49*, 671–683.

Hutchinson, C. F. (1991). Uses of satellite data for famine early warning in Sub-Saharan Africa. *International Journal of Remote Sensing 12*, 1405–1421.

Itten, K. I. and P. Meyer (1993). Geometric and radiometric correction of TM data of mountainous forested areas. *IEEE Transactions on Geoscience and Remote Sensing 31*(4), 764–770.

Iverson, L. R., R. L. Graham, and E. A. Cook (1989). Applications of satellite remote sensing to forested ecosystems. *Landscape Ecology 3*, 131–143.

Jeon, B. and D. A. Landgrebe (1990). Spatio-temporal contextual classification of remotely sensed multispectral data. In *1990 IEEE International Conference on*

Systems, Man, and Cybernetics: Conference Proceedings, pp. 1819–1822. IEEE.

Kasten, F. and E. Raschke (1974). Reflection and transmission terminology by analogy with scattering. *Applied Optics 13*, 460–464.

Kato, Z., J. Zerubia, and M. Berthod (1992). Satellite image classification using a modified metropolis dynamics. In *Multidimensional Signal Processing / 1992 International Conference on Acoustics, Speech and Signal Processing*, Volume 3, pp. 573–576. IEEE.

Kaufman, Y. J. (1984). Atmospheric effect on spatial resolution of imagery. *Applied Optics 23*, 3400–3408.

Kaufman, Y. J. (1985). The atmospheric effect on the separability of field classes measured from satellites. *Remote Sensing of Environment 18*, 21–34.

Keller, M. and M. Köhl (1996). Optimisation of the Swiss National Forest Inventory (NFI) with remote sensing techniques. In P. J. Kennedy (Ed.), *Application of Remote Sensing in European Forest Monitoring: International Workshop Proceedings*, pp. 161–166. Institute of Surveying, Remote Sensing, and Land Information, University of Agricultural Sciences (BOKU).

Kelly, P. A., H. Derin, and K. D. Hartt (1988). Adaptive segmentation of speckled images using a hierarchical random field model. *IEEE Transactions on Acoustics, Speech, and Signal Processing 36*, 1628–1641.

Kent, B., D. Johnston, and W. Frayer (1979). The application of three-phase sampling for stratification to multi-resource inventories. In W. Frayer (Ed.), *Forest Resource Inventories: Proceedings of a Workshop*, pp. 993–1000. CSU, SAF, IUFRO.

Kent, J. T. and K. V. Mardia (1988). Spatial classification using fuzzy membership models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 10*,

659–671.

Kettig, R. L. and D. A. Landgrebe (1976). Classification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Transactions on Geoscience Electronics GE-14*, 19–26.

Khazenie, N. and M. M. Crawford (1990). Spatial-temporal autocorrelated model for contextual classification. *IEEE Transactions on Geoscience and Remote Sensing 28*, 529–539.

Kiiveri, H. and N. Campbell (1992). Allocation of remotely sensed data using Markov models for image data and pixel labels. *Australian Journal of Statistics 34*, 361–374.

Kimes, D. S. and J. A. Kirchner (1981). Modeling the effects of various radiant transfers in mountainous terrain on sensor response. *IEEE Transactions on Geoscience and Remote Sensing GE-19*(2), 100–108.

Kindermann, R. and J. L. Snell (1980). *Markov Random Fields and their Applications*. American Mathematical Society.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi (1983). Optimization by simulated annealing. *Science 220*, 671–680.

Kittler, J. (1983). Image processing for remote sensing. *Philosophical Transactions of the Royal Society of London A 309*, 323–335.

Kittler, J. and J. Föglein (1984). Contextual classification of multispectral pixel data. *Image Vision and Computing 2*, 13–29.

Kittler, J. and D. Pairman (1985). Contextual pattern recognition applied to cloud detection and identification. *IEEE Transactions on Geoscience and Remote Sensing 23*, 855–863.

Kittler, J. and D. Pairman (1989). Global performance criteria and corresponding algorithms for contextual classification of multispectral images. In S. R. Brooks (Ed.), *Mathematics in Remote Sensing*, Number 21 in The Institute of Mathematics and its Application Conference Series, pp. 241–256. Clarendon Press.

Klein, R. and S. J. Press (1992). Adaptive Bayesian classification of spatial data. *Journal of the American Statistical Association 87*, 844–851.

Kleinn, C. (1996). Multiphase sampling for forest monitoring. In P. J. Kennedy (Ed.), *Application of Remote Sensing in European Forest Monitoring: International Workshop Proceedings*, pp. 167–177. Institute of Surveying, Remote Sensing, and Land Information, University of Agricultural Sciences (BOKU).

Kleman, J. (1987). Directional reflectance factor distributions for two forest canopies. *Remote Sensing of Environment 23*, 83–96.

Kriebel, K. T. (1976). On the variability of the reflected radiation field due to differing distributions of the irradiation. *Remote Sensing of Environment 4*, 257–264.

Kriebel, K. T. (1978). Measured spectral bidirectional reflectance properties of four vegetated surfaces. *Applied Optics 17*, 253–259.

Kriebel, K. T. (1996). On the limited validity of reciprocity in measured BRDFs. *Remote Sensing of Environment 58*, 52–62.

Labovitz, M. L. (1986). Issues arising from sampling designs and band selection in discriminating ground reference attributes using remotely sensed data. *Photogrammetric Engineering and Remote Sensing 52*, 201–211.

Landgrebe, D. (1976). Computer-based remote sensing technology. *Remote Sensing of Environment 5*, 229–246.

Latham, P. A. (1997). Forest structural class development for the upper columbia

river basin. General technical report, USDA, Forest Service, Pacific Northwest Research Station.

Lauer, D. T., J. E. Estes, J. R. Jensen, and D. D. Greenlee (1991). Institutional issues affecting the integration and use of remotely sensed data and geographic information systems. *Photogrammetric Engineering and Remote Sensing 57*, 647–654.

Lee, J.-H. and W. D. Philpot (1991). Spectral texture pattern matching: A classifier for digital imagery. *IEEE Transactions on Geoscience and Remote Sensing 29*, 545–554.

Leprieur, C. E., J. M. Durrand, and J. L. Peyron (1988). Influence of topography on forest reflectance using Landsat Thematic Mapper and digital terrain data. *Photogrammetric Engineering and Remote Sensing 54*, 491–496.

Leversee, D. (1993). Cooperative forest management: The Sierra Club of western Canada and British Columbia Ministry of Forests use multiple GIS data sources to assess logging effects on Vancouver Island. *Earth Observation Magazine 2*(7), 42–46.

Li, W.-H. (1997). *Significance of Multiple Scattering in Remotely Sensed Images of Natural Surfaces*. Ph. D. thesis, University of Washington, Department of Geology.

Li, W.-H., R. Weeks, and A. R. Gillespie (1996). Multiple scattering in the remote sensing of natural surfaces. In *Eleventh Thematic Conference and Workshops on Applied Geologic Remote Sensing*, pp. I–205–I–215. Environmental Research Institute of Michigan.

Li, X. and A. H. Strahler (1986). Geometric-optical bidirectional reflectance modeling of a conifer forest canopy. *IEEE Transactions on Geoscience and Remote Sensing 24*, 906–919.

Li, X. and A. H. Strahler (1992). Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: Effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing 30*, 276–292.

Li, X. and A. H. Strahler (1996). A knowledge-based inversion of physical BRDF model and three examples. In *Remote Sensing for a Sustainable Future / 1996 International Geoscience and Remote Sensing Symposium*, pp. 2173–2176. IEEE.

Li, X., A. H. Strahler, and C. E. Woodcock (1995). A hybrid geometric optical-radiative transfer approach for modeling albedo and directional reflectance of discontinuous canopies. *IEEE Transactions on Geoscience and Remote Sensing 33*, 466–480.

Liang, S. and A. H. Strahler (1994). Retrieval of surface BRDF from multiangle remotely sensed data. *Remote Sensing of Environment 50*, 18–30.

Lillesand, T. M. (1993). The "grayware" required to deal with global change issues. *Photogrammetric Engineering and Remote Sensing 59*, 961–968.

Lillesand, T. M. and R. W. Kieffer (1994). *Remote Sensing and Image Interpretation*. Wiley.

Lowell, K. E., G. Edwards, and C. M. Gold (1992). Considerations for adapting conventional forest management methodologies to a spatial framework using a GIS: Data, statistical techniques, and errors. In G. Wood and B. Turner (Eds.), *Integrating Forest Information Over Space and Time*, pp. 429–439. IUFRO.

Lowman, P. D. (1965). Space photography. *Photogrammetric Engineering 31*, 76–86.

Lund, H. G., R. Päivinen, and S. Thammincha (Eds.) (1992). *Remote Sensing and Permanent Plot Techniques for World Forest Monitoring*. IUFRO.

MacEachren, A. M. (1994). *Some Truth with Maps: A Primer of Symbolization and Design*. Association of American Geographers.

Mack, P. E. (1990). *Viewing the Earth: The Social Construction of the Landsat Satellite System*. MIT Press.

Malaret, E., L. A. Bartolucci, D. F. Lozano, P. E. Anuta, and C. D. McGillem (1985). Landsat-4 and Landsat-5 Thematic Mapper data quality analysis. *Photogrammetric Engineering and Remote Sensing 51*, 1407–1416.

Marceau, D. J., D. J. Gratton, R. A. Fournier, and J.-P. Fortin (1994). Remote sensing and the measurement of geographical entities in a forested environment. 2. The optimal spatial resolution. *Remote Sensing of Environment 49*, 105–117.

Marceau, D. J., P. J. Howarth, and D. J. Gratton (1994). Remote sensing and the measurement of geographical entities in a forested environment. 1. The scale and spatial aggregation problem. *Remote Sensing of Environment 49*, 93–104.

Markham, B. L. (1985). The Landsat sensors' spatial responses. *IEEE Transactions on Geoscience and Remote Sensing 23*, 864–875.

Markham, B. L. and J. L. Barker (1985). Spectral characterization of the Landsat Thematic Mapper sensors. *International Journal of Remote Sensing 6*, 697–716.

Marroquin, J., S. Mitter, and T. Poggio (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association 82*, 76–89.

Mather, P. M. (1989). Problems in passive remote sensing of the Earth's land surface. In S. R. Brooks (Ed.), *Mathematics in Remote Sensing*, Number 21 in The Institute of Mathematics and its Application Conference Series, pp. 17–35. Clarendon Press.

McCarter, J. B., J. S. Wilson, P. J. Baker, J. L. Moffett, and C. D. Oliver (1998).

Landscape management through integration of existing tools and emerging technologies. *Journal of Forestry 96*(6), 17–23.

McDonald, R. M. (1995). CORONA: Success for space reconnaissance, a look into the Cold War, and a revolution of intelligence. *Photogrammetric Engineering and Remote Sensing 61*, 689–720.

McGillem, C. D., P. E. Anuta, and E. Malaret (1985). Spatial resolution estimation of Landsat-4 Thematic Mapper data. In J. L. Barker (Ed.), *Landsat-4 Science Characterization Early Results*, Volume III, Part 2 of *NASA Conference Publication 2355*, pp. 527–535. NASA Goddard Space Flight Center.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machines. *Journal of Chemistry and Physics 21*, 1087–91.

Metzler, M. D. and W. A. Malila (1985). Characterization and comparison of Landsat-4 and Landsat-5 Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing 51*, 1315–1330.

Meyer, M. and L. Werth (1990). Satellite data: Management panacea or potential problem? *Journal of Forestry 88*, 10–13.

Mitchell, K. (1975). Dynamics and simulated yield of Douglas fir. *Forest Science 21*(4), 39 pp. Supplemental Monograph 17.

Moffett, J. L. and J. Besag (1996). Spatial and probabilistic classification of forest structures using Landsat TM data. In H. T. Mower, R. L. Czaplewski, and R. H. Hamre (Eds.), *Spatial Accuracy Assessment in Natural Resources and Environmental Sciences: Second International Symposium*, pp. 493–500. USDA, Forest Service, General Technical Report, RM-GTR-277.

Moffett, J. L., J. E. Besag, S. D. Byers, and W.-H. Li (1997). Probabilistic clas-

sification of forest structures by hierarchical modelling of the remote sensing process. In F. Preteux, J. L. Davidson, and E. R. Dougherty (Eds.), *Statistical and Stochastic Methods in Image Processing*, Volume 3167 of *Proceedings of SPIE*, pp. 118–129. SPIE.

Mohn, E., N. J. Hjort, and G. O. Storvik (1987). A simulation study of some contextual classification methods for remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing 25*, 796–804.

Morrison, P. (1992, 4). Ecosystem mapped. *Earth Observation Magazine 1*, 28–34.

National Academy of Sciences (1970). *Remote Sensing: With Specific Reference to Agriculture and Forestry*. Committee on Remote Sensing for Agricultural Purposes, Agricultural Board, National Research Council, National Academy of Sciences.

National Academy of Sciences (1976). *Resource and Environmental Surveys from Space with the Thematic Mapper in the 1980's*. Committee on Remote Sensing Programs for Earth Resource Surveys, Commission on Natural Resources. National Research Council, National Academy of Sciences.

Nicodemus, F. E. (1970). Reflectance nomenclature and directional reflectance and emissivity. *Applied Optics 9*, 1474–1475.

Nicodemus, F. E., J. C. Richmond, and J. J. Hsia (1977). Geometrical considerations and nomenclature for reflectance. NBS Monograph 160, USDC, National Bureau of Standards.

Norheim, R. A. (1996). Is there an answer to mapping old growth? An examination of two projects conducted with remote sensing and GIS. Master's thesis, University of Washington.

North, P. R., S. E. Plummer, D. W. Deering, and M. Leroy (1996). Validation of

a BRDF model for boreal forest. In *Remote Sensing for a Sustainable Future / 1996 International Geoscience and Remote Sensing Symposium*, pp. 1654–1656. IEEE.

North, P. R. J. (1996, July). Three dimensional forest light interaction model using a Monte Carlo method. *IEEE Transactions on Geoscience and Remote Sensing 34*(4), 946–956.

Odell, P. L. and A. H. Feiveson (1966). A numerical procedure to generate a sample covariance matrix. *Journal of the American Statistical Association 61*, 199–203.

O'Hara, K. L., P. A. Latham, P. Hessburg, and B. G. Smith (1996). A structural classification for inland northwest forest vegetation. *Western Journal of Applied Forestry 11*, 97–102.

Oliver, C. D. and B. C. Larson (1990). *Forest Stand Dynamics*. McGraw-Hill.

Oliver, C. D., A. Osawa, and A. Camp (1996). Forest dynamics and resulting animal and plant population changes at the stand and landscape levels. *Journal of Sustainable Forestry 6*, 281–311.

Owen, A. (1984). A neighborhood-based classifier for Landsat data. *Canadian Journal of Statistics 12*, 191–200.

Palmer, J. M. (1984). Effective bandwidths for Landsat-4 and Landsat-D' multispectral scanner and Thematic Mapper subsystems. *IEEE Transactions on Geoscience and Remote Sensing 22*, 336–338.

Peddle, D. and S. E. Franklin (1991). Image texture processing and data integration for surface pattern discrimination. *Photogrammetric Engineering and Remote Sensing 57*, 413–420.

Peytavin, L., F. Dansaert, and C. Rhin (1995). Multi-sources classification: Application to temporal refinement of forest cover using SPOT and ERS/SAR data.

In *Image and Signal Processing for Remote Sensing 2*, Volume 2579, pp. 42–51. SPIE Proceedings Series.

Philipson, W. R. (1986). Problem-solving with remote sensing: An update. *Photogrammetric Engineering and Remote Sensing 52*, 109–110.

Potts, R. B. (1952). Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society 52*, 106–109.

Prince, S. D. and C. O. Justice (1991). Editorial: Special issue on coarse resolution remote sensing of the Sahelian environment. *International Journal of Remote Sensing 12*, 1127–1146.

Ranson, K. J., J. R. Irons, and D. L. Williams (1994). Multispectral bidirectional reflectance of northern forest canopies with the Advanced Solid-State Array Spectroradiometer (ASAS). *Remote Sensing of Environment 47*, 276–289.

Richards, J. A. (1986). *Remote Sensing Digital Image Analysis: An Introduction*. Springer-Verlag.

Rosema, A., W. Verhoef, H. Noorbergen, and J. J. Borgesius (1992). A new forest light interaction model in support of forest monitoring. *Remote Sensing of Environment 42*, 23–41.

Rowe, J. S. (1971). Why classify forest land? *The Forestry Chronicle 47*(6), 144–148.

Rowe, J. S. and J. W. Sheard (1981). Ecological land classification: A survey approach. *Environmental Management 5*, 451–464.

Russell, C. A., J. R. Irons, and P. W. Dabney (1996). Estimating hemispherical reflectance and selected biophysical parameters for boreal forest canopies using spectral bidirectional reflectance data acquired by ASAS. In *Remote Sensing*

for a Sustainable Future / 1996 International Geoscience and Remote Sensing Symposium, pp. 1417–1419. IEEE.

Ryerson, R. (1989). Image interpretation concerns for the 1990's and lessons from the past. *Photogrammetric Engineering and Remote Sensing 55*, 1427–1430.

Sader, S. A., T. A. Stone, and A. T. Joyce (1990). Remote sensing of tropical forests: An overview of research and applications using non-photographic sensors. *Photogrammetric Engineering and Remote Sensing 56*, 1343–1351.

Sæbø, H. V., K. Bråten, N. L. Hjort, B. Llewellyn, and E. Mohn (1985). Contextual classification of remotely sensed data: Statistical methods and development of a system. Technical Report 768, Norwegian Computing Center.

Sample, V. A. (1994). Realizing the potential of remote sensing and GIS in ecosystem management planning, analysis, and policymaking. In V. A. Sample (Ed.), *Remote Sensing and GIS in Ecosystem Management*, pp. 346–352. Island Press.

Schaaf, C. B., X. Li, and A. H. Strahler (1994). Topographic effects on bidirectional and hemispherical reflectances calculated with a geometric-optical canopy model. *IEEE Transactions on Geoscience and Remote Sensing 32*(6), 1186–1193.

Schott, J. R. (1997). *Remote Sensing, the Image Chain Approach*. Oxford University Press.

Schowengerdt, R. A., C. Archwamety, and R. C. Wrigley (1985). Landsat Thematic Mapper image-derived MTF. *Photogrammetric Engineering and Remote Sensing 51*, 1395–1406.

Schröder, M., H. Rehrauer, K. Seidel, and M. Datcu (1998). Spatial information retrieval from remote-sensing images–Part II: Gibbs-markov random fields. *IEEE Transactions on Geoscience and Remote Sensing 36*, 1446–1455.

Schultz, R. R. and R. L. Stevenson (1995). Stochastic modeling and estimation of multispectral image data. *IEEE Transactions on Image Processing 4*, 1109–1119.

Schwaller, M. R. and B. T. Dealy (1986). Landsat and GIS: Two friends grow up together. *Journal of Forestry 84*(9), 40–41.

Settle, J. J. (1989). The use of spatial context in the classification of remotely sensed data. In S. R. Brooks (Ed.), *Mathematics in Remote Sensing*, Number 21 in The Institute of Mathematics and its Application Conference Series, pp. 381–389. Clarendon Press.

Sims, R. A. and P. Uhlig (1992). The current status of forest site classification in Ontario. *The Forestry Chronicle 68*(1), 64–77.

Slater, P. N. (1979). A re-examination of the Landsat MSS. *Photogrammetric Engineering and Remote Sensing 45*, 1479–1485.

Slater, P. N. (1980). *Remote Sensing, Optics and Optical Systems*. Addison-Wesley Publishing Company.

Smith, A. and G. Roberts (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B 55*, 3–23.

Smith, J. A., T. L. Lin, and K. J. Ranson (1980). The Lambertain assumption and Landsat data. *Photogrammetric Engineering and Remote Sensing 46*, 1183–1189.

Smith, M., D. Roberts, J. Hill, W. Mehl, B. Hosgood, J. Verdebout, G. Schmuck, C. Koechler, and J. Adams (1994). A new approach to quantifying abundances of materials in multispectral images. In *Surface and Atmospheric Remote Sensing: Technologies, Data Analysis, and Interpretation / 1994 International Geo-*

science and Remote Sensing Symposium. IEEE.

Smits, P. C. and S. G. Dellepiane (1997). Discontinuity adaptive MRF model for remote sensing image analysis. In *Remote Sensing – A Scientific Vision for Sustainable Development*, pp. 907–909. IEEE International Geoscience and Remote Sensing Symposium.

Smreker, M. S. and N. Khazenie (1990). Contextual classification based on a Markov random field and a spectral spatial-temporal correlation model. In *Remote Sensing Science for the Nineties / 1990 International Geoscience and Remote Sensing Symposium*, pp. 1629–1632. IEEE.

Solberg, A. H. S. and T. Taxt (1994). Classification of multisource satellite imagery based on a Markov random field model. In *Image and Signal Processing for Remote Sensing*, Volume 2315, pp. 322–331. SPIE Proceedings Series.

Story, M. and R. G. Congalton (1986). Accuracy assessment: A user's perspective. *Photogrammetric Engineering and Remote Sensing 52*, 397–399.

Strahler, A. H. and D. L. B. Jupp (1990). Modeling bidirectional reflectance of forests and woodlands using boolean models and geometric optics. *Remote Sensing of Environment 34*, 153–166.

Strain, P. and F. Engle (1992). *Looking at Earth*. Turner Publishing.

Swain, P. H., S. B. Vardeman, and J. C. Tilton (1981). Contextual classification of multispectral image data. *Pattern Recognition 13*, 429–441.

Switzer, P. (1980). Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. *Mathematical Geology 12*, 367–376.

Switzer, P. (1983). Some spatial statistics for the interpretation of satellite data. In *Bulletin of the International Statistical Institute: Proceedings of the 44th Session*, Volume L, Book 2, pp. 962–971.

Syrén, P. (1994). Reflectance anisotropy for nadir observations of coniferous forest canopies. *Remote Sensing of Environment 49*, 72–80.

Takemata, K., T. Yonekura, M. Asae, and Y. Kawata (1996). Estimation of bidirectional reflectance distribution function from land surfaces using airborne polder data. In *Remote Sensing for a Sustainable Future / 1996 International Geoscience and Remote Sensing Symposium*, pp. 1666–1668. IEEE.

Teillet, P. M. (1986). Image correction for radiometric effects in remote sensing. *International Journal of Remote Sensing 7*, 1637–1651.

Teply, J. and K. Green (1991). Old growth forest: How much remains? *Geo Info Systems 1*(4), 23–31.

Thapa, K. and J. Bossler (1992). Accuracy of spatial data used in geographic information systems. *Photogrammetric Engineering and Remote Sensing 58*, 835–841.

Therrien, C. W. (1983). An estimation-theoretic approach to terrain image segmentation. *Computer Vision, Graphics and Image Processing 22*, 313–326.

Tilton, J. C., S. B. Vardeman, and P. H. Swain (1982). Estimation of context for statistical classification of multispectral image data. *IEEE Transactions on Geoscience and Remote Sensing 20*, 445–452.

Tjelmeland, H. and J. Besag (1998). Markov random fields with higher-order interactions. *Scandinavian Journal of Statistics 25*, 415–433.

Tubbs, J. D. (1980). Effect of autocorrelated training samples on Bayes' probabilities of misclassification. *Pattern Recognition 12*, 351–354.

Tubbs, J. D. and W. A. Coberly (1978). Spatial correlation and its effect upon classification results in Landsat. In *Proceedings of the International Symposium*

on Remote Sensing of Environment, pp. 775–781. Environmental Research Institute of Michigan.

Veregin, H. (1989). Error modelling for the map overlay operation. In M. F. Goodchild and S. Gopal (Eds.), *Accuracy of Spatial Databases*. Taylor and Francis.

Veregin, H. (1996). Error propagation through the buffer operation for probability surfaces. *Photogrammetric Engineering and Remote Sensing 62*, 419–428.

Walsh, S. J. (1987). Variability of Landsat MSS spectral responses of forests in relation to stand and site characteristics. *International Journal of Remote Sensing 8*, 1289–1299.

Walsh, S. J. (1989). User considerations in landscape characterization. In M. Goodchild and S. Gopal (Eds.), *Accuracy of Spatial Databases*. Taylor and Francis.

Walter-Shea, E. A., M. A. Mesarch, and L. Chen (1996). Conifer shoot bidirectional scattering: Methodology and preliminary results. In *Remote Sensing for a Sustainable Future / 1996 International Geoscience and Remote Sensing Symposium*, pp. 1908–1910. IEEE.

Wang, L. and D. C. He (1990). A new statistical approach for texture analysis. *Photogrammetric Engineering and Remote Sensing 56*, 61–66.

Williams, D. L. (1991). A comparison of spectral reflectance properties at the needle, branch, and canopy level for selected conifer species. *Remote Sensing of Environment 35*, 79–93.

Wilson, J. D., P. J. Green, and I. S. Weir (1994). A Bayesian analysis of remotely sensed data using a hierarchical model. Technical report, Statistics group, Department of Mathematics, University of Bristol.

Woodcock, C. E., J. B. Collins, S. Gopal, V. D. Jakabhazy, X. Li, S. Macomber, S. Ryherd, V. J. Harward, J. Levitan, Y. Wu, and R. Warbington (1994).

Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model. *Remote Sensing of Environment 50*, 240–254.

Woodcock, C. E., J. B. Collins, V. D. Jakabhazy, X. Li, S. A. Macomber, and Y. Wu (1997). Inversion of the Li-Strahler canopy reflectance model for mapping forest structure. *IEEE Transactions on Geoscience and Remote Sensing 35*, 405–414.

Wrigley, R. C., D. H. Card, C. A. Hlavka, J. R. Hall, F. C. Mertz, C. Archwamety, and R. A. Schowengerdt (1984). Thematic Mapper image quality: Registration, noise, and resolution. *IEEE Transactions on Geoscience and Remote Sensing 22*, 263–271.

Wrigley, R. C., C. A. Hlavka, and D. H. Card (1985). Evaluation of Thematic Mapper interband registration and noise characteristics. *Photogrammetric Engineering and Remote Sensing 51*, 1417–1425.

Wu, H.-H. P. and R. A. Schowengerdt (1993). Improved estimation of fraction images using partial image restoration. *IEEE Transactions on Geoscience and Remote Sensing 31*, 771–778.

Wu, Y. and A. H. Strahler (1994). Remote estimation of crown size, stand density, and biomass on the Oregon transect. *Ecological Applications 4*, 299–312.

Yamazaki, T. and D. Gingras (1995). Image classification using spectral and spatial information based on MRF models. *IEEE Transactions on Image Processing 4*, 1333–1339.

Yu, T.-S. and K.-S. Fu (1983). Recursive contextual classification using a spatial stochastic technique. *Pattern Recognition 16*, 89–108.

Zhang, M., R. Haralick, and J. Campbell (1990). Multispectral image context classification using stochastic relaxation. *IEEE Transactions on Systems, Man, and Cybernetics 20*, 128–140.

Appendix A

ORIGINAL TRAINING DATA HISTOGRAMS

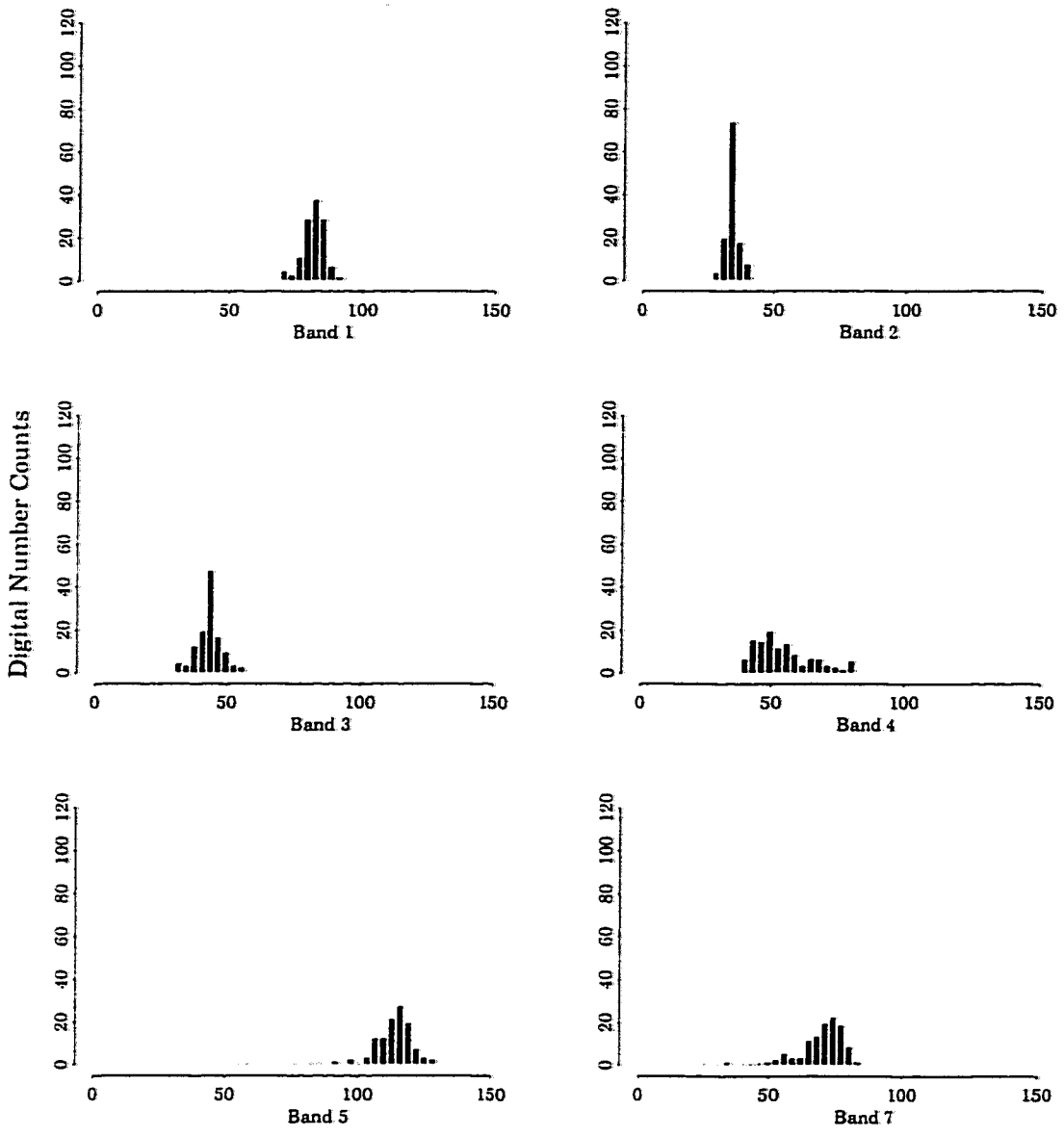


Figure A.1: CC histograms for July.

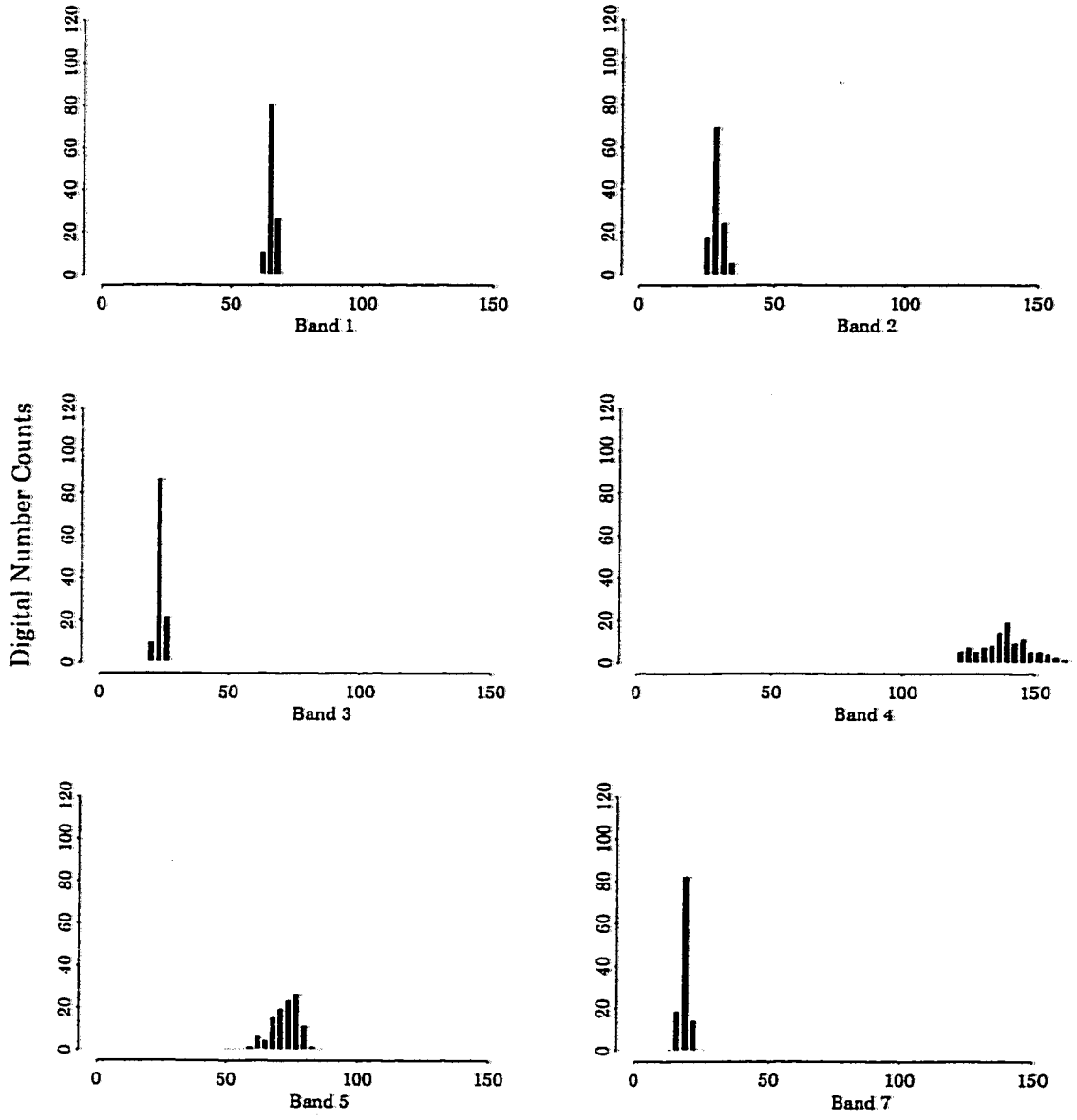


Figure A.2: HW histograms for July.

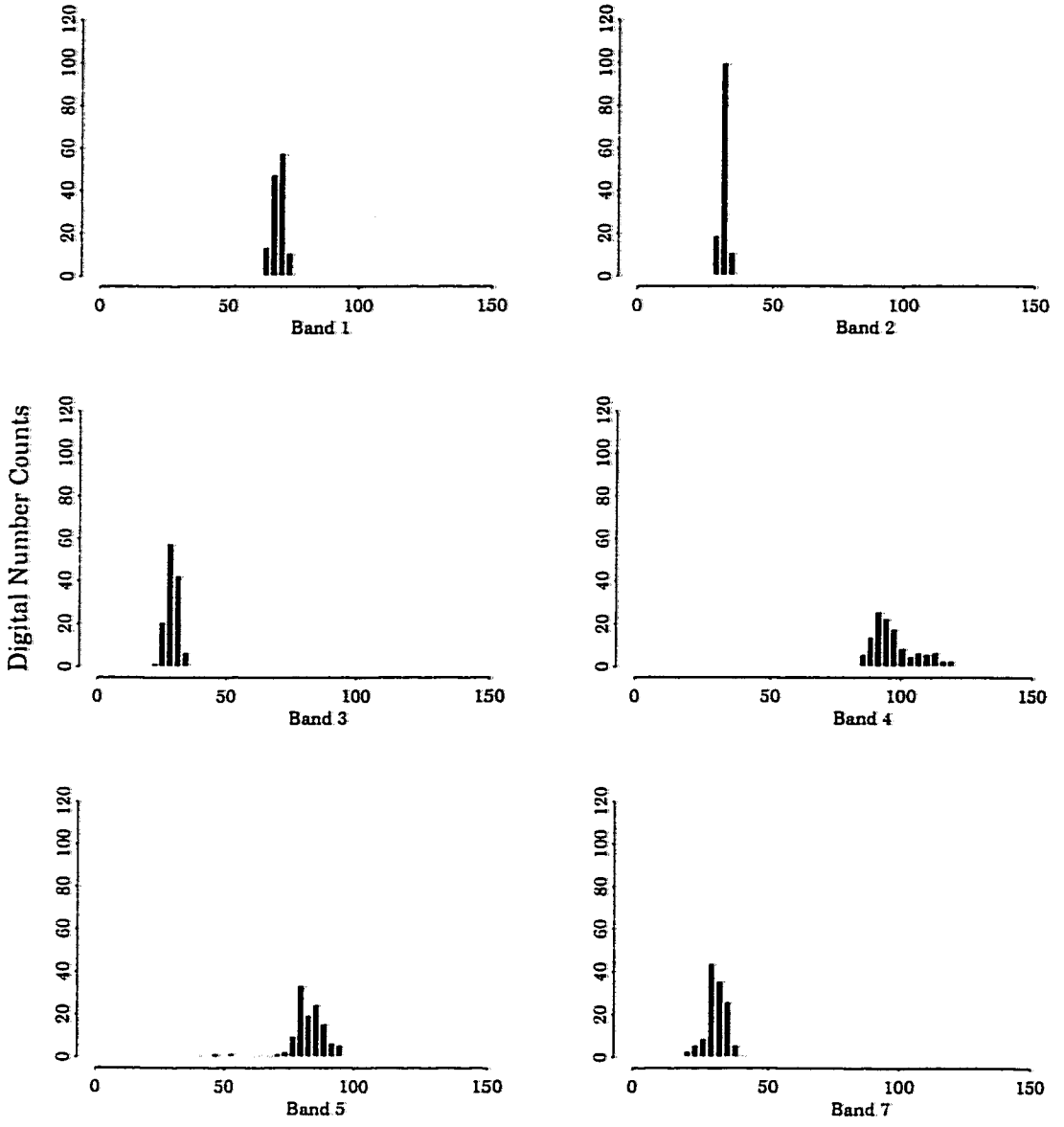


Figure A.3: AI histograms for July.

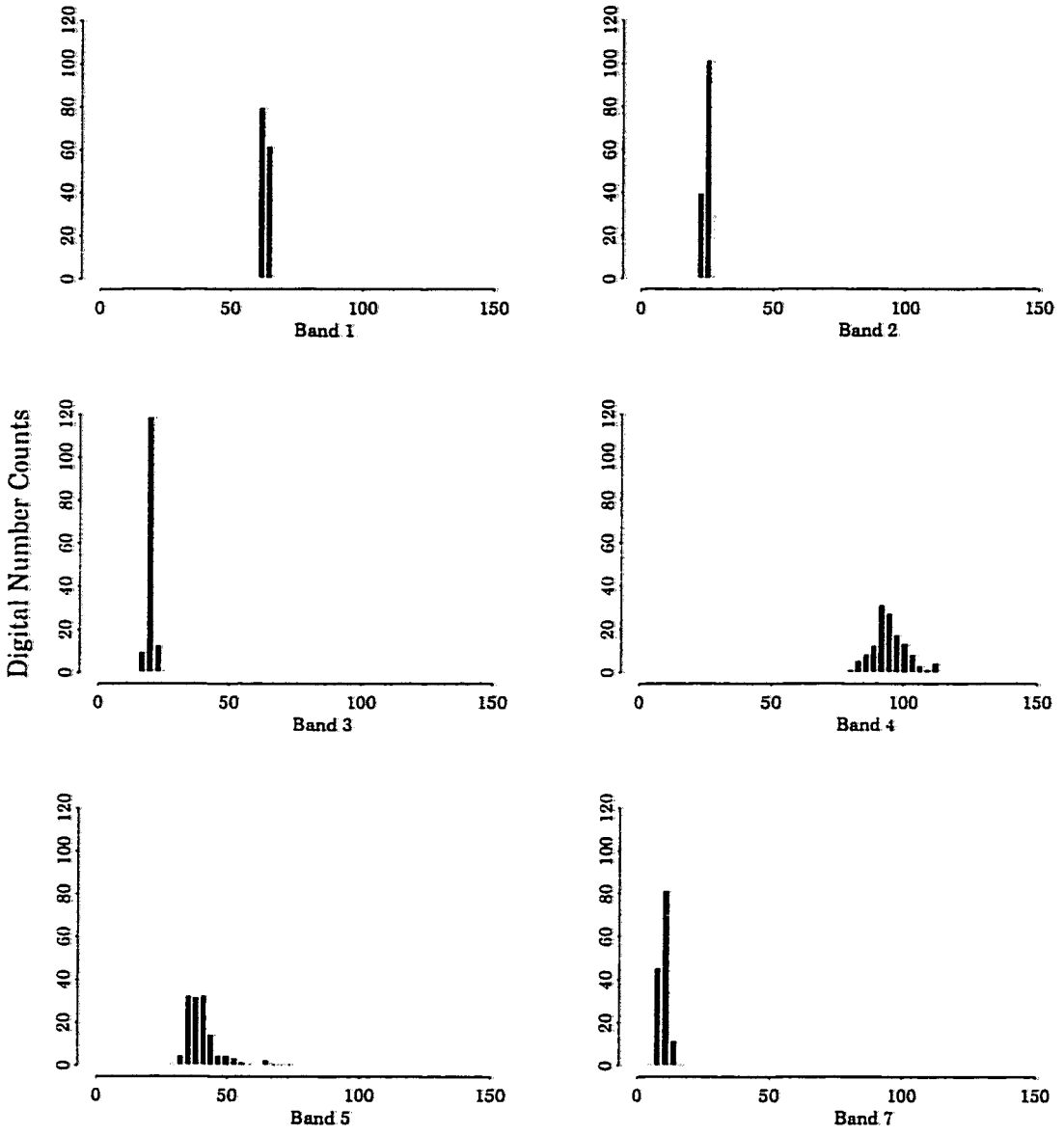


Figure A.4: EE histograms for July.

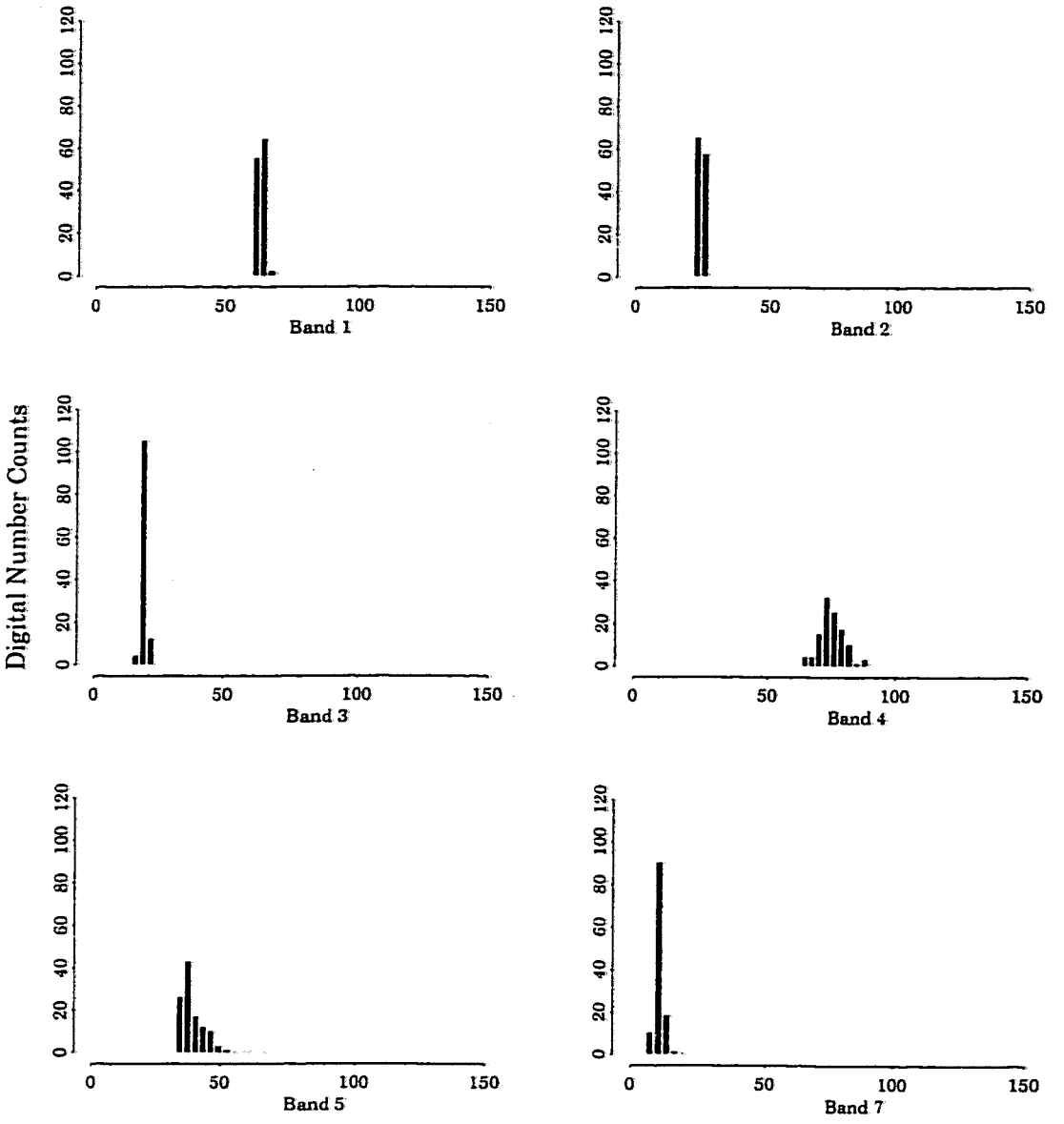


Figure A.5: LE histograms for July.

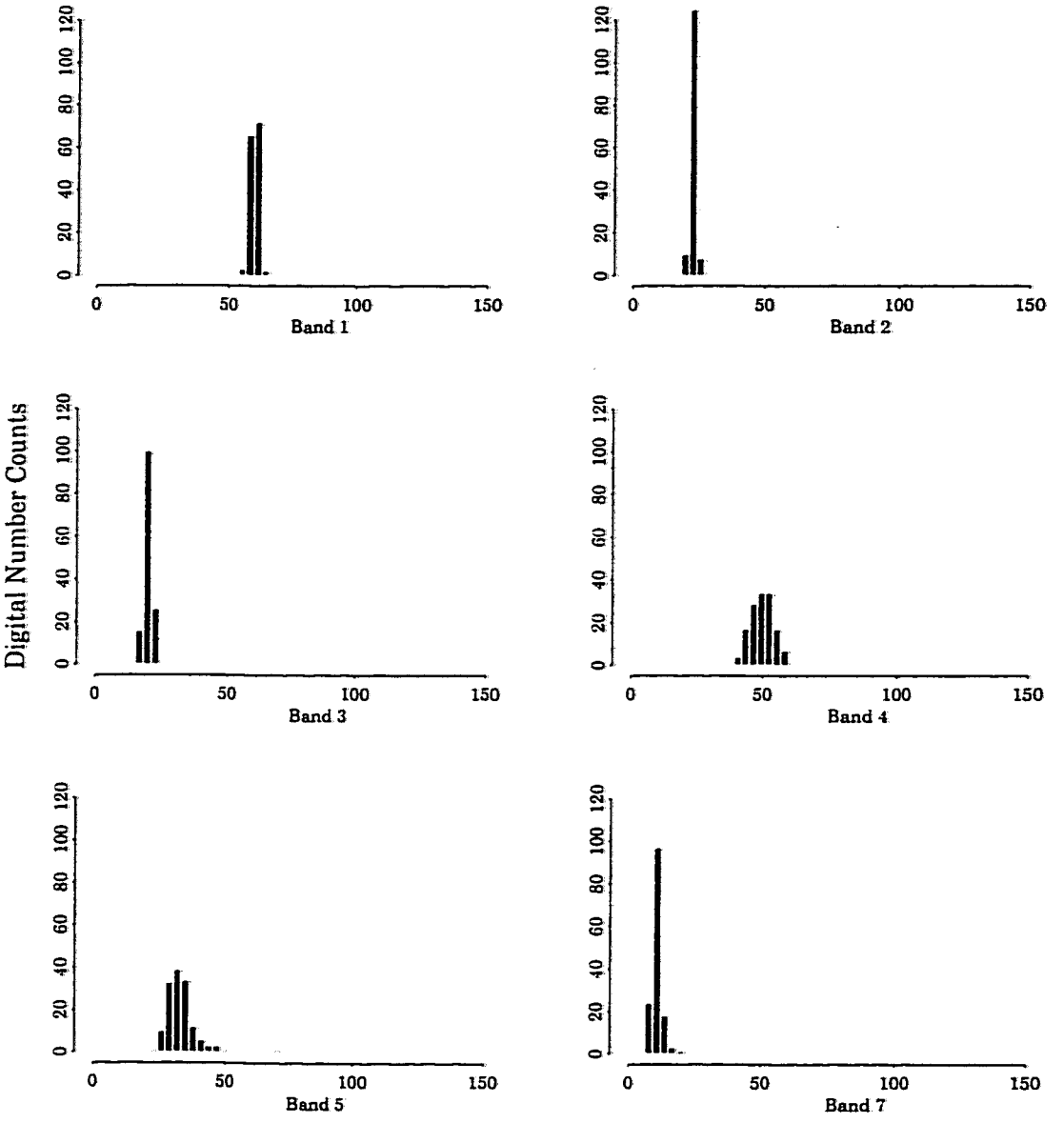


Figure A.6: UR histograms for July.

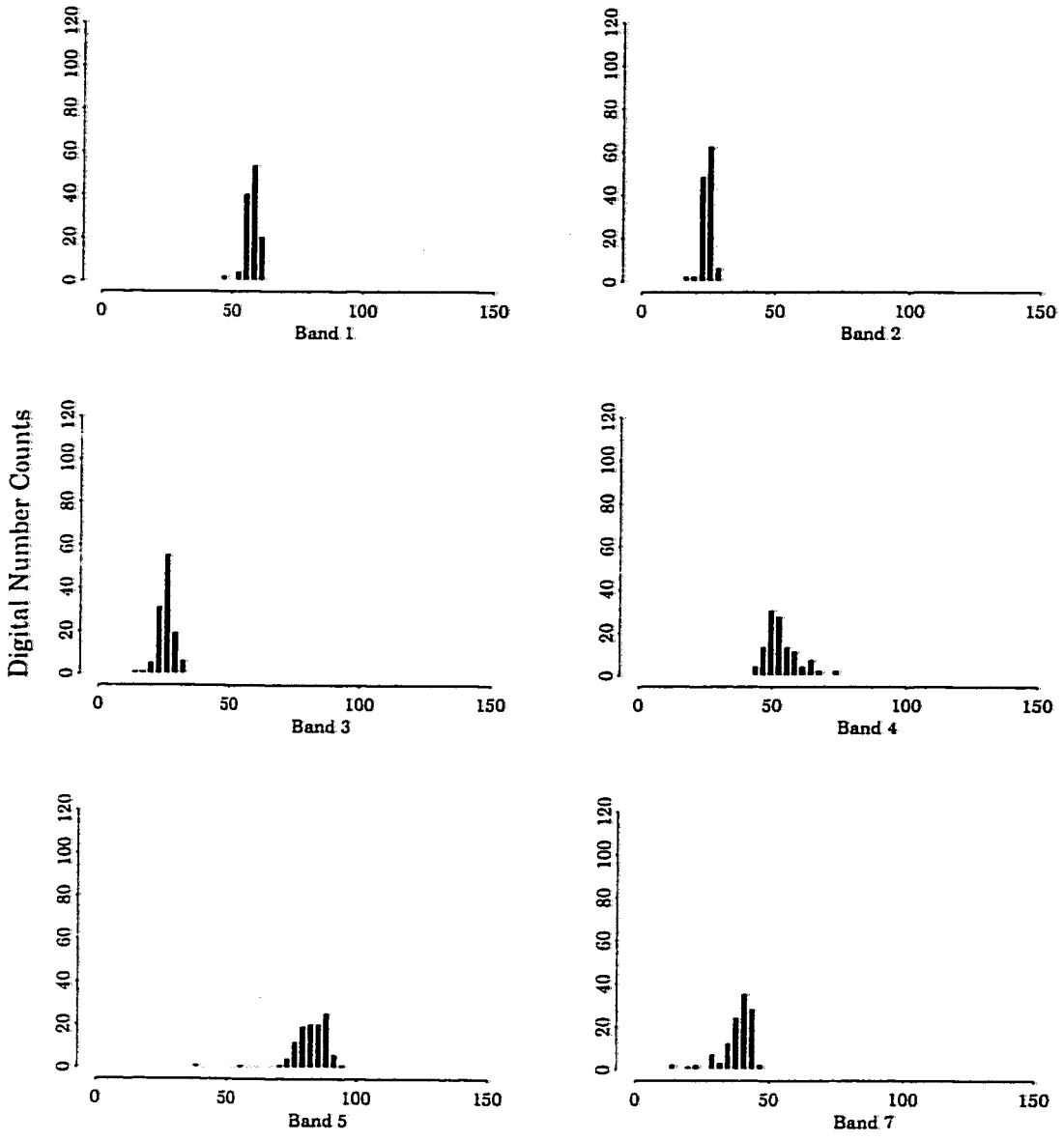


Figure A.7: CC histograms for August.

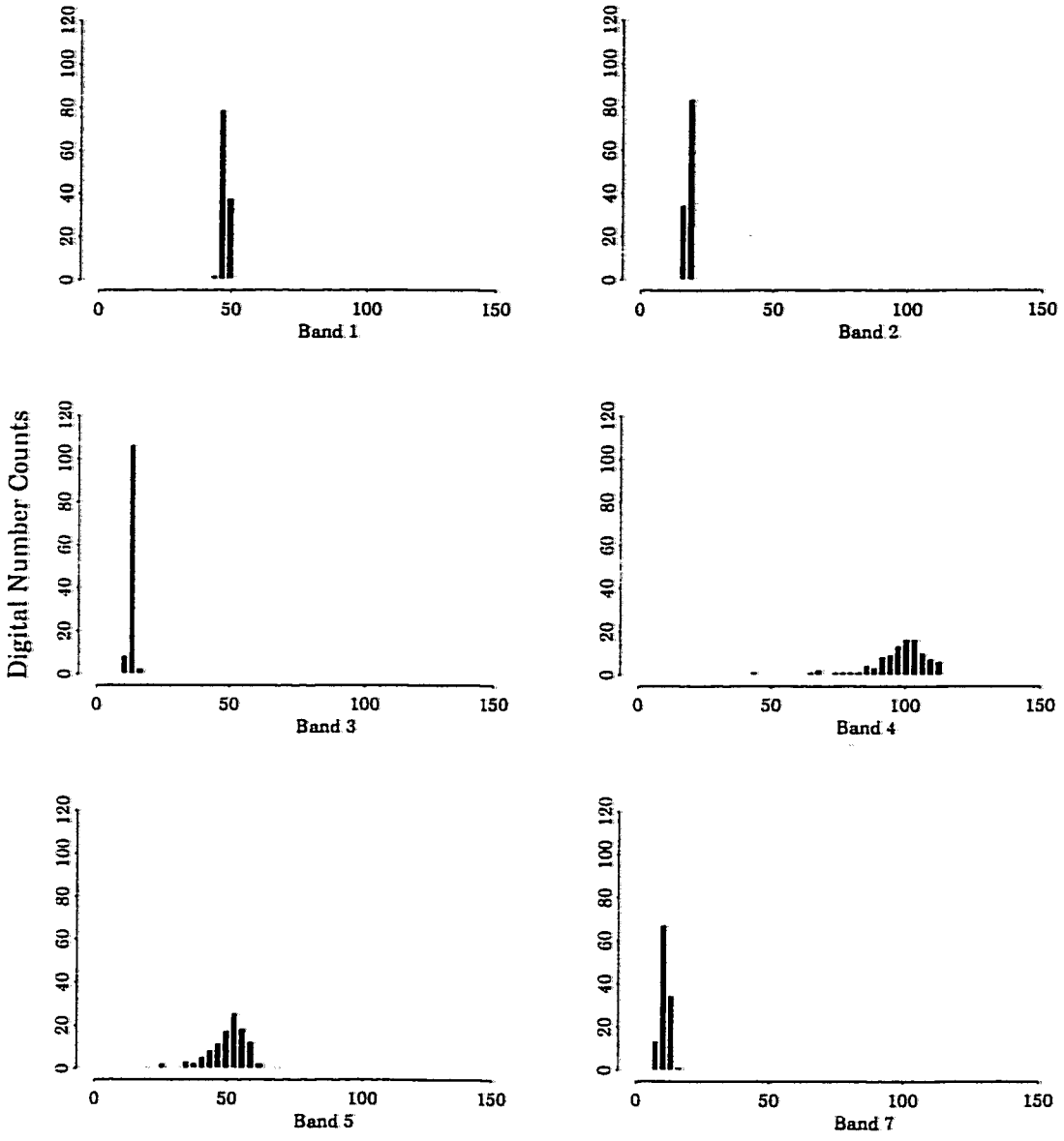


Figure A.8: HW histograms for August.

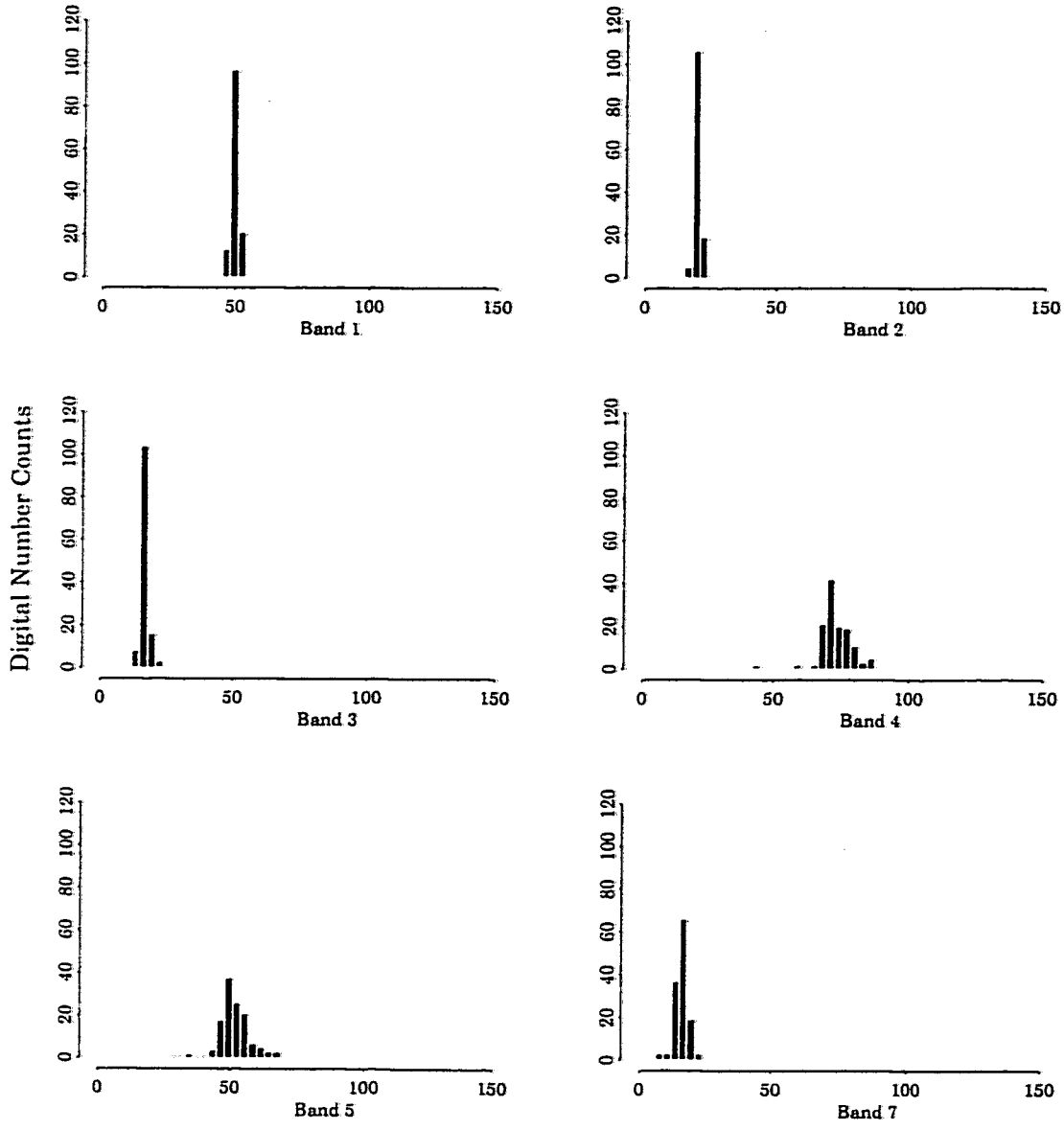


Figure A.9: AI histograms for August.

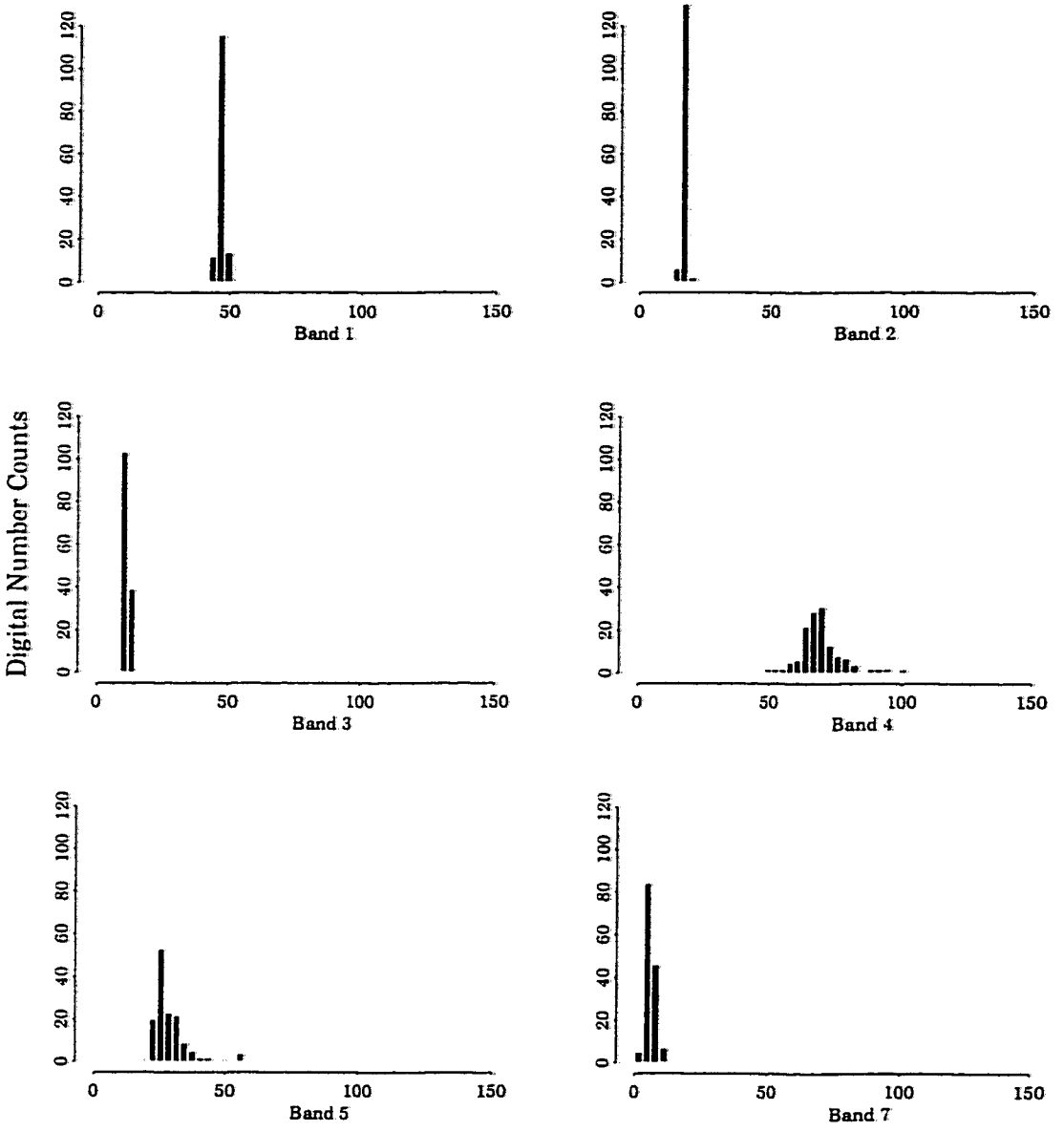


Figure A.10: FE histograms for August.

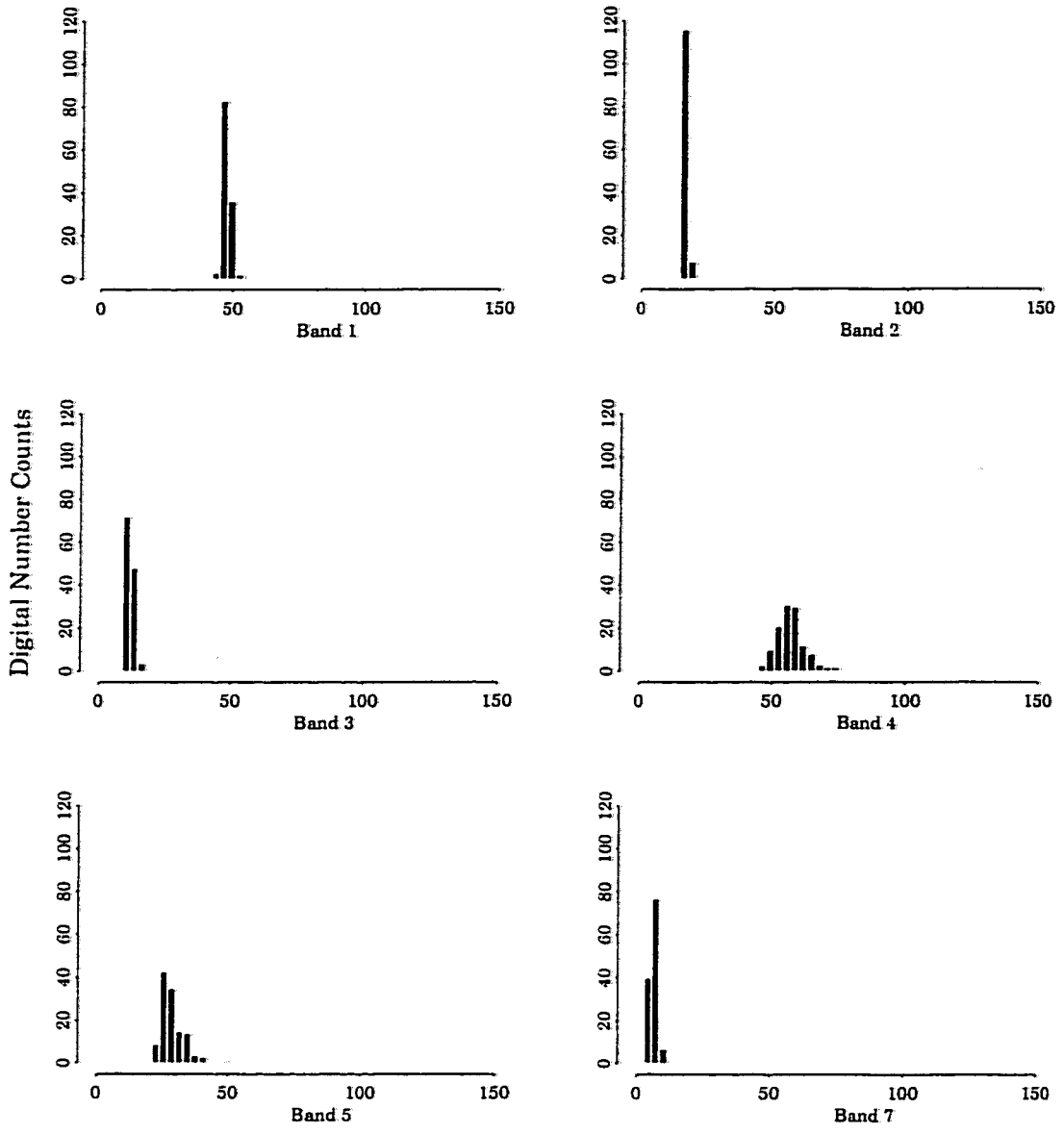


Figure A.11: LE histograms for August.

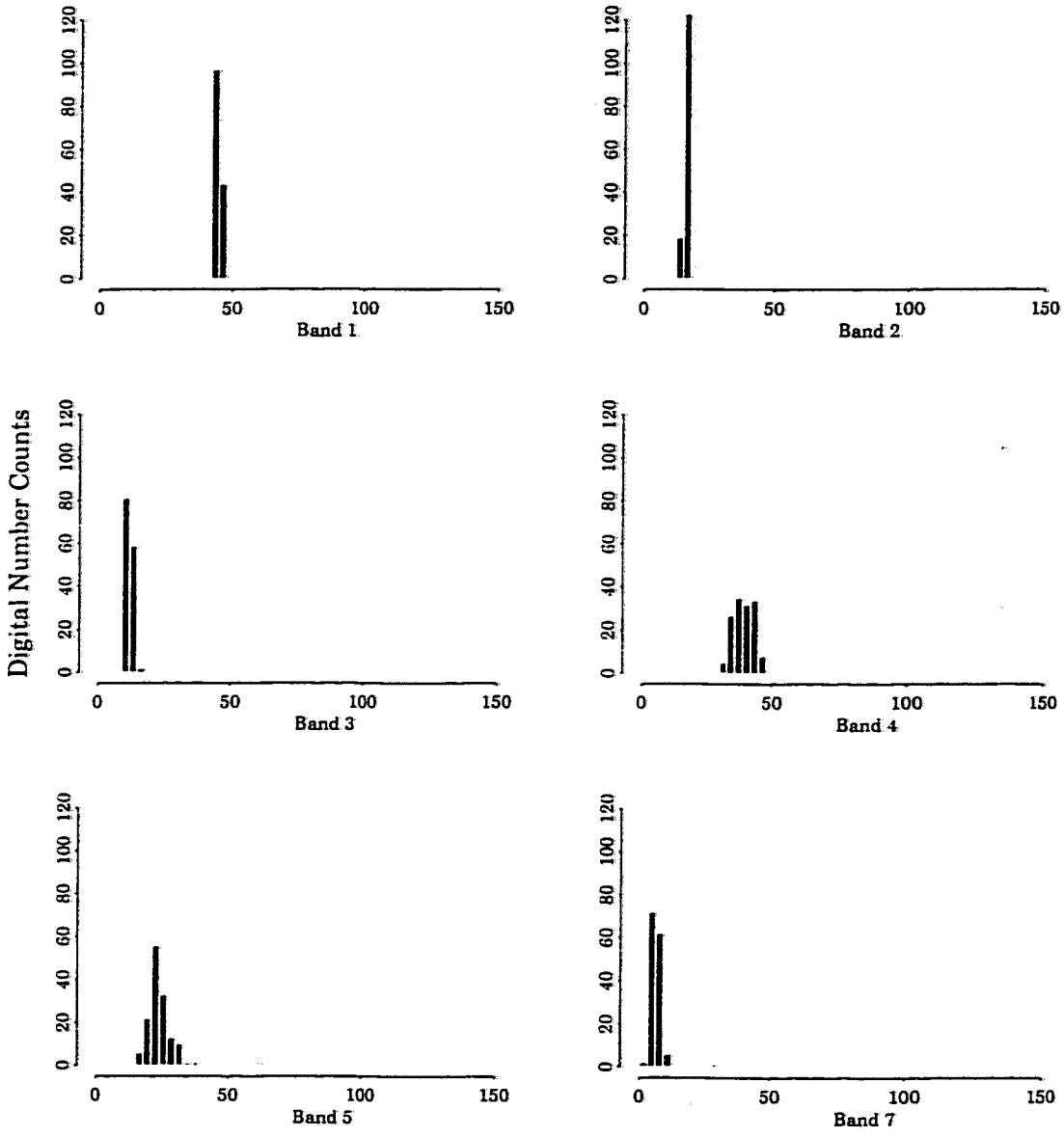


Figure A.12: UR histograms for August.

Appendix B

SUBSAMPLED NORMALIZED TRAINING DATA HISTOGRAMS

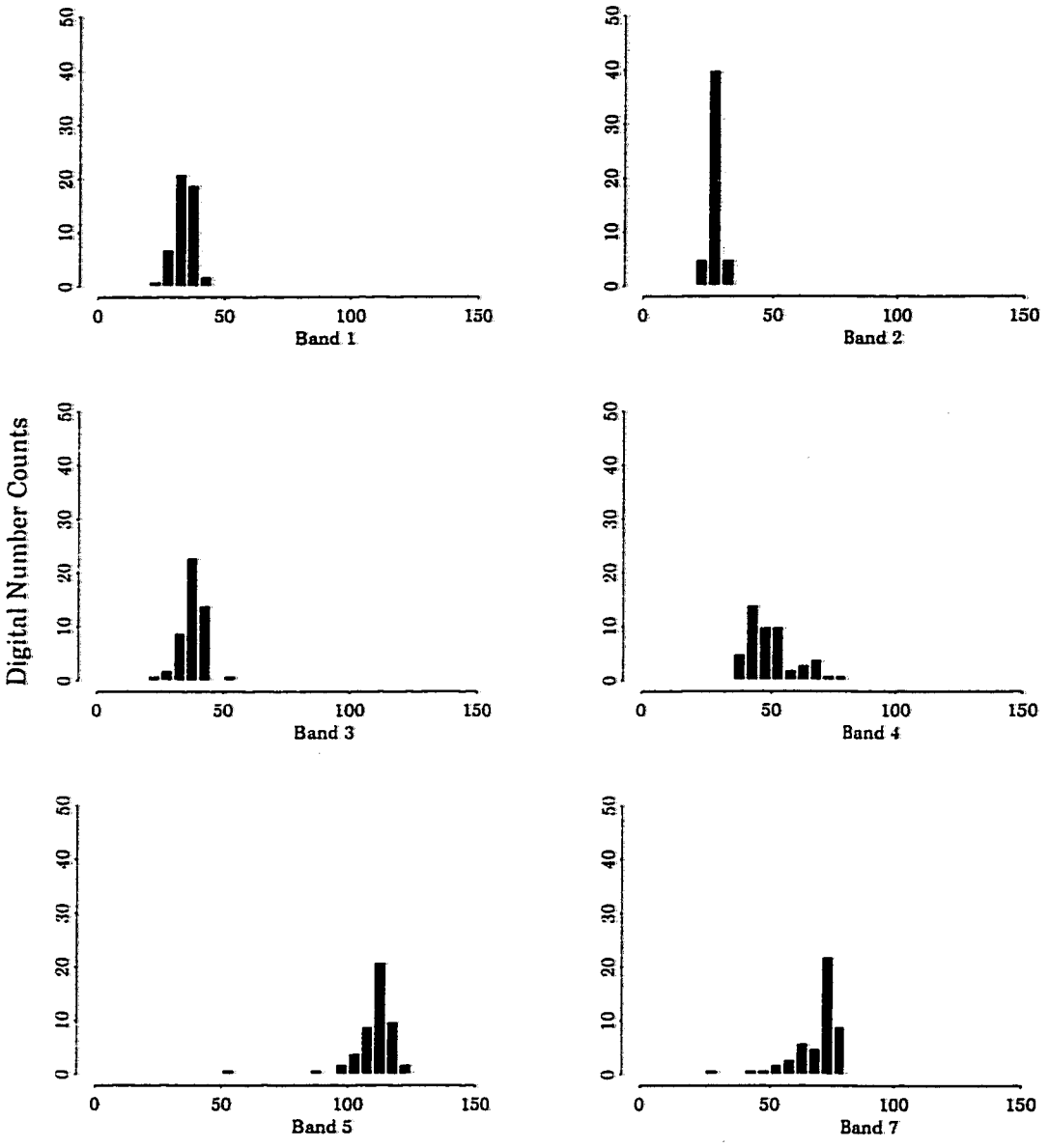


Figure B.1: CC histograms for July image.

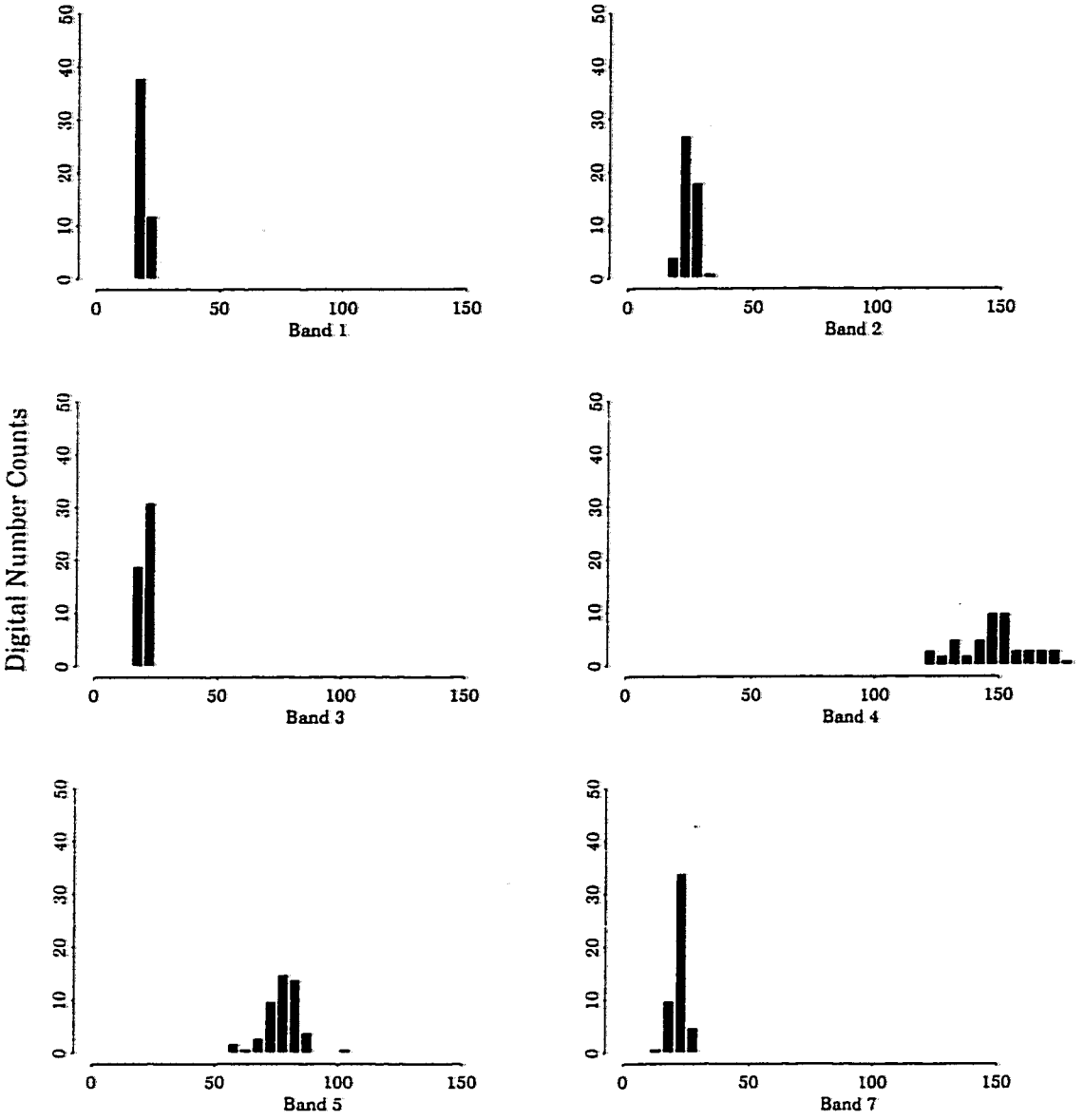


Figure B.2: HW histograms for July image.

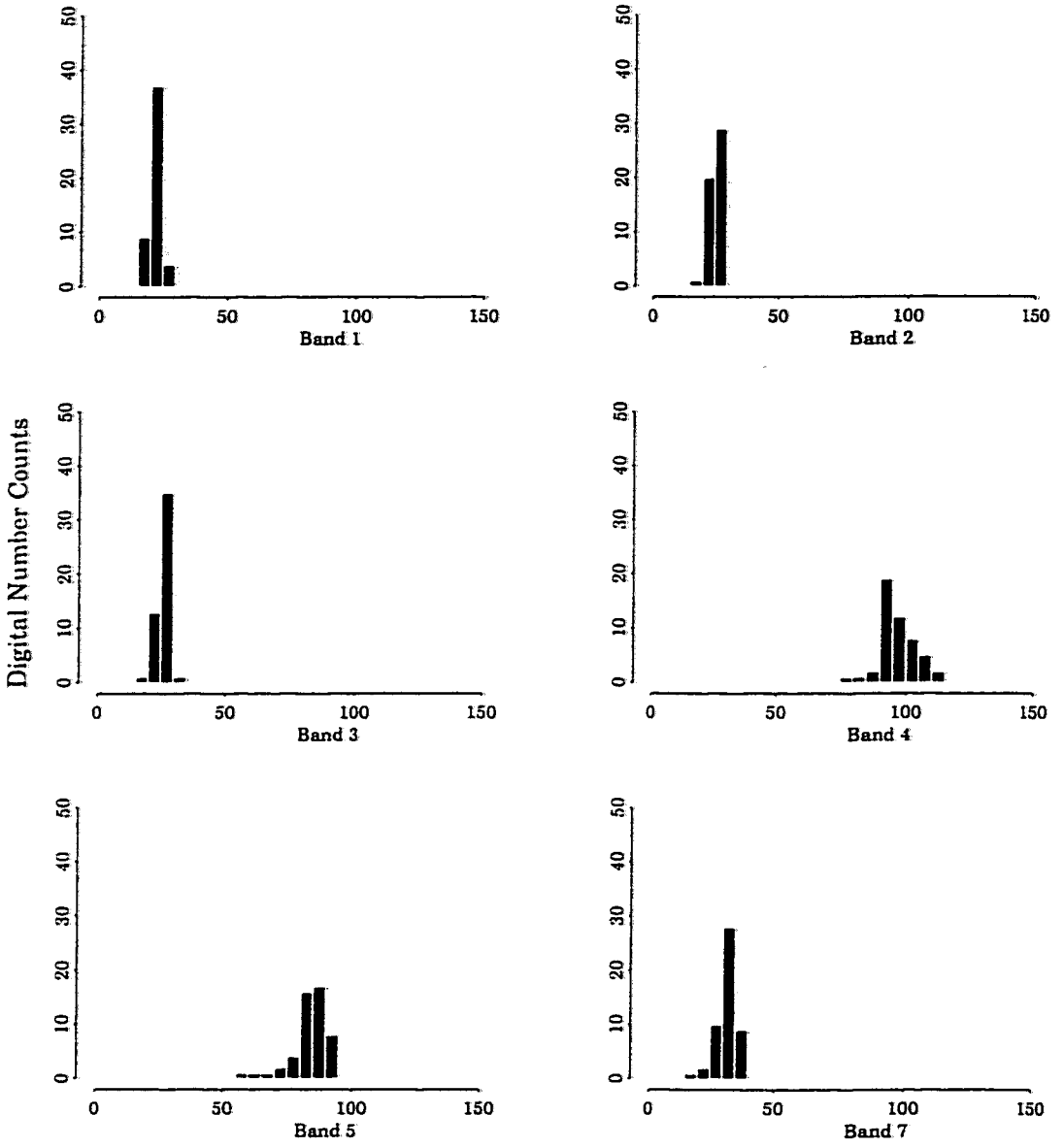


Figure B.3: AI histograms for July image.

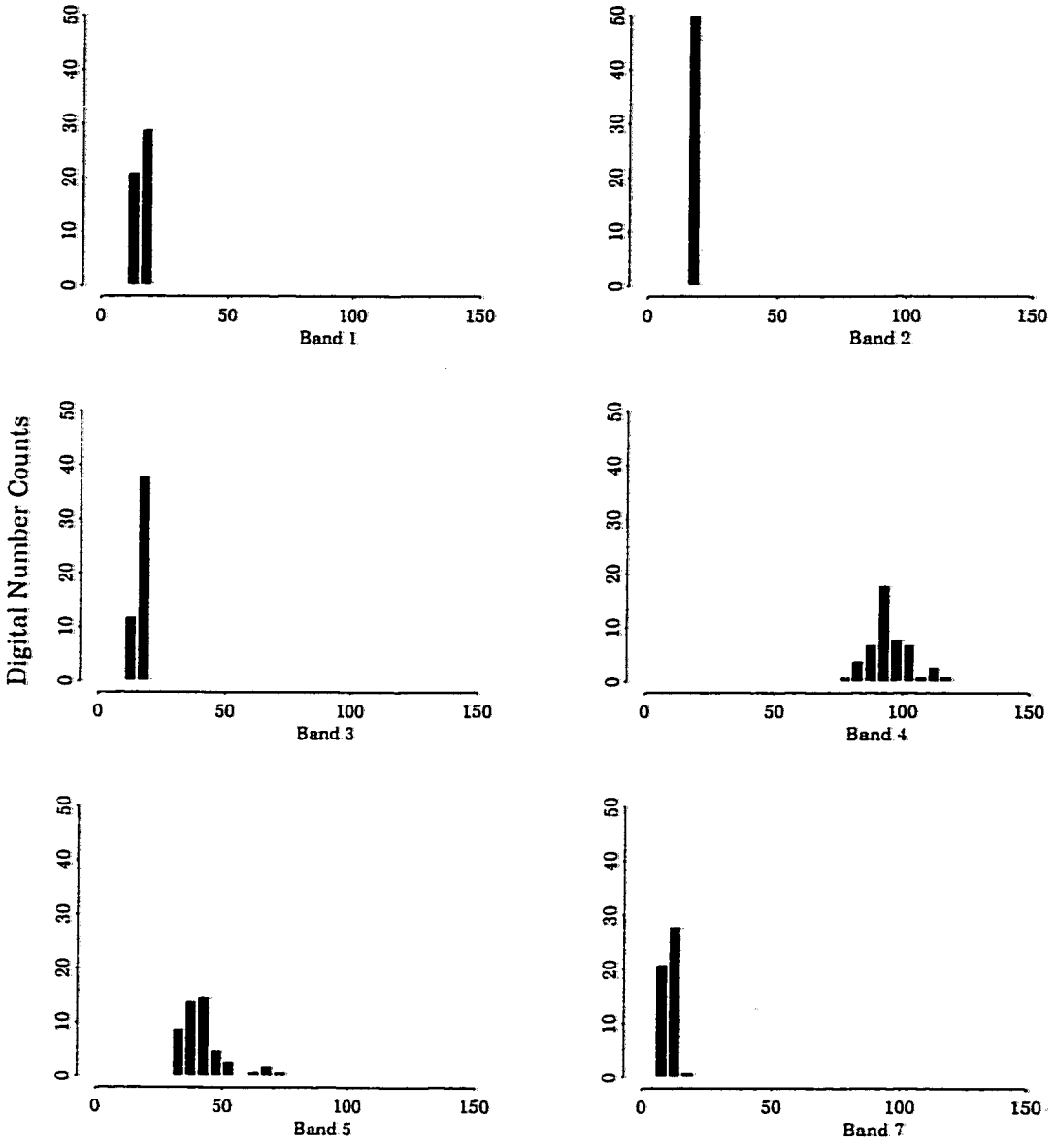


Figure B.4: EE histograms for July image.

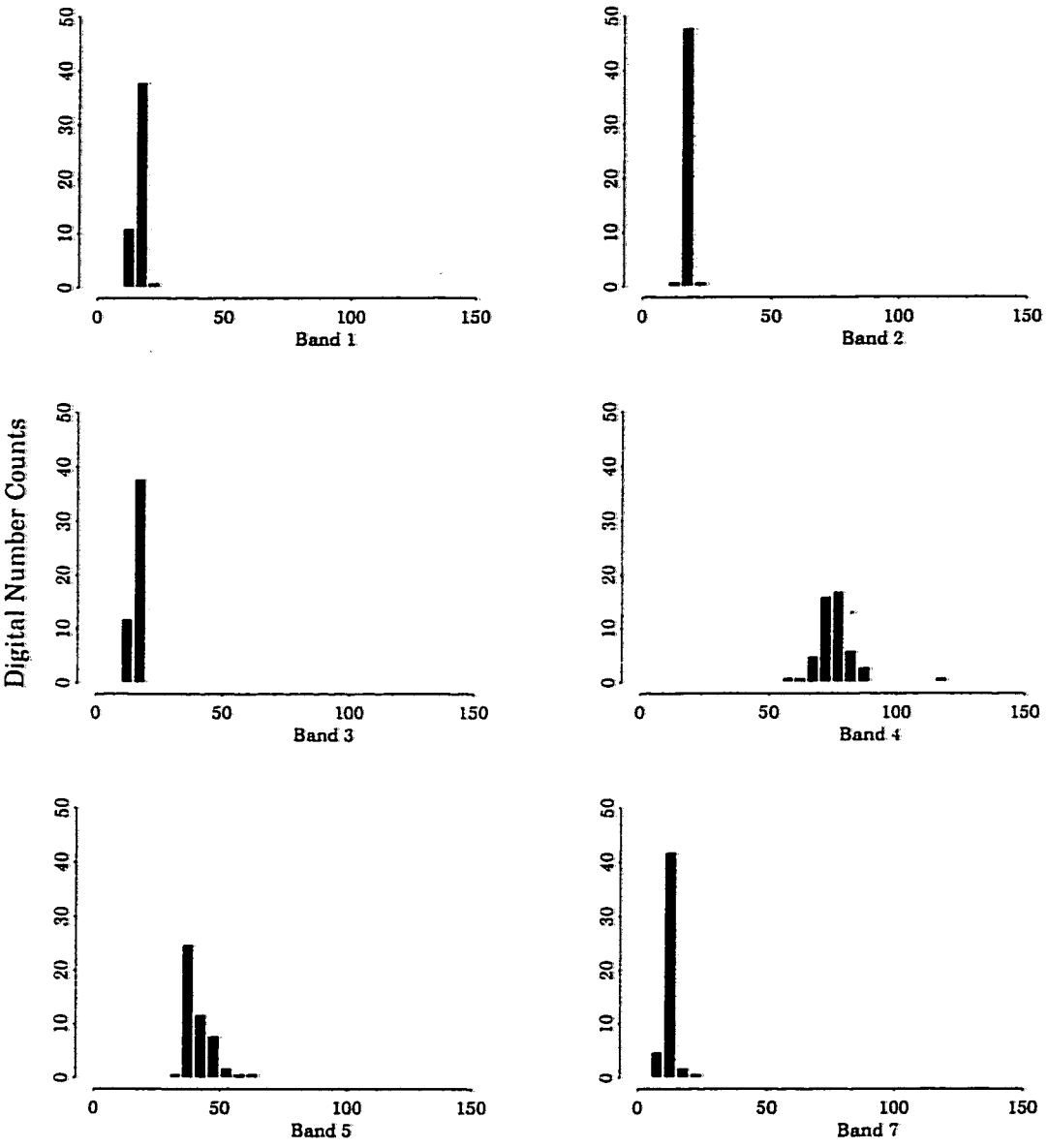


Figure B.5: LE histograms for July image.

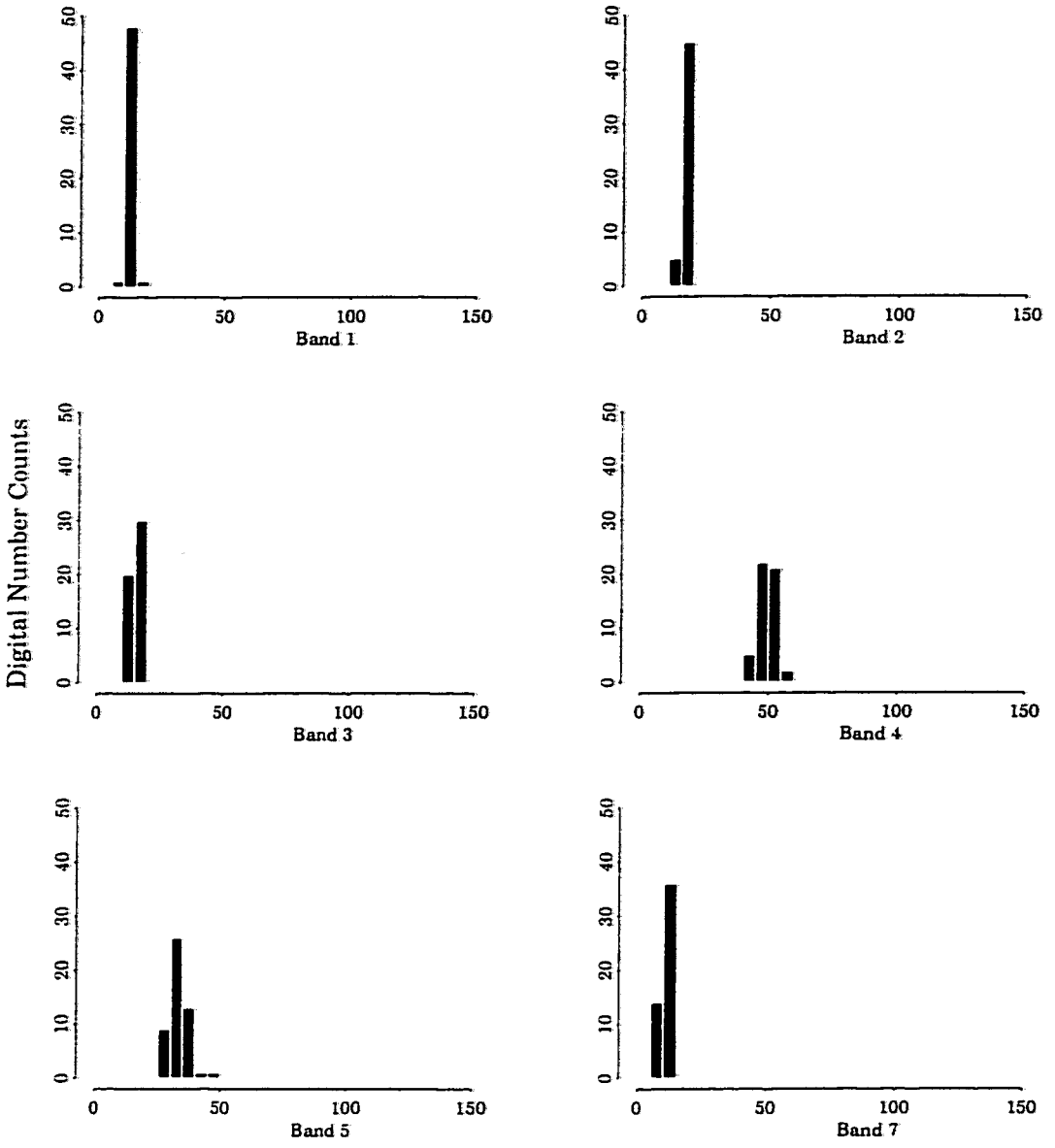


Figure B.6: UR histograms for July image.

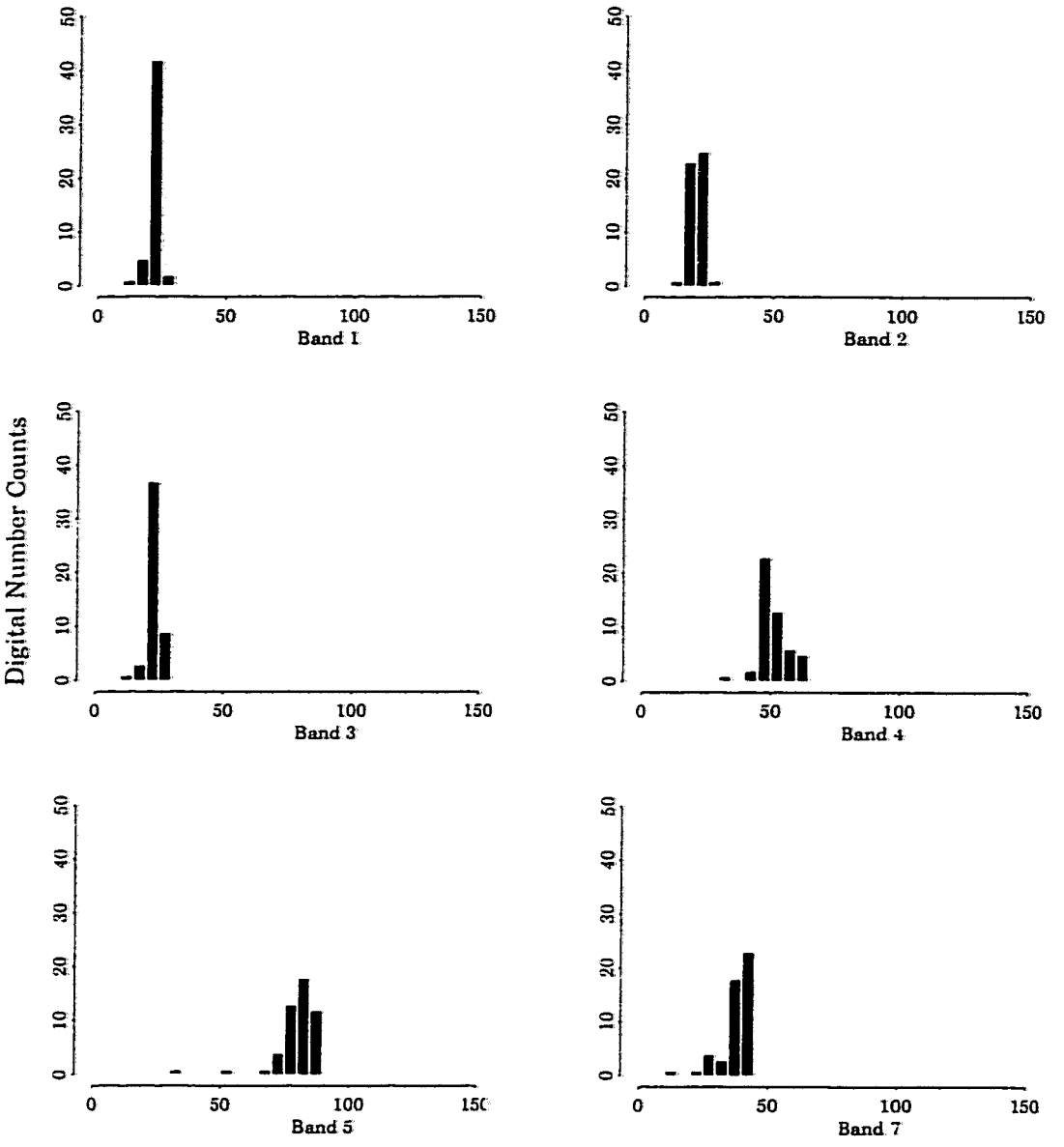


Figure B.7: CC histograms for August image.

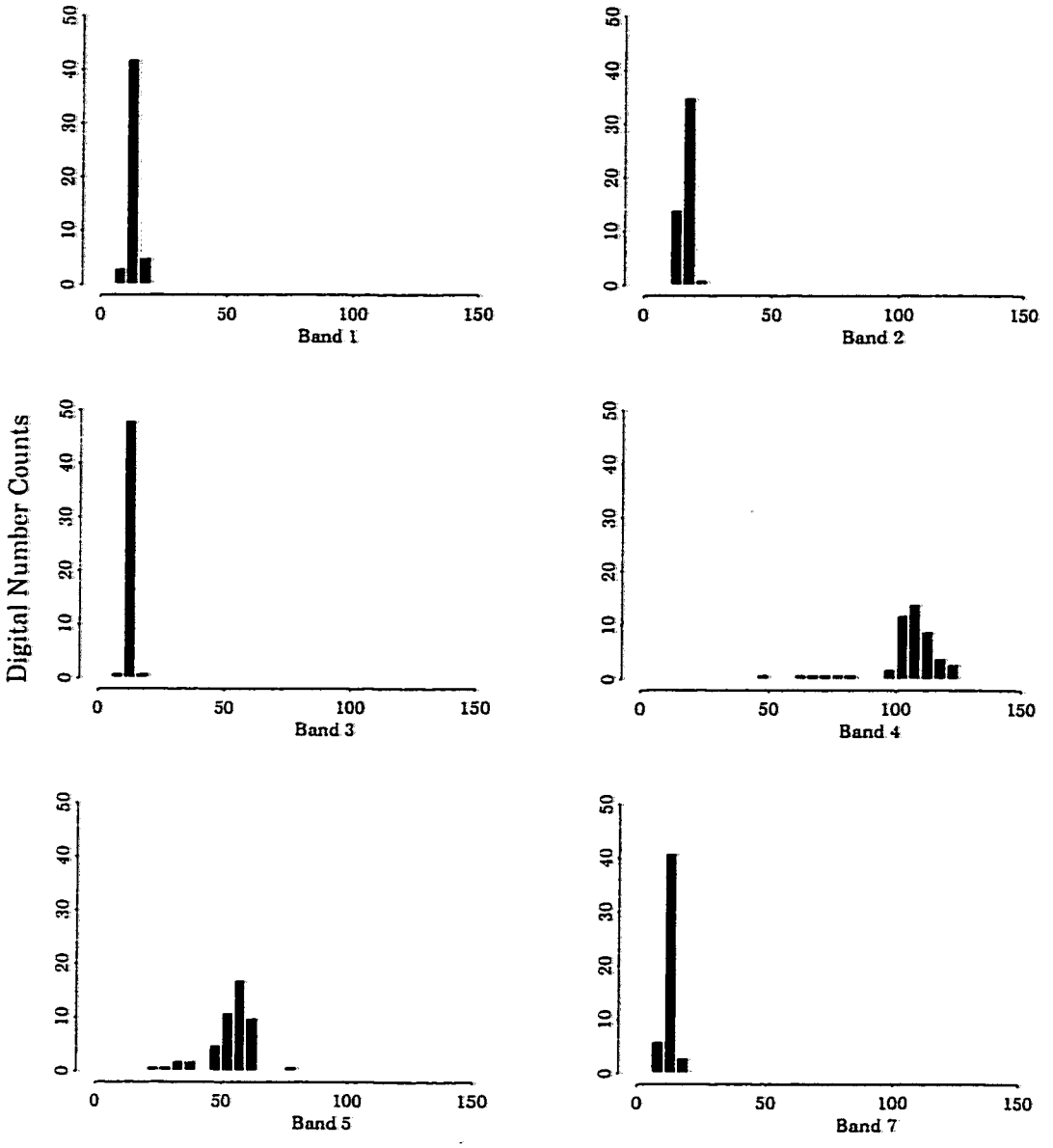


Figure B.8: HW histograms for August image.

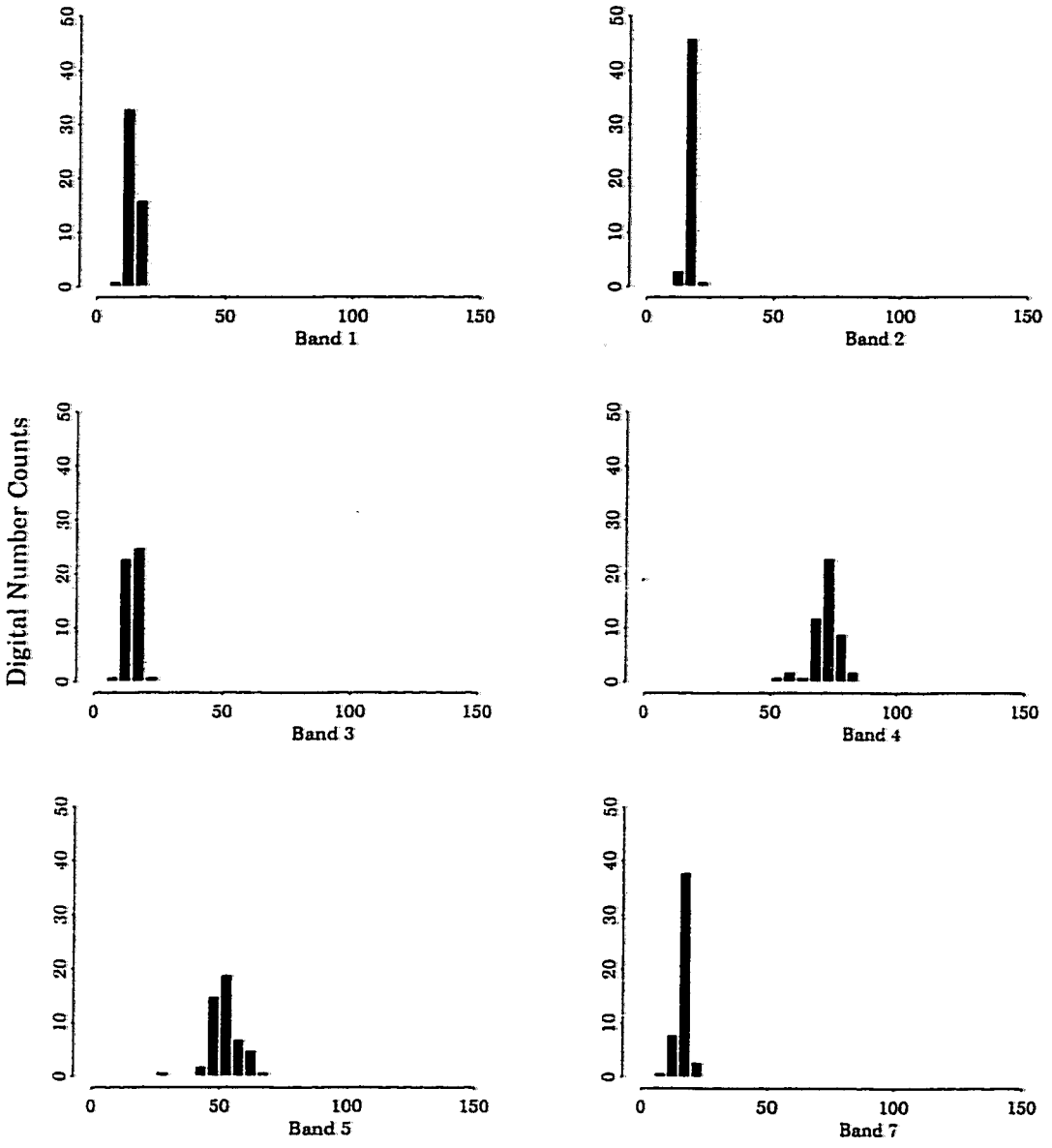


Figure B.9: AI histograms for August image.

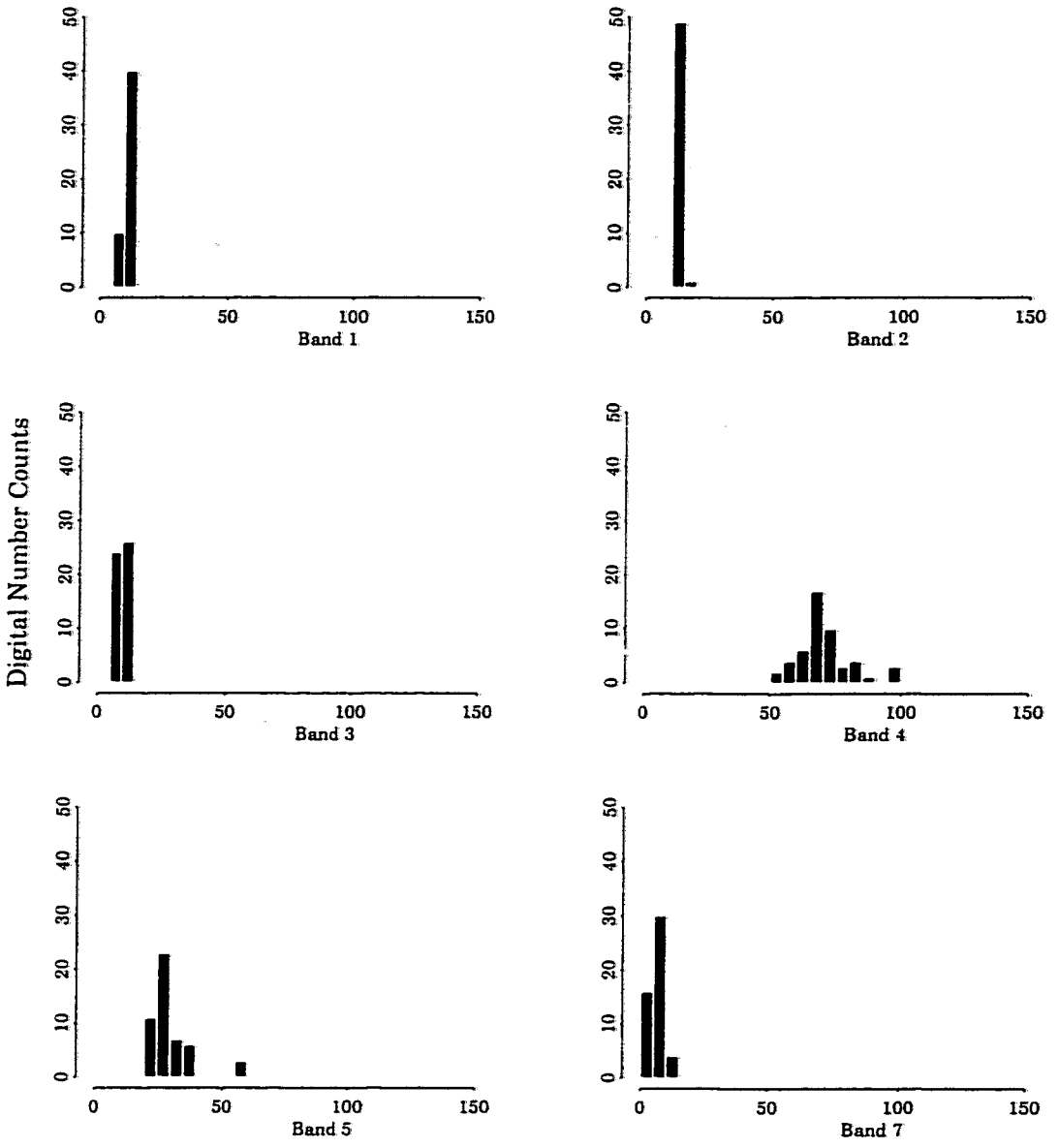


Figure B.10: EE histograms for August image.

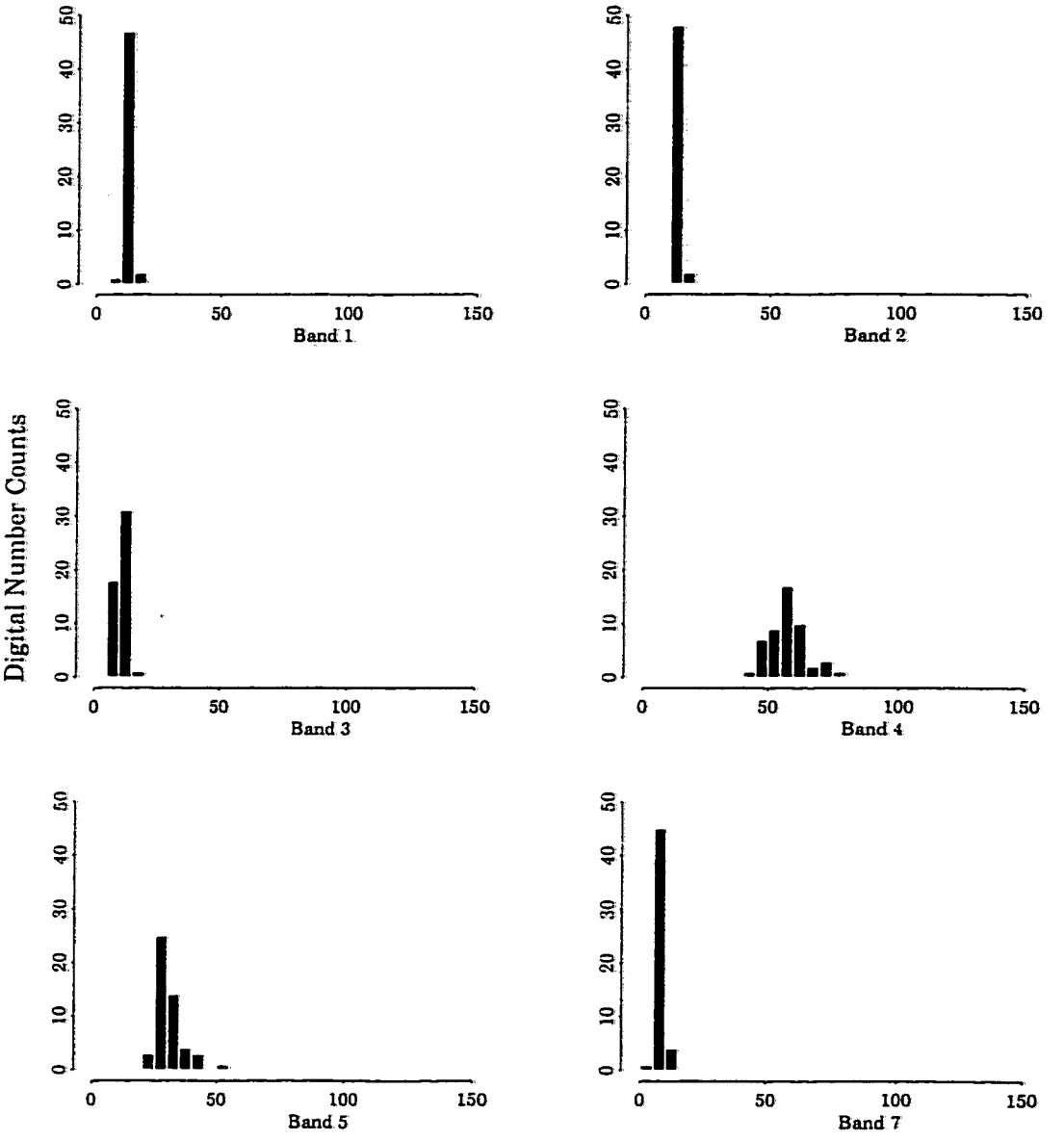


Figure B.11: LE histograms for August image.

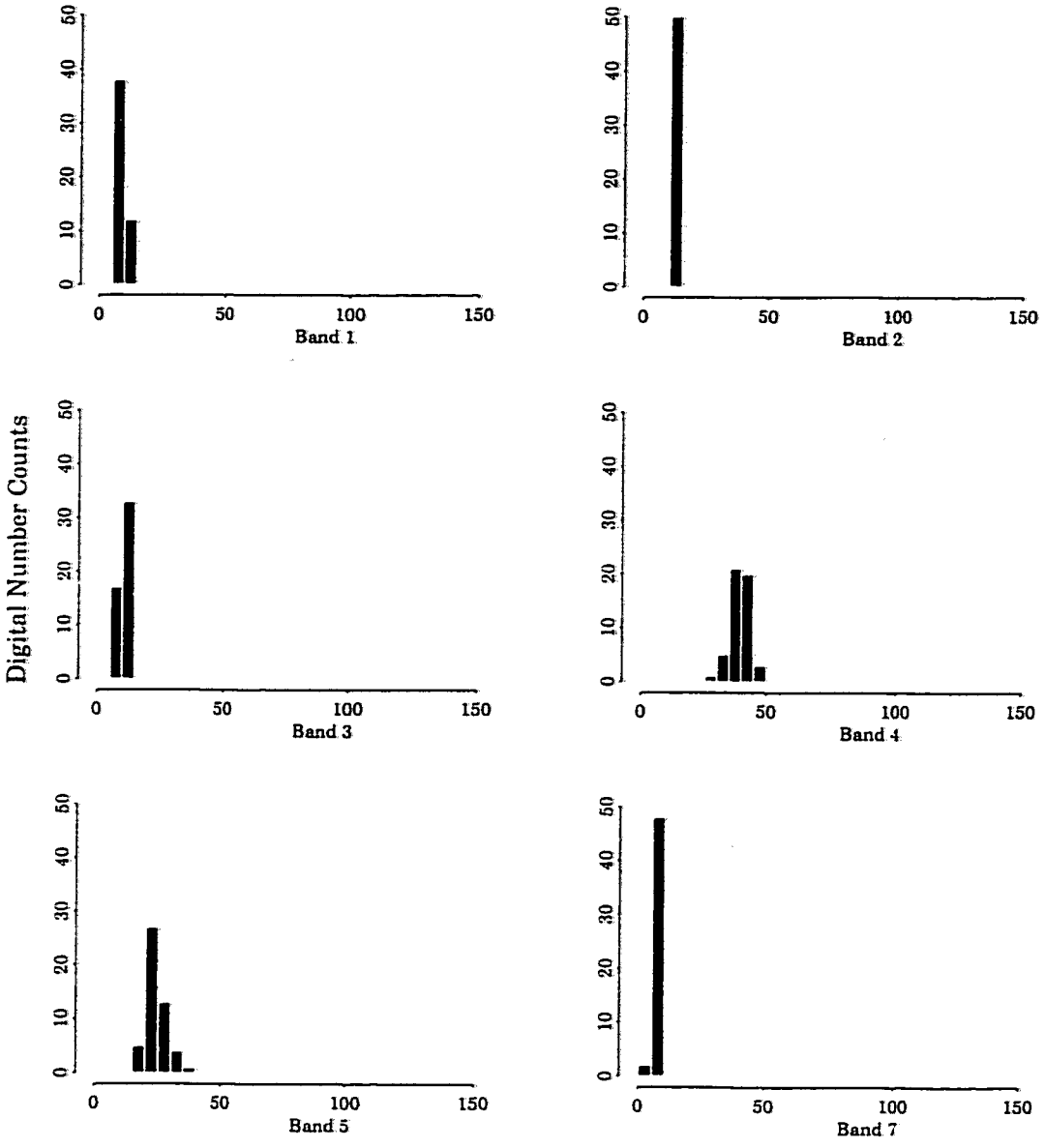


Figure B.12: UR histograms for August image.

Appendix C

ESTIMATED LINE SPREAD FUNCTIONS

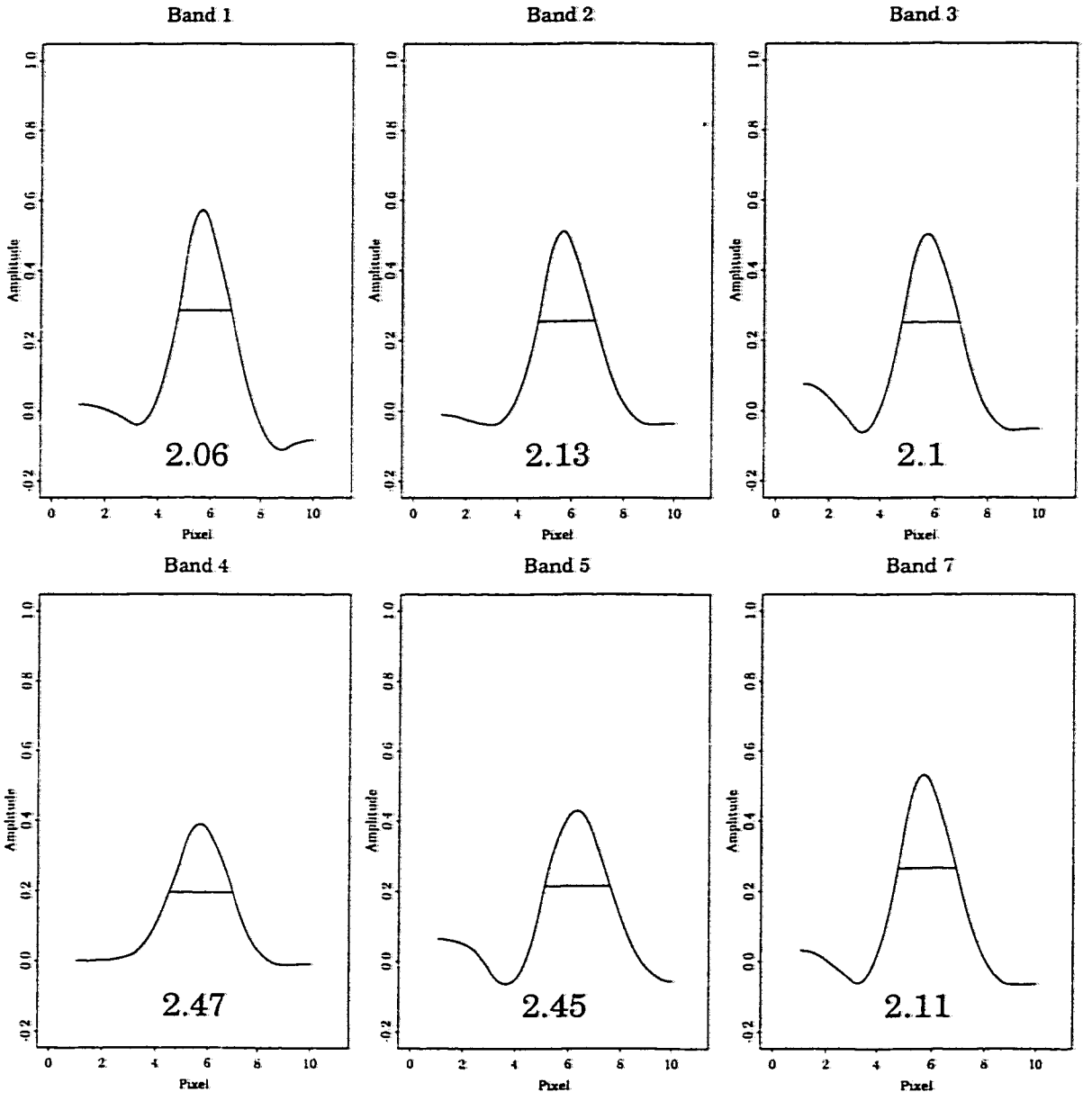


Figure C.1: Horizontal LSFs estimated for the July image.

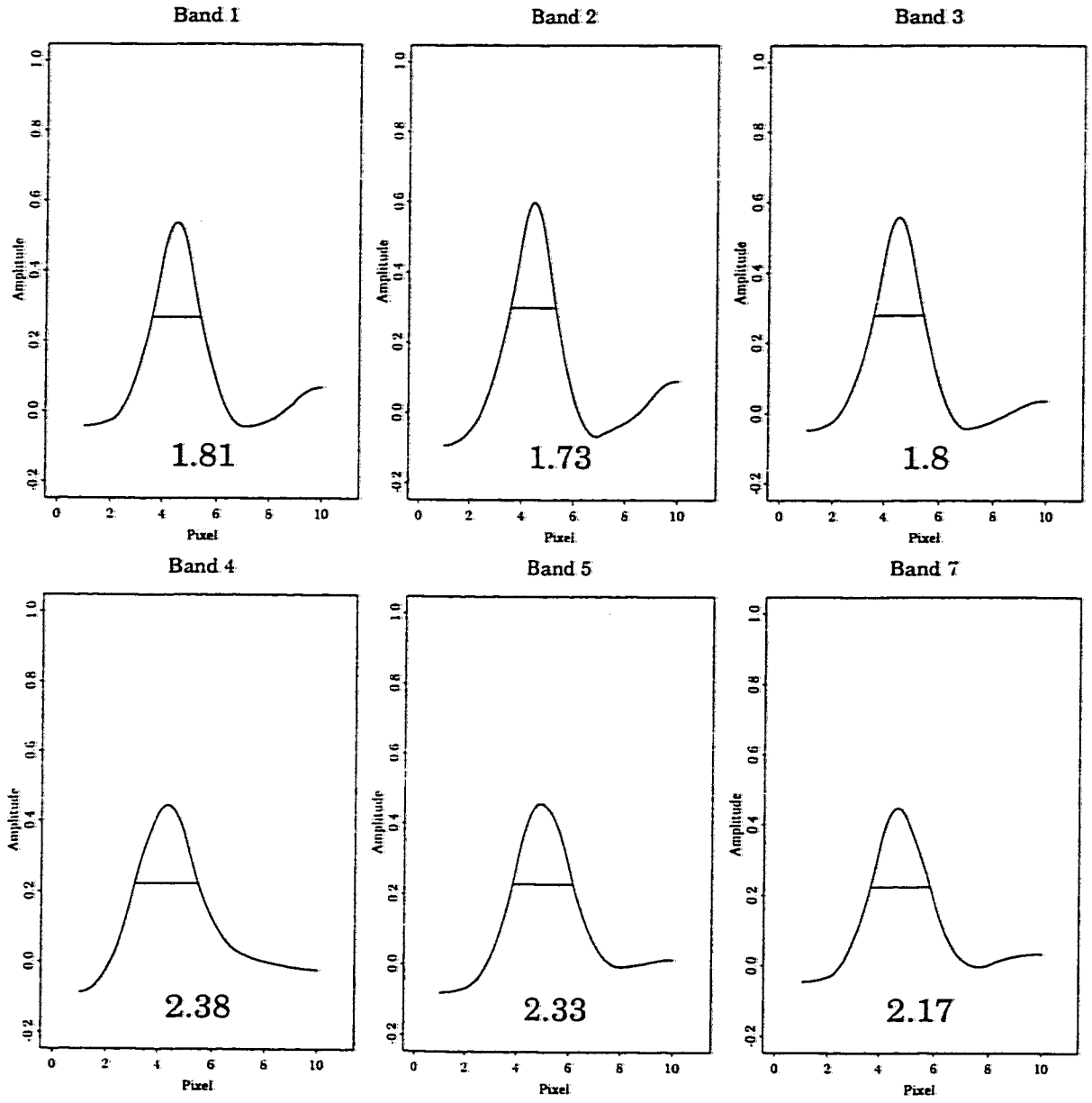


Figure C.2: Vertical LSFs estimated for the July image.

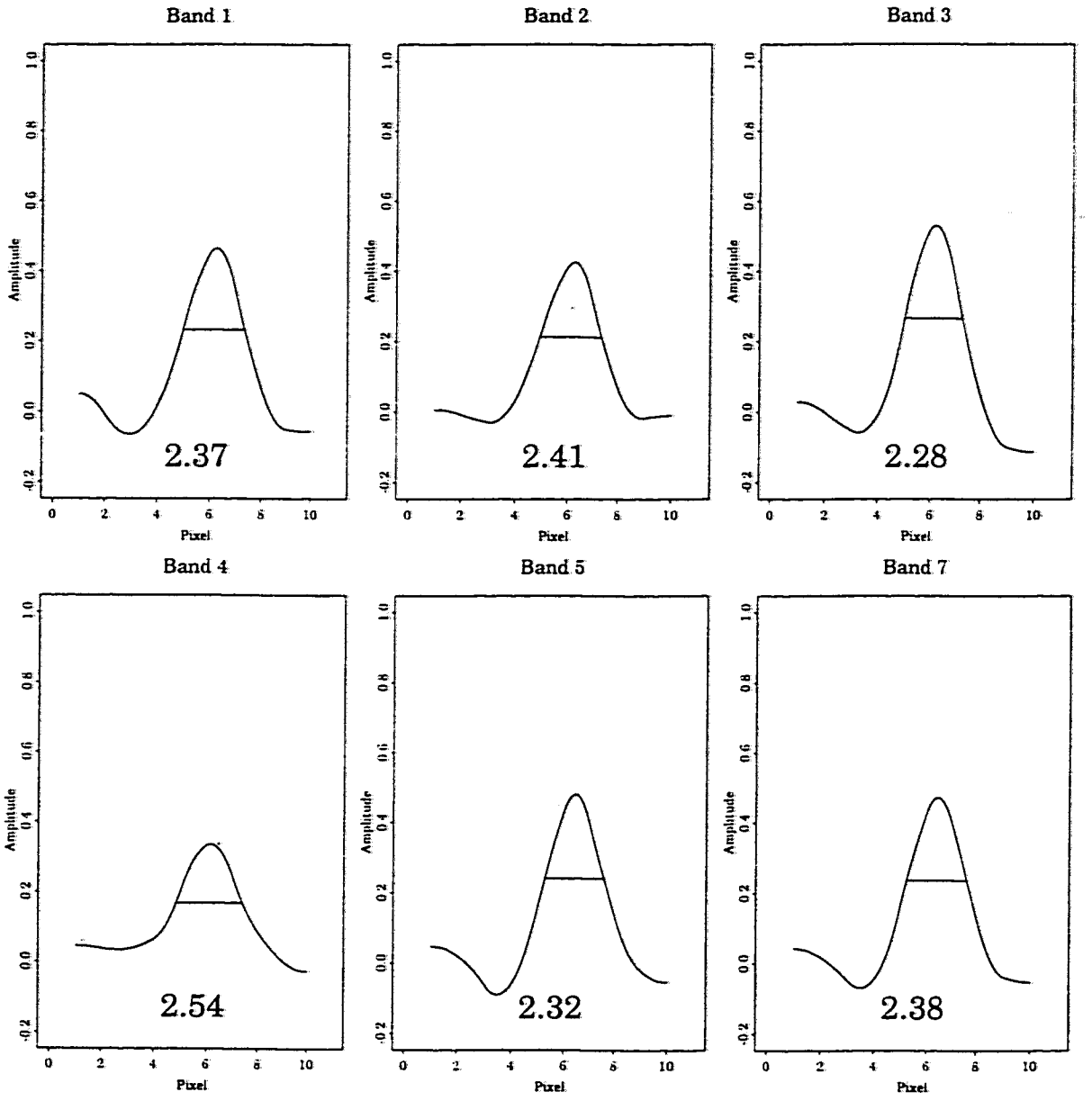


Figure C.3: Horizontal LSFs estimated for the August image.

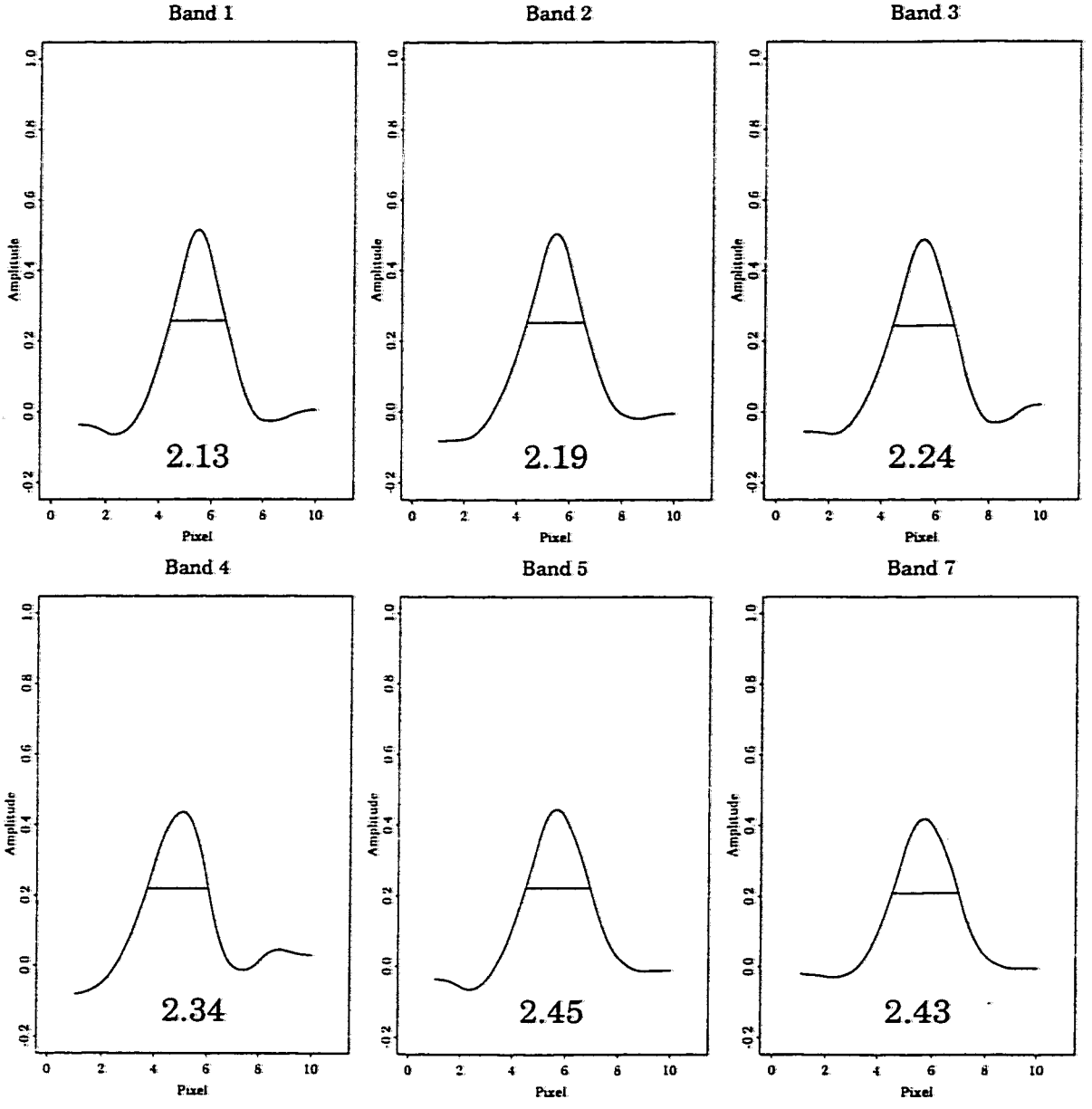


Figure C.4: Vertical LSFs estimated for the August image.

VITA

Education

Doctor of Philosophy 1998, Quantitative Resource Management
University of Washington, Seattle

Master of Science 1993, Forest Econometrics, University of Washington, Seattle
Thesis title: Wood/Non-Wood Substitution: An Econometric Anaysis

Bachelor of Arts 1987, Religion (with Honors) and Economics
Trinity College, Hartford, Connecticut
Senior essay title: Provisions for Compassion in Mahayana Buddhism
Thesis title: Deforestation in Nepal: A Social Crisis

Intensive Nepali language 1986, University of Wisconsin, Madison, Wisconsin

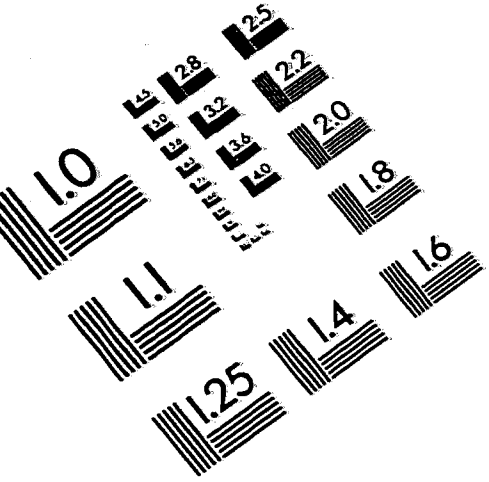
Experience

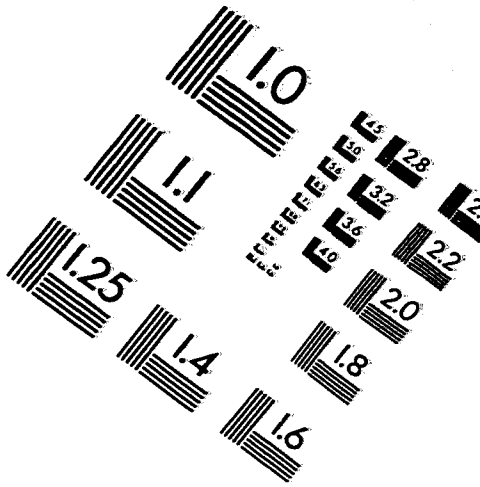
Field consultant, High altitude forest research project, Khumbu, Nepal, Spring 1996
Remote sensing analyst, Western Resources Analysis, Summer 1993

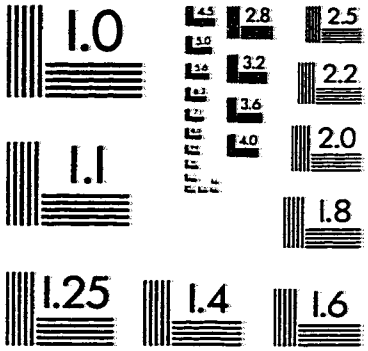
Papers

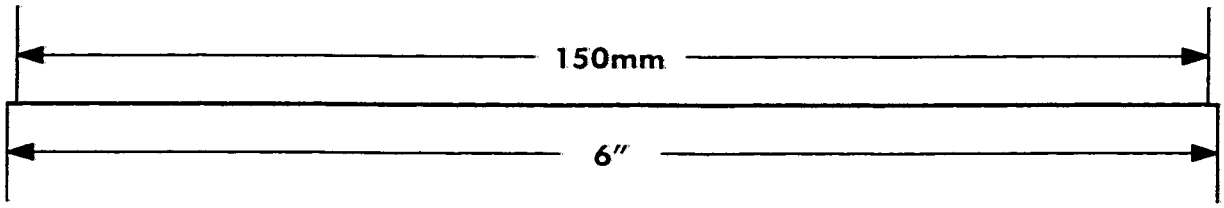
McCarter, Wilson, Baker, Moffett, and Oliver (1998)
Moffett, Besag, Byers, and Li (1997)
Moffett and Besag (1996)

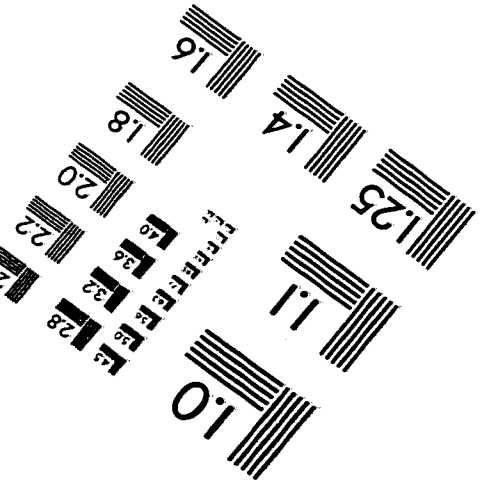
IMAGE EVALUATION TEST TARGET (QA-3)

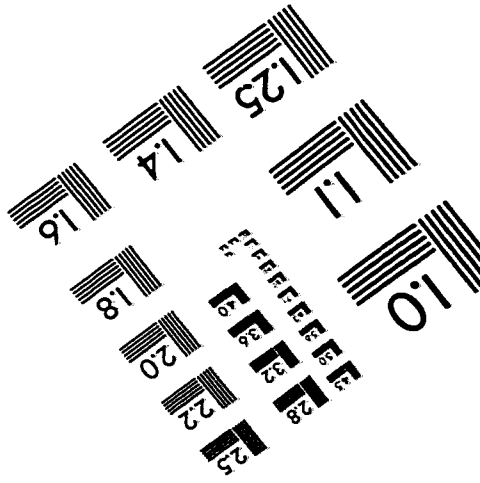












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989