THE UNIVERSITY OF CALGARY

Integrating Remote Sensing and GIS Techniques With Ecological Models to Map Biological Diversity in Boreal Forest

by

Anthony James Warren

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA
DECEMBER, 1999

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

0-612-49691-0

Abstract

The primary objective of this work was to map biological diversity at the southern extent of the boreal forest in Prince Albert National Park, Saskatchewan, Canada. This was accomplished by using remote sensing and GIS techniques to spatially estimate the four input variables of an ecological model able to predict biological diversity. The variables of interest were (1) the distance from a forest stand to a watershed ridgeline, (2) the time since the last forest fire, (3) the canopy species type and (4) the canopy stem density. The methods used to map each variable are discussed in detail. The data used to estimate these variables included spaceborne imagery (electro-optical and synthetic aperture radar) and vector format elevation contours, streams and lakes. Close attention was paid to estimating the uncertainty associated with each input variable. The results are presented in the form of three maps of biological diversity in The Park. These maps include predicted biodiversity as well as an upper and lower bound map based on the propagation of all quantified uncertainties. The results show that combining spatially estimated input parameters with such a model was reasonably successful and is an innovative use of remote sensing and GIS.

Acknowledgements

My project would not have been possible without the support of so many people. The supervision of Dr. Michael Collins has made this work possible through his valuable experience and the countless hours of attention he has given to my project. I thank Dr. David Clausi for pulling me through some of the rough times when things looked grim. In addition, they have both provided funding for this work.

Many others deserve my gratitude: Dr. Edward Johnson and his graduate student Sylvia Chipman of the Ecology Division of the Department of Biological Sciences. They have been an invaluable resource of information and data that has made this project possible; Dr. Michael Chapman for his encouragement along the way; Dr. Peter Ehlers of the Department of Mathematics and Statistics for his invaluable advice. Also, I thank Brian Christiensen and Jamie Coultier at Weyerhaeuser Canada for assisting with the processing of GPS data.

My thanks go out to a number of graduate students in my department: Peter Tate, Bruce Chaplin, Jonathan Weibe and Steven Adam for their programming expertise and a multitude of other advice.

Finally, I cannot thank my family and friends enough for their support and encouragement during both the good and bad times. It was always there when I needed it.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The boreal forest spans the northern hemisphere through Canada, Russia and Alaska and plays an important role in our society. In addition to its vital role in earth-atmosphere interactions, the boreal forest is of great economic significance as a renewable resource. It provides us with valuable and numerous pulp, paper and lumber products that are used by all facets of our society. It is in our interest to sustain this resource and it is, therefore, the subject of much scientific research geared toward understanding the mechanisms and processes of which it is a part.

Fire has always played an important role in the proper functioning of the boreal forest. The serotinous pine cones of the *Pinus banksiana* tree species, for instance, will not open and release their seeds without exposure to extremely high temperatures and therefore rely on fire for regeneration (Cameron, 1953; Johnson, 1992). Although the boreal forest is a dynamic ecosystem with continual disturbance by fire, it has been capable of maintaining itself in a relatively stable state. With this in mind, the goal of harvesting practices is seek to mimic this form of natural disturbance. However, only by understanding the processes that govern natural mortality and regeneration, can this goal be realised.

Biological diversity (or *biodiversity*) is one important aspect of this research and it is the subject of this study. Biodiversity in the boreal forest is not a dynamic process itself; rather, it is the product of many processes. These include internal processes operating within the ecosystem as well as external processes operating on the ecosystem.

Currently, ecologists are able to measure and map biodiversity of flora in the boreal region by collecting various ground sampled measurements of vegetation in forested stands. Although this method is by far the most accurate, it is a very costly, time consuming and labour intensive process. From these ground sampled measurements, ecologists can also study and characterize the *processes* operating in the boreal forest. If the underlying processes which shape biodiversity can be determined and parameterized in a mathematical model, they can potentially be used to *predict* biodiversity.

However, whether biodiversity is estimated directly from the ground sampled data or predicted with mathematical models from ground measured input variables, we are left with only stand-based estimates of diversity. Our understanding of biodiversity over the landscape is therefore limited by these sparse data. To improve our understanding, there is a need to *map* diversity over an entire landscape. The question then arises, "how can this be accomplished?"

Remote sensing and GIS techniques are concerned with handling and manipulating spatial data. This spatial data might include remotely sensed imagery or a digital elevation model. Given these spatial datasets, GIS and remote sensing techniques can potentially be used to measure and map the variables which ecologists use to predict biodiversity for a single point, over a large area. And given that a biodiversity prediction model can be created and that the input variables can be spatially characterised over large areas, a map of biodiversity can be produced.

The **primary objective** of this work is to map the spatial distribution of herbaceous plant biodiversity along with an estimate of its uncertainty, in the southern boreal mixwood forest in Prince Albert National Park.

Given this goal and the premises that have been set above, two general questions can be defined: first, what are the ecological processes that can be used to predict biodiversity and how can they be measured and parameterized in a mathematical model? Second, how can remote sensing techniques and GIS be used with spatial data to map these parameters over an area of the boreal forest. The focus of this thesis will be on the second question. However, the reader will also be presented with a discussion of an ecological model developed by Chipman (1999) which is able to predict biological diversity in the southern mixwood boreal forest. In conjunction with spatial input parameters derived from this work, the model will be used to map diversity over the landscape.

As a framework for discussing this research, background information on topics relative and pertinent to this work will be presented in the remaining sections of this chapter. First, biodiversity will be discussed in terms of how it is described and quantified (§1.1) for this study.

Chapters 3 through 6 discuss the four input variables of the ecological model needed to predict diversity. Within each of these chapters, the reader will find an introduction to the topic, a discussion of any computer algorithms developed, the methods used to map each variable (spatially) and an assessment of the uncertainty in each map created. Chapter 7 links Chapters 3 through 6 together by discussing how each input map was used in the implementation of the ecological model to map biodiversity. The resulting maps of biodiversity estimates and uncertainty are presented and discussed.

1.1 Defining and Quantifying Biological Diversity

In the broadest sense, biological diversity is *natural variation*. This natural variation can occur at the level of the molecule, gene and species. At these levels, biological diversity can be described across many scales such as the forest stand or over landscape units such as a drainage basin (Huston, 1994).

Ecologists generally use two components to describe biological diversity or *biodiversity*. The first is species richness (S) which is simply a count of the number of different species present in a given area. The second is known as evenness which is also referred to as species relative abundance and describes the proportions of species in the area. A natural ecosystem is usually composed of a few species that are very abundant within an area and many species which are much less abundant. A measure of evenness accounts for this aspect of diversity (Magurran, 1988; May, 1975; Whittaker, 1977). Although both components are accepted representations of diversity, relying strictly on a measure of richness to describe diversity is potentially dangerous because it does not describe the distribution of species in the area.

A note on biological diversity in the context of forests: trees usually contribute a greater quantity of usable natural resources to a site than herbaceous plants. High diversity areas, therefore, are not necessarily beneficial. For example, shortly after an area has been harvested, herbaceous plants and shrubs that were once starved for light and competing for nutrients and moisture with the mature canopy above would be free to grow. In addition, invasive species that have migrated in from nearby areas will be able to grow. Within a short time span (which could be years or months), the number of species in the area would be very high but not necessarily beneficial to the organisms that use the area (Harper and Hawksworth, 1995; Bormann and Likens, 1979).

1.2 Ecological Modeling For Mapping Biodiversity

As was previously mentioned, biological diversity is shaped by many different processes operating at different spatial and temporal scales. In the context of this research, the major processes shaping diversity are of a geomorphological and ecological nature. Although it is difficult to describe and quantify these process over an entire landscape, often there are surrogates that we can measure and use to obtain relevant and meaningful information. In the boreal mixwood forest, research has indicated that four surrogate variables can be used to describe biological diversity. These include (1) canopy stem density and (2) canopy type which are associated with light interactions in the canopy, (3) time-since-fire which is a disturbance process and (4) distance from a ridgeline which is related to the processes operating on a hillslope.

The common thread between these variables is that they are related to the use or manipulation of plant resources. Principally, the resources are light, soil nutrients and water (soil moisture). Although perhaps meaningless in their own right, these variables can be shown to be related to geomorphological and ecological processes – the very processes that shape biological diversity. Subsections 1.2.1 to 1.2.3 describe the processes that are thought to influence biodiversity in the boreal forest.

1.2.1 Light-Canopy Interactions

Canopy stem density (CSD) is simply a measure of the number of canopy tree stems or trunks in an area (stems/ha). For this study, a canopy tree is defined as one that is greater than 10m tall with a trunk diameter at breast height (DBH) of 10 cm or greater. The CSD measure has a number of inter-related implications. First, almost all species are competing with each other for resources within the ecosystem. If the canopy stem density is relatively high, there will be many trees competing for the same resources. Sunlight is one of these resources which all plants need for photosynthesis. Specifically, one can expect that with an increase in canopy stem density, there should be a relative increase in leaf (or needle) area from canopy trees. This increased leaf area will decrease the amount of light reaching the subcanopy vegetation since the canopy trees will use and reflect relatively more light than in a low leaf area canopy. Increased light competition between subcanopy vegetation can result in some species experiencing local extinction. Lieffers et al. (1998) states that understory development is inversely proportional to canopy development. Some species are shade tolerant and can compete effectively with low levels of direct light. Others which are not tolerant to these low light levels will die. In a mature forest with a well developed canopy, it was not uncommon to find very little herbaceous vegetation below the canopy (personal observation). This observation is supported by Halpern and Spies (1995) and Smith and Huston (1989).

The type of canopy present will also influence light transmission to the forest floor. For instance, a dense conifer (white spruce for example) canopy can cause an almost total absence of shrubs on the forest floor due to low light levels. Trembling aspen canopies, on the other hand, are more voluminous but less dense and allow direct sunlight to intermittently reach the forest floor. These small sunlight patches are called *sun flecks* (Lieffers et al., 1998).

1.2.2 Disturbance Processes

A disturbance is a process that causes mortality of plants due to an external condition. It has been shown that the intensity, frequency, timing, area affected and effect on resources are useful properties for describing disturbance processes (Huston, 1994). Of interest in this work is how a disturbance works to shape the diversity of a landscape. Even with accurate parameterisation of the above properties, there is no single rule for predicting diversity. Huston (1994) identifies that the initial state of the ecosystem, population growth dynamics and species competition dynamics are also important factors to consider when examining the effects of disturbance on biological diversity. Also important to consider are the effects of disturbance on diversity at different time scales. In the short-term, for example, species mortality could result in local extinction and therefore, a reduction in diversity. In the long-term, there could be evolutionary changes in species that result in adaptation and therefore, resilience to disturbance. And within these time scales, changes in resource availability caused by disturbance can have a slow affect on diversity in terms of species growth, reproduction and competition strategies (Huston, 1994).

Fire is a major disturbance process operating on the boreal forest. Although some are caused by humans, both accidently and purposely (prescribed burning techniques), most are lightning caused and have been a part of the ecosystem for thousands of years. Lightning caused fires account for 90% of the area burned. Fire in the boreal region has produced patchy or mosaic patterns of diversity across the landscape (Bridge, 1997; Johnson, 1992; Weir, 1996).

In §1.1 there was brief mention of the effect of forest harvesting on the vegetation composition. It is known that immediately after a disturbance, the number of species in an area will rapidly increase. Although there is great competition for resources, the absence of a canopy dramatically increases the availability of light and nutrients. Eventually, certain species will dominate and a canopy will form. The mortality of shade intolerant species under the canopy will increase and the vegetation underneath the canopy will thin out (Aber and Milillo, 1991; Bormann and Likens, 1979; Halpern and Spies, 1995). Ideally, the function of diversity with time can be used to predict the diversity of a given forest stand at a certain point in time.

1.2.3 Hillslope Processes

A landscape can be conceptualized as a wire framework of ridgelines and valley bottoms on which hillslopes are hung. Landscape topography, through the force of gravity, exerts a major influence on the movement of moisture and nutrients. Generally, moisture and nutrients will move from higher elevations to lower elevations through pathways governed by the shape of the landscape. This simple notion says something important about where we should expect to find high levels of nutrients and moisture versus lower levels. Moving into the realm of ecology, nutrients and moisture are important resources for the growth of plants. We can expect, therefore, to find relatively higher amounts of resources in valley bottoms versus the tops of hills or on ridgelines. As a result, the general pattern that emerges on the landscape is one of higher plant diversity on hilltops and lower diversity in valley bottoms

(Bridge, 1997).

The fact that biological diversity will usually be higher on relatively higher positions on a hillslope is somewhat counter-intuitive since the quantity of moisture and nutrients will be relatively lower. It is very intuitive to think that there should be more plants and a greater variety of them in the valley bottoms where resources are relatively high. How do we explain this? An ecological theory of *competitive exclusion* may explain this phenomenon. Essentially, the argument states that if two or more organisms are competing for similar resources, one will ultimately become more successful than the others and crowd the others out. The result would be local extinction of the unsuccessfully competing species and ultimately an expected reduction in species diversity (Huston, 1994).

Ideally, given the preceding discussion, an estimate of soil moisture and nutrients available to the vegetation can be determined for any point in a drainage basin. However, basin hydrology and nutrient concentrations are affected by complex interactions between vegetation, geology, geomorphology and climate which make such estimates difficult and sometimes impractical to obtain over large areas. Furthermore, it has been shown by Bridge (1997) that vegetation patterns in the southern boreal mixwood can be predicted from the surrogate measurement of relative position on a hillslope. Chapter 3 discusses the methods used in this work to map distance from a ridgeline.

1.2.4 A Biodiversity Prediction Model

The preceding discussions offered a mixture of theoretical and empirical explanations of processes that influence the shaping of biodiversity in the boreal forest. By combining our knowledge of processes that we understand, there is the potential to create a mathematical model based upon empirical observations that can be used to predict biodiversity at a given point in space and time. Given the spatial distribution of these surrogate variables, such a model can be used to predict and *map* biological diversity.

A model was developed by Chipman (1999) using the ground sampled observations discussed in §2.4 which is able to predict species richness of herbaceous plants in the southern mixwood boreal forest. This model was adopted for this study. Therefore, this work will be limited to mapping the species richness component of biodiversity. The model combines inputs of time-since-fire, distance from a ridgeline, canopy stem density and canopy type into a prediction of biodiversity. This work focuses on mapping the four model components within a section of the boreal forest. Chapter 7 will offer a more complete discussion of the form of the model and how it was implemented.

Chapter 2

Study Site and Data Description

2.1 Regional Description

The study site is part of the boreal mix-wood forest located in and around Prince Albert National Park (PANP), central Saskatchewan (53° 35' N to 53° 20' N and from 106° 0' W to 106° 47' W). PANP is located on the southern fringe of the boreal forest which gives way to an expansive agricultural region to the south (Figure 2.1). The following physical description of the area has been summarized from Bridge (1997).

The geomorphology of the area has been defined primarily by the glacial events of the Pleistocene ($\approx$ 12 000 years ago). Although glacial tills dominate the area, there are also organic glaciofluvial (of glacial river origin) and glaciolacustrine (of glacial lake origin) deposits of significance. The topography is composed of rolling hills with an elevation range of 500 to 800 m above sea level.

Long cold winters and short cool summers are characteristic of the regional climate. Between 400 and 500 mm of precipitation is received by the area with approximately 70% falling as rain.

There are two major disturbance regimes at work in the area: forest fires and forest harvesting. The first includes both human induced and natural lightning caused fires although the latter make up the majority. Weyerhaeuser Canada operates a Forest Management License Area (FMLA) in the region and is responsible for the

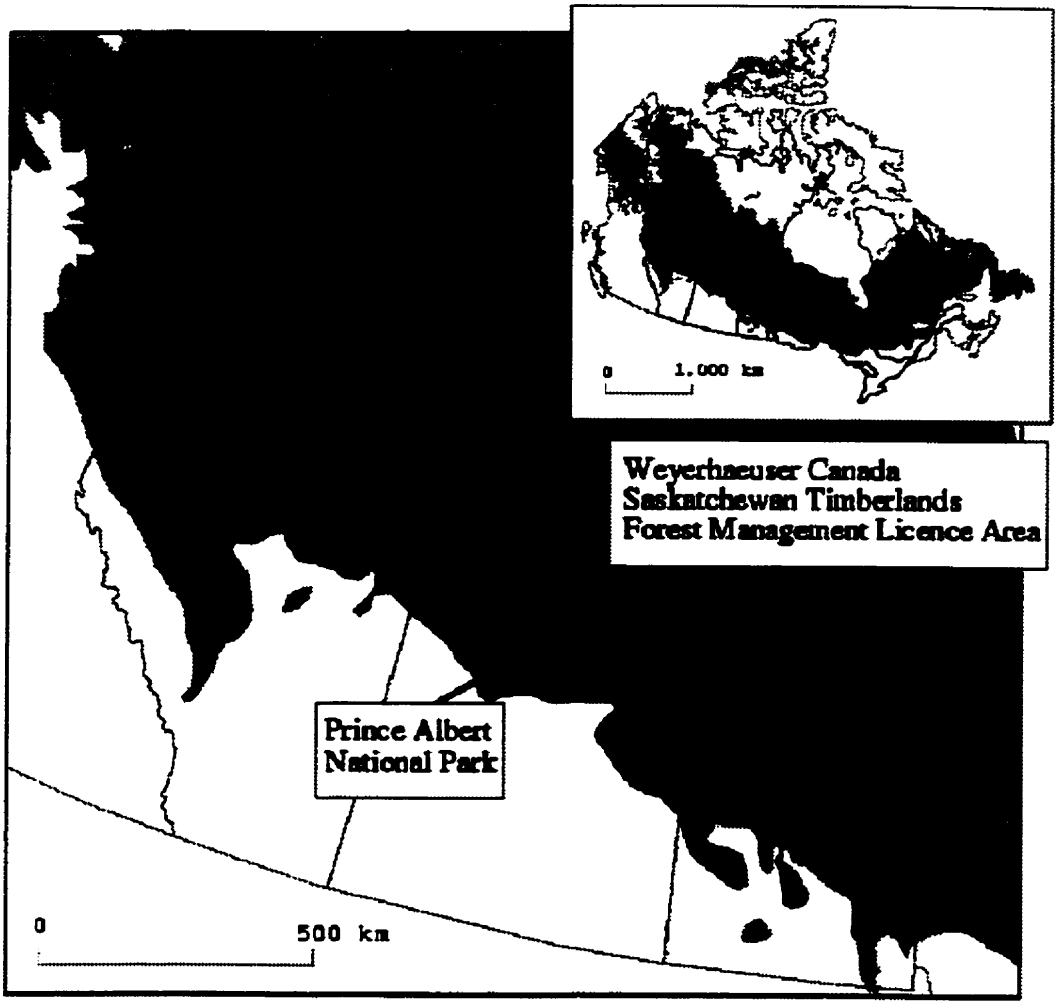


Figure 2.1: Location of the study area. The boreal region is shown by the gray shaded areas with the southern mix-wood boreal forest delineated by the darker grey belt. This unpublished figure has been used with the permission of Dr. E.A. Johnson of the Department of Biology, The University of Calgary.

regeneration of harvested areas. However, within PANP there is no harvesting and the only source of significant natural disturbance is fire.

Eight tree species dominate the area which are comprised of both coniferous and deciduous species. The coniferous species are *Picea glauca* (Moench) Voss, *Picea mariana* (Mill.) B.S.P., *Pinus banksiana* Lamb., *Abies balsamea* (L.) Mill., *Larix laricina* (DuRoi) K. Kock. *Populous tremuloides* Michx., *Populus balsamifera* L. and *Betula papyrifera* Marsh. make up the deciduous species. There are also many herbaceous shrubs and ground cover plants that contribute to the vegetation composition.

2.2 Image Data Acquired From Space

Two Thematic Mapper images were obtained for this research. They are both roughly centered over Prince Albert National Park and they were acquired on June 10, 1996 and August 29, 1996. Both images are virtually cloud free and each include all seven spectral bands at 30m ground resolution. Processing in the form of precise geometric correction was performed on the two 1996 images by Radarsat International. Both images were referenced to the WGS-84 ellipsoid in a UTM projection system. This processing resulted in resampling the image pixels to 25m ground sample spacing (Appendix D).

In addition to electro-optical data, two L-band SIR-C SAR and two C-band SIR-C SAR images (Figure 2.2) acquired by the NASA Space Shuttle on October 4 and 6, 1994 were obtained from the NASA Jet Propulsion Lab. The ground resolution for both images is 12.5 metres. Together, the two images cover a substantial portion of the PANP study site (Appendix D).

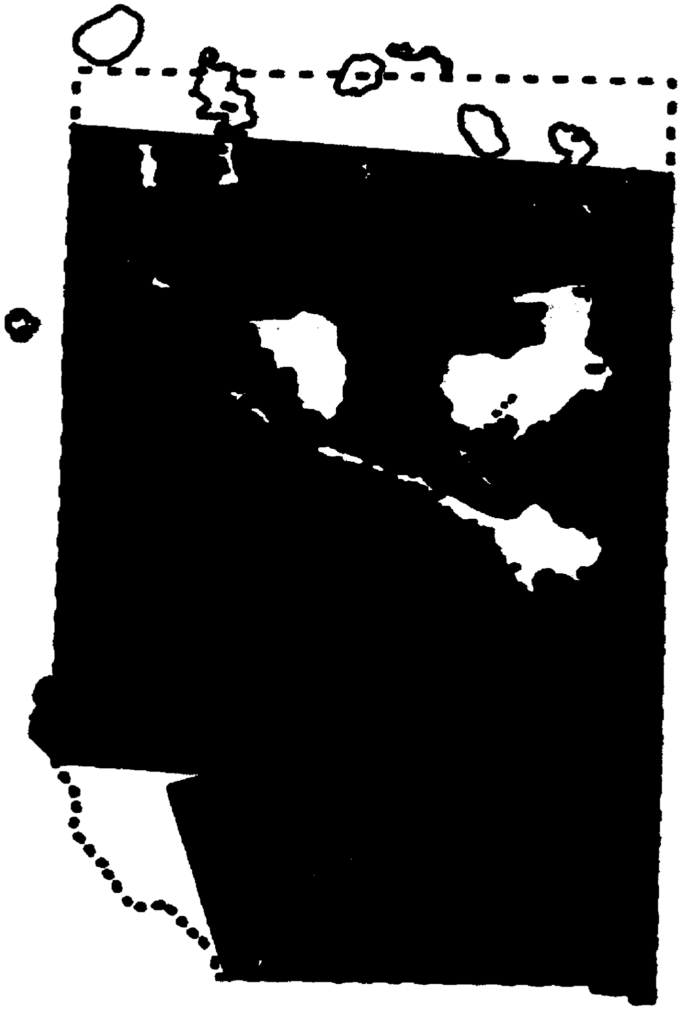


Figure 2.2: SIR-C SAR image coverage map. The dashed line is the PANP boundary. The light gray areas have SAR coverage while the dark gray areas show image overlap. Lakes have been included for geographic reference.

The original SAR imagery was filtered using a gamma-gamma filter in order to remove some of the speckle noise (Lopes et al., 1993). Close inspection of the imagery prior to filtering revealed great local variation in backscatter amplitude. This was especially apparent in the L-band imagery. After filtering, this variation had been removed.

2.3 GIS Data

ArcInfo GIS (ARC/INFO, 1997) vector line coverages of elevation contours, rivers and lakes were made available by Parks Canada for PANP and surrounding adjacent lands. These were used to interpolate an elevation model (DEM) of the area using ArcInfo software. The contour interval of the topographic vectors was 8m. These vector GIS files were not provided with any metadata so that errors of unknown type and quantity exist within these files (Appendix D). The DEM interpolation process is discussed in greater detail in Chapter 3.

A raster map of time-since-fire was obtained from The University of Calgary, Biology Department, Ecology Division. This map was derived from data records collected and maintained by Weir (1996). Its development is discussed further in Chapter 4.

2.4 Ground Sampled Data

The ground data includes approximately one hundred and fifty sampled forest stands in and around PANP. The sampling method used was the point centered quarter method as described by Cottam and Curtis (1956). Each forest stand was point-sampled 15 times along a U-shaped transect. In some instances, the shape of the sampling transect was modified to accommodate unsuitable sampling terrain. Each sample point is divided into four equal sections or quadrats. At each point on the transect, four measurements were taken (one for each quadrat) which included: selected tree diameters *at breast height* (DBH) for canopy and understory and tree distances from the sample point (if less than or equal to 10 m), selected shrub distances and base diameters, tree seedling counts in two, one by one metre opposing quadrats (centered on the sample point) and herbaceous species and moss counts in two 25 x 25 cm opposing quadrats (centered on the sample point). The data were collected by Bridge (1997) in the summers of 1993 and 1994 and Chipman (1999) and this author in the summer of 1997.

For each forest stand sampled, three to five GPS positions were measured (depending on the transect shape) for georeferencing purposes. Generally, these points were collected at the corners of the transect. All points were post-differentially processed and corrections applied. These sites were used as training and testing sites for all classifications performed.

2.5 Data Preparation

2.5.1 SAR Imagery Extraction

The SIR-C SAR imagery was delivered on 8mm tape and was extracted to harddisk using CEOS Tape Reader software (NASA Jet Propulsion Lab, 1993; Vuu et al., 1995) which was compiled from C source code. The data was extracted to a raw format which consists of a set of leader, trailer and image files. The imagery at this point was stored in the form of a Stoke's matrix. Data Compression software (NASA Jet Propulsion Lab, 1994; Chapman, 1995) compiled from Fortran source code was then used to synthesize the Stoke's matrix into the required imagery.

First, the Stoke's Matrix for each image was multilooked by two in both the azimuth and range direction. Since the October 4 image was quad-polarized, a total of four images were extracted. HH, HV, VH and VV images were created and imported into PCI software. The October 6 image was only dual-polarization and therefore was synthesized into two images. HH and HV images were created and also imported into PCI software.

2.5.2 Georeferencing Spatial Data

The ArcInfo vector line coverages were obtained in the NAD 27 datum and a UTM projection system. Using datum conversion routines designed specifically for Canada, within ArcInfo, all coverages were converted to the NAD 83 datum.

Using the GCPWorks module of PCI software and 1:50 000 NTS map sheets, approximately 40 ground control points were collected for each SAR image for georeferencing. The images were transformed into the WGS-84 ellipsoid with an RMS error of approximately one pixel (25 m).

Chapter 3

Mapping Distance From a Ridgeline

Section 1.2 examined the role of moisture and nutrients and their importance in explaining biodiversity. This chapter outlines the development of a map of distance to ridgeline for use in the biodiversity model. In addition, attention is paid to the uncertainty of the distance map.

3.1 Problem Definition

To map distance from a ridgeline (DFR), it was necessary to develop an algorithm that was able to determine the distance from each image pixel to its respective ridgeline. The details of this development are discussed in §3.2.1. In this application, the *respective* ridge is that which would contribute water and therefore, nutrients (via overland flow or groundwater movement) to the pixel in question. Since water, if unimpeded, will flow in the direction of greatest slope gradient (aspect), it is necessary to find the path back to the ridge area from which the water originated. To accomplish this task, two important assumptions were incorporated into the computer algorithm: (1) it is assumed that for each image pixel, the direction of maximum gradient will govern the path direction of all overland water movement down a hillslope. And (2) it is assumed for each pixel that a straight line path in the direction of its local aspect will lead to the correct region of the ridgeline.

The first assumption is questionable since water flow (surface and subsurface) is governed not only by aspect but also by interactive factors such as surface slope, basin shape, geology, vegetation and basin meteorology and climate (Briggs et al., 1989; Tuttle, 1980). Nevertheless, the definition of surficial drainage networks as well as subsurface hydrology modeling using DEMs has become widely practiced (in Beven (1997); Beven and Kirkby (1979); Kirkby (1007); Jenson and Dominique (1998); O'Loughlin (1986); Wigmosta et al. (1994) for instance). The reasoning behind its widespread use is that despite these other influences, surface topography remains a major influence on the movement of water. This is confirmed by work of Bridge (1997) as discussed in §1.2.3.

Figure 3.1 addresses the second assumption by illustrating three possible methods for distance determination. For Pixel A, Method 2 provides a good distance approximation relative to Method 3. However, for Pixel B, Method 2 provide a gross underestimate of the actual distance. Clearly, the assumptions will not always hold true but, given the available data inputs, they will usually provide a reasonable approximation of the path leading to the source of moisture for a given pixel. It is also noteworthy to consider that Pixel B will receive moisture inputs from a *section* of ridgeline and it could be argued that neither Methods 2 or 3 provide the absolutely correct distance to the ridge. Rather, the distance should be a median or average value that lies somewhere in between.

Given the shortcomings that these assumptions introduce into the algorithm, it is worthwhile to examine some other approaches used in similar applications. Bridge (1997) used straight-line shortest distance to the nearest ridge. Algorithmically, this is a simple measurement but it lacks the crucial ability to account for the actual movement direction of water flow through a basin. Flowing water will usually have a

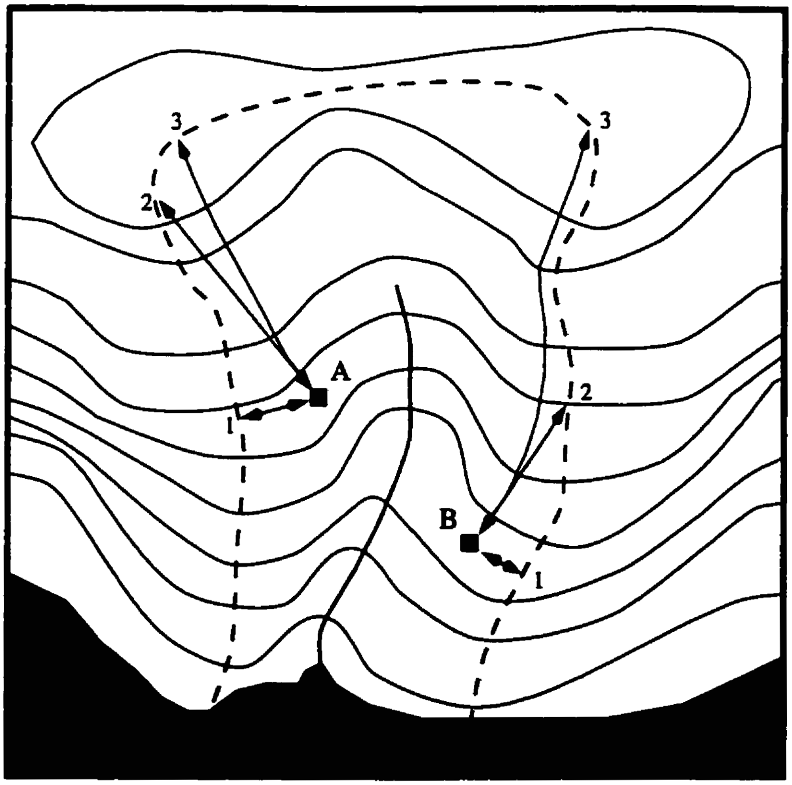


Figure 3.1: This contour map illustrates three possibilities for measuring distance from two pixels (A and B) to a basin ridgeline. Method 1 (used by Bridge (1997)) finds the closest ridge (horizontal distance). Method 2 (from this work) finds the closest ridge (horizontal distance) based on a path direction calculated from local aspect. The third method traces a path which cuts perpendicularly through contour lines and is considered to be the best estimate.

movement direction component that points toward the outflow area of the basin and thus, the closest ridgeline will usually be a gross underestimate of true distance (to the proper ridgeline). This is illustrated in Figure 3.1 by comparing the distances from pixel to ridgeline for Method 1 with Methods 2 and 3.

The approach presented here offers an improvement on the work of Bridge (1997) by attempting to follow a path opposite to water movement as shown with method 2. In basins with a regular shape (such as those which generally occur in the study site of this work), this path will usually closely approximate the distance given by a path which cuts perpendicularly through contour lines. The latter is obviously more accurate but is problematic in terms of algorithm development. Initially, the algorithm implemented in this work was designed to follow a path up to the ridgeline such that the aspect of each pixel in the path would be used to determine a travel direction (opposite to the aspect). The path would be mapped out on a pixel by pixel basis until the ridgeline was reached and the distance determined by the number of horizontal, vertical and diagonal pixel movements made. Figure 3.2 illustrates this idea.

However, in some cases small local maxima existed in the DEM. At these maxima, the aspects of adjacent pixels can define travel directions which point toward each other. This means that the path would become trapped between these pixels. In these situations, an attempt was made to take an average aspect direction from a window surrounding these pixels and continue the path. This method solved the problem most of the time, but situations still arose in which the path became trapped. Unfortunately, the effects of these traps were often far reaching. Since each image pixel must be assigned a distance to the ridge, often the paths from other pixels

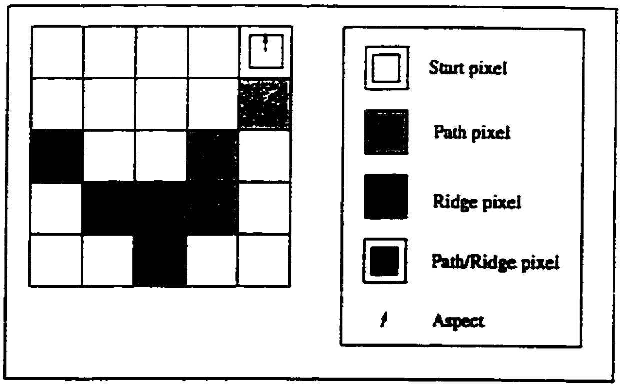


Figure 3.2: The path from pixel to ridge is defined on a pixel by pixel basis. The opposite direction of the aspect of each pixel points to the next pixel in the path until the ridge is reached.

around these local maxima also lead into the trap. If these pixels were ignored, large patches in the image would not be assigned a distance to the ridge.

A more challenging issue was the possibility of loops occurring in the pixel paths. In relatively flat areas, pixel paths would sometimes loop back onto themselves, to be trapped in an endless path. The only way to determine if a pixel path was trapped in an endless loop was to track its path and check (after each pixel movement) to see if it returned to a past position. This solution was computationally intensive. Furthermore, the problem of how to resolve these looping situation was not adequately solved. An averaging window (as discussed above) was used with some success but not all cases could be resolved. These problems lead to the development of the alternative algorithm implemented in this work.

At this point, we examine some alternative strategies for determining distance to the nearest ridgeline:

In an application involving mapping of surface saturation zones in drainage basins, O'Loughlin (1986) determined a trajectory from a given point up to a ridgeline using contour lines rather than a cell-based aspect approach. This was accomplished by minimizing the distance between each successive contour line as the trajectory moved up the slope. This ensured that the path of maximum gradient (and therefore the path of most likely water flow) would be found. Why was this potential improvement not used in this work? The primary reason for this choice is because the ***solution-space*** of this work is in the raster or cell-based domain and not the vector domain. This method offered by O'Loughlin (1986) requires contour line vectors of elevation. The algorithm developed for this work was geared toward solving a problem in the raster domain. Although these vectors were available for this project, this may not always be the case. For instance, DEMs derived from SAR interferometry (Atlantis Scientific) or stereo SPOT imagery are raster based.

Skidmore (1990) used a digital elevation model to map the ***topographic position*** of pixels relative to a valley bottom and ridgeline. This was performed by calculating the shortest straight-line distance to the nearest valley and to the nearest ridgeline. The position is then calculated by dividing the Euclidean distance to the nearest valley bottom by the sum of the distances to valley bottom and ridgeline. Although this measure gives us a sense of the ***relative*** location of a pixel in a drainage basin with respect to the ridgeline and valley, it would not be practical to use in this model. Using the Euclidean distance, we again ignore that fact that moisture movement will be largely influenced by the aspect of the terrain and not the shortest path to the ridge. This approach is similar to that used by Bridge (1997).

3.2 Methods

3.2.1 Computer Algorithm Development

The distance from ridgeline program (DFR) relies on three sources of data to calculate distance. The first is an image of pixel slope aspect. Aspect defines the direction (in the horizontal plane) that a sloping surface faces and is taken perpendicularly from the line of steepest slope on the surface. The input image pixel aspects must be indexed starting from north at 0 degrees and moving clockwise to 360 degrees. Pixels with no slope (and therefore no aspect) must have a value of 510. The second is an image of drainage basin ridgelines. Ridges must have a value of 255 with a value of 0 assigned to all non-ridge pixels. The DFR program traces a path from each pixel up to a ridgeline. The path direction from pixel to ridge is defined as the opposite direction of the aspect of the starting pixel. The user can specify a window size centered on the start pixel whose aspects are averaged and used to determine a path direction to the ridge. Averaging the angles was accomplished by converting the angles into polar coordinates, taking the mean of the vertical and horizontal components and converting back to an angle (in degrees). Increasing the window size to a 3×3 or 5×5 reduced the sensitivity of travel direction to local variation in pixel aspect. The third data source is an image of path barriers such as lakes and rivers. Its role is examined below in the paragraph discussing error handling.

Movement occurs one pixel at a time in one of eight directions as shown in Figure 3.3. These movement directions are limited by increments of 45 degrees (starting at 0 or 360 degrees for north movement) which will ultimately result in over- or under-movement in one or both of the horizontal and vertical directions. To avoid this potential problem, after each pixel move the **desired** travel direction (the precise direction in which you would like to move) was subtracted from the **actual** angle travelled (the direction in which you are forced to move due to the constraints of a grid). This difference was used to make adjustments to the pixel movement directions so that the overall path travelled closely followed the desired direction of travel. The following two equations generalize this iterative adjustment procedure:

$$TD_{actual} - TD_{desired} = \Delta D \tag{3.1}$$

$$TD_{desired} - \Delta D = TD_{new} \tag{3.2}$$

where TD is the travel direction and ΔD is travel direction adjustment angle. After the first iteration, the *new* travel direction becomes the *desired* travel direction. The *actual* travel direction will be re-evaluated at each iteration (pixel movement) and adjusted when necessary, according to the movement constraints outlined in Figure 3.3.

The user is able to specify the size of an image pixel in the horizontal and vertical directions in the desired units. The program counts all horizontal and vertical pixel movements and then calculates the actual straight line distance from pixel to ridge based on this information. Figure 3.4 shows an example of defining a path to the ridge for a given pixel. Notice that the actual path of travel differs from the desired travel path due to pixel movement constraints. However, the starting and end points are very similar so that the dx and dy components can be used to calculate total path distance.

Three error situations can occur during this procedure:

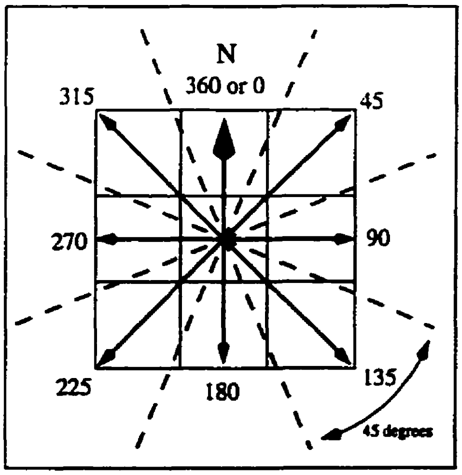


Figure 3.3: *Actual* Pixel movement direction is defined by a 45 degree range of angle in each of 8 directions from a central pixel. The dashed lines indicate the range of angle for each movement direction.

1. Pixel-to-ridge paths may lead off the image so that a distance cannot be calculated. This is likely to occur with pixels near the image edge. These pixels were flagged with a value of -999 so that the user can remove them if necessary. If not removed, parts of the output image will not have valid distance values.

2. In some areas, there was no aspect (slope = 0). These pixels were easily identified because they had been previously assigned a value of 510. First, a window of user specified size (centered on the slopeless pixel) was searched for other pixels with aspect (slope > 0). If such pixels did exist, their mean aspect was assigned to the central pixel and the algorithm was able to continue. If all pixels had a value of 510, the central pixel was assigned an error flag value of -1 and written to the output image.

3. Where small, local elevation maxima occur in a basin, it is possible that the

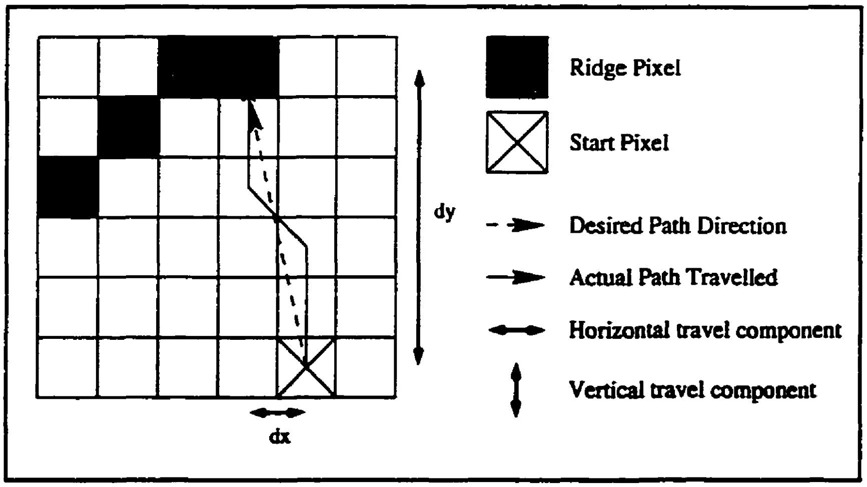


Figure 3.4: An example of tracing a path from a pixel to its respective ridgeline.

direction of path travel will point toward the opposite ridgeline. In this case, the path of travel will first move down to a valley bottom and then up to the opposing ridge which is not a desirable measurement. To prevent this problem, the algorithm checks to see if the path encounters any barriers such as lakes or rivers along the way. If it does, the start pixel is assigned an error flag value of -2 and is written to the output image.

Error flag values were used so that the user could evaluate potential areas of error and to speed up the algorithm for pixels that were problematic. In many cases, the areas of error will be small so that median or averaging filters could be used to assign values to the pixels in error.

3.2.2 Preliminary Data Processing

The primary data source for mapping distance from a ridgeline was a DEM. It was necessary to interpolate a DEM from contour lines. Using ArcInfo GIS software (ARC/INFO, 1997) and the vector line files of elevation contours, lake outlines and rivers, a DEM was interpolated. The TOPOGRID function was used which interpolates a grid from line features. This function allows for the incorporation of lake information and drainage information (rivers). Due to hardware memory constraints, the smallest size of grid cell that could be interpolated was 30m.

Around some of the lakes it was found that the interpolation algorithm had produced a number of thin linear depressions. After close inspection, it was determined that these were primarily the result of including island polygons within the lakes which did not have contour lines within them. Deleting some of the smaller islands and adding contour lines to the larger ones removed most of these undesirable lines after re-interpolation. In areas in which these features persisted, they were locally filtered out using median filters of varying sizes. The size of the filter used depended on the severity of the errors. Although simplifications were introduced to the filtered areas of the DEM, the trade-off was necessary to maintain a connected drainage network. Furthermore, the primary use of the DEM was to produce a map of basin ridgelines and the topographic details within the basins were of lesser importance.

Within PCI EASIPACE software, the DEM was then imported and resampled to 25m pixels using a nearest neighbor algorithm and co-registered with the existing imagery in reference to the WGS-84 ellipsoid. Next, using terrain analysis programs within PCI (1997) software, a map of basin ridgelines was created. These programs are based on the work of Jenson and Dominique (1998) and are able to extract geomorphological features from DEMs. In all, five PCI programs were needed to create the ridgeline map. First, DWCON (Drainage Watershed Conditioning) was used to create a depressionless DEM. This function removes small *sinks* that may impede the flow of water over the surface. On most landscapes, sinks are likely due to imperfections in a DEM and if not removed, may result in discontinuous water flow paths. DWCON also produces maps of water flow direction, flow accumulation and delta values. The delta value represents the increase in flow accumulation in the flow direction.

The initial results of DWCON were problematic. There were ten areas for which a flow direction could not be defined. These areas were large *sinks* in which water did not have a natural outflow path and all water in the surrounding area flowed into the lake. A number of compounding factors caused these problems: First, the area is relatively flat and the contour data used in the interpolation had a fairly wide interval (8 m). Interpolation of these flat areas where contours are sparse can result in small imperfections (such as peaks) in the DEM which can act as dams to impede flow. Second, lake polygons and river vectors were included in the interpolation as natural breaklines. Some of the river vectors were not complete and thus, flow paths to and from lakes were often broken. During the interpolation process, lakes can be interpolated as slight depressions. If a river is not present to force a drainage path from the lake, the lake may become a sink. To remedy this situation, the DEM was manually altered. Lake levels were increased to the lowest level of surrounding land such that they would have an outflow path. This procedure mimics that used by computer algorithms for sink removal.

Next the SEED program was executed which automatically places starting or *seed* points at the outflow points of watersheds (where there are major tributaries). A threshold value must be specified by the user. If the area draining into a tributary (as calculated from the flow accumulation and delta values) is greater than the threshold value, a seed point will be placed at the tributary fork. If the area draining into a tributary is smaller than the threshold, the basin will not be seeded and it will become incorporated into a larger basin. In other words, the threshold value will globally govern the smallest size of basin that will be delineated. Following Bridge (1997) (whose work was in the same area), a threshold value of 3000 pixels was chosen.

The WTRSHED (Make Watersheds) program was then executed which delineates watersheds from the cell flow directions and seed points. The resulting image was a map of filled polygons representing watersheds. In order to delineate only the ridgelines, a raster to vector conversion was performed using the RTV program and then GRDVEC to *burn* the vectors into a raster based image.

3.2.3 Map Creation

Using the DFR EASI program discussed in §3.2.1, a map of DFR was then created. In order to use the DFR program, the first step involved calculating an image of aspect from the DEM which was accomplished with the ASP (Aspect) routine in PCI. Next, a map of ridgelines was created using GRDVEC (see above). Ridges were assigned a value of 255 and all other areas were assigned a zero value. The barrier map was created in a similar fashion except that river vectors and lake polygons were burned into an image with a value of 255. An aspect averaging window size of 5x5 was specified and the DFR program executed.

As discussed in §3.2.1, in some cases it was not possible for the program to determine a distance because the path to the ridge encountered a barrier such as a lake or river. To determine distances for these areas, the PCI GRDINT program was used to interpolate values from the output image from the DFR program.

3.3 Results

Figure 3.5 shows a sample of the resulting image after interpolation. Notice that as a path cutting perpendicularly through the contours is followed from a ridge to the center of a valley, distance to the ridge increases (as shown by lighter shaded pixels).

Figure 3.5: This small subsample of the final image shows pixel distances to ridgelines created from the DFR program. Light shades represent distances far from the ridgeline relative to darker shades. The black polygons are lakes. Superimposed onto the image are contour lines (thin dark lines) and basin ridgelines (light dashed lines). Histogram equalization has been performed on this image to improve contrast.

3.3.1 Assessing the Accuracy of the Distance From a Ridgeline Map

Calculation of DFR was subject to many sources of uncertainty. For the DEM, these sources include uncertainty from the contour lines used for interpolation, the errors produced during interpolation and resampling the grid cell size. For the basin and ridge delineation, these sources include uncertainty from the flow direction, flow accumulation and delta value calculations. Uncertainty was also introduced in calculating the distance from the ridgelines. In some situations, there were no river vectors to ensure that a basin was properly divided such that distances would only be measured from a pixel to its respective ridge.

The quantity of uncertainty present in these sources was mostly unknown and given the nature of the software used, it was not practical to propagate the known uncertainty. To propagate known uncertainty through all processing steps would have required re-writing these algorithms (to accommodate uncertainty) and this was not considered a viable option for the scope of this project. The assessment of uncertainty for the DFR map, therefore, was of an empirical nature. The formulation of the biodiversity model used manually map measured distance from ridgeline estimates for the ground sampled forest stands. Therefore, uncertainty in the map of DFR was assessed by comparing manually measured values to computer calculated pixel values. Forest stands were first located on the 1:50000 topographic maps. By visually assessing the terrain around a forest stand, the nearest ridge was located on the map and the distance measured. To clarify, the measured transect running from stand up to ridge always cut perpendicularly through the contour lines. This was repeated for 31 forest stands such that a full range of distances were sampled.

Figure 3.6 shows a scatterplot of the collected points represented by hollow circles.

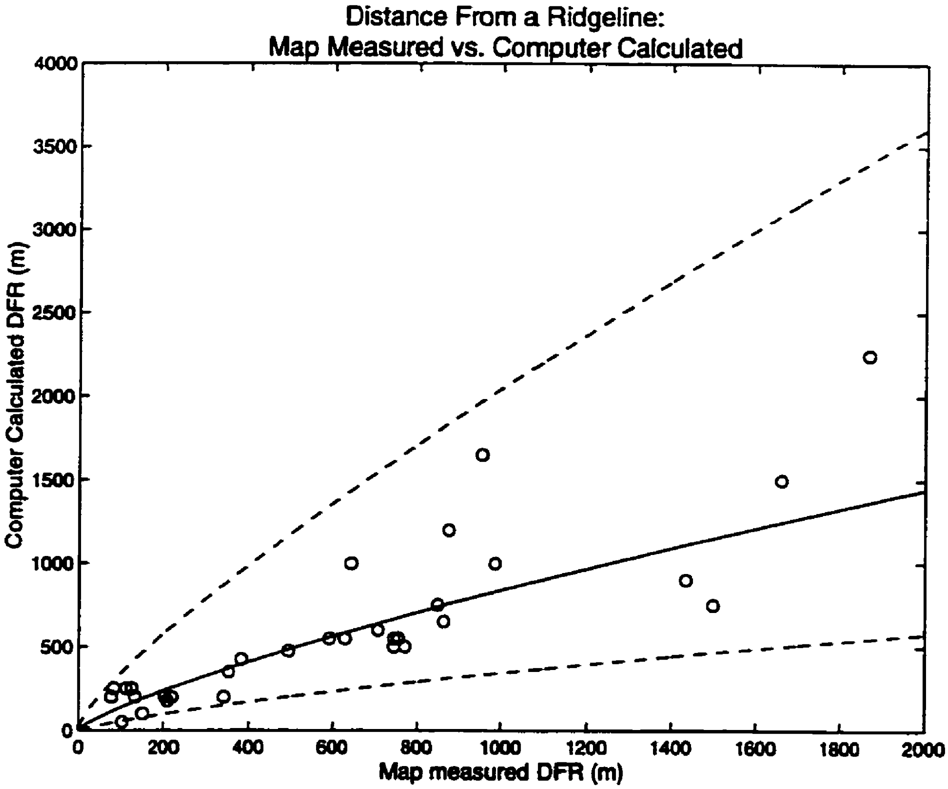


Figure 3.6: This scatter plot shows the correlation between the two methods used to determine DFR for 31 ground sampled forest stands. For the manually calculated method, 1:50000 topographic maps were interpreted and basin ridgelines were traced out. The distance was measured from forest stand to ridge by cutting a path perpendicularly through contours. Included on the plot is the loglinear regression line and error bars (dotted lines) based on a 95% confidence level.

Ideally, there would be a one-to-one relationship between hand measured distances and computer calculated distances such that the slope of a fitted line would equal one. It is clear that the relationship breaks down and the points become more spread out. Why does this happen? In some cases, very large basins can usually be subdivided into smaller basins (*sub-basins*) and so on. The DEM spatial resolution will ultimately limit the divisibility of basins (Goodchild and Mark, 1987). However in this case, the threshold parameter used in the watershed seeding program is used to globally define the scale at which a basin is defined. Increasing the threshold will increase the smallest size of a basin defined. The *actual* closest ridge to a given pixel will, therefore, not necessarily coincide with the closest ridge defined by the computer algorithms. The manual measurements taken from topographic maps were aimed at finding the closest ridge to a stand. Situations arose in which computer defined basins were larger than the manually (visually interpreted) defined basins and this resulted in a drastic difference between the two measurements. In a few cases, the difference in distance exceeded 500m.

It was clear that there was a need to account for these discrepancies. This need was met by fitting a model to the collected data points. In order to fit a regression line to these data points, both measured and calculated distances were natural log transformed. After plotting the transformed data, the relationship became tighter and the spread of points became more even. Linear regression was then performed on the data and 95% confidence level error bars were determined. Figure 3.6 shows the model prediction line (solid line) and the error bars (dashed lines) after undoing the log-log transformation. It was evident that the computer calculated distance usually overestimated the map measured distance such that the slope of the prediction line was less than one. For this reason, the regression model was used as a *correction* function on the computer calculated distances as well as a mechanism for estimating the uncertainty associated with the new estimate. A PCI program was written to apply the model to the original DFR image. The result was a corrected image of DFR. Another PCI program was written to produce upper and lower bound images of DFR based on the 95% confidence interval of the correction model.

3.4 Discussion

The algorithm developed in this chapter is conceptually very simple; yet it serves a specific purpose by providing necessary data for the biodiversity model. It is not surprising that the literature on algorithm development for this kind of use is very limited. It is not the intention of this work to critique the biodiversity prediction model being used here but clearly, alternative variables could be suggested. A simple model for which input parameters can be easily obtained is usually a very desirable model (given that it is still able to produce reasonable results). In this sense, the simplicity of the distance from ridgeline variable is an asset of the biodiversity prediction model. However, in §1.2.3 it was suggested that *distance from a ridgeline* was a surrogate variable for some very important underlying earth processes; namely the movement and concentration of moisture and nutrients. Furthermore, there is a reasonable pool of literature to draw from which deals with characterizing the moisture status of a given point in a landscape. For instance, the use of a *topographic index* was implemented in TOPMODEL and was developed by Beven and Kirkby (1979). The index is calculated as: $\kappa = a/\tan\beta$, where a is the area draining through a point from upslope and $\tan\beta$ is the local slope angle. The index can be calculated for each cell of a DEM. High values of the index indicate areas that will saturate first (also see Beven (1997) and Kirkby (1007)). This index could offer an improvement over distance from ridgeline because it offers a quantitative measure of moisture availability at any given point in a drainage basin.

Chapter 4

Mapping Time-Since-Fire

Section 1.2 introduced the role of disturbance processes in shaping biodiversity. Since this study site has been limited to land within the PANP boundary, the primary disturbance of interest is fire. Of particular interest is the mosaic of forest fire ages over the landscape. A map of time-since-fire already existed from previous research which will be discussed shortly. Recall that part of the primary objective of this study was to account for uncertainty in the biodiversity estimates. It was therefore, necessary to develop a mechanism for estimating uncertainty in the time-since-fire map. This chapter will discuss the methodology used to perform this uncertainty assessment on the time-since-fire map. The results section (§4.3) will offer a discussion on how these uncertainty measurements will be used in the final model.

4.1 Problem Definition

Weir (1996) produced a map of time-since-fire (TSF) polygons for PANP. This was accomplished by first identifying preliminary fire boundaries on 1:12 500 scale air-photos. The initial results yielded a map of 3168 polygons of approximately 2 to 5 ha in area. Extensive field reconnaissance was then used to check the validity of the photo-interpreted fire boundaries. Boundaries were adjusted where necessary. It was then necessary to determine a fire date for each polygon. The dates were determined from field evidence from fire scarred trees, remnant trees (unscarred fire survivors) and forest canopy ages. The sample data included 520 disks from scarred trees, 400 increment cores from remnant trees and 15 000 from canopy trees. Adjacent polygons with the same time-since-fire were merged and a final 1:50 000 scale map of 1249 polygons resulted. A more detailed discussion of the map derivation is offered by Weir (1996).

Through personal communication with Weir (1996), uncertainty estimates for polygon fire ages were obtained and are summarized in Table 4.1. Where a range of uncertainty was specified, the upper limit of the range is used in order to be conservative.

Fire age (yrs)	Estimated Uncertainty (yrs)
0 - 149	± 1
150 - 200	± 1 - 5
> 200	± 15 - 20

Table 4.1: Expert opinion estimates of time-since-fire measurement uncertainty.

In addition to the measurement uncertainty associated with the fire ages, there is also spatial uncertainty present in the location of the polygon boundaries. This spatial uncertainty manifests itself in the form of additional uncertainty in the ages of the polygons. The time-since-fire variable presents us with a slight paradox which does no allow uncertainty to be represented using traditional methods. Taylor (1982) discusses various methods such as using a symmetric interval centered on a measured best estimate of the variable in question. Alternatively, a variable could be treated as random such that it is interpreted as the mean (best estimate) of a normal distribution. The distribution standard deviation and a confidence interval could be used to access the uncertainty in the mean (or best estimate) of the variable. However, from the preceding discussion, it will be made clear that these methods cannot be used.

Consider first that time-since-fire is a continuous variable that can be any value greater than or equal to zero. The paradox, however, is that it must be treated as *ordinal* data. That is, the fire ages within each polygon are essentially ranked, discrete attributes. To illustrate, consider a point on the *exact* boundary between two adjacent polygons of ages 10 and 205 years. This point **must** take on the fire age of one of the two polygons (either 10 or 205 years) and cannot take on an intermediate value. If there is any uncertainty in the location of the fire polygon boundaries, there must be uncertainty attached to the ages assigned to points which are close to or on these boundaries. The magnitude of this uncertainty will be expected to increase with an increase in the magnitude of the difference in adjacent polygon fire ages. For instance, the border point between a very old TSF polygon neighbouring a very young one will have a high age uncertainty relative to a border point between two polygons of similar age.

A question that must still be answered is *what constitutes a boundary point?* As one follows a straight line transect from a shared polygon boundary to the center of the polygon, the expected uncertainty will decrease such that the uncertainty estimate will be partly a function of distance. However, this uncertainty-distance relationship is not known so a conservative *worst case* approach to uncertainty assessment must be adopted. At some point along this transect (still moving toward the center), there will be a very high probability that the TSF is correct. It is here that we assume that the spatial uncertainty in the polygon border no longer has an effect on the estimated TSF. The distance from the boundary to this point along the transect will define the width of a buffer zone within the polygon as shown in Figure 4.1. The buffer zones represent areas in which the amount of uncertainty is in question. In the cross-section view, this is shown with the rectangle labeled as the *zone of unknown uncertainty.* Since we do not know what the uncertainty is at specific points within this zone, the uncertainty interval is taken as the *lower* and *upper limits* regardless of distance from the boundary.

If we consider the point near the shared polygon boundary in Figure 4.1, the age *could* be as old as 225 years or as young as 9 years. This gives rise to asymmetric uncertainty intervals such that a point labeled as 205 could be +20 or -196 years uncertain:

$$205 - 196 \text{ years} = 9 \text{ years} = \textit{lower limit}$$

$$205 + 20 \text{ years} = 225 \text{ years} = \textit{upper limit}$$

using this conservative approach, only an estimate of the width of the buffer zone is needed. The discussion turns now to the implementation of these ideas on the TSF map.

4.2 Methods

From the preceding discussion in §4.1, two ideas can be identified in terms of approaches to uncertainty assessment:

1. For areas inside a polygon in which we are reasonably sure that the TSF is correct (areas with attribute uncertainty but **not** spatial uncertainty), the un-

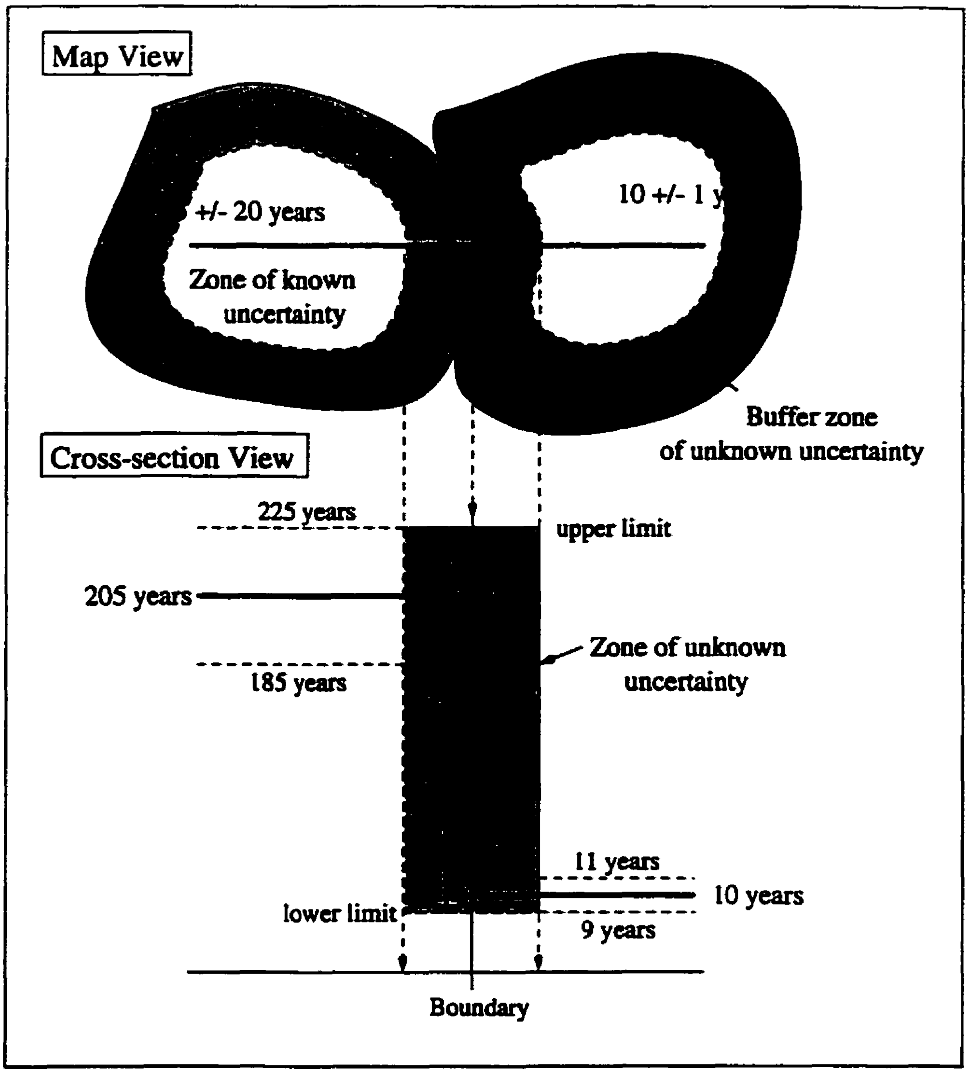


Figure 4.1: Illustrated here is the idea that the location of a boundary line between time-since-fire polygons is *fuzzy*. The inner polygons represent areas for which the uncertainty can be reasonably estimated. The shaded buffer zones within the perimeter of the polygons represent areas for which uncertainty cannot be accurately defined. There is some unknown probability that the point within the 205 year polygon should actually belong within the 10 year polygon which as a result, manifests itself in a wide range of uncertainty as shown by the *upper* and *lower limits*.

certainty can be represented with a traditional interval based on the attribute uncertainties from Table 4.1. For example, a TSF of 170 years can be represented by 170 ± 5 years or alternatively by a lower and upper limit, (165, 175) respectively.

2. For the buffer zones (areas with attribute **and** spatial uncertainty), the uncertainty can be represented by a lower and upper limit. These limits are defined by the youngest and oldest possible ages of any neighbouring polygons to which a point might belong. This is illustrated by the example from §4.1. Notice that these limits include the known attribute uncertainty from Table 4.1. In the event that more than two polygons intersect, only the youngest and oldest TSFs will be of interest in defining the uncertainty bounds.

These two ideas are the basis for estimating and mapping the uncertainty in the time-since-fire map and are addressed in the development of a computer program called TSF.

4.2.1 Computer Algorithm Development

The TSF program runs in a PCI (1997) EASI command line environment. It requires two input images: (1) an image of TSF polygons and (2) an image of TSF uncertainty as illustrated with Figure 4.2. The preparation of these images will be discussed in the next section.

The user must specify a global buffer zone size (in pixels) to be applied to each TSF polygon. This buffer size is multiplied by 2 and then 1 is added to it to give a search window size. This search window is passed over the time-since-fire image pixel by pixel. At each pixel, it is determined whether or not the window is homogeneous or whether it contains lake pixels. Lake pixels are identified by their value of zero. The boundary between lake polygons and fire polygons is not subject to the same spatial uncertainty that the boundary between two neighbouring fire polygons share. It is assumed that these boundaries can been mapped with reasonable accuracy. If the window is homogeneous (lake pixels excepted), the central pixel value in the window is written to an output image of time-since-fire and its associated uncertainty (taken from the input uncertainty image) is written to a new uncertainty image. However, if it is not homogeneous, an output value of -1 is written to both the time-since-fire and uncertainty images. The -1 values indicate buffer zone pixels which will be located at the outskirts of polygons. The two output images address first requirement for uncertainty estimation (from §4.2).

For each buffer zone pixel (-1 value), the second uncertainty estimation procedure must be addressed. This is accomplished by determining the oldest and youngest fire age (from the input TSF image) in the search window. The uncertainty associated with the oldest pixel is added to the TSF value to produce an *upper limit*. The uncertainty associated with the youngest pixel (lake pixels excepted) is subtracted from the TSF value to produce a *lower limit*. These were both written to separate output images. On these images, all non-buffer zone pixels were assigned a -1 value. These two output images address the second requirement for uncertainty estimation (from §4.2).

4.2.2 Map Creation

The time-since-fire map obtained from the Department of Biology at The University of Calgary, was a pixel map of polygons with each cell having a TSF value. Lakes had been logically assigned a value of zero. Four sources of spatial uncertainty in the map were identified:

1. Since fire polygons were hand drawn from 1:12 500 scale airphotos, a 1 mm ambiguity in the tracing of the polygon borders would result in a spatial uncertainty of ± 12.5 m. Through personal communication with Weir (1996), it was established that 1 mm (on the 1:25,000 paper map) was the greatest amount of spatial uncertainty that could be expected. This quantity translated to 25m on the ground which was equal to almost 1 pixel (of 30 x 30 m size) on the raster map.

2. Uncertainty in digitizing the line work. Digitizing was performed by GAIA Consultants, Calgary, Alberta. GAIA was contacted for this information but their records of the work did not include digitizing error. Consequently, this source of unknown uncertainty was not included in the analysis.

3. The map had been previously derived from a file of vector based polygons which were encoded into pixels of 30 x 30m in size. Immediately, this step introduces spatial uncertainty in the location of all fire polygon boundaries. Although it is intuitive to assign a ±0.5 pixel spatial uncertainty to the location of the rasterized vector (the worst case scenario), it can be shown mathematically that the actual error is only ±0.289 pixels (Chapman, 1988). This source of error must be included in the specified buffer zone width.

4. In §4.1 it was mentioned that the smallest polygons mapped were between 2 and 5 ha. Since 2 ha is the very smallest polygon that will be found on the map, there will be larger polygons on the map that *may* have smaller polygons wrapped into them that have not been delineated. These small polygons may be located within larger polygons or on their borders. In either case, their frequency of occurrence and locations cannot be determined from the digital map and thus, we are left with a source of spatial uncertainty for which we cannot account.

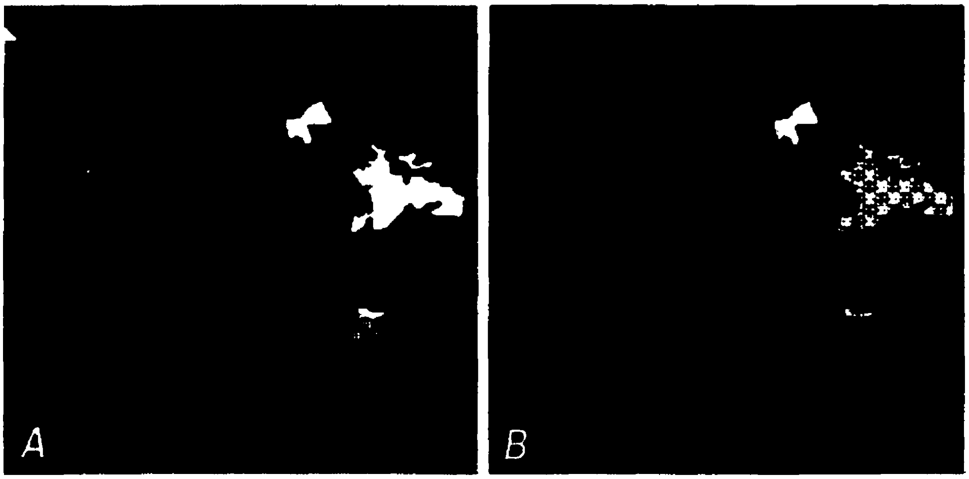


Figure 4.2: These are small subsamples of the input images. Image *A* is the time-since-fire and *B* is the starting uncertainty estimate image. In both images, the lakes are represented by the black polygons.

Given this map and estimates of the uncertainty associated with the ranges of fire ages, the first step required producing a map of uncertainty. A simple PCI EASI script was written to do this. The script essentially treats Table 4.1 as a look-up table. The value of each pixel in the TSF image is determined and the appropriate uncertainty value is written to a separate image in the corresponding pixel location. Next, these two images were used as inputs into the TSF computer program and the output maps were produced.

4.3 Results

A sample of the output imagery from the TSF program is shown in Figure 4.3. Older polygons are represented by the lighter shades in A. Likewise, higher uncertainty is represented by the lighter shades in B. As one would expect, there are even width buffer zones surrounding all polygons which represent the uncertain boundary locations. On these images, the buffer zones are a total of 4 pixels wide because each polygon has its own buffer zone of 2 pixels in width. Images C and D illustrate the uncertainty assessment for the buffer zones for the upper and lower limits respectively. These buffers provide uncertainty coverage for the undefined areas from A and B. Notice that buffer zones do not appear around lake polygons.

At this point, it is not entirely obvious how these intermediate results will be used in the final biodiversity model. Essentially, we must tackle this problem using two approaches: the first will use the information from images A and B and the second will use images C and D, respectively. Figure 4.4 illustrates the relevance of the two approaches in terms of their use in the biodiversity model.

The main estimate of biodiversity is a function of time-since-fire, canopy stem density, canopy type and distance from a ridgeline. For this estimate, the entire image of TSF is used for input into the model. Uncertainty estimation is treated in a slightly different manner. Referring to Figure 4.4, Pixel 1 falls within the *inner fire*

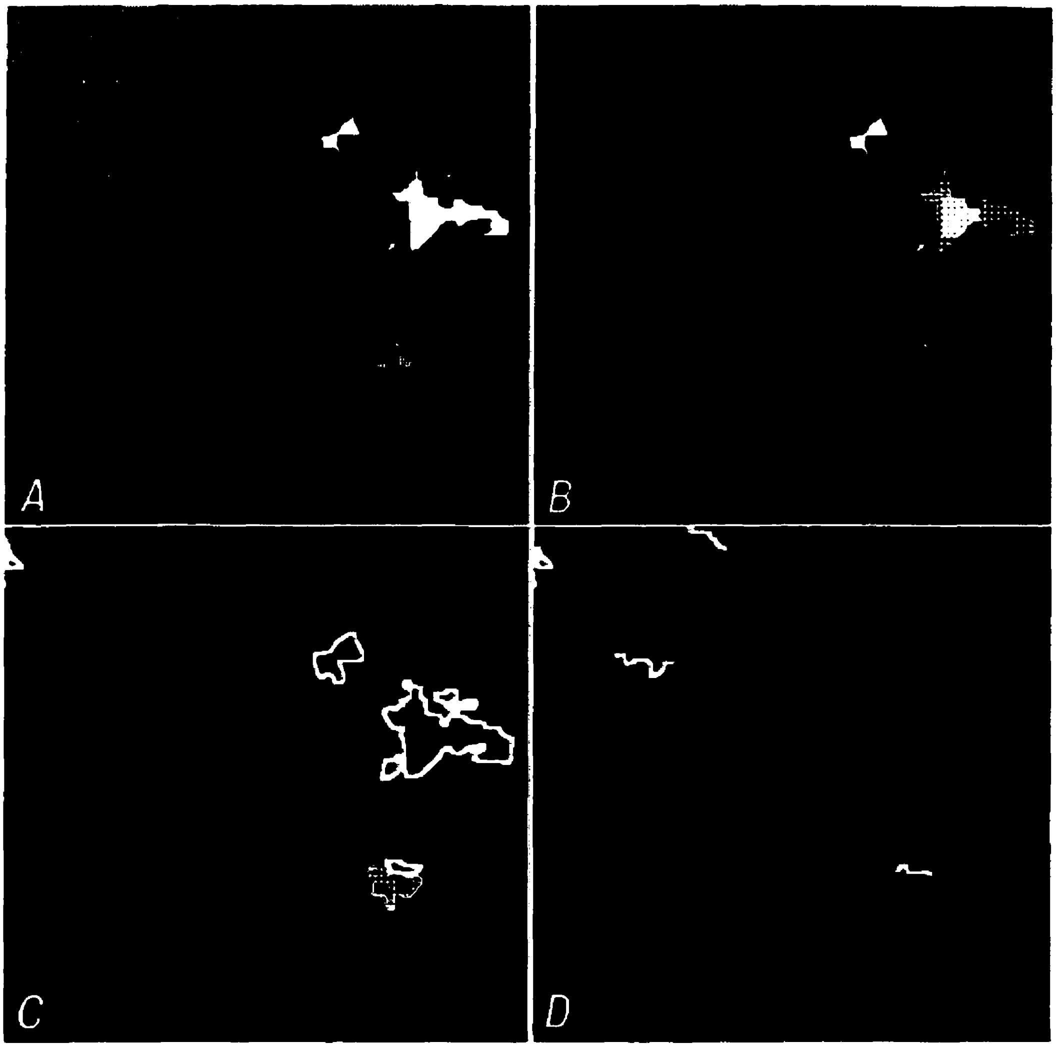


Figure 4.3: These small subsamples of the final images illustrate the output from the TSF program. *A* is the time-since-fire estimate. *B* is the uncertainty estimate. The black rings around the polygons in these images are the buffer zones. *C* and *D* are the upper and lower limits, respectively, of uncertainty in the buffer zones.

polygon. To estimate the upper and lower bounds of uncertainty, images A and B are used. For the upper bound, B is added to A and for the lower bound, B is subtracted from A. Pixel 2 falls outside of the inner polygon and is considered to be within the *buffer zone* of unknown uncertainty. Its upper and lower bounds are expressed by the imagery of C and D respectively. When calculating an upper or lower bound on biodiversity, the location of the image pixel (for which the calculation is being performed) is essential. If an upper bound on biodiversity was being calculated, the upper bound images for canopy stem density, distance from a ridgeline would be input into the model as well as a canopy type. The choice of imagery for time-since-fire will depend on the pixel location. For Pixel 2, for example, image C would be used to calculate the upper bound of biodiversity. A detailed discussion of how the data inputs are handled in the biodiversity model will be offered in Chapter 7.

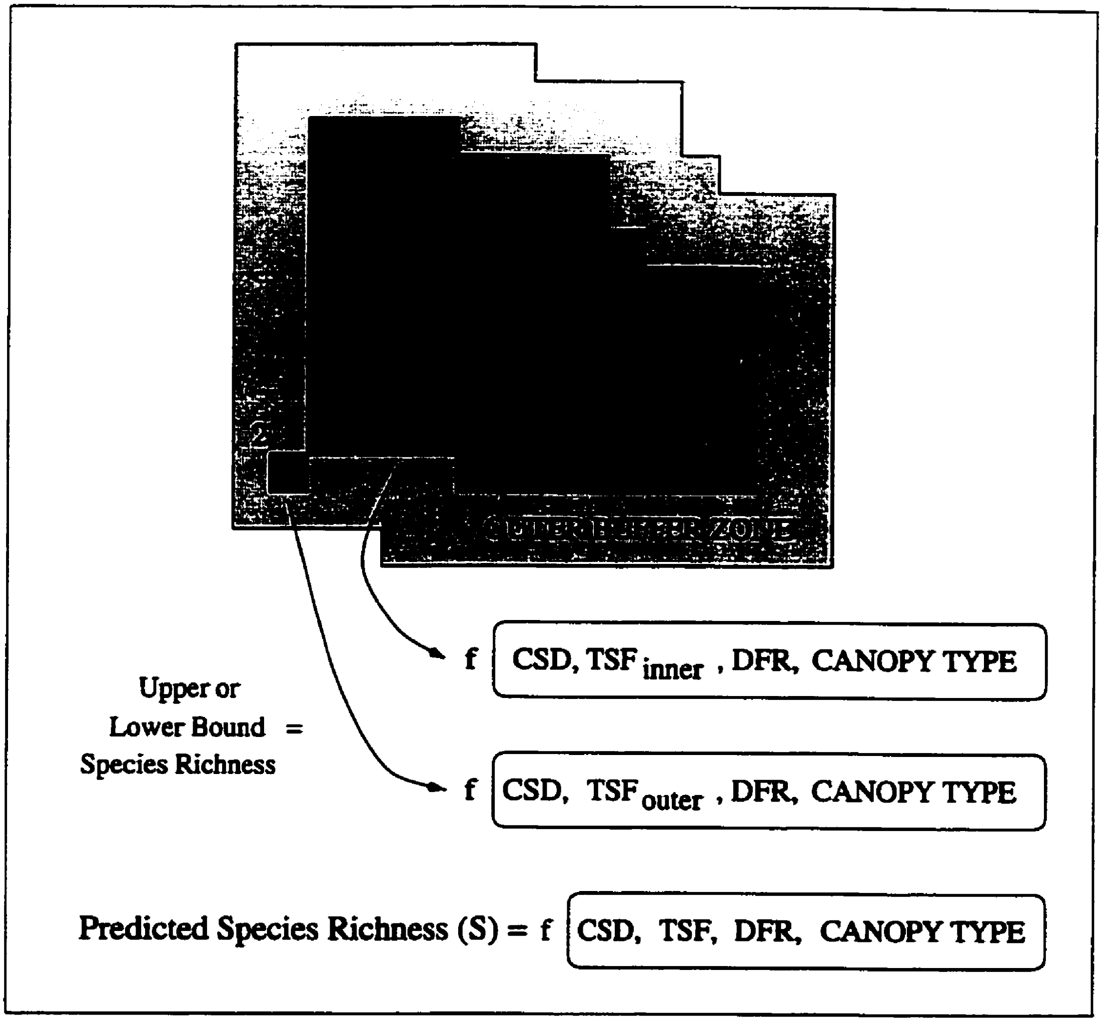

2
OUTER BUFFER ZONE
Upper or Lower Bound = Species Richness
f CSD, TSF inner , DFR, CANOPY TYPE
f CSD, TSF outer , DFR, CANOPY TYPE
Predicted Species Richness (S) = f CSD, TSF, DFR, CANOPY TYPE

Figure 4.4: The upper and lower bounds on predicted biodiversity depend on pixel location on the image. In this example, Pixel 1 uses symmetric uncertainty intervals but pixel 2 uses independent upper and lower bound image of TSF for input into the biodiversity model.

Chapter 5

Mapping Canopy Type

This chapter addresses the need to map dominant forest canopy types for the study area which is an important component of the biological diversity prediction model. As with distance from a ridgeline and time-since-fire, the uncertainty in our canopy type prediction must also be determined for propagation through the model. This chapter describes the steps taken to accomplish these tasks using the available data.

5.1 Problem Definition

Given the nature of the solution, two problems were encountered:

1. There were poor classification results from the initial data classifications. The results were investigated with an in-depth examination of the training site data in an attempt to improve the accuracy of canopy determination.

2. As a byproduct of the images classifications which resulted from this work, the question arose: how can the outputs of two or more classifications be combined in order to improve the overall results?

One method that has received recent attention is the use of the Dempster-Shafer Theory of Evidence (Dempster, 1967; Shafer, 1976) in the context of spatial data fusion and classification (referred to herein as evidential reasoning (ER)). Section

5.3 pursues the use of ER data fusion and how it was implemented in this work to solve the second problem.

5.2 Canopy Type Classification Methodology

5.2.1 Training and Testing Site Preparation

Using the GPS data collected from each ground sample site, a vector database of all site transects was built in ARC/INFO (1997). An attribute table of site IDs was also constructed and linked to the vectors. The database was then imported into a PCI (1997) database. Using the classification tools of the Imageworks module of PCI, the pixels that fell directly underneath the transect lines of each site were encoded into an image plane (as training/testing pixels for that site). Approximately 150 individual sites were defined this way.

Each site was then labeled with its appropriate forest type category. The forest classes used by Chipman (1999) are jack pine (JP), black spruce (BS), trembling aspen (TA) and a mixed class of trembling aspen, white spruce and balsam fir (MIX). A description of how these classes were chosen is offered in Appendix C. For each of the four categories, the sites that fell within each were randomly divided into two halves to provide two sets of sites – one for training and one for testing. The image-encoded sites were then converted to bitmaps and aggregated into their respective groups. The result was 4 testing site and 4 training site datasets; a testing and training set for each class. In addition to the forest classes, water (WAT), anthropogenic (ANT), and wetland (WET) classes were also defined (each with a training and testing set of data). The anthropogenic class includes towns, roads, recent harvests as well as other man-made features. A total of 14 datasets resulted.

5.2.2 Choice of a Classification Algorithm

As a pre-classification step, the Kolmogorov-Smirnov (K-S) test was used to determine if the training site data fit a normal distribution. The K-S test was especially suited to this data set because it can be used with very small sample sizes (Lilliefors, 1997). The test compares the cumulative density function (CDF) of an input dataset to that of a hypothetical CDF parameterized with the mean and standard deviation of the input data. The test statistic is calculated by determining the absolute maximum vertical distance between the two distributions. The null hypothesis for the test is that the two distributions (input and hypothetical) are the same. Small probability values indicate that the two distributions are significantly different such that the null hypothesis is rejected. A probability value of 1.0 is likely to result when the distributions are identical – that is, the input data fits a perfectly normal distribution (Siegal and Castellan, 1988; Press et al., 1992).

	JP	BS	TA	MIX
TM1	0.0000	0.6032	0.0002	0.0160
TM2	0.0000	0.0016	0.0000	0.0000
TM3	0.0000	0.3219	0.0000	0.0014
TM4	0.2943	0.7046	0.7622	0.0031
TM5	0.0001	0.2331	0.5155	0.0026
TM6	0.0000	0.0000	0.0000	0.0000
TM7	0.0000	0.0288	0.0000	0.0002

Table 5.1: Probability values from the Kolmogorov-Smirnov test for normality. Tested were the training data for the four canopy types of interest for each of the seven TM bands.

In all but a few cases, the probability values from the K-S test were close to 0 meaning that overall, the training data were not normally distributed (Table 5.1). This preliminary assessment of the data immediately eliminated the option of using the maximum likelihood parametric image classifier (MLC).

The MLC requires that the input data for training sites fit a Gaussian distribution with known means and covariances. In addition, the classifier requires that the sample sizes of the training data be quite large so that the covariance parameters can be reasonably estimated for multidimensional surfaces (Richards, 1993; Schowengerdt, 1997). The training data failed to meet both of these requirements. Instead, the ***k-nearest neighbours*** (kNN) classifier was chosen. The kNN classifier provides a non-parametric supervised approach to data classification that has been shown to provide pixel assignment accuracies similar to parametric methods such as the MLC (Hardin, 1994). Since it is non-parametric, the kNN classifier is also more suited toward the use of small training data sample sizes.

5.2.3 Computer Algorithm Development

Since the propagation of uncertainty throughout all components of the biodiversity prediction model is an integral part of this work, it was necessary to devise a method of uncertainty estimation for the classification. The logical choice of an uncertainty measure was the use of class assignment probabilities for each pixel. For instance, if a pixel had a 0.5 probability of being labeled as jack pine and a 0.48 probability of being labeled as black spruce, the uncertainty in the final decision to label the class as jack pine would be very high since the probabilities are so similar. The following paragraphs detail the derivation of these probabilities.

Although a kNN classifier already existed within PCI, the algorithm did not allow for the output of probability values for pixel assignments into each possible class. This need was addressed in the development of an alternative kNN algorithm for use in the PCI (1997) EASI command line environment. The user is able to specify an input channel containing training sites, channels to be classified, output channels for the classified image and probability values, an integer value for k and a maximum number of training samples per class. Training data is first extracted under each training site and stored in a two dimensional matrix. For an unclassified image pixel, the Euclidean distance (in multispectral space) is calculated to each training pixel. A list of the k shortest distances is then examined for the most frequently occurring training class which is chosen as the winning pixel label. If there is a tie between two or more potential winners, the class with the training pixel with the shortest distance to the unclassified pixel is chosen.

The number of *votes* (out of the total K-nearest neighbours) that each class received was also used to calculate an assignment probability for each class by simply dividing the frequency of votes for each class by the total number of nearest neighbours, k. Using the example from Figure 5.1, the probabilities for jack pine, black spruce and trembling aspen would be 0.6, 0.3 and 0.1, respectively. Logically, the winning class has the highest probability. These values are then output into images; one for each class. The use of these uncertainty images will be addressed in Chapter 7 which addresses the implementation of the biodiversity model.

These steps are repeated for each image pixel. An obvious drawback to the kNN classifier is the heavy computational requirement which results in long execution time. The imagery in this project measures 2611 by 3785 pixels. To classify 7 TM

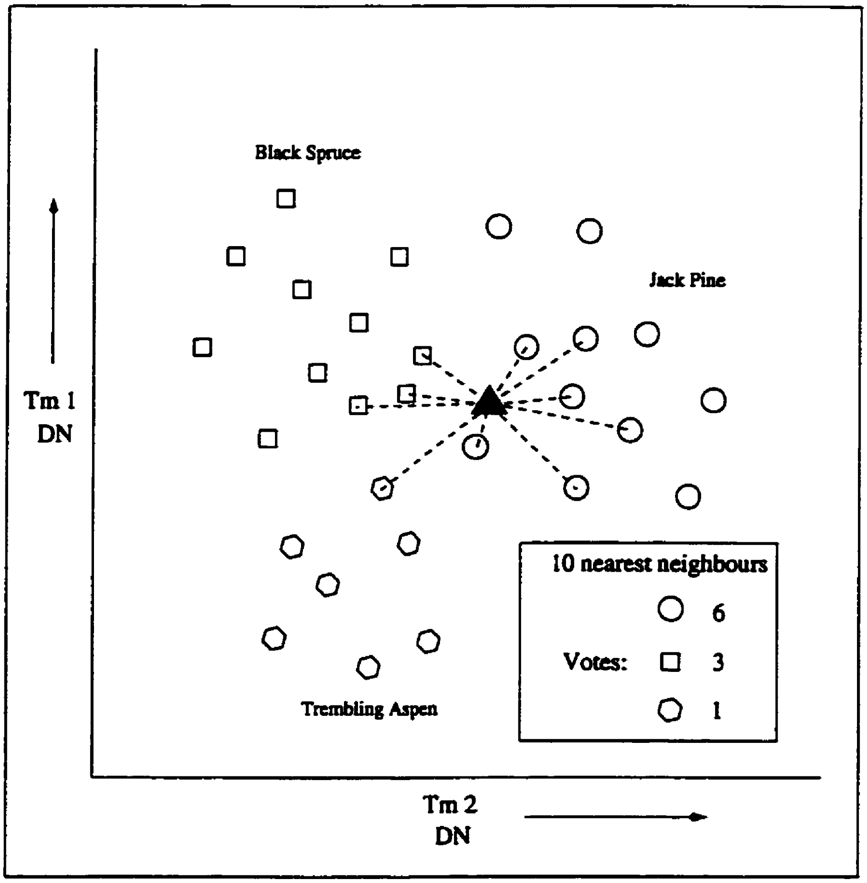


Figure 5.1: Example of kNN-based class assignment for a single pixel (black triangle) in a two-dimensional feature space for 10 nearest neighbours. The Euclidean distance is measured to each training pixel but only the closest 10 distances are kept. The jack pine class would be chosen as the winner based on the number of *votes* each class received.

bands with 10 nearest neighbours and 1000 training pixels (in total), the algorithm would need to calculate approximately 6.9 × 10^{11} distances. In some cases, these execution times make the kNN classifier an impractical choice. However, for this project, a MATLAB (1997) program was available which creates a kNN classification confusion matrix using the training and testing datasets without classifying the entire image. The use of this program tremendously cut down preliminary processing time. The program was written by Dr. Michael Collins of the Department of Geomatics Engineering at the University of Calgary, Alberta.

5.2.4 Classification

The June and August, 1996 images were both classified with the Matlab program to yield preliminary classification accuracy results. The initial kNN classification using the June, 1996 TM imagery yielded the best overall results for the four forest type classes. However, in both classifications, the black spruce and mixed classes are heavily confused with the other classes resulting in very poor classification accuracy for those classes. The June results are shown in Table 5.2 as well as Appendix B for comparison with the August results.

From these results, it was thought that the small sample size of training and testing sites might be negatively influencing the classification. Notice that for the small pixel totals (sample sizes) in Table 5.2, the classification accuracies are quite low. For larger samples, the accuracy tends to increase.

To test this hypothesis and hopefully improve the accuracy of the classification, each of the 100 ground sampling sites was examined with the intent of expanding their area on the imagery (ie. increase the number of pixels per sample site). Forest

class	JP	BS	TA	MIX	WAT	ANT	WET	Percent
JP	53	0	12	8	0	4	8	62.35
BS	8	22	1	7	0	0	0	57.89
TA	9	4	151	14	0	2	0	83.89
MIX	12	30	31	31	0	6	1	27.93
WAT	0	0	0	0	199	0	0	100.00
ANT	1	0	1	1	0	163	33	81.91
WET	0	0	0	0	0	1	199	99.50
TOTAL	83	56	196	61	205	170	241	

Table 5.2: kNN Classification results for testing sites using 7 bands of the June, 1996 TM imagery. The classification accuracies of the black spruce and mixed class are unacceptably low.

inventory maps, 1:50 000 scale topographic maps and the raw TM imagery were used to determine if a site could be expanded. For many sites, the forest type patterns on the inventory maps matched with the patterns seen on the TM imagery. If a site was clearly contained within an area that was homogeneous with respect to forest type, its extent was expanded within the area by a number of concentric rings around the original site. However, expansion only occurred within the homogeneous areas. The largest individual training site included 71 pixels; about seven times the original number of pixels for that site. In some cases sites could not be safely expanded.

The new sites were then aggregated as before and classification was performed again on the June and August images. For the black spruce and trembling aspen classes, the results were even poorer than those offered by the unexpanded sites of which the results were poor to begin with. Table 5.3 shows a slight improvement in the jack pine and mixed class. The results do not support the hypothesis that the small sample sizes of the training sites were negatively influencing the performance of the kNN classifier.

class	JP	BS	TA	MIX	WAT	ANT	WET	Percent	Change
JP	183	5	26	10	0	5	35	69.32	+6.97
BS	17	64	0	51	0	0	0	48.48	-9.41
TA	47	8	413	27	0	0	0	83.43	-0.46
MIX	26	27	44	46	0	6	1	30.67	+2.74
WAT	0	0	0	0	199	0	0	100.0	0
ANT	3	0	2	1	0	160	33	80.40	-1.51
WET	0	0	0	0	0	1	199	99.50	0

Table 5.3: kNN Classification results for expanded testing sites using 7 bands of the June, 1996 TM imagery. The classification accuracies of the black spruce and mixed class are unacceptably low. The last column shows the change in accuracy from the unexpanded sites classification results.

The classifications were not limited to electro-optical TM data. Two other classifications were performed. The first used four bands of SIR-C SAR data; LHH, LHV, CHH and CHV. The second used four DEM derived, geomorphometric variables including elevation, slope, aspect and distance from a ridgeline. The accuracies resulting from the geomorphometric variable classification offered no improvement over the June TM classification and in some instances were considerably worse. The SAR data showed only slight improvement in the black spruce class but a considerable decline in accuracy in all other forest classes (Table 5.4).

The poor results of all classifications performed were puzzling and prompted further investigation as to why the classification accuracies were not higher. Examination of the canopy species compositions (Appendix A) of the ground sampled sites provided a strong clue toward an explanation. For only a few forest stands, the species compositions measured as *pure*. For instance, Site 48 was recorded as

class	JP	BS	TA	MIX	WAT	ANT	WET	Percent	Change
JP	108	18	83	4	0	18	33	40.91	-28.41
BS	33	68	17	7	0	2	5	51.52	+3.04
TA	132	14	289	54	0	5	1	58.38	-25.05
MIX	21	30	60	30	0	9	0	20.00	-10.67
WAT	0	0	0	0	200	0	0	100.00	0
ANT	28	6	13	4	1	120	27	60.30	-20.1
WET	0	0	0	0	0	0	199	100.00	+0.5

Table 5.4: kNN Classification results for expanded testing sites using SAR LHH, LHV, CHH and CHV polarizations. The classification accuracies of the jack pine and mixed classes in particular, are unacceptably low. The last column shows the change in accuracy from the expanded June TM sites classification results.

100% jack pine. However, most sites were extremely mixed and in some cases, there is no dominant canopy type even though the site was being used to represent one in the classification. For instance, Site 81 was classed as a jack pine site for the classification training yet its composition is 53% jack pine and 47% trembling aspen. Further examination of Appendix A reveals that there are many sites for which a dominant canopy cannot be reasonably distinguished. It seemed likely that the natural mixture of species in the training sites was the cause of great confusion in the classifications. For this reason, a further in-depth investigation of the imagery data extracted from the ground sampled sites (training and testing data) was undertaken.

To accommodate this investigation, a new program was written based on the kNN classifier methodology from §5.2.3. Its purpose was to test the validity of the training and testing site data by treating each site, one at a time, as an unclassified image. The program iterates through all sites and classifies each one using all other sites as training data. A table of the number of pixels falling into each class for each site is output. It is similar to a classification confusion matrix except that the pixels being classified are all *known* to be of one class. For example, at one iteration, the program might attempt to classify a black spruce site using all other sites as training data. Since we know what the site should be classified as, we calculate the percent accuracy in classifying the site as black spruce.

Initially, all sites in each class were included in the kNN site testing. The resulting output tables are presented in Appendix B. The resulting classification accuracies were then used to identify potential problem sites. The double lined columns indicate the class of interest in each table. In each row, the top number indicates the number of pixels that fell into that class and the bottom number indicates the same in a percentage of the total number pixels for that site. These percentages can be used as classification accuracies since each site is considered homogeneous. For instance, in Table B the jack pine site (etrs3) was classified with 97.8 percent (or 45/46 pixels) accuracy. Moving down the column, it is easy to identify potential problem sites with low classification accuracies. Overall, the individual accuracies for the jack pine and trembling aspen site tests reflect the reasonable, overall kNN classification test accuracies for each class (Table 5.3). Most of the black spruce and mixed sites performed extremely poorly which reflected their low classification accuracies.

The individual sites were examined more closely in terms of their age (time-since-fire), canopy stem density and under- and overstory species composition (as determined from ground sampling estimates). The main goal here was to explain the level of classification accuracy of the sites. First, the canopy composition (Appendix A) was examined to see if the canopy species distributions for a site could explain the deviations. If this failed, the canopy stem density was examined. If the stand was of low density, the understory species composition was also examined. If the deviation could be qualitatively explained within reason, it was removed from the training/testing dataset. Particular attention was paid to the sites with low classification accuracy. In general, the heuristic rules used to determine if a site should be kept or removed were applied equally throughout the set of sites for each class. For instance, in some cases sites with high classification accuracy were also questioned and removed because canopy compositions did not indicate that a high level of accuracy should be expected. Since the black spruce class consisted of very few sites (and therefore few pixels), none were removed. Tables 5.5 though 5.7 address the justification for removing training sites from the classification.

Class ID	Justification for removal of **jack pine** sites
etrs7	Canopy composition includes 50% trembling aspen which coincides closely with kNN test results
etrs9	Canopy composition is very mixed with high proportions of white spruce and trembling aspen.
etrs76	Canopy composition includes less than 30% jack pine which indicates a non-dominant jack pine canopy
etrs81	Canopy composition includes 47% trembling aspen which coincides closely with kNN test results
etrs85	According to the 1:50 000 topographic maps, the site is located in a relatively wet area (in terms of ground moisture). This observation coincides with 97% of the pixels being classified as wetland

Table 5.5: Training sites removed for the jack pine class.

Once the identified problem sites had been removed, the classification was performed again on the June TM data using the Matlab test program. The results are presented in Table 5.8. Note that three of the four forest classes showed an increase in classification accuracy. In particular, there was substantial improvement in the

Class ID	Justification for removal of **trembling aspen** sites
etrs5	Canopy composition includes less than 12% trembling aspen which indicates a non-dominant trembling aspen canopy
etrs13	Although this site was classified reasonably well at 64% accuracy, its canopy composition consists of 53% white spruce
etrs14	Although this site was classified well at 86% accuracy, its canopy composition consists of 46% white spruce
etrs41 etrs42	These sites are very mixed in terms of canopy composition
etrs46	Although this site was classified well at 100% accuracy, its canopy composition consists of less than 6% trembling aspen
etrs54	Although this site was classified well at 70% accuracy, its canopy composition is very mixed
etrs57	Although this site was classified well at 100% accuracy, its canopy composition is very mixed with only 36% trembling aspen
etrs60	Although this site was classified well at 71% accuracy, its canopy composition is dominated by 66% white spruce with only 26% trembling aspen
etrs97	Although this site was classified well at 82% accuracy, its canopy composition is very mixed with only 37% trembling aspen

Table 5.6: Training sites removed for the trembling aspen class.

Class ID	Justification for removal of mixed sites
etrs4	Canopy composition includes 52% black spruce which coincides well with classification accuracy. This dominance of black spruce was the reason the site was placed into the black spruce class
etrs15	Site classified at 100% trembling aspen yet the canopy composition indicates only 24%. However, forest inventory maps indicate that trembling aspen is a dominant canopy type
etrs17	Canopy composition includes 54% trembling aspen indicating that the site is not very mixed. This coincides with high trembling aspen in the classification test result
etrs19	Site classified at 97% trembling aspen yet the canopy composition indicates only 24%. However, forest inventory maps indicate that trembling aspen is the canopy species for the area
etrs21	Site classified at 94% trembling aspen yet the canopy composition indicates only 36%. However, forest inventory maps indicate that trembling aspen is the canopy species for the area
etrs30	The site classified at 78% jack pine yet the canopy composition indicates 52% black spruce. The result cannot be explained but neither of these should be represented so heavily within the site indicating that it is a problem site.
etrs56	This site is located at the edge of a lake and on the imagery it appeared spectrally different from the forest. This observation likely explains the misclassification of pixels into the anthropogenic class.
etrs64	The canopy composition consists of 63% white spruce and 37% black spruce indicating that it is should not be in the mixed category.

Table 5.7: Training sites removed for the mixed class.

trembling and mixed classes. Black spruce accuracy declined considerably and was much lower than desired. It is interesting to note that as the mixed class accuracy increased, the black spruce class accuracy declined.

class	JP	BS	TA	MIX	WAT	ANT	WET	Percent	Change
JP	173	2	26	15	0	0	3	79.0	+9.68
BS	14	43	0	75	0	0	0	32.6	-15.88
TA	6	0	335	9	0	0	1	95.4	+11.97
MIX	10	16	13	61	0	0	0	61.0	+30.33
WAT	0	0	0	0	2018	0	0	100.0	0
ANT	1	0	32	0	0	397	50	82.7	+2.3
WET	19	4	27	0	0	117	544	76.5	-23.0

Table 5.8: Summary of kNN results for testing sites using 7 band TM **after** site expansion and selective site removal. The change is relative to the June TM classification before site removal.

class	JP	BS	TA	MIX	WAT	ANT	WET	Percent	Change
JP	67	25	74	4	0	49	0	30.6	-10.3
BS	31	68	6	6	0	15	6	51.5	-0.02
TA	55	25	204	17	0	11	4	64.6	+6.22
MIX	18	37	6	30	0	9	0	30	+10.0
WAT	0	0	0	0	788	0	0	100	0
ANT	35	41	39	11	0	336	17	70.2	+9.9
WET	0	4	1	0	0	113	593	83.4	-16.6

Table 5.9: Summary of kNN results for testing sites using SIR-C SAR LHH, LHV, CHH and CHV polarizations **after** site expansion and selective site removal. The change is relative to the SAR classification before site removal.

Notice that for the SAR classification, the classification accuracies are generally quite low. However, the black spruce accuracy is considerably higher than the TM accuracy for the same class. This result was the motivation for the next section of this chapter in which the evidential reasoning combination algorithm is implemented on the classified produced so far. At this point, the kNN classification algorithm was implemented on both sets of imagery (TM and SAR). The result was two classified images of forest canopy type and two sets of probability imagery of seven images each (one image per class).

5.3 Evidential Reasoning Combination

The following sections introduce *evidential reasoning* and how it can be used to combine classifier outputs. Specifically, the combination of the SAR and TM probability images is also examined.

5.3.1 Introduction to Evidential Reasoning

ER is similar in concept to the boolean algebra based 'overlay' operation found in image processing and GIS packages. However, in the typical overlay, a resulting pixel or polygon is labeled with *hard* decision criteria such as $C{=}A{\cup}B$ or $C{=}A{\cap}B$.

The Theory of Evidence combination methodology is based on the concept of *evidential mass*. Each pixel in each data set is assigned a mass of evidence which represents our *belief* in terms of a probability value that the pixel in question belongs to a certain label class. Since probabilities are being used, the total evidential mass for a pixel must always sum to unity with each component representing a proportion of mass in favour of a certain class label.

To illustrate, consider the following example of a mass distribution function from Richards (1993):

$$m(<A,B,C>) = <0.5, 0.25, 0.25> \tag{5.1}$$

where m is the mass distribution for a given data source and $<A,B,C>$ is the set of possible label classes for a pixel with their respective probabilities of assignment. Notice that the sum of all evidence is 1.0. Recall from §5.2.3 on the kNN computer algorithm development that class assignment probabilities are output for each pixel in the image. These probabilities provide us with a convenient source of *evidential mass* which can be used in evidential reasoning. Also, recall that since they are proportions, they sum to 1.0.

A method was formulated by Dempster (1967) for combining multiple sets of mass distribution functions into a single set. It is referred to as Dempster's Orthogonal Sum and will be elaborated upon in §5.3.4. The method allows for evidential mass to be combined or integrated from different data sources. A decision rule is then applied to the combined set of evidence for each pixel and a class label is determined (see §5.3.5 for a more complete discussion).

5.3.2 Incorporating Uncertainty & Representing Ignorance

A great asset of the Theory of Evidence is that it allows for uncertainty estimates to be included in the total evidential mass for a pixel at each level of the classification. The simplest incorporation of uncertainty into the mass distribution function can be illustrated with an example.

Suppose that one was only 80% confident that the correct pixel label was in fact one of the possible classes *A, B or C.* In other words, what if there were other possibilities that we have not taken into account? By multiplying each element in the set from Equation 5.1 by 0.80, the mass distribution would become:

$$m(< A, B, C, \theta >) =< 0.4, 0.2, 0.2, 0.2 > \tag{5.2}$$

where θ denotes the uncertainty. The uncertainty can be viewed as our ignorance in choosing a comprehensive set of label classes. The higher the uncertainty, the less confidence we have that a given pixel will accurately be represented by the possible labels we have chosen. In this example, there is a probability of 0.2 that none of the possible labels is the correct one.

Again, the kNN classification provides us with a convenient mechanism for determining the uncertainty: From a classified image, we are able to determine the accuracy of each class (see for example, Table 5.8). These accuracies can be used as certainty measurements such that the uncertainty for an individual piece of evidence in the set is equal to $1 - (accuracy\%/100)$. Since we have an accuracy for *each* class, we can be more specific about the quantity of uncertainty associated with each piece of evidence than was illustrated in Equation 5.2. In this case, each element is multiplied by its associated uncertainty. The uncertainty term in the set then becomes equal to $1 - \sum evidence$.

5.3.3 Evidential Measures

Mathematically, the elements of the evidential mass distribution are described with three measures: ***support***, ***plausibility*** and ***evidential interval***. The first, support, is a measure of the minimum amount of evidence that supports the labeling of a pixel as a certain class. The second, plausibility, is a measure of the maximum amount of evidence that supports the labeling of a pixel and is calculated as one minus the total support for all other label possibilities. In the example above, each possible class is assigned a support and a plausibility. The evidential interval is the difference between the support and plausibility. Using the example above for label class A, the notation is given as:

$$s(A) = 0.4 \tag{5.3}$$

$$p(A) = 1 - s(B) - s(C) = 1 - 0.2 - 0.2 = 0.6 \tag{5.4}$$

$$u(A) = p(A) - s(A) = 0.2 \tag{5.5}$$

where s, p and u are the support, plausibility and evidential interval respectively for label class A. A graphical representation is shown in Figure 5.2. The evidential interval can be interpreted as the amount of imprecision in the mass allocated to a certain label class. A very tight or small interval indicates a relatively precise estimate that the amount of mass assigned to a given label class is correct. A relatively wide interval obviously indicates that we cannot precisely determine what the mass allocation should be.

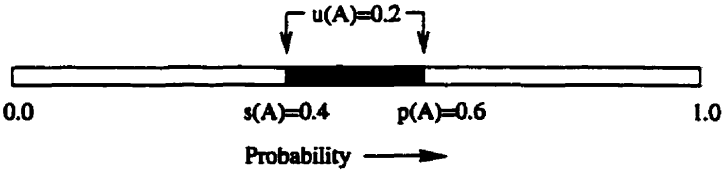


Figure 5.2: A linear depiction of the *evidential interval* as represented by the shaded segment. The evidential interval is a measure of how certain one is about the evidence assigned to the mass function in support of a given pixel labeling.

5.3.4 Combining Evidence

In order to combine classifier outputs, it is necessary to combine the mass of evidence distributions for each source. The evidence for each data source is combined using Dempster's orthogonal sum (Dempster, 1967) which is described mathematically below with an example using two sets of labels, Y and Z from mass distribution functions m_1 and m_2:

$$m_{12}(X) = \kappa^{-1} \sum_{Y \cap Z = X} m_1(Y) m_2(Z) \tag{5.6}$$

where m_{12} represents the combined mass distribution for the sets of labels, Y and Z. The new set of labels is denoted by X. It would be illogical to combine evidence for two different label classes. For example, the mass from one data set in favour of labeling a pixel as black spruce would not be combined with the mass in favour of labeling a pixel as trembling aspen from the other data set. Thus, a null or *empty* set denoted by $\emptyset$ is defined which represents the discarded *contradictory* label classes. The constant, κ restores the total probability mass to 1.0 and is calculated as follows:

$$\kappa = 1 - \sum_{Y \cap Z = \emptyset} m_1(Y) m_2(Z) \tag{5.7}$$

Once a new evidential mass distribution is calculated, it can be combined with another. This process is repeated until all sets of evidence are combined. The order and grouping of mass distribution combinations is irrelevant since the orthogonal sum has commutative and associative properties (Garvey et al., 1981; Moon, 1990; Richards, 1993; Shafer, 1976).

Figure 5.3, adapted from Richards (1993) and Garvey et al. (1981), shows a

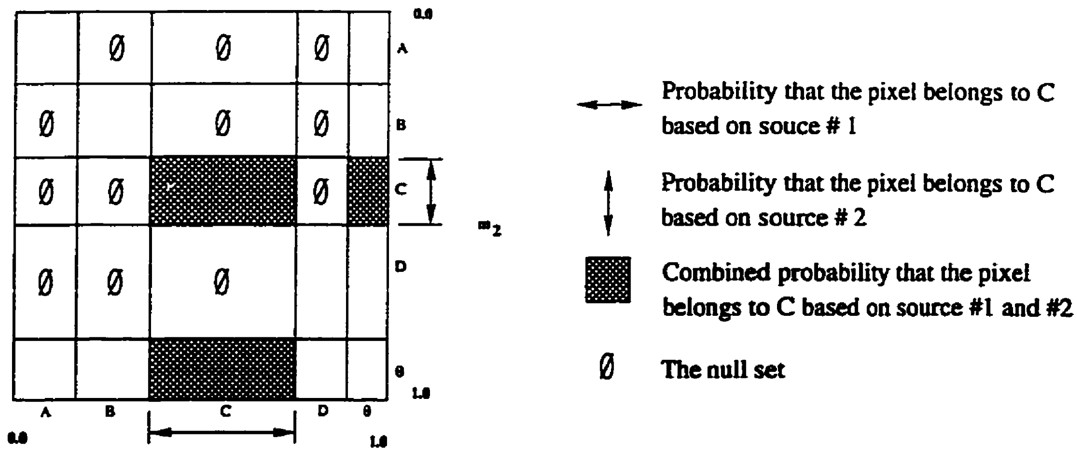

0.0
A
B
C
D
θ
1.0
m_2
A
B
C
D
θ
0.0
1.0
∅
Probability that the pixel belongs to C based on souce # 1
Probability that the pixel belongs to C based on source # 2
Combined probability that the pixel belongs to C based on source #1 and #2
The null set

Figure 5.3: A conceptual *Unit Square* for combining evidence from two data sources, m_1 and m_2.

conceptual unit square for two data sets. The horizontal lines partition the four label class probabilities (A, B, C, D) for a pixel from the first data source giving a total probability of 1.0. Similarly, the vertical lines partition the identical four label class probabilities for the second data source. The area of a single box defined by two intersecting *identical* classes $(m_1(C) \cap m_2(C)$ for example) represents the combined evidence (probability) for a pixel label class from the two evidential mass distributions m_1 and m_2. Boxes defined from the intersection of conflicting label classes $(m_1(A) \cap m_2(D)$ for example) are assigned to the null set as denoted by $\emptyset$. The inclusion of uncertainty is indicated by the θ partitions (§5.3.2).

Figures 5.2 and 5.3 are closely related to one another. The distance between label class divisions on one axis in Figure 5.3 represents the magnitude of support in favour of that label from a data source. Notice that the total quantity of evidence, when all label class support magnitudes are summed for a single source, equals 1.0.

5.3.5 The Decision Rule

Once a final set of evidence has been derived (through successive combinations), a decision rule must be applied to actually classify the data. Various decision rules are used in the literature with no consensus on which is the most correct.

1. The label class with the ***highest support*** is chosen (Le Hégarat-Mascle et al., 1998; Kim and Swain, 1995; Lee et al., 1987; Peddle, 1993, 1995).
2. The label class with the ***highest plausibility*** is chosen (Le Hégarat-Mascle et al., 1998; Lee et al., 1987; Kim and Swain, 1995).
3. The label class with the highest support AND plausibility is chosen (The ***maximum support and plausibility*** rule). (Lee et al., 1987)
4. The label class with the ***highest sum of support and plausibility*** is chosen (Le Hégarat-Mascle et al., 1998; Peddle, 1993, 1995).
5. The label class with support that exceeds the plausibilities for all other possible labels is chosen (The ***absolute*** rule). (Lee et al., 1987)

Lee et al. (1987) note that if there are no union subsets ($A \cup B$ for example) for the possible classes, the first three rules will result in the same decision. This is evident in Figure 5.4 where it can be seen that for decision rules 1, 2 and 3 applied to the mass distribution, Class A will always be chosen.

If only a maximum support decision rule (1) is chosen, Richards (1993) argues that a labeling decision should be considered risky if the plausibility for the next most likely class is higher than the support for the class with the greatest support.

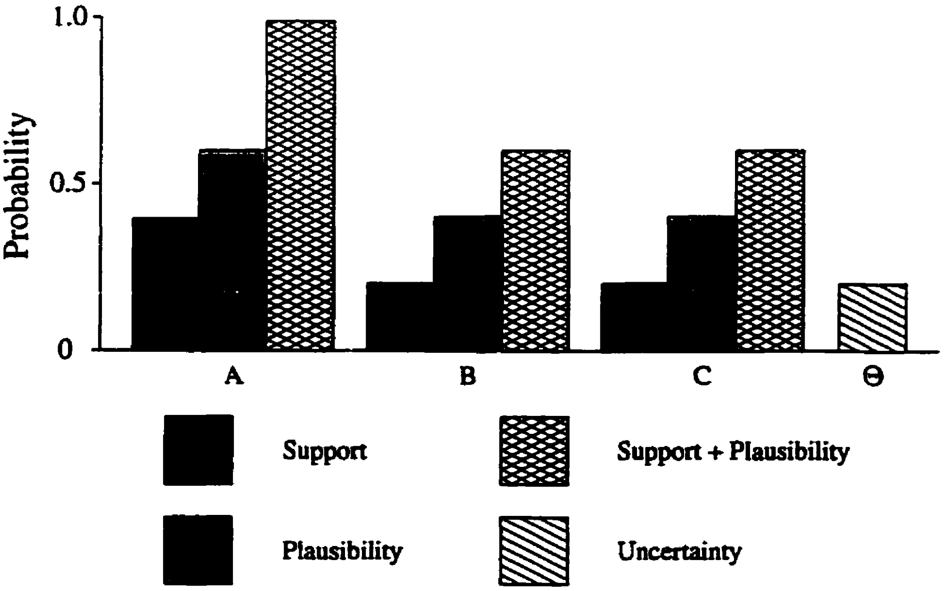


Figure 5.4: A graphic example of a set of evidence represented in terms of *support* and *plausibility*. The label classes for the mass distribution function are A, B and C.

The *absolute rule* proposed by Lee et al. (1987) would account for this situation but as they note, the decision rule can result in a situation in which no class will be chosen. This will occur when the uncertainty is greater that the difference in support of the two most likely classes.

The use of supports and plausibilities for decision rules can be useful for determining the weakness or strength of a particular labeling. For example, in a geological mapping application, Moon (1990) produced separate maps for plausibility and support to aid in the interpretation of the final labeling decision. A user would then be able to associate a confidence level with any information extracted from such a map.

5.3.6 Computer Algorithm Development

Two programs were developed to implement evidential reasoning combination on the data sets: (1) an Evidential Reasoning Combination (ERC) program and (2) a Classification of Evidential Reasoning Probabilities (CERP) program. Both run in a PCI (1997) EASI command line environment.

The primary function of the ERC program is to combine evidential mass derived from two sources of image data using Dempster's orthogonal sum (§5.3.1). The program takes as inputs, two sets of images of which the pixel values must be probabilities. For each pixel location on a set of imagery, an evidential mass distribution function can be extracted. The user must specify these two sets of probability images to be combined as well as a set of output channels to house the combined probability values where each set represents a single data source. Both input sets must have an equal number of elements (imagery). The user is able to optionally specify two sets of values which represent the uncertainties for each individual classes for each data source.

The algorithm operates on the sets of imagery one pixel at a time. At a given pixel location, a set of probability values is extracted from a set of imagery and stored in an array. The set of evidence is then prepared for combination. If the user does specify label class uncertainties, the algorithm ensures that the sum of all mass is 1.0 by applying the following rule to each set of evidence:

1. All elements are multiplied by their respective uncertainty values.

2. The total mass for the set of evidence is calculated.

3. If total mass $= 0$: It is assumed that there are no probabilities and a flag is set for later use in the algorithm.

4. If $0 <$ total mass $<= 1$: Any remaining difference (1 - total mass) is assigned to the uncertainty term.

5. If total mass > 1 : The mass total is normalized to 1.0 such that all elements are divided by the total number of possible classes.

If uncertainty values were **not** specified, only rules 2 to 5 were applied. It should be noted that for the kNN classifier, the total mass will never be greater that one. However, with other classifiers such as the MLC, the probabilities can sum to greater than one. These steps are repeated for the second set of imagery. The algorithm then proceeds to the combination stage. Before combination begins, the algorithm checks for the *mass = 0 flag*. If the flag has been set, combination does not occur for that pixel set. The combination stage uses Dempster's orthogonal sum to combine the two arrays of probability values. These are stored in a third array and written to the specified output probability channels. This entire process is repeated for each pixel location on the imagery until a complete set of combined probability imagery is created. The program can be re-run if additional sets of evidence must be combined. This can be facilitated with a PCI EASI macro script such that the output from one program execution becomes the input (for one set of imagery) for the next. As with ERC, the CERP program also runs in PCI (1997) EASI command line environment. Its function is to determine a winning pixel label from a set of probabilities. These probabilities are ideally extracted from the output imagery of ERP.

class	JP	BS	TA	MIX	WAT	ANT	WET	Percent	Change
JP	164	6	36	9	0	1	3	74.9	-4.1
BS	17	61	2	40	0	6	6	46.2	+13.6
TA	1	0	312	1	0	0	2	98.7	+3.3
MIX	10	20	17	50	0	3	0	50	-11
WAT	0	0	0	0	788	0	0	100	0
ANT	1	0	39	0	0	408	32	85	+2.3
WET	0	1	12	0	0	141	557	78.3	+1.8

Table 5.10: Summary of classification results after SAR and TM classifications were combined using evidential reasoning. The change is relative to the *highest* accuracy of the pre-combination TM results.

5.3.7 Image Combination

The ERC program was implemented using the probability imagery of the SAR and TM data as inputs. The classification accuracies were used as uncertainty measurements for each label class as discussed in §5.3.2. The result was a new set of combined class labeling probabilities which were used in CERP to produce a classified image of forest types. The *maximum support* decision rule was applied.

5.4 Assessing the Accuracy of Canopy Type

Referring to Table 5.10, the jack pine and trembling aspen classes performed well in the classification assessment with 75% and 99% accuracy, respectively. The black spruce and mixed class were still considerable poorer than desired. However, the combination of SAR and TM imagery improved the black spruce accuracy considerably. At the same time, it lowered the accuracy of the mixed class. If we compare these results back to the *original* TM classification, we can see that there are both gains and losses. In general, however, there have been reasonable gains made in classification accuracy. The mixed class, for instance, jumped from a very low 28% accuracy to a more reasonable 50%. The black spruce class was the only one that really suffered a notable reduction in accuracy.

Trembling aspen and black spruce appeared to be responsible for much of the misclassification in the mixed class. This observation is not unreasonable because trembling aspen appears as both a category on its own and within the mixed class. Confusion with the black spruce class is more difficult to explain. The poor performance of the classification of the black spruce category is possibly due to the heavy presence of white spruce in the mixed stands. The spectral signatures of the two species cannot be examined due to lack of necessary data. However, it is possible that they are similar and therefore, classification confusion has resulted.

Lastly, it should be noted that although the jack pine, black spruce and trembling aspen sites are representing the dominance of their respective species, it is not likely that *pure* stands will be found in nature. Examination of the canopy composition tables in Appendix A tells us that all stands are to some degree, mixed stands. Furthermore, if we look to the understory species, it is unlikely that it will be the same as the canopy species, nor pure itself. Now consider that each tree in a site will contribute some energy (from reflection) to the sensor which will manifest itself in the pixel value found on an electro-optical remotely sensed image. It can be conjectured then, that the confusion arising in the mixed and black spruce classes can be explained with the preceding observations.

It also is possible that some of the so-called *pure* sites are isolated forest stands within larger *mixed* areas (ie. they are part of a patchy landscape). The coarse resolution of the satellite imagery and the forest inventory maps may not delineate these small stands and thus, if the site was unduly expanded, confusion might result.

Chapter 6

Mapping Canopy Stem Density

The role of canopy trees in explaining biodiversity was examined in §1.2. This chapter discusses the development of a map of canopy stem density for PANP. Again, close attention is paid to uncertainty propagation.

6.1 Problem Definition

Mapping canopy stem density with satellite imagery proved to be the most difficult task encountered. Reports of canopy stem density estimates for boreal forest from spaceborne imagery are sparse in the literature. Many studies demonstrate the use of SAR to predict forestry parameters such as *species composition*, *biomass*, *LAI* or *DBH* (Ranson et al., 1995; Ranson and Sun, 1994; Le Toan et al., 1992; Wilson, 1996). However, within this literature there is little or no attention given to the estimation of stem density. In the few instances in which it is given mention, we are left without a summary of results. It seems plausible that estimation of stem density has been largely unsuccessful by the scientific community and thus, reporting of results has not followed. Given this premise, success seemed unlikely. Nevertheless, this chapter discusses the approach used to estimate stem density for the PANP study site. The reader should note that the results are not without error; the errors, however, do not go unreported.

6.2 Modeling Canopy Stem Density

The following subsections discuss (1) the flow of the model design, (2) provide an explanation as to how the model was developed and (3) the development of the mathematical relationships which were used to predict stem density from remotely sensed imagery. The development of the final model was not a straight forward exercise and it encompassed many failed approaches. The proceeding discussions will offer an insight into these failures and how they influenced the final model. To begin the discussion, it is instructive to first examine the flow of the model.

6.2.1 Model Design and Flow

Figure 6.1 shows the flow of the initial and final models used to predict canopy stem density. In the initial three-part model, HV polarized backscatter is used to estimate biomass for each image pixel (Step 1). The biomass estimate represents the amount of biomass of a single *representative* tree in a pixel. In other words, if a random tree was selected from within a given pixel, its biomass would be characterized by the pixel biomass estimate. Although this is a highly questionable assumption, it is an artifact of the coarse resolution of the image data.

The biomass estimate is then used to predict DBH (Step 2). Again, the DBH estimate represents the diameter of a *representative* tree for the pixel. Finally, the DBH estimate is used to predict canopy stem density (Step 3). Given the previous assumptions, the estimate implies that if we sampled the density of trees at any given area within a pixel, our results would be similar to the estimate of CSD for the pixel. Common sense tells use that this is unlikely to be the case if we actually

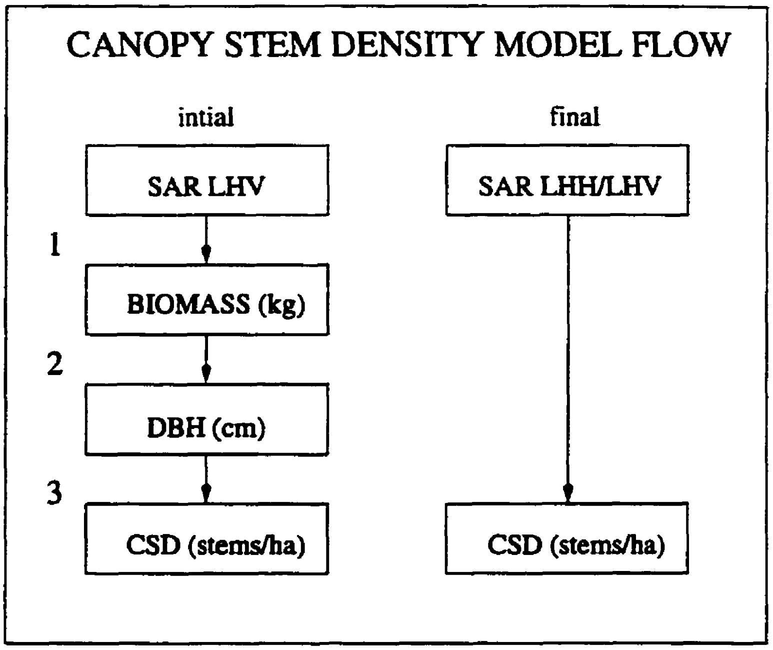


Figure 6.1: Initial and final versions of the canopy stem density model

put it to the test with field reconnaissance. However, since the estimates are pixel based, there will inevitably have to be some generalization.

Ultimately, circumstances dictated that it would be more suitable to estimate CSD directly from SAR imagery (Figure 6.1). Although this seems simple and straight forward, arriving at this solution was not trivial. The assumptions made in the initial model and the interpretation of the result remains the same with the exception of the biomass estimation and its relation to DBH. The following subsection will discuss how this model was developed.

6.2.2 Background of the Model Development

The first question that was asked was what forestry parameters have been successfully estimated with remotely sensed imagery? The most logical place to start was with stem density itself since this was the desired parameter. However, as was alluded to in §6.1, the literature concerning this kind of estimate was sparse. Kurvonen et al. (1999) estimated stem volume using L-band JERS-1 and ERS-1 SAR. However, their prediction equations required the use of vegetation and ground moisture parameters. Their approach did not seem practical for this work.

Scatter plots of all four SAR bands and all seven TM bands versus stem density measurements (from the ground sampled sites) were constructed. None of the plots appeared to show any semblance of useful relationship between the remotely sensed inputs and stem density. This suggested that a *surrogate* variable, which could be estimated, would have to be used and then related to stem density.

Based on previous work by Ranson et al. (1995); Ranson and Sun (1994); Le Toan et al. (1992) and Wilson (1996), biomass appeared to be a parameter that could be estimated with reasonable success for boreal forest using SAR imagery. In particular, a good linear correlation between SAR LHV backscatter and biomass (kg/m^2) had been reported. Le Toan et al. (1992) reported even stronger correlations between P-band SAR and biomass. Unfortunately such data was not available for this study site. Ranson et al. (1995) linearly related the log of forest biomass to SIR-C SAR LHV backscatter with a coefficient of determination (r^2) of 0.846. Their study site was part of the BOREAS project and was located adjacent to PANP. Their results seemed extremely promising for this research.

It was noted that Ranson et al. (1995) calculated biomass for their training sites using DBH measurements as inputs into allometric equations (Singh, 1982). Singh empirically derived a set of third-order polynomial equations for relating DBH (cm) to biomass (kg) for major tree species in the prairie provinces of Canada which include relationships for individual boreal species. The r^2 values for these relationships ranged between 0.96 and 0.99. These DBH-biomass relationships were important because biomass measurements did not exist in the ground sampled dataset for this project. They provided a means by which biomass could be determined since DBH measurements *did* exist for the ground sampled data. To predict DBH from biomass, a straight forward inversion of the equations was carried out.

The sampling method discussed in §2.4 used 15 sample points of which each was divided into four quadrats. Therefore, the maximum number of canopy trees sampled was 60 (15 points times 4 quadrants). If a canopy tree was further than 10 m from the sample point, it was not measured so in some instances there were less than 60 sample trees. From the population of sample trees for a given stand, mean DBHs were calculated for each species. The mean DBH for a tree species in the stand was input into the appropriate biomass prediction equation (for the species of interest). This was repeated for each species in the stand. A weighted average of biomass was then calculated such that species with greater abundance in the stand would have a greater influence on the biomass estimate. This yielded a predicted estimate for a single *representative* tree for a stand.

Scatter plots of all four SAR bands versus the estimated biomass were then constructed. Linear regression was performed on the SAR LHV polarization which was the only independent variable yielding anything resembling a useful relationship.

The adjusted r^2 value was very low at approximately 0.3.

In parallel to the work of the preceding discussions, it was determined that DBH can be related linearly to the natural logarithm of stem density such that it will increase as stem density decreases (Husch et al., 1972; Oliver and Larson, 1990). Intuitively this makes sense because as trees get bigger, they need more space to grow. As a preliminary check of the validity of this relationship for the study site of this research, DBH and stem density data for the southern boreal mixwood forest of Saskatchewan published by Ranson et al. (1995) (from the BOREAS project) were plotted as described above. The resultant correlation coefficient for the linear regression was -0.94. This prompted further investigation to find such a relationship with the data used in this research. After plotting the canopy stem densities versus the DBH values from the ground sampled sites, a very similar relationship was revealed with an r^2 of approximately 0.6.

Before continuing, now is a convenient time to recapitulate: from the SAR imagery, we are able to estimate biomass. Using Singh's equations, we can relate biomass to DBH. DBH can then be related to CSD. Using these relationships, we can logically piece together a method for stem density estimation. However, it should be evident to the reader that this method is marred by the weak relationship between SAR imagery and biomass estimation. The poor correlation introduces a large amount of uncertainty into the CSD prediction. In addition, each intermediate model needed to arrive at a CSD estimate also introduces its own uncertainty which is compounded as the results of one are used as input into another. Simplification was the most logical answer for reducing uncertainty.

When testing the solution with input biomass estimates, it was noted that the output DBH values for different species were quite similar. The biomass to DBH functions were then plotted and it was evident that within the range of useful input biomass estimates, the functions were all nearly linear. This implied that perhaps DBH could be directly estimated from the SAR imagery. Plotting SAR LHV versus mean DBH revealed that the relationship was very similar to that of SAR LHV and biomass except that there were a number of major outliers at the lower and upper range of DBH values. If the input SAR was limited to pix values within 156 and 255 (on an 8-bit scale), a linear relationship could be defined with an r^2 of approximately 0.3. The relationship, although still weak, allowed part two of the initial model to be eliminated. This simplified the model as well as reduced the introduction of additional uncertainty associated with the biomass-DBH models.

It seemed plausible that the relationship between the three variables (SAR LHV, DBH and CSD) could be further simplified. Recall that scatter plots constructed for CSD versus the four SAR bands revealed no apparent relationship. However, by limiting the range of SAR values used in the model and reconstructing the scatter plots, a weak relationship emerged. Could a simple SAR LHV-ln CSD relationship be used with the exclusion of DBH (and biomass)? Some sample SAR data was run though the original set of models to determine CSD values. Then, using the simplified relationship of LHV to the natural logarithm of CSD, the same data was input. The results were remarkably similar. This relationship, although weak, was a good approximation of the original set of models. The advantages to using this model were twofold. First it is simpler to implement and second, the uncertainty is reduced.

Using a sub-sample of the SAR LHV imagery, the model was implemented with a PCI program. Immediately it was evident that model was problematic. Due to the limited range of valid SAR values that could be input into the model, CSD could not be predicted for large patches of the image which was highly unacceptable. This prompted the search for an alternative SAR variable to relate to CSD; one in which a complete range of SAR values would be valid as input into the model. Ranson et al. (1995) and Ranson and Sun (1994) used SAR band ratios to improve correlation with biomass. They proposed that using a ratio of SAR bands (as opposed to unratioed bands) may increase signal dynamic range and thus, improve correlation. Collins and Livingston (1996) used SAR polarization ratios (same band) for mapping thin sea ice. These techniques were examined for this research: regressions were performed using all possible combinations of SAR polarizations and band ratios versus the natural logarithm of canopy stem density. Of these many combinations, the LHH/LHV showed a reasonable correlation with CSD. Utilizing this ratio enabled a model to be constructed for which SAR inputs were not limited by a specific range. The data used to construct the model is shown plotted in Figure 6.2.

6.3 Methods

6.3.1 Construction of the Model Equations

Although the actual model was simplified down to a single linear regression equation, its final mathematical development was heavily influenced by the nature of the data from which it was constructed. Canopy stem density measurements were provided by Chipman (1999). CSD is estimated with the equation, $CSD = 1/mean\ distance^2$. The mean distance is calculated by summing the distances from tree to sample point of all trees sampled in the stand and dividing by the total number of trees. The error on the mean distance for a stand was estimated using the standard deviation on the mean distance, $\sigma_{\overline{x}} = \sigma_x/\sqrt{N}$. Using standard error propagation techniques, the error on CSD was then determined. Likewise, the SAR backscatter used in the model was a mean value pooled from the pixels which fell within the sampled stand. The standard deviation of the mean was used as an error estimate. Both sources of error were significant and needed to be incorporated into the regression model.

In response to this need, a weighted least squares technique was used to estimate the model parameters as well their errors in terms of variances and covariances. The technique, which differs from standard regression techniques, is able to estimate parameters based on input measurements with errors in **both** variables (a complete treatment of these methods is given in Krakiwsky and Gagnon (1987)). This was ideal for the data being used to build the model. The basic form of the model relationship is:

$$\ln CSD = m \times (LHH/LHV) + b \tag{6.1}$$

where m and b are the slope and intercept respectively. Then letting

$$\ln CSD = CSD' \tag{6.2}$$

so that a linear solution can be obtained using

$$CSD' = m \times (LHH/LHV) + b \tag{6.3}$$

An additional benefit of the least squares approach is that it allows for the calculation of error on any prediction using the variance-covariance matrix of the model parameter estimates. The benefit here is that the errors from both the SAR and CSD original measurements (used to construct the model) are incorporated into the error on a prediction.

The least squares solution to Equation 6.3 was programmed in MATLAB (1997). From the solution, the slope and intercepts were obtained as well as the variance-covariance matrix. Figure 6.2 shows the relationship as the solid line amid the measured data points.

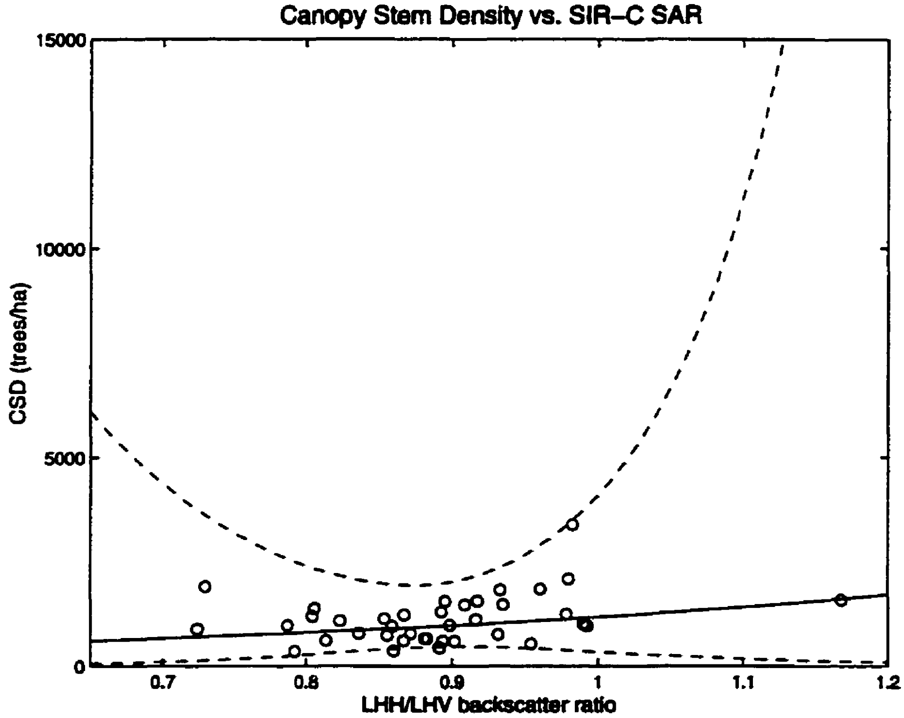


Figure 6.2: Measurement data and the linear regression model used to predict canopy stem density from the SAR backscatter ratio. The solid line is the model prediction line (of best fit) and the dashed lines represent error bars two standard deviations from the prediction line.

6.3.2 Computer Algorithm Development

In order to predict canopy stem density for the entire image area, it was necessary to automate the process. A PCI program was written to both predict CSD values as well as propagate uncertainty from start to finish. The program was executed using the SAR LHH/LHV ratio as input data. An estimate image of CSD and an upper and lower bound image were output.

6.4 Results

Figure 6.3 shows a sample of the output estimated CSD image. Light patches represent higher canopy stem densities relative to darker patches. There are some noisy areas apparent on the image. These are a result of wetland areas and will ultimately be filtered out in the final maps of species richness.

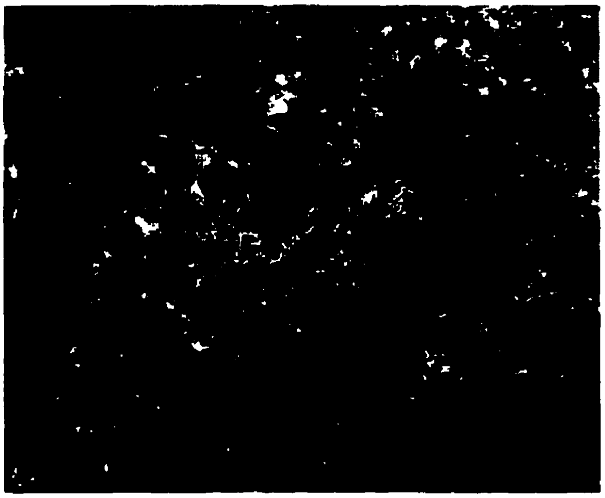


Figure 6.3: An image subsample of estimated canopy stem density using a SIR-C SAR LHH/LHV polarization ratio. Dark areas represent lower CSD than lighter areas. Rivers have been superimposed onto the image.

6.4.1 Assessing the Accuracy of Canopy Stem Density

Equation 6.3, although linear, predicts the natural logarithm of canopy stem density. This is true for the error as well. In order to determine the error interval, the ln(error) was added and subtracted from the ln(CSD) to gain the upper and lower error bounds on the prediction. The exponential of each was then calculated which produced the error in terms of CSD. These bounds are plotted in Figure 6.2 as the dashed lines. An important feature to notice is that the error bounds are not symmetric about the prediction value. This is a product of using log transformed data in the model development. SAR input values below 0.8 and above 1.0 are prone to extremely high uncertainty.

Chapter 7

Mapping Biodiversity Using An Ecological Model

This chapter draws together the results of Chapters 3 though 6 into the biodiversity prediction model. The culmination of this work is presented as a set of three predicted species richness images at the end of this chapter.

7.1 The Model

The general form of the species richness prediction equation developed by Chipman (1999) is:

$$S = \beta_0 + \beta_1\, TSF + \beta_2\, DFR + \beta_3\, CSD + \delta_1\, JP + \delta_2\, BS + \delta_3\, TA \qquad (7.1)$$

where β_0 is the intercept, β_{1-3} are parameter estimates for the variables TSF, DFR and CSD respectively, and δ_{1-3} are parameter estimates for the categorical *dummy* variables jack pine (JP), black spruce (BS) and trembling aspen (TA), respectively. The categorical forest type variables are binary. Only one of these variables may have a value of *one* (1) at any one time while all others must have a zero value. The mixed class is not included in the equation but is represented when all other species have a value of 0. The effect of these binary variables is to change the intercept of the model. The parameter estimates are listed in Table 7.1

Term	Estimate
Intercept (β_0)	25.995261
TSF (β_1)	-0.036611
DFR (β_2)	-0.000926
CSD (β_3)	-0.004021
JP (δ_1)	0.8042173
BS (δ_2)	-4.295932
TA (δ_3)	4.7861226

Table 7.1: Parameter estimates for the species richness equation (Eq.7.1).

7.2 Methods

7.2.1 Computer Algorithm Development

Two PCI programs were written to apply the model to the input imagery. The first, called MODPRED (for model prediction) takes input data for each variable and predicts species richness. The second, called MODERR (for model error) performs the same function as MODEST but was designed to predict the upper and lower error bound on the prediction of species richness.

It is appropriate now, to summarize the sources of data that will be used in the model. From Chapter 3, an image of distance from ridgeline was produced. In addition, upper and lower estimate images of DFR were also produced based on a 95% confidence interval. Chapter 4 took an existing image of field measured time-since-fire and produced two sets of upper and lower bound TSF imagery based upon the highest expected measurement error. One set covered the inner part of the fire polygons and the other covered the outer rings of the fire polygons (at the borders). In Chapter 5, a classified image of four forest types was produced, each corresponding to a category in the model. In addition, the probability that a given pixel should be classified as each forest type was recorded into a set of four images (one for each class). Lastly, Chapter 6 produced an image of canopy stem density as well as upper and lower estimate images of CSD.

The MODPRED program accepts the images of DFR, TSF, CSD, canopy type and the four images of pixel assignment probabilities. The program extracts the set of input variables for a single pixel location. The DFR, TSF and CSD values are input directly into the model. The canopy type was treated in a slightly different manner because it is of a categorical nature. Initially, classification probabilities for an individual pixel were summed. If the sum was less than a user defined threshold value, the model was **not** implemented. Rather, the pixel was assigned the class from the canopy type image. Remember that in addition to the four canopy type classes, this image included a water, wetland and anthropogenic class. In most instances, very low probability sums were the result of the pixel (under consideration) belonging to one of the non-forest classes. If the probability sum exceeded the threshold, the model was implemented four times for an individual pixel. On each implementation, the DFR, TSF and CSD inputs remained constant but the canopy type was changed. The four estimates of species richness (one for each canopy type) were then multiplied by their respective classification probabilities to produce a *weighted* estimate. The four estimates were then summed and the final value for the pixel was then written to an output image.

The MODERR program acted identically to MODPRED with two exceptions. First, the program was designed to accept either the upper **or** lower estimates of DFR, TSF and CSD. Second, recall that there were **two** sets of upper and lower

$$
\begin{aligned}
S_{JP} &= f\ \boxed{\text{CSD}, \text{TSF}_{\text{inner}} \cup \text{TSF}_{\text{outer}}, \text{DFR}, \text{JP}} \times \boxed{p(\text{JP})} \\
+\ S_{BS} &= f\ \boxed{\text{CSD}, \text{TSF}_{\text{inner}} \cup \text{TSF}_{\text{outer}}, \text{DFR}, \text{BS}} \times \boxed{p(\text{BS})} \\
+\ S_{TA} &= f\ \boxed{\text{CSD}, \text{TSF}_{\text{inner}} \cup \text{TSF}_{\text{outer}}, \text{DFR}, \text{TA}} \times \boxed{p(\text{TA})} \\
+\ S_{MIX} &= f\ \boxed{\text{CSD}, \text{TSF}_{\text{inner}} \cup \text{TSF}_{\text{outer}}, \text{DFR}, \text{MIX}} \times \boxed{p(\text{MIX})} \\
\hline
&= \textbf{Predicted Species Richness (S)}
\end{aligned}
$$

Figure 7.1: Schematic of the method used to calculate predicted species richness. Species richness is calculated four times for each canopy type and then multiplied by its respective canopy classification probability. The results are summed to give a weighted average predicted species richness. For each intermediate prediction, the TSF variable input into the model will depend on whether the pixel is in the inner fire polygon or in the outer buffer zone area.

estimates for the TSF variable. When the upper and lower bound images were created, a value of -1 was assigned to image pixels outside the area of interest for each image. For instance, for the inner polygon uncertainty image, the outer rings were given a value of -1. These flag values were used to determine which uncertainty image should be used in the model for a particular pixel. Once, determined, the implementation of the model proceeded as described in the preceding paragraph.

7.2.2 Mapping Biodiversity

Using MODPRED and the images of TSF, DFR, CSD and canopy type classification probabilities, an image of predicted species richness was produced. MODERR was then run twice using using the upper and then lower bound image sets to produce an upper and lower bound on the prediction of species richness. The three slope estimates in Equation 7.1 were all negative. This means that if the input variables TSF, DFR and CSD are relatively large, the diversity estimate will be relatively small. Therefore, the input upper bound uncertainty images produced the lower bound of predicted species richness. The reverse was true for the upper bound of predicted species richness. The images were colour coded using dark green to represent high diversity and light green to represent low diversity.

7.2.3 Performance of the Model and Uncertainty Propagation

Much attention was paid to quantifying and propagating uncertainty throughout the preceding four chapters of this work. These efforts culminate here as all estimations of uncertainty in the input variables are propagated though the biodiversity prediction model.

Using the training and testing sites discussed in §5.2.1, *predicted* diversity values were extracted from the final images. It should be noted that the testing data set was used in neither the canopy stem density estimation nor the canopy type estimation. The distance from ridgeline estimation did include *some* of the testing sites because there were difficulties in measuring the distances from all sites to their respective ridgelines using the topographic maps.

Figures 7.2 shows the actual correlations between species richness measured in the field at the ground sampled sites with species richness predicted by the model using image inputs (for the same sites). It is readily apparent that the correlations are neither strong nor free of error. Care must be taken when interpreting these error bars: the uncertainty estimates derived for each input variable (TSF, DFR and CSD) were quite conservative or what can be termed as the *worst case* error. Consider now that it is quite improbable that for a given image pixel, the worst case error will be present for all input variables simultaneously. There is a much higher probability that two variables will have small error and that only one is subject to a large error. These error bars represent the worst possible case in which all input variables are subject to the maximum possible error.

In order to assess how well the model can predict species richness, a residual analysis was performed. The collected field data was used as the *observed* dataset. Plots of the residuals were constructed and are shown in Figure 7.3. Notice that both plots exhibit a linearly increasing trend. If a regression line was plotted on the testing residual data, it would cross the x axis around the value 14. This implies that predicted species richness values greater than or equal to 14 have been underestimated by the model. Values less than 14 have been overestimated. For the training data, this value is approximately 16. Potentially, a correction could be applied to the model for the input data used in this project so that these over and under predictions could be minimized. However, the unmodified model used here generally produces conservative estimates of species richness from an ecological point of view. Both training and testing data yielded RMS error values of 6.8 species.

Thus far, we examined how well the model was able to predict species richness compared to what was actually observed in the field. Although this is what we are ultimately interested in predicting, it is not an entirely fair evaluation of the model when used with the image based variable inputs. The model itself was constructed using field measurements of time-since-fire, canopy stem density and canopy type and

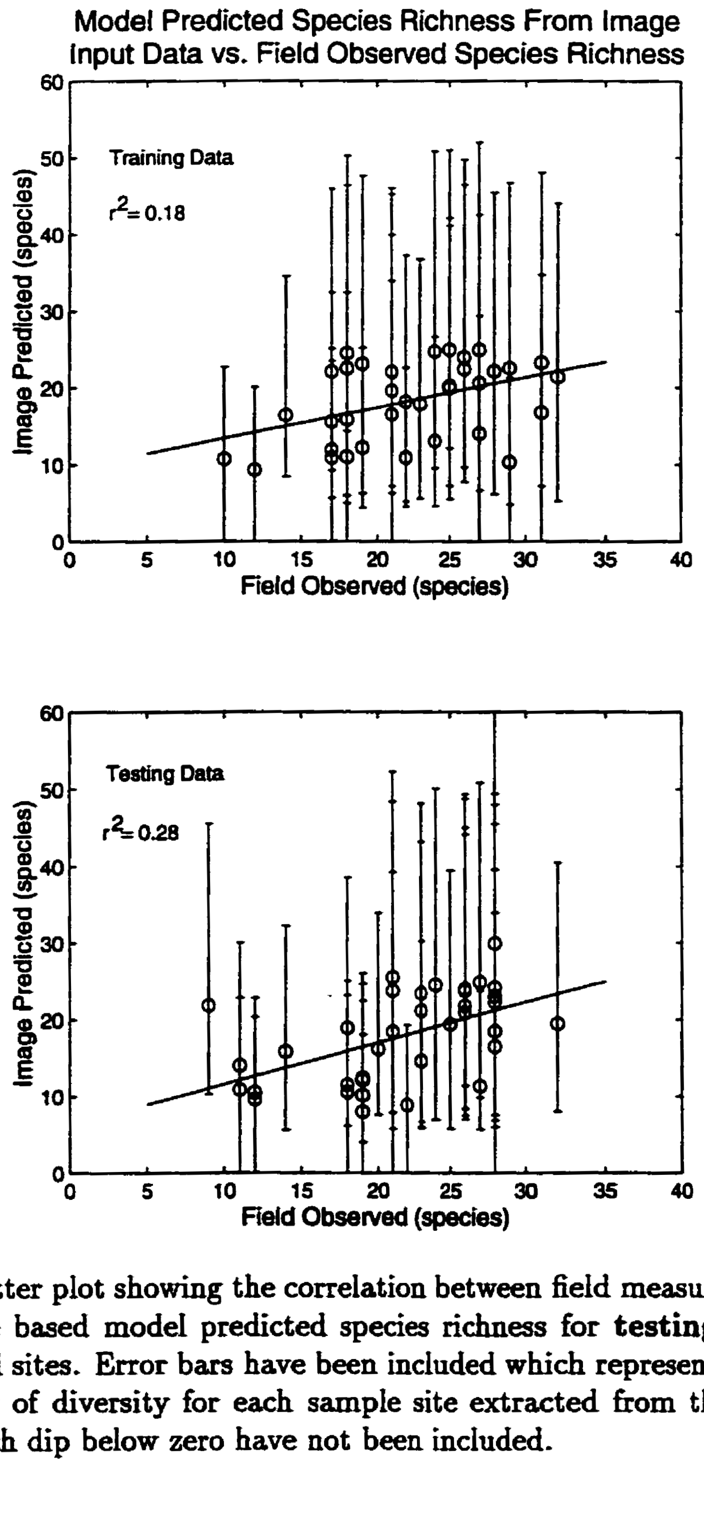


Figure 7.2: Scatter plot showing the correlation between field measured species richness and image based model predicted species richness for **testing** and **training** ground sampled sites. Error bars have been included which represent the upper and lower estimates of diversity for each sample site extracted from the final images. Error bars which dip below zero have not been included.

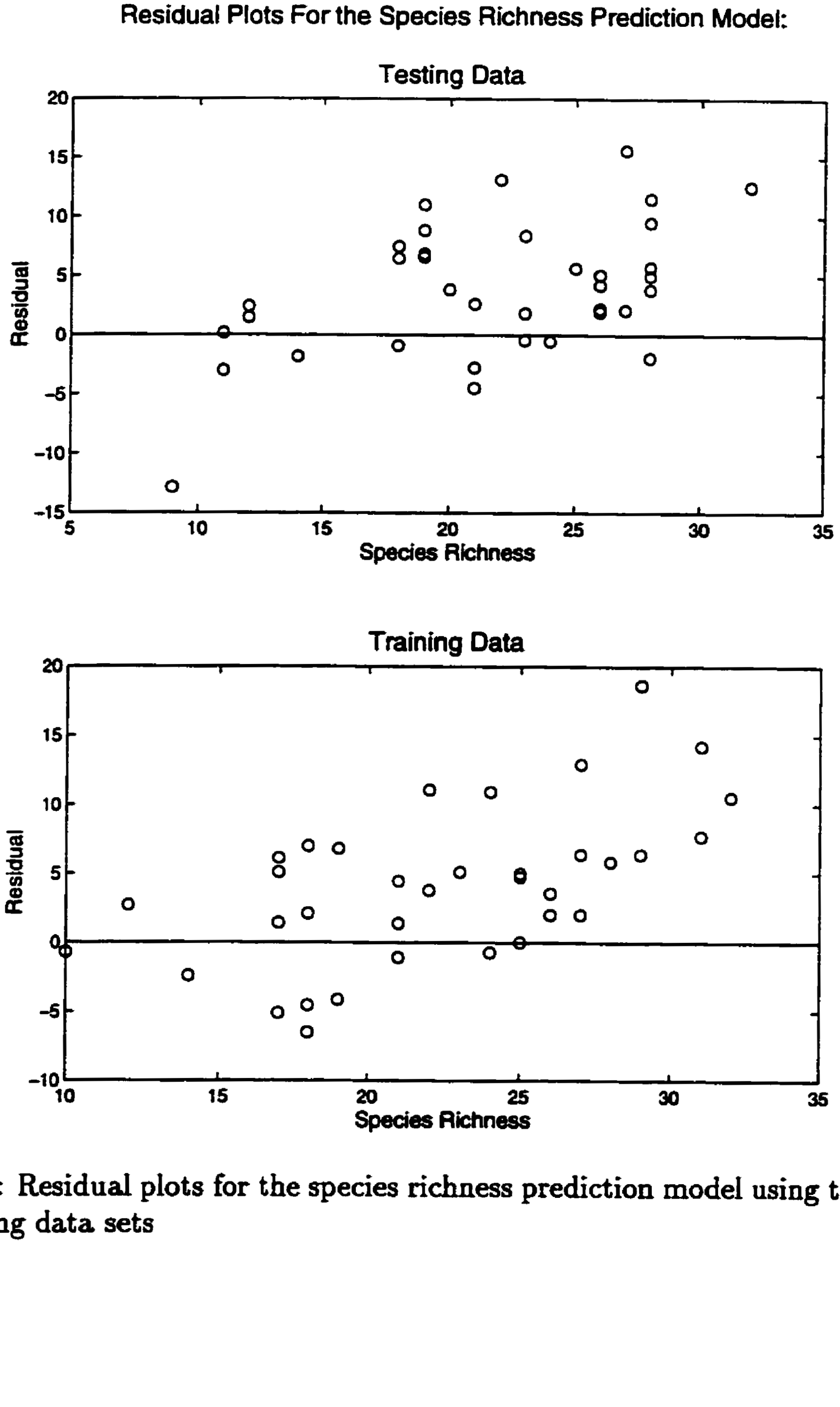


Figure 7.3: Residual plots for the species richness prediction model using the testing and training data sets

map measurements of distance from a ridgeline. If these field measured variables are input into the model, species richness can be predicted. Presented now is a comparison between species richness predicted with image derived inputs and species richness predicted with field based inputs.

The upper plot in Figure 7.4 shows the correlation between species richness predicted by image derived inputs and field measured inputs. Clearly there is a much stronger correlation than is exhibited in Figure 7.2. The lower plot shows the residuals when comparing the two prediction methods. The systematic underestimation is no longer apparent and the residuals have been reduced compared to Figure 7.3. The RMS error of the residuals is 3.84.

Finally, it is worthwhile to examine how well the model predicts species richness with field collected inputs compared to the field observed species richness values. In other words, the data used to construct the model is used within the model. Figure 7.5 shows the correlation between observed species richness and the predicted species richness using field observed input data. It is clear from the residual plot that the model itself predicts species richness with a large amount of associated error. This observation serves to strengthen our assessment of the apparently weak image-based parameter estimation (Figure 7.2). However, it weakens our confidence in the validity of the final biodiversity maps.

7.2.4 Sensitivity Analysis

In order to get an idea of the stability of the equation when subject to changes in input variables, a simple sensitivity analysis was performed. For each of the model variables, a maximum, minimum and mean value were chosen. The mean values

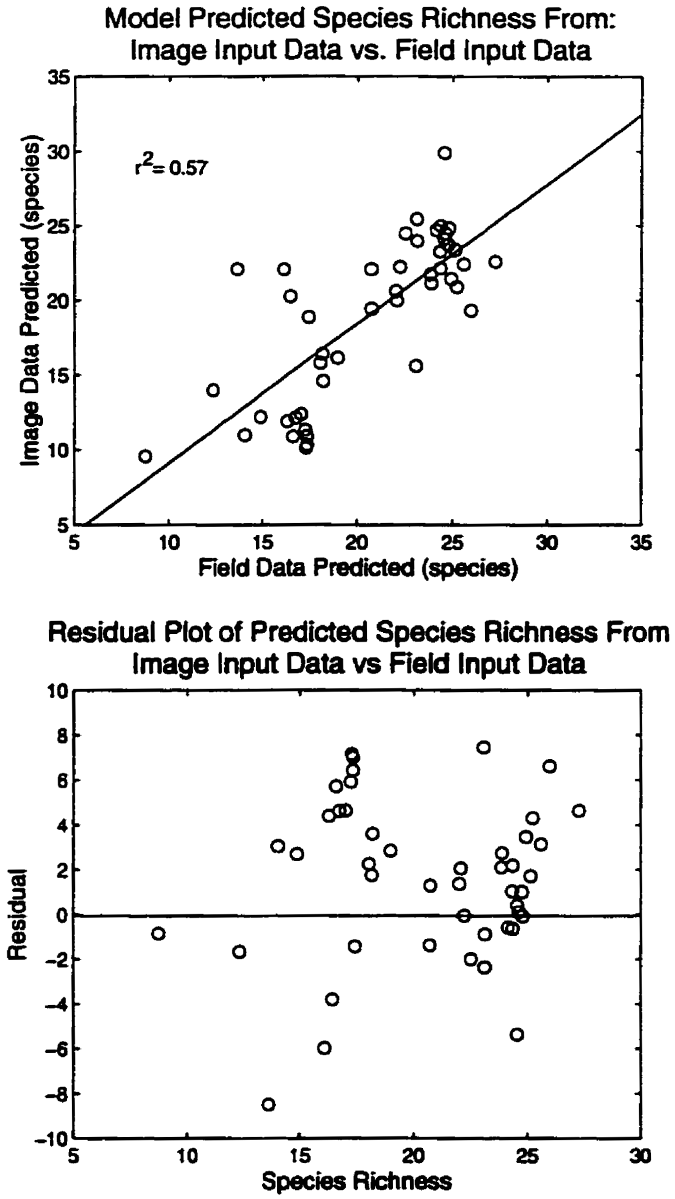


Figure 7.4: The upper plot shows the correlation between species richness predicted by image derived inputs and field measured inputs. The lower plot shows the residuals (*field predictedvs.image predicted*) for the same data. Error bars have been omitted for clarity.

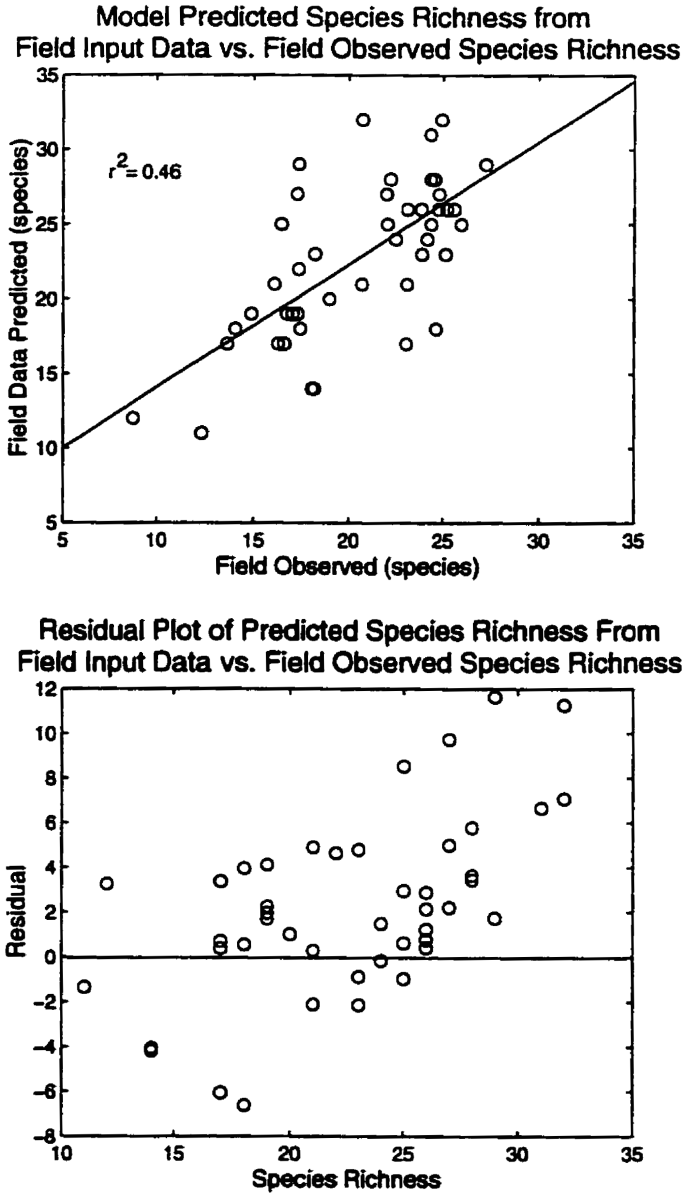


Figure 7.5: The upper plot shows the correlation between species richness predicted by image derived inputs and field measured inputs. The lower plot shows the residuals (*field observedvs.field predicted*) for the same data.

represent a typical forest pixel.

Variable	Minimum	Maximum	Mean
TSF	0	170	50
DFR	0	5000	300
CSD	0	7300	1000

Maximums for time-since-fire (TSF) and distance from a ridgeline (DFR) were chosen based on the expected upper limit of the value of the input variable. Canopy stem density (CSD) was chosen because it represents the approximate upper limit of valid values for use in the species richness equation.

Species richness was repeatedly predicted by substituting the minimum and maximum, one at a time for each variable, into the equation while holding all others constant with their respective mean value. Each forest type was treated separately and a mean for all forest types was calculated. These results are shown in Figure 7.6. The model was also used to predict species richness using the mean value for each variable (labeled as *Mean* on each Figure). For each of the sets of bars for TSF, DFR and CSD, all **other** variables were held constant with the mean values. For instance, when reading the *maximum* bar for TSF, species richness was calculated using the maximum value for TSF and the mean values for DFR and CSD.

As expected from the form of the model, the general patterns of species richness due to variable sensitivity are the same between species graphs. Due to the nature of the great uncertainty in canopy stem density estimation, the model is very sensitive to that variable. It is evident from the graphs that a value of greater than 7000 trees/ha in magnitude predicts negative species richness which is not a valid possibility. In many instances, the upper limit of CSD reached well beyond this value.

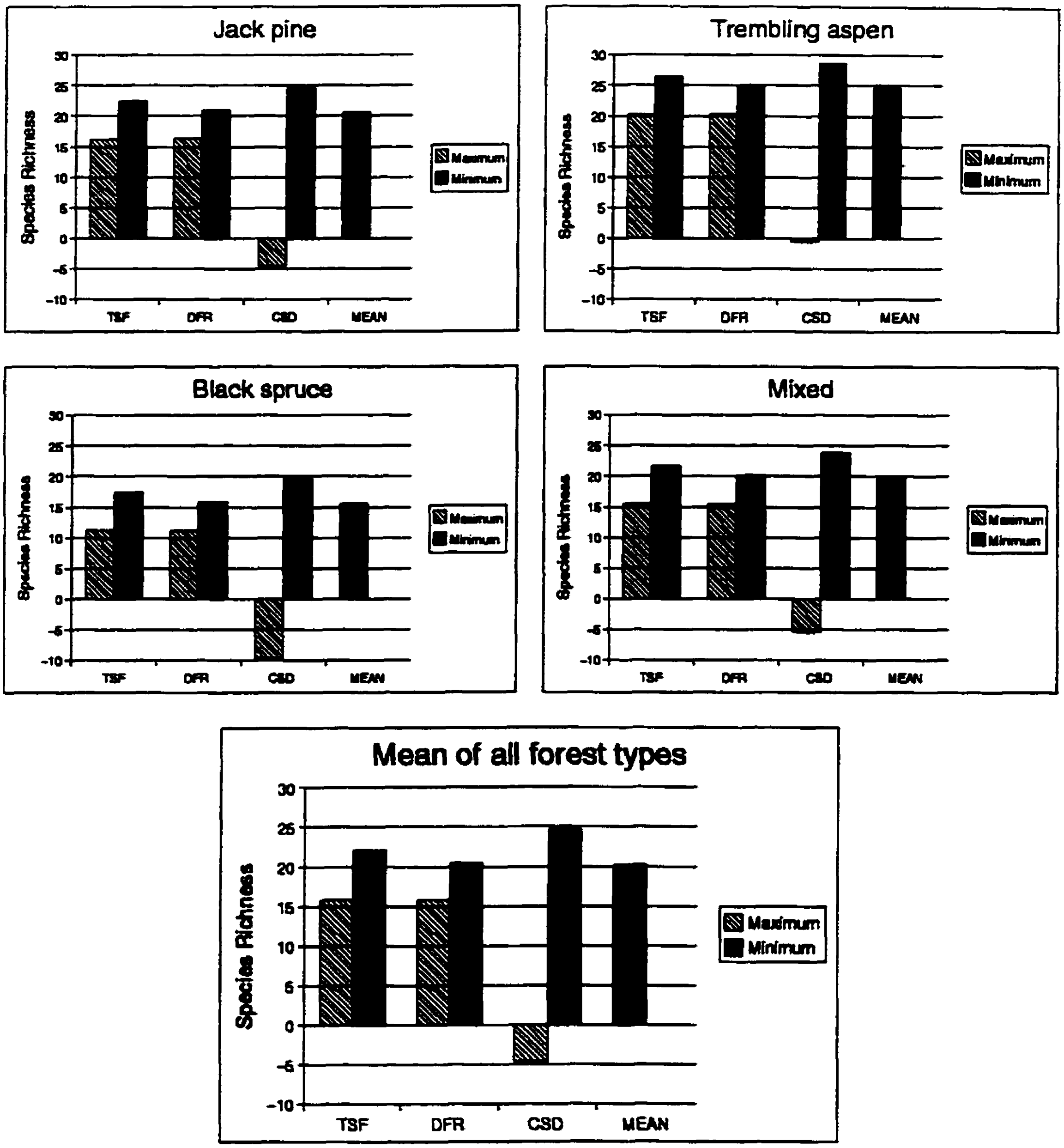


Figure 7.6: Sensitivity of species richness using the model with four different canopy types and the mean of all types.

7.3 Results

Figure 7.7 shows predicted species richness maps and upper and lower bounds on predicted species richness for Prince Albert National Park. The reader should note that the upper bound image was derived from the lower bounds of variable input uncertainty for reasons discussed in §7.2.2. The same is true for the lower bound image.

The wetter areas of The Park tend to have lower diversity. This observation coincides with the reasoning for the inclusion of the distance from a ridgeline variable into the model. However, the actual patterns in the distance from a ridgeline occur at a finer scale than on the species richness map and therefore, there are no apparent spatial correlations between the two maps. The same is true for canopy stem density. In general, areas which were classified as trembling aspen are highly correlated with relatively high biodiversity. Lower biodiversity areas tend to coincide with the spatial distribution of jack pine and the mixed class. The lowest biodiversity coincides with black spruce. There was no apparent spatial correlation between the patterns of species richness and the time-since-fire map. Since fire is suppressed within the park, there tends to be very little young forest area. This might account for the lack of apparent variation in species richness due to this variable.

Lower Bound of Predicted Species Richness

Prince Albert National Park

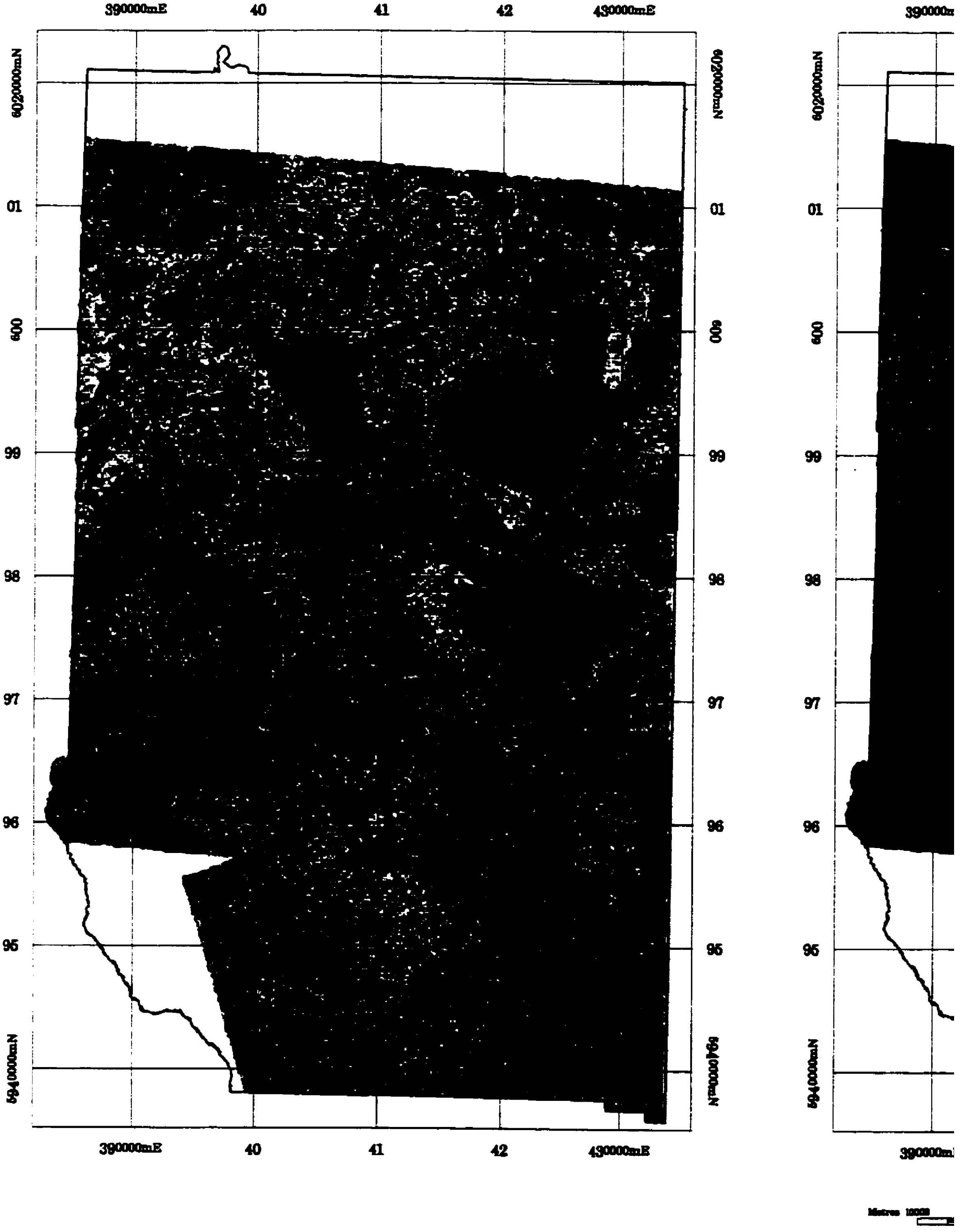


Figure 7.7: Species richness maps

Species Richness

□ 1-8

▨ 9-10

■ 11-12

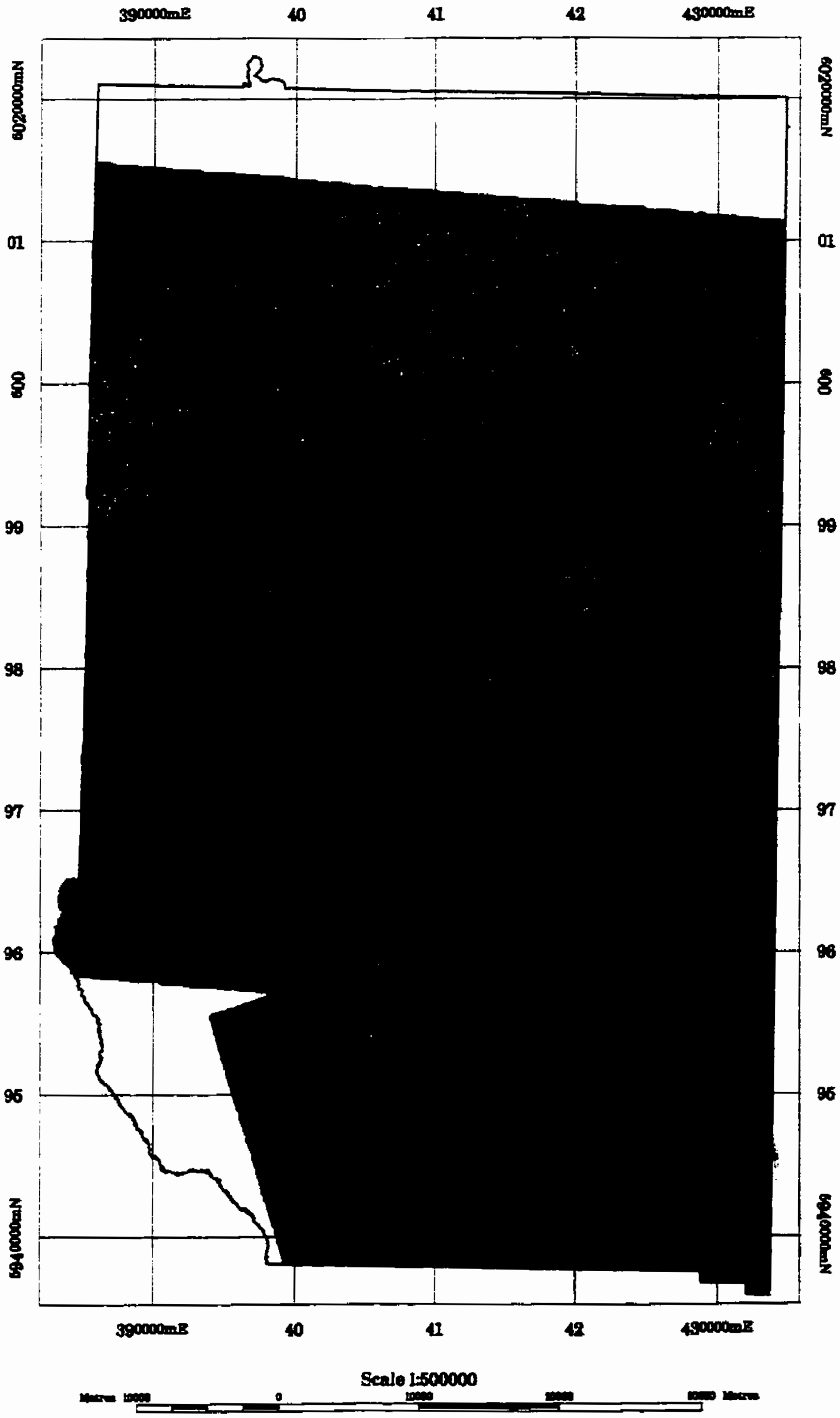
■ 13-14

■ 15-16

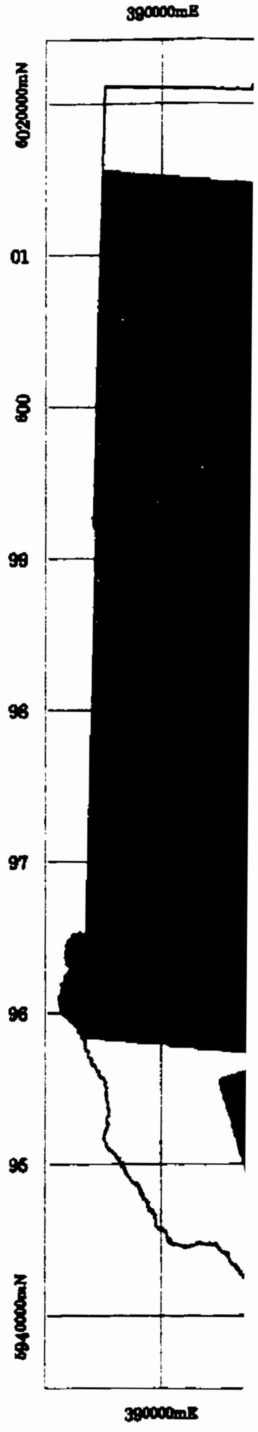
Projection: UTM Zone 13U Ellipsoid: GR

Predicted Species Richness

Prince Albert National Park, SK



Upper Bou

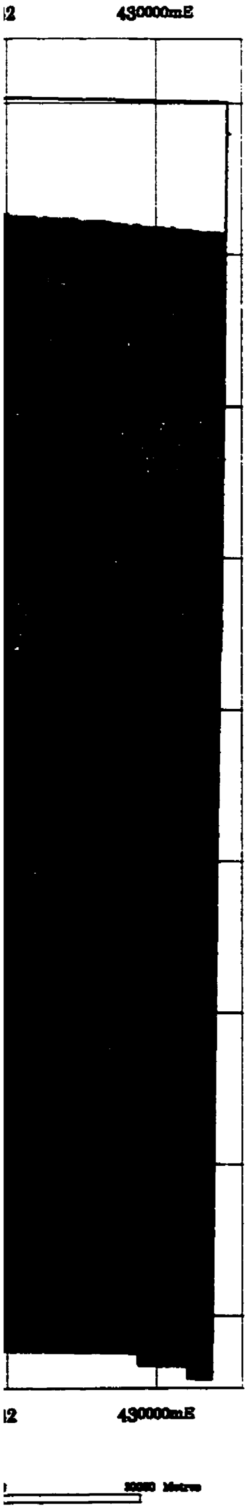


Legend

■ 11-12	■ 17-18	■ 23-24	■ 29-30	Land Class
■ 13-14	■ 19-20	■ 25-26	■ .... Model Invalid	■ .... Open
■ 15-16	■ 21-22	■ 27-28		■ .... Anthrop
ne 13U Ellipsoid: GRS 1980			—— .... PANP Boundary	■ .... W

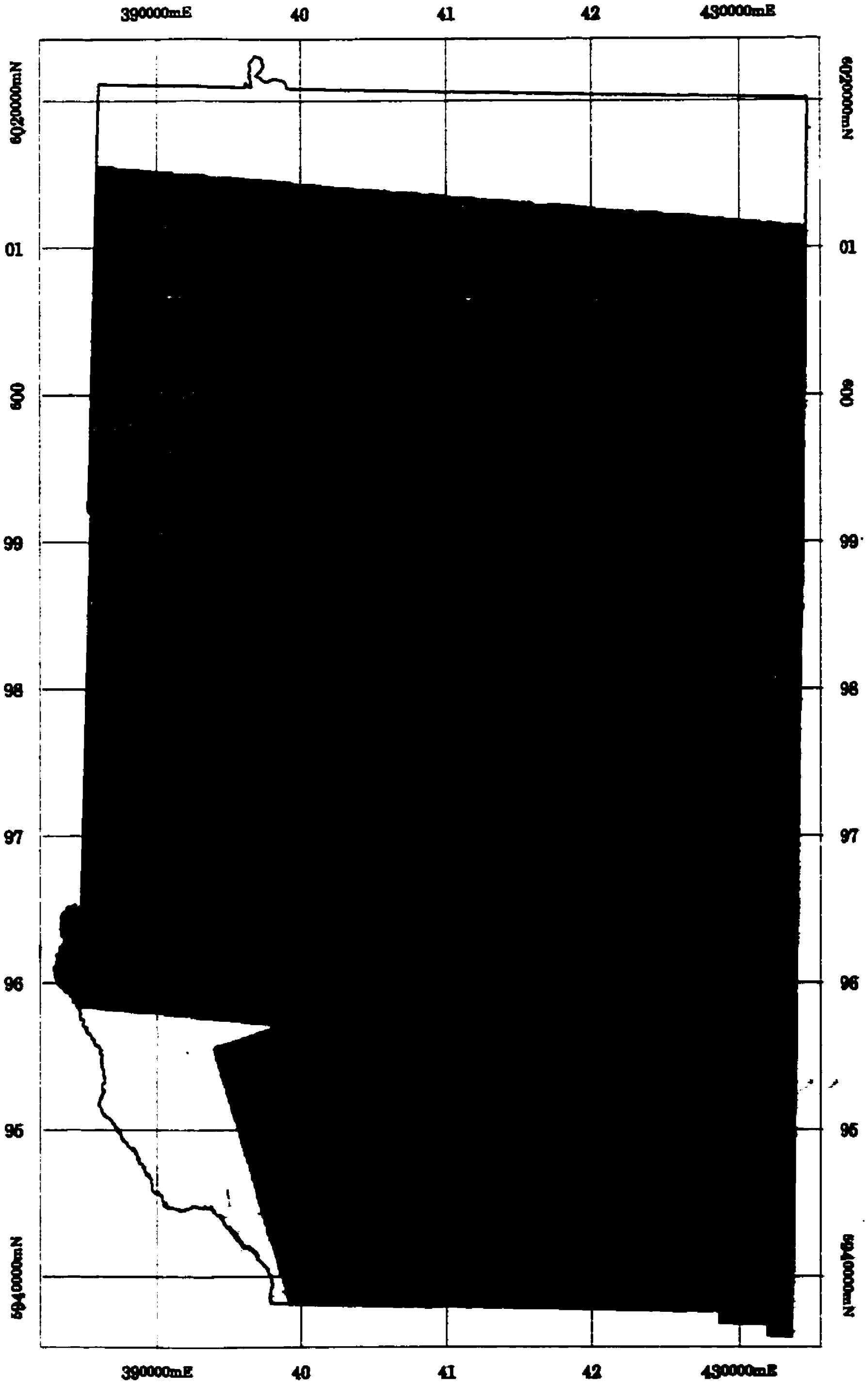
ness

K



Upper Bound of Predicted Species Richness

Prince Albert National Park



Chapter 8

Conclusions and Recommendations

8.1 Conclusions

In this authors opinion, the success of this work should be judged on the merits of three components which extend from the primary objective put forth in Chapter 1:

1. Success of the image based estimation of the species richness model parameters.
2. The estimation of uncertainty throughout all facets of both the parameter estimation and final biodiversity prediction.
3. Performance of the image based parameter estimates for predicting biodiversity.

Regarding the first: distance from ridgeline estimation with the new algorithm met with good success. The new algorithm developed for estimating the distance for image pixels to the ridge improved upon the past work of Bridge (1997). This was accomplished by incorporating local terrain aspect into determining the correct ridge to measure the distance to. The new canopy type estimation algorithm met with reasonable success. The classification accuracies were not as high as desired but considering the nature of the training sites (they were far from *pure*), the overall result was very acceptable. A highlight of the work was the use of *evidential reasoning* to combine classifiers and improve the accuracies of some of the classes. The new canopy stem density estimation algorithm was not a great success. Although a result was obtained, spaceborne imagery was unable to explain most of the variation associated with stem density for the ground sampled forest stands.

One of the highlights of this work was the estimation of uncertainty associated with each parameter image and the final result. It is common in the literature to find results reported without an estimate of the associated uncertainty. For the distance from ridgeline and canopy stem density estimates, the associated uncertainties are in many instances, more than 100% of the estimate itself. However, it is the opinion of this author that the uncertainty provides as much information as the prediction itself. Results are almost worthless if one does not know how good they are. Uncertainty estimation allows us to manage uncertainty and ultimately pinpoint areas which need improvement. In this sense, the uncertainty estimation was very successful.

The performance of the image based parameter estimates as inputs into the model was reasonably successful. In the comparison of the species richness predictions made from field measured inputs versus those made from image-based inputs, this work is quite successful overall. The performance of the image-based inputs for predicting species richness compared to the actual field measured (*observed*) species richness was not as satisfactory. However, given that the model used for this work is still under development, the results shown here are not without promise of improvement.

Overall, the primary objective of mapping the spatial distribution of biodiversity in boreal forest was met. Based on the criteria set forth at the beginning of this chapter, this work should be considered a success in almost all aspects.

8.2 Recommendations

Given the resources and option to extend this work, the following recommendations would be made:

1. The estimation of canopy stem density was highly problematic. The use of the SAR P-band might improve the correlation with stem density because the longer wavelength energy can penetrate deeper into the canopy where it can have more intense backscatter interactions with the tree trunks. Failing this, an alternative variable for canopy stem density would be suggested (for the biodiversity model). Since both canopy stem density and canopy type are surrogate variables for light transmission to the forest floor, an estimate of leaf area index (LAI) might prove more effective in the biodiversity model. Much work has been done in the realm of LAI estimation for forest canopies and it may provide a good alternative. This suggestion, however, would require revisiting the ground sampled forest stands with a light meter in order that correlations be made between LAI and spaceborne imagery.

2. The classification of canopy type suffered tremendously from the nature of the ground sampled training sites. In the boreal mixwood forest, stands which are 100% species pure are virtually non-existent so it is naive to expect that ground sampled training sites should be pure either. However, the mixture of species within the training sites proved to be the largest obstacle in the classification process. Classification improvement probably lies in redefining the canopy type categories used in the model to reflect the species mixtures found in nature. This variable undoubtedly is correlated with canopy stem

density and could also potentially be removed from the model and substituted with an LAI estimate. Failing this, the size and number of ground sampled training sites should be expanded in order to better represent each canopy class in the image classification training procedure. Recall that in order to increase the number of training pixel's, the sites were expanded. If the sites were revisited and circumnavigated with a GPS receiver, we would have a very accurate representation of the actual spatial bound of each stand. In this way, sites could be expanded without fear of including other stands within them.

3. As was noted in Chapter 7, the upper and lower bounds of uncertainty on the final predicted species richness map represented the *worst case* scenario, in which the maximum amount of expected error from each input variable was propagated through the model. Common sense tells us that this situation is highly improbable so we are left with an ultra-conservative representation of uncertainty. Ideally, a method would be devised which takes into account the probability of the worst case error occurring for each input variable for each species richness prediction. The estimated uncertainty would surely decrease and the results would become more meaningful.

4. The distance from ridgeline algorithm could potentially be improved by incorporating contour information or the aspect information of *each* pixel in determining the pathway to the ridge. The first would require a solution which encompassed both the raster and vector domain. A pixels location in raster space could be converted to a point in vector space. For the point, a path which minimizes the distance between each contour as a path is drawn to the

ridge would be found. The distance of this path would then be recored for the pixel. This would be a very computationally intensive solution.

Bibliography

Aber, J. and Milillo, J. (1991). *Terrestrial Ecosystems.* Saunders College Publishing, Philadelphia.

ARC/INFO (1997). Environmental Systems Research Institute. Version 7.2.1.

Beven, K. (1997). TOPMODEL: a critique. *Hydrological Processes*, 11:1069–1085.

Beven, K. and Kirkby, M. (1979). A physically based variable contributing area model of catchment hydrology. *Hydrol. Science Bulletin*, 24:43–49.

Bormann, F. and Likens, G. (1979). *Patterns and Process in a Forested Ecosystem.* Springer-Verlag, New York, NY.

Bridge, S. (1997). The landscape scale spatial distribution of vegetation gradients in a mixedwood boreal forest: Linking ecological patterns to geomorphic processes across scales. Master's thesis, Department of Biology, Univeristy of Calgary, Calgary, AB.

Briggs, D., Smithson, P., and Ball, T. (1989). *Fundamentals of Physical Geography.* Copp Clark Pitman Ltd.

Cameron, H. (1953). Melting point of the bonding material in lodgepole pine and jack pine cones. Silavculture Leaflet 86, Canada Department of Resources and Development, Forestry Branch.

Chapman, B. (1995). *SIR-C Data Compression Software User Guide.* NASA Jet Propulsion Lab, Pasadena, California.

Chapman, M. (1988). *Processing GPM Dense Digital Elevation Models.* PhD thesis, Laval University.

Chipman, S. (1999). Comparing the relative contribution of geomorphic and disturbance processes to landscape plant diversity. Unpublished Research Proposal.

Collins, M. and Livingston, C. (1996). On the dimensionality of multiparameter microwave image data from thin sea ice in the Labrador Sea. *IEEE Transactions on Geoscience and Remote Sensing*, 34(1):114–136.

Cottam, G. and Curtis, J. (1956). The use of distance measures in phyosociological sampling. *Ecology*, 37(3):451–460.

Dempster, A. (1967). Upper and lower probabilities induced by multi-valued mapping. *Annals of Mathematical Statistics*, 38:325–339.

Garvey, T., Lowrance, J., and Fischler, M. (1981). An inference technique for integrating knowledge from disparate sources. In Drinan, A., editor, *Proceedings of the 7th International Conference on Artificial Intelligence*, volume 1, pages 219–325, Vancouver, BC.

Goodchild, M. and Mark, D. (1987). The fractal nature of geographic phenomena. *Annals of the Association of American Geographers*, 77(2):265–278.

Halpern, C. and Spies, T. (1995). Plant species diveristy in natural and managed forests of the pacific northwest. *Ecologicl Applications*, 5(4):913–933.

Hardin, P. (1994). Parametric and nearest-neighbour methods for hybrid classification: A comparison of pixel assignment accuracy. *Photogrammetric Engineering and Remote Sensing*, 60(12):1439–1448.

Harper, J. and Hawksworth, D. (1995). *Biodiversity Measurement and Estimation*, chapter Preface. The Royal Society and Chapman & Hall, London, UK.

Husch, B., Miller, C., and Beers, T. (1972). *Forest Mensuration.* The Ronald Press Company, New York.

Huston, M. (1994). *Biological Diversity The coexistence of species on changing landscapes.* Cambridge University Press, Cambridge.

Jenson, S. and Dominique, J. (1998). Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing*, 54(11):1593–1600.

Johnson, E. (1992). *Fire and Vegetation Dynamics.* Cambridge University Press.

Kim, H. and Swain, P. (1995). Evidential reasoing approach to multi-data classification in remote sensing. *IEEE Transactions on Systems, Man and Cybernetics*, 25(8):1257–1265.

Kirkby, M. (1007). TOPMODEL: a personal view. *Hydrological Processes*, 11:1087–1097.

Krakiwsky, E. and Gagnon, P. (1987). Least squares adjustment. In Krakiwsky, E., editor, *Papers for the CISM adjustment and analysis seminars*, pages 108–149. Canadian Institute of Geomatics, 2nd edition.

Kurvonen, L., Pulliainen, J., and Hallikainen, M. (1999). Retrieval of biomass in boreal forests from multitemporal ERS-1 and JERS-1 SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1):198–205.

Le Hégarat-Mascle, S., Bloch, I., and Vidal-Madjar, D. (1998). Introduction of neighborhood information in evidence theory and applications to data fusion of radar and optical images with partial cloud cover. *Pattern Recognition*, 31(11):1811–1823.

Le Toan, T., Beaudoin, J., and Guyon, D. (1992). Relating forest biomass to SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2):403–411.

Lee, J., Richards, J., and Swain, P. (1987). Probabilistic and evidential approaches for multisource data analysis. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(3):283–293.

Lieffers, V., Messier, C., Gendron, F., Comeau, P., and Stadt, K. (1998). Predicting and managing light in the understory of boreal forests. Paper under review for publication.

Lilliefors, H. (1997). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *American Statistical Association Journal*, pages 399–402.

Lopes, A., Nezry, E., Touzi, R., and Laur, H. (1993). Structure detection and statistical adaptive speckle filtering in SAR images. *International Journal of Remote Sensing*, 14(9):1735–1758.

Magurran, A. (1988). *Ecological Diversity and Its Measurement.* Princeton University Press, Princeton, NJ.

MATLAB (1997). The Mathworks Inc. Version 5.3.

May, R. (1975). *Ecology and Evolution of Communities*, chapter Patterns of Species Abundance and Diversity. Harvard University Press, Cambridge, Mass. Cody, M.L. and Diamond, J.M., editors.

Moon, W. (1990). Integration of geophysical and geological data using evidential belief function. *IEEE Transactions on Geoscience and Remote Sensing*, GE-28(4).

NASA Jet Propulsion Lab (1993). SIR-C CEOS Tape Reader v2.3.

NASA Jet Propulsion Lab (1994). SIR-C Data Compression.

Oliver, C. and Larson, B. (1990). *Forest Stand Dynamics.* McGraw-Hill Inc., New York.

O'Loughlin, E. (1986). Prediction of surface saturation zones in natural catchments by topographic analysis. *Water Resources Research*, 22(5):794–804.

PCI (1997). PCI remote sensing corportation. Software Version 6.3.

Peddle, D. (1993). An empirical comparison of evidential reasoning, linear discriminant analysis and maximum likelihood algorithms for alpine land cover classification. *Canadian Journal of Remote Sensing*, 19(1):31–44.

Peddle, D. (1995). Mercury: An evidential reasoning image classifier. *Computers and Geosciences*, 21(10):1163–1176.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical Recipes in C.* Cambridge University Press, Cambridge, second edition.

Ranson, K., Saatchi, S., and Sun, G. (1995). Boreal forest ecosystem characterization with SIR-C/XSAR. *IEEE Transactions on Geoscience and Remote Sensing*, GE-33(4):867–876.

Ranson, K. and Sun, G. (1994). Mapping biomass of a northern forest using multifreqency SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2):388–395.

Richards, J. (1993). *Remote Sensing Digital Image Analysis.* Springer-Verlag, Berlin, Germany, second edition.

Schowengerdt, R. (1997). *Remote Sensing Models and Methods For Image Processing.* Academic Press, San Diego, 2nd edition.

Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, NJ.

Siegal, S. and Castellan, N. (1988). *Nonparametric Statistics For the Behavioural Sciences.* McGraw-Hill Book Co., New York, 2nd edition.

Singh, T. (1982). Biomass equations for ten major tree species of the prairie provinces. Technical Report NOR-X-242, Northern Forest Research Centre, Canadian Forestry Service.

Skidmore, A. (1990). Terrain position as mapped from a gridded digital elevation model. *International Journal of Geographical Information Systems*, 4(1):33–49.

Smith, T. and Huston, M. (1989). A theory of spatial and temporal dynamics of plant communities. *Vegetatio*, 83:49–69.

Taylor, J. (1982). *An Introduction to Error Analysis.* Oxford University Press.

Tuttle, S. (1980). *Landforms and Landscapes.* WM. C. Brown Company Publishers, 3rd edition.

Vuu, C., Wong, C., and Barret, P. (1995). *SIR-C CEOS Tape Reader User's Guide.* NASA Jet Propulsion Lab, Pasadena, California.

Weir, J. (1996). The fire frequency and age mosaic of a mixedwood boreal forest. Master's thesis, Department of Biology, Univeristy of Calgary, Calgary, AB.

Whittaker, R. (1977). *Evolutionary Biology*, volume 10, chapter Evolution of Species Diversity in Land Comminities, pages 1–67. Plenum.

Wigmosta, M., Vail, L., and Lettenmaier, D. (1994). A distributed hydrology-vegetation model for complex terrain. *Water Resources Research*, 30(6):1665–1679.

Wilson, B. (1996). Estimating coniferous forest structure using SAR texture and tone. *Canadian Journal of Remote Sensing*, 22(4):382–389.

Appendix A

Canopy Species Compositions For Ground Sites

Table A.1: Ground site canopy compositions in percentages. This table has been compiled from the work of Bridge (1997). Note that sites 82, 83 and 91 were recently burned and thus, consist of understory trees only.

Site	BF	PB	WS	BS	JP	BP	TA	CSD
1	50.36	2.71	42.80					602.2
2		2.55	2.62				94.82	610.4
3			2.63		94.71		2.66	613.7
4			28.15	51.53			20.32	1251
5	7.63	13.85	41.80			25.23	11.50	526.3
6			16.41		2.70	14.92	65.97	731.1
7					50.10		49.90	1899.2
9		8.63	35.57	3.41	34.44		17.95	1198.7
10		2.03	19.93	6.38			71.66	663
11			33.59	55.32		8.87	2.31	2149
12		1.63	27.06	45.98	4.66		20.68	2412
13		8.85	53.36				37.78	778.7
14	3.05	1.99	46.37				48.28	983.7
15		52.63	19.22		4.35		23.79	1478.3
16			28.01	2.87	64.54		4.58	1191.1
17		3.35	27.45	3.42	11.98		53.79	1143.9
18	69.53		21.21				8.15	591.5
19		22.84	51.86	2.15			23.15	2091.19
20	6.42		67.22			3.19	23.17	998.8

BF - Balsam fir PB - Paper birch
WS - White spruce BS - Black spruce
JP - Jack pine BP - Balsam poplar
TA - Trembling aspen CSD - Canopy stem density (trees/ha)

Continued on next page....

Site	BF	PB	WS	BS	JP	BP	TA	CSD
21	20.89	29.87	13.47				35.77	1103.7
22			57.13				42.87	1474.4
23			64.25	8.73			27.03	1190.7
24		11.37				2.48	86.15	817.2
25		28.93	13.19			11.63	46.24	367.6
26		15.92	21.15				62.93	1231.9
27							100.00	1084.7
28	2.05	4.73	16.78				76.44	947.9
29				81.31		11.20	7.50	3390.7
30			13.15	51.77		1.74	33.34	1280.1
31	2.26	11.49	34.50			7.99	43.76	1008
32		6.82	30.81			2.12	60.25	1233
33		2.67				9.61	87.71	824.5
34		2.59		8.17	89.24			1559.4
35			6.24	93.76				3002.4
36			10.51	47.83	24.25	8.31	2.55	1307.7
37			28.85		4.50	33.90	32.76	782.9
38		6.42					93.58	1400.1
39			5.11			15.97	78.92	974.3
40	12.43	16.46	30.23			3.73	37.15	1406.4
41		2.33	31.81	12.71			53.15	585
42	1.65	14.92	46.04				37.38	660.8
43			88.20			5.57	6.23	1357.5
44		2.37	47.59		2.16		47.88	446.9
45		13.34	44.45		2.49	8.51	31.20	785.7
46		44.54	39.51			10.04	5.91	1194.1
47		7.48	11.61	3.36	49.27		28.28	1183.2
48					100.00			1586.9
49			7.81		69.42		22.21	572
50		2.28	22.55				75.16	231.1

Continued on next page....

Site	BF	PB	WS	BS	JP	BP	TA	CSD
51		31.70	4.24			8.75	55.31	505.7
52		1.89	63.11	2.50		24.33	8.17	927.1
53	70.07	5.14	24.80					1844.6
54	7.90	15.37	39.77				36.95	1135.5
55		34.13	65.87					1343.6
56		5.46	48.99	11.13	7.77		26.65	1209.9
57		24.34	39.26				36.40	1384.5
58		15.04	31.01			3.87	50.09	354.7
59		15.00	85.00					1489.6
60		6.09	65.53			2.20	26.18	610.7
61				63.41	33.77		2.82	966.2
62			9.12	27.49	46.95		16.44	1361.6
63	18.48		42.88				38.64	764.9
64			62.54	37.46				1625.9
65	3.65	15.81	51.06		6.15		23.33	938.2
66		2.64	88.82	2.49	3.35		2.69	1210.5
67	9.12		42.78	48.10				2901.4
68	54.02	5.15	40.82					279.9
69	1.77		29.74	3.95			64.54	757.9
70		4.36	77.77	8.90			8.97	1824.7
71		1.98	25.97				72.05	1384.7
72		2.208	13.64	3.276	3.331		77.54	390.4
73A					93.04		6.961	3900.6
74	59.31	8.621	18.2				13.86	372.8
75			10.76			7.146	82.1	734.1
76	4.821		60.33		29.85		4.993	5.8
77					100			738.1
78		2.803		7.19	90.01			3075
79			81.53				18.47	159.7
80			2.816			5.519	91.67	886.9

Continued on next page....

Site	BF	PB	WS	BS	JP	BP	TA	CSD
81					52.75		47.25	437.6
82								0
83								0
84						41.6	58.4	879.5
85					100			202.7
86						7.304	92.7	1569.7
87						5.989	94.01	957.6
88					100			1934.7
89				20.93	73.73	2.258	3.081	2982.8
90	35.55	1.771	41.76			3.448	17.48	1456.1
91								0
92				3.553	92.93		3.516	200.9
93		8.732	9.47				81.8	749.6
94		5.781	37.57			7.016	49.64	976.4
95	18.44	15.05	20.98	45.53				892.8
96	3.445	17.1	37.8				41.66	1157.2
97		12.69	44.53			6.14	36.63	736.4
98						3.052	96.95	1119.8
99			7.62		92.38			531.6
100					96.95		3.049	955.9
101			46.22	20.3	26.49	6.989		1722.9
102			12.58	72.22	7.879	7.323		2322.9

Appendix B

Classification Test Results

	JP	BS	TA	Mix†	Water	Anthro	Wetland
etrs3	45 97.8	0 0.0	1 2.2	0 0.0	0 0.0	0 0.0	0 0.0
*etrs7	3 37.5	1 12.5	4 50.0	0 0.0	0 0.0	0 0.0	0 0.0
*etrs9	9 52.9	1 5.9	2 11.8	5 29.4	0 0.0	0 0.0	0 0.0
etrs16	5 18.5	2 7.4	8 29.6	12 44.4	0 0.0	0 0.0	0 0.0
etrs34	57 83.8	8 11.8	1 1.5	2 2.9	0 0.0	0 0.0	0 0.0
etrs47	2 10.5	0 0.0	16 84.2	1 5.3	0 0.0	0 0.0	0 0.0
etrs48	25 100.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs49	23 74.2	0 0.0	7 22.6	1 3.2	0 0.0	0 0.0	0 0.0
etrs73	53 98.1	1 1.9	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
*etrs76	0 0.0	0 0.0	6 75.0	2 25.0	0 0.0	0 0.0	0 0.0
etrs77	51 96.2	1 1.9	0 0.0	0 0.0	0 0.0	0 0.0	1 1.9
etrs78	19 61.3	2 6.5	5 16.1	5 16.1	0 0.0	0 0.0	0 0.0
*etrs81	14 42.4	0 0.0	18 54.5	0 0.0	0 0.0	0 0.0	1 3.0
*etrs85	0 0.0	0 0.0	1 2.7	0 0.0	0 0.0	0 0.0	36 97.3
etrs88	39 100.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs92	6 100.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs99	47 94.0	0 0.0	1 2.0	0 0.0	0 0.0	0 0.0	2 4.0
etrs100	34 97.1	1 2.9	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0

† Mix = Trembling Aspen, White Spruce, Balsam Fir

*etrs = Site was subsequently removed from the classification

Table B.1: kNN classification testing of Jack Pine ground sampled sites. For each site, the first row shows the number of pixels classified into each class. The second row indicates the percentage of pixels classified into each class.

	JP	BS	TA	Mix†	Water	Anthro	Wetland
etrs11	0 0.0	12 52.2	0 0.0	11 47.8	0 0.0	0 0.0	0 0.0
etrs29	0 0.0	9 52.9	0 0.0	8 47.1	0 0.0	0 0.0	0 0.0
etrs35	0 0.0	8 53.3	0 0.0	7 46.7	0 0.0	0 0.0	0 0.0
etrs36	7 15.2	6 13.0	11 23.9	22 47.8	0 0.0	0 0.0	0 0.0
etrs61	17 68.0	8 32.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs62	14 19.7	28 39.4	0 0.0	29 40.8	0 0.0	0 0.0	0 0.0
etrs101	0 0.0	19 82.6	0 0.0	4 17.4	0 0.0	0 0.0	0 0.0
etrs102	3 9.7	21 67.7	0 0.0	7 22.6	0 0.0	0 0.0	0 0.0

Table B.2: kNN classification testing of Black Spruce ground sampled sites.

	JP	BS	TA	Mix†	Water	Anthro	Wetland
etrs2	0 0.0	0 0.0	26 92.9	2 7.1	0 0.0	0 0.0	0 0.0
*etrs5	2 8.3	2 8.3	10 41.7	10 41.7	0 0.0	0 0.0	0 0.0
etrs6	0 0.0	0 0.0	23 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs10	0 0.0	0 0.0	33 100.0	0 0.0	0 0.0	0 0.0	0 0.0
*etrs13	0 0.0	0 0.0	6 85.7	1 14.3	0 0.0	0 0.0	0 0.0
*etrs14	0 0.0	0 0.0	14 63.6	8 36.4	0 0.0	0 0.0	0 0.0
etrs24	0 0.0	0 0.0	8 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs25	0 0.0	0 0.0	23 95.8	1 4.2	0 0.0	0 0.0	0 0.0
etrs26	0 0.0	0 0.0	20 95.2	1 4.8	0 0.0	0 0.0	0 0.0
etrs27	0 0.0	0 0.0	24 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs28	0 0.0	0 0.0	28 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs32	0 0.0	0 0.0	8 88.9	1 11.1	0 0.0	0 0.0	0 0.0
etrs33	0 0.0	0 0.0	30 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs37	2 11.1	2 11.1	12 66.7	2 11.1	0 0.0	0 0.0	0 0.0
etrs38	0 0.0	0 0.0	12 92.3	1 7.7	0 0.0	0 0.0	0 0.0
etrs39	0 0.0	0 0.0	28 100.0	0 0.0	0 0.0	0 0.0	0 0.0
*etrs41	7 23.3	2 6.7	16 53.3	5 16.7	0 0.0	0 0.0	0 0.0
*etrs42	0 0.0	0 0.0	30 96.8	1 3.2	0 0.0	0 0.0	0 0.0
etrs44	6 14.3	0 0.0	36 85.7	0 0.0	0 0.0	0 0.0	0 0.0
etrs45	2 4.4	0 0.0	40 88.9	3 6.7	0 0.0	0 0.0	0 0.0
*etrs46	0 0.0	0 0.0	24 100.0	0 0.0	0 0.0	0 0.0	0 0.0
*etrs54	0 0.0	0 0.0	33 70.2	14 29.8	0 0.0	0 0.0	0 0.0
*etrs57	0 0.0	0 0.0	21 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs58	0 0.0	0 0.0	26 89.7	3 10.3	0 0.0	0 0.0	0 0.0
*etrs60	4 12.9	0 0.0	23 74.2	4 12.9	0 0.0	0 0.0	0 0.0
etrs69	1 2.9	0 0.0	33 97.1	0 0.0	0 0.0	0 0.0	0 0.0
etrs71	0 0.0	0 0.0	40 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs72	3 42.9	0 0.0	4 57.1	0 0.0	0 0.0	0 0.0	0 0.0
etrs75	1 4.2	0 0.0	22 91.7	1 4.2	0 0.0	0 0.0	0 0.0
etrs80	0 0.0	0 0.0	8 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs82	0 0.0	0 0.0	8 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs83	1 7.1	0 0.0	10 71.4	0 0.0	0 0.0	2 14.3	1 7.1
etrs84	0 0.0	0 0.0	35 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs86	0 0.0	0 0.0	47 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs87	0 0.0	0 0.0	27 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs91	0 0.0	0 0.0	7 100.0	0 0.0	0 0.0	0 0.0	0 0.0
etrs96	0 0.0	0 0.0	23 100.0	0 0.0	0 0.0	0 0.0	0 0.0
*etrs97	3 10.7	0 0.0	23 82.1	2 7.1	0 0.0	0 0.0	0 0.0
etrs98	0 0.0	0 0.0	29 100.0	0 0.0	0 0.0	0 0.0	0 0.0

Table B.3: kNN classification testing of Trembling Aspen ground sampled sites.

	JP	BS	TA	Mix†	Water	Anthro	Wetland
etrs1	1 16.7	1 16.7	0 0.0	4 66.7	0 0.0	0 0.0	0 0.0
eetrs4	0 0.0	9 45.0	0 0.0	11 55.0	0 0.0	0 0.0	0 0.0
etrs12	8 44.4	0 0.0	6 33.3	4 22.2	0 0.0	0 0.0	0 0.0
*etrs15	0 0.0	0 0.0	26 100.0	0 0.0	0 0.0	0 0.0	0 0.0
*etrs17	0 0.0	0 0.0	5 62.5	3 37.5	0 0.0	0 0.0	0 0.0
etrs18	1 4.0	5 20.0	3 12.0	16 64.0	0 0.0	0 0.0	0 0.0
*etrs19	0 0.0	0 0.0	31 96.9	1 3.1	0 0.0	0 0.0	0 0.0
etrs20	3 10.7	1 3.6	14 50.0	10 35.7	0 0.0	0 0.0	0 0.0
*etrs21	0 0.0	0 0.0	15 93.8	1 6.2	0 0.0	0 0.0	0 0.0
etrs22	0 0.0	4 36.4	0 0.0	7 63.6	0 0.0	0 0.0	0 0.0
etrs23	0 0.0	0 0.0	4 50.0	4 50.0	0 0.0	0 0.0	0 0.0
*etrs30	7 77.8	0 0.0	0 0.0	2 22.2	0 0.0	0 0.0	0 0.0
etrs31	0 0.0	0 0.0	20 95.2	1 4.8	0 0.0	0 0.0	0 0.0
etrs43	0 0.0	4 28.6	1 7.1	9 64.3	0 0.0	0 0.0	0 0.0
etrs53	0 0.0	2 20.0	0 0.0	8 80.0	0 0.0	0 0.0	0 0.0
etrs55	1 9.1	6 54.5	0 0.0	4 36.4	0 0.0	0 0.0	0 0.0
*etrs56	1 11.1	0 0.0	2 22.2	0 0.0	0 0.0	6 66.7	0 0.0
*etrs64	0 0.0	4 50.0	0 0.0	4 50.0	0 0.0	0 0.0	0 0.0
etrs65	1 9.1	0 0.0	0 0.0	10 90.9	0 0.0	0 0.0	0 0.0
etrs66	0 0.0	1 11.1	0 0.0	8 88.9	0 0.0	0 0.0	0 0.0
etrs70	8 36.4	4 18.2	5 22.7	5 22.7	0 0.0	0 0.0	0 0.0
etrs74	0 0.0	1 8.3	0 0.0	11 91.7	0 0.0	0 0.0	0 0.0
etrs79	15 57.7	6 23.1	0 0.0	5 19.2	0 0.0	0 0.0	0 0.0
etrs90	6 12.8	22 46.8	1 2.1	18 38.3	0 0.0	0 0.0	0 0.0

* = site subseqently moved to black spruce class

Table B.4: kNN classification testing of Mixed ground sampled sites.

June TM - unexpanded sites

class	JP	BS	TA	MIX	WAT	ANT	WET	Percent
JP	53	0	12	8	0	4	8	62.35
BS	8	22	1	7	0	0	0	57.89
TA	9	4	151	14	0	2	0	83.89
MIX	12	30	31	31	0	6	1	27.93
WAT	0	0	0	0	199	0	0	100.00
ANT	1	0	1	1	0	163	33	81.91
WET	0	0	0	0	0	1	199	99.50

August TM - unexpanded sites

class	JP	BS	TA	MIX	WAT	ANT	WET	Percent
JP	37	2	10	19	0	15	0	44.58
BS	3	2	2	34	0	0	0	4.88
TA	23	0	118	14	0	0	0	76.13
MIX	26	7	31	36	0	0	0	36.00
WAT	0	0	0	0	149	0	0	100.00
ANT	5	0	7	0	0	95	51	60.13
WET	8	0	5	0	0	1	156	91.76

June TM - expanded sites

class	JP	BS	TA	MIX	WAT	ANT	WET	Percent
JP	318	24	21	21	0	0	7	81.33
BS	33	59	10	68	0	0	2	34.30
TA	38	1	402	33	0	5	0	83.92
MIX	48	12	77	58	0	6	0	28.86
WAT	0	0	0	0	233	0	0	100.00
ANT	17	2	19	1	0	407	34	84.79
WET	8	3	0	0	0	0	167	93.82

August TM - expanded sites

class	JP	BS	TA	MIX	WAT	ANT	WET	Percent
JP	245	13	30	34	0	13	2	72.70
BS	31	67	15	59	0	0	0	38.95
TA	68	3	358	22	0	0	0	79.38
MIX	60	15	76	24	0	0	0	13.71
WAT	0	0	0	0	149	0	0	100.00
ANT	7	0	5	0	0	96	50	60.76
WET	13	0	7	0	0	1	149	87.65

Table B.5: Summary tables of kNN results for testing sites using 7 band TM.

Appendix C

Choosing Classes For Canopy Classification

The forest classes used in this study were determined by Bridge (1997) and subsequently adopted by Chipman (1999). Bridge performed a principal components analysis (PCA) on a *habitat matrix*. First, a matrix with dimensions of forest sites or stands (n) by species (s) with the cells containing species abundance scores is formed. The matrix is then analyzed to remove the effect of time-since-fire (temporal variation) and a new matrix, the *habitat matrix*, is formed. The first two components contained most of the variation and in identifying each stand by a dominant canopy type Bridge was able to form five distinct tree species clusters on a two dimensional PCA plot (of the first two principle components). These were jack pine, black spruce, white spruce, trembling aspen and balsam fir.

The position of each stand was located on a map of geomorphology and the surficial material (glaciofluvial or glacial till) and hillslope position on which the stand fell was determined. The centroids of the topographic positions were plotted on the first two components of the *habitat matrix* PCA by surficial material type. These centroids showed a remarkable affinity to the location of the dominant species type clusters. Bridge hypothesized that the clusters could be explained by environmental information such as moisture and nutrient gradients and treated the first two principal components as such. Ultimately, this hypothesis led to his theory that vegetation composition can be explained by the relative position of a stand on a hillslope.

The dominant species classes were then adopted by Chipman (1999) for her biological diversity prediction model. Since species richness was very similar for balsam fir, white spruce and some of the trembling aspen stands, Chipman formed a new *mixed* class. These four classes (black spruce, jack pine, trembling aspen and mixed) were the basis for the image classification of this study.

Appendix D

Data Summary For The Project

This appendix summarises the data discussed in Chapter 2.

	Satellite	Sensor	Bands	Date Acquired
1	Landsat 5	Thematic Mapper	1-7	June 10, 1996
2	Landsat 5	Thematic Mapper	1-7	August 29, 1996

Table D.1: Summary of available electro-optical satellite image data.

	IA range (deg)	Ground Res.	Data Type	Polarization	Date Acquired
1	38.05 - 43.33	12.5m	MLC	HH,HV,VH,VV	Oct 4, 1994
2	55.87 - 60.30	12.5m	MLC	HH,HV	Oct 6, 1994

(IA = Incidence Angle , MLC = Multi-look Complex)

Table D.2: Summary of available SIR-C SAR image data (L and C band).

	Description	Type*	Source
1	Elevation Vectors	V, L	Parks Canada
2	Lake Polygons	V, P	Parks Canada
3	River Vectors	V, L	Parks Canada
4	Time since fire	R, P	The University of Calgary

*(R = raster, V = vector, P = polygon, L = line)

Table D.3: Summary of available GIS data in ArcInfo format.

Appendix E

System Hardware and Software Specifications

Machine:	SUN SPARC Ultra 1 Sparc
	200 MHz CPU, 128 MB RAM
OS:	SUN Solaris 2.6
Compiler:	GCC - Gnu C/C++ Compiler (v2.8.1)
	G77 - Gnu Fortran Comiler (v0.5.23)
Commercial Software:	PCI EASI/PACE (v6.3), MATLAB (v5.3),
	SIR-C CEOS Tape Reader (v2.3), SIR-C Data Compression
Project Software*:	TSF, DFR, ERC, CERP, CSD, MODPRE, MODERR

Machine:	Bull RISC System/6000
OS:	IBM AIX (v4.1)
Commercial Software:	ARC/INFO (v7.2.1), ArcView (v3.1)
	SPSS (v6.1)

*All code developed for this project was written in C++. Many of these programs are executed in a PCI EASI command line environment. By linking the C++ code to the PCI C language libraries, the programmer and the user are able to make use of PCI image processing functionality.