An Investigation Into the Use of Artificial Neural Networks and Landsat Thematic Mapper Imagery for Vegetation Classification in Southwestern British Columbia

by

Joanne Cheryl White

B.Sc., University of Victoria, 1994

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Geography

We accept this thesis as conforming to the required standard

[redacted]

Dr. K.O. Niemann, Supervisor (Department of Geography)

[redacted]

Dr. M.W. Sondheim, Departmental Member (Department of Geography, Adjunct)

[redacted]

Dr. B.J. Hawkins, Outside Member (Department of Biology)

[redacted]

Dr. K.G. Stewart, External Examiner (Department of Economics)

SUPERVISOR: DR. K.O. NIEMANN

ABSTRACT

The goal of this research was the experimentation with and characterization of neural networks for the classification of vegetation from multi-spectral remotely sensed imagery. A literature survey was undertaken to provide a theoretical context for the use of neural networks in remote sensing image classification and to determine the extent of previous research on this topic. The research described in this thesis combines the spatial analysis capabilities of a geographical information system with advanced multi-spectral image analysis procedures. Existing forest inventory information is utilized along with digital elevation data to provide training data and ancillary information useful to vegetation classification. Neural network classification algorithms were used to pre-stratify a Landsat Thematic Mapper image into categories corresponding to five levels of the new hierarchical vegetation inventory system being implemented in British Columbia. Emphasis was placed on the configuration of the neural network and on comparisons with conventional maximum likelihood estimates. Limitations to the success of this approach included the restricted availability of reliable training and reference data for all of the vegetation classes as well as the disparity between spectral classes found in the image data and the desired informational classes of the classification hierarchy.

Examiners:

[redacted]

Dr. K.O. Niemann, Supervisor (Department of Geography)

[redacted]

Dr. M.W. Sondheim, Departmental Member (Department of Geography, Adjunct)

[redacted]

Dr. B.J. Hawkins, Outside Member (Department of Biology)

[redacted]

Dr. K.G. Stewart, External Examiner (Department of Economics)

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

ACKNOWLEDGEMENTS

I am extremely grateful to those who have contributed their time and resources to this research. First and foremost I would like to thank my supervisor, Dr. Olaf Niemann for his support, encouragement and guidance throughout the life of this project. I would also like to extend my appreciation to Dr. Barbara Hawkins and Dr. Mark Sondheim for their welcomed participation as members of my thesis committee. Special thanks to Michal Lodin and John Wakelin, Ministry of Forests, Resources Inventory Branch for providing funding and the Landsat Thematic Mapper data. The technical advice and assistance provided by Tim Salkeld (Ministry of Forests, Resources Inventory Branch) and Dan Sirk (Ministry of Forests, Vancouver Region) regarding the forest inventory data was greatly appreciated as was Thomas Piekutowski's guidance in completing the atmospheric corrections. Last but not least, the technical prowess of Richard Sykes must also be acknowledged with gratitude. On a personal note, I would like to acknowledge the contribution of my family, without whose steadfast moral support, none of this would have been possible. The opportunities they afforded me ultimately enabled the completion of this work, and to them I shall be eternally grateful.

DEDICATION

To my father, Robert John White.

"NOLITE TE BASTARDES CARBORUNDORUM"

CHAPTER ONE

INTRODUCTION

1.1 Overview

The importance of vegetation classification and mapping is clearly evidenced by the number of works completed on this subject (Kuchler and Zonneveld, 1988). The possibility of using satellite data for land cover classification and vegetation mapping has been extensively researched since the beginning of the 1980's (Norwine and Greegor, 1983) and it is generally acknowledged that remote sensing presents a potential source of valuable information in the inventory (at various scales) of vegetation covering the earth's surface. Unfortunately, the application of conventional, statistically based classification methods for the purposes of land cover discrimination have proven to be ineffective in achieving the desired levels of classification accuracy.

The poor classification results generated by these conventional methods are attributable to a number of factors which include limitations in the spectral and spatial resolutions of existing sensors, coupled with unrealistic expectations for the classification output given these aforementioned limitations. Furthermore, land cover is an extremely complex target to detect and classify, primarily due to the heterogeneity of land cover types and the fuzzy boundaries which exist between discrete cover classes. The development of alternative classification techniques that increase the accuracy of image classification (particularly the consistency of accuracy estimates) and improve the discrimination of land cover types is a necessary prerequisite for the production of quality image products and useful data sources for vegetation management.

The British Columbia Ministry of Forests is expanding its forest inventory program to include non-forest and non-commercial vegetation, thereby supplementing the traditional forest inventory with the collection of a wide variety of additional ecological attributes. This new comprehensive vegetation inventory will involve the amassing of significantly more information than previously was considered necessary for management purposes. This expanded vegetation inventory will create new pressures on the Ministry of Forests to effectively collect and store the large volumes of data which will be generated by such an inventory. In addition, because the new inventory is designed to provide data to a greater number of users than the present inventory system, there will be increased demands for the frequent update and quick dissemination of inventory information. The British Columbia Land Cover Classification System was developed to provide the framework for the new vegetation inventory (Resources Inventory Committee, 1997).

Traditionally, 1:40000 air photos have been used as the primary data source for inventory (Resources Inventory Branch, 1996). While this method provides the Ministry of Forests with the level of detail required for its operational inventory, it is not without fault. One of the main concerns has been the consistency of the photo interpretation (Salkeld, 1998). Inventory audits conducted by the Ministry of Forests have shown a wide range in the accuracy of the forest attributes collected from the air photos (Gilbert, 1998). An air photo based inventory is also costly, averaging greater than three dollars per hectare (Resources Inventory Committee, 1996b). Satellite imagery has previously been used by the Ministry of Forests for updating cutblock information, suggesting a possible role for satellite remote sensing in the new

vegetation inventory, specifically with regards to improving the consistency of interpretation and/or reducing costs.

The new vegetation inventory will be completed for each of the seven thousand 1:20000 mapsheets which cover the Province of British Columbia. It would seem logical that the Ministry of Forests would attempt to move towards an inventory system which took advantage of every opportunity for automation. It was initially suggested that the use of medium resolution (30 metre pixel) satellite imagery as a tool for pre-stratifying the vegetation of large land areas into the broader vegetation classes could be one potential component of such a digital, automated inventory system. The vegetation polygons resulting from the pre-stratification process could then be combined with the technology of digital orthophotos. Digital image processing techniques in conjunction with more traditional air photo interpretation methods could then be used to refine the classification, resulting in a more efficient and more consistent method of vegetation classification.

There are a number of problems associated with this rather idealistic conceptualized interaction of satellite and air photo technology. The primary concern is one of spatial accuracy because the resolution differences between the satellite image and the air photo would logically preclude the placement of the coarse satellite derived boundaries on the finer resolution air photos. Furthermore, the pre-stratification of vegetation into the initial broad classes of the classification hierarchy will not alleviate the inconsistencies in the photo interpretation, since the inconsistencies in interpretation occur primarily in the identification of tree species and more detailed attributes such as tree height, crown closure and basal area - not in the discrimination of broad classes

such as vegetated versus non-vegetated or treed versus non-treed. The detailed attributes are essential for an operational forest inventory and clearly could not be discriminated at the thirty metre resolution offered by the satellite imagery. Other more pragmatic suggestions have been put forward for improving the consistency and accuracy of air photo interpretation – most of which involve the use of technology to facilitate the interpretation process by increasing the ease with which supplemental information is provided to the interpreter (Befort, 1988; Leckie et al., 1996).

Having acknowledged the difficulties in incorporating satellite imagery into the operational-scale vegetation inventory, it is important to clarify that satellite data may still provide a useful source of information for the new inventory, particularly in a strategic planning context. In this context, satellite imagery may have potential uses for large area land cover characterization and for documenting historical changes in land cover classes. The use of satellite imagery to provide information to the new vegetation inventory may therefore be useful in some capacity, providing that there is an understanding of the practical limitations inherent in the data and that its use is not extended beyond these limitations.

1.2 Thesis Objectives

There are two primary objectives of this thesis. The first is to assess the information content of Landsat Thematic Mapper imagery relative to the new British Columbia Land Cover Classification System. This will be achieved through an evaluation of the interpretive depth which the classification of Landsat Thematic Mapper Data can consistently achieve. It is hypothesized that, with the addition of ancillary data (slope,

aspect, elevation), consistent classification results could be achieved to at least level three (landscape position) and in some instances to level four (vegetation type). It is important to acknowledge that the interpretive depth attained is this assessment will be inextricably linked to the image processing and classification methods used and that other methods may provide a different assessment of the relevant BCLCCS information content of the Landsat TM.

Relatively poor classification results have been achieved with conventional, statistical based classifiers and this has spawned extensive research into alternative classification techniques. The second primary objective of this thesis will therefore be to compare the classification results of one of these alternative techniques, the neural network classifier, to a standard maximum likelihood classifier. Given the complex spectral response of vegetation, it is hypothesized that the neural network will provide superior results to the maximum likelihood classifier in all of the classification trials.

The configuration of neural networks for image classification is an extremely complex task and this complexity has seriously impeded the operational implementation of neural networks for this purpose. Empirical research has led to the development of several different heuristics for network architecture and a secondary objective of this thesis will be to determine the usefulness of two specific heuristics developed by Paola (1994) and by Kanellopoulos and Wilkinson (1997) and thereby determine an optimal neural network configuration for classifying each of the five levels of the BCLCCS hierarchy within the project study area.

1.3 Organization Of The Thesis

The thesis is organized with the aforementioned objectives in mind and hence, Chapter 2 characterizes the target objects (vegetation) and the satellite remote sensing instrument (Landsat Thematic Mapper) and provides a literature review of the conventional and alternate classification methods which are currently utilized by the remote sensing community. The purpose of the literature review is to provide a broad theoretical context for the use of artificial neural networks as a mechanism for remote sensing image classification. Chapter 3 focuses specifically on artificial neural networks and their application to remote sensing image classification, providing a definition and review of the salient elements of neural network architecture. An emphasis is placed on identifying the advantages and disadvantages of artificial neural networks relative to other techniques, in particular the maximum likelihood classifier. A detailed description of the specific algorithm used for the research reported in this thesis, the multi-layer perceptron, is also provided in Chapter 3.

The research methodology used in developing the image classification trials is outlined in Chapter 4, including a description of the project study area and the specific remotely sensed images used. The British Columbia Land Cover Classification System is explained in this chapter, as are the image processing procedures. The extensive results of the classification trials are presented in Chapter 5 - coupled with a discussion of these results in relation to the theory outlined in Chapter 3 and the heuristics outlined in Chapter 4, Section 4.8.1.1. Comparisons are made between the classification results achieved by trials performed with different neural network architectures and between the neural networks and the maximum likelihood classifier. The presentation of the

results in Chapter 5 are preceded by a rationale for the methods used in the correspondence assessment along with a description of the selected measures of correspondence used to assess the results of the image classification trials. Finally, Chapter 6 will present the conclusions drawn from the classification trials and relate these to the thesis objectives outlined in Chapter 1. An assessment of the information content of the Landsat Thematic Mapper data relative to the new vegetation inventory will be provided and the capabilities, problems and potential future use of neural networks as a tool for discriminating vegetation cover from remotely sensed imagery will be discussed.

CHAPTER TWO

VEGETATION, REMOTE SENSING AND IMAGE CLASSIFICATION

The objective of this and the following chapter is to provide a theoretical framework for the research presented in later sections of this document. The discussion in this chapter will focus on: (1) the spectral properties of the research objects (vegetation); (2) the spectral and spatial properties of the imaging instrument (Landsat Thematic Mapper); and (3) the present status of classification techniques and approaches, including both conventional methods and more recent innovations in remote sensing image classification. From this discussion, the reader should discover that vegetation has a complex spectral response pattern which cannot be readily classified by parametric methods where a normal distribution is assumed and that Landsat Thematic Mapper has been used extensively for vegetation classification applications - with varying degrees of success. Furthermore, it should be apparent that there are wide variety of classification methodologies which are currently being employed by the remote sensing community, both in a research context and in an operational context.

2.1 Spectral Properties of Vegetation

Green vegetation has a spectral reflectance which is distinctive and highly variable at different wavelengths (Hoffer, 1978). Figure 2.1 illustrates the typical

spectral response of healthy green vegetation over the visible, near-infrared and mid-infrared portions of the electromagnetic spectrum.

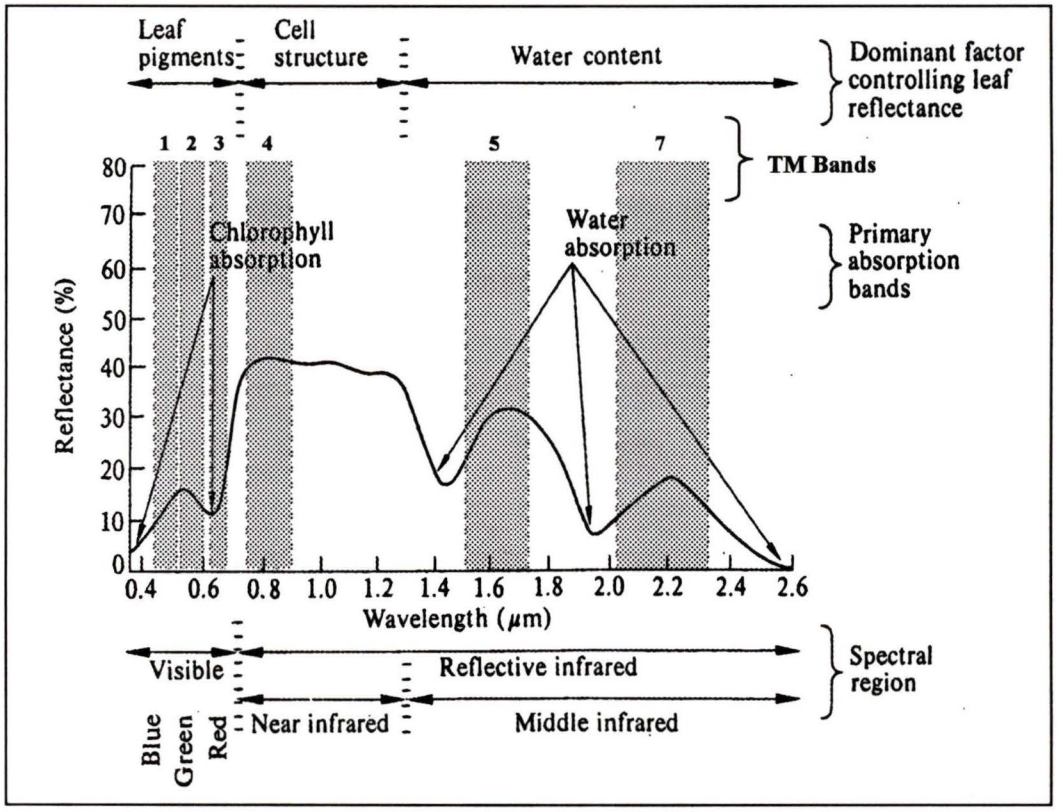


Figure 2.1 Significant Spectral Response Characteristics of Green Vegetation.
(Source: Adapted from Hoffer, 1978)

Incident energy (I_λ), as a function of wavelength (λ), can be described by three components: reflection (R_λ), absorption (A_λ) and transmittance (T_λ), as presented in the energy balance equation (Hoffer, 1978) in Equation 2.1:

$$I_\lambda = R_\lambda + A_\lambda + T_\lambda \qquad \textit{[Equation 2.1]}$$

This equation can be used to account for the spectral response of vegetation at various regions of the electromagnetic spectrum. In the visible portion of the spectrum, pigments in the vegetation are essential in determining the spectral response of

vegetation where the bulk of the incident energy is absorbed and the remainder reflected. Absorption occurs primarily at the blue (0.45 μm) and red (0.65 μm) wavelength regions of the visible portion of the spectrum and these regions are referred to as the chlorophyll absorption bands (Curran, 1980). Relatively less absorption occurs at the green wavelengths (0.54 μm) of the visible portion, resulting in stronger reflectance in this region and the characteristic green colour associated with healthy vegetation (Gates et al., 1965). The ideal spectral region for sensing chlorophyll absorption characteristics are 0.45 μm to 0.52 μm (Knipling, 1970; Tucker, 1977a; Tucker 1977b) and 0.63 μm to 0.69 μm (Wooley, 1971; Tucker, 1978).

Vegetation which is experiencing stress, disease or senescence will produce less chlorophyll, reducing absorption in the chlorophyll absorption bands and resulting in increased reflectance, particularly from the red portion of the visible spectrum (Murtha, 1978). Under these conditions of low chlorophyll production, reflectance in the blue portion of the spectrum will remain modest due to the absorption properties of yellow pigments such as carotenes and xanthophylls which also absorb incident energy in the blue wavelengths. The yellow colour associated with chlorotic vegetation is attributable to changes in the relative levels of these various plant pigments (Tucker, 1978). (In healthy vegetation, the relative absorption of incident energy ascribed to these yellow pigments is negligible in comparison to the level of chlorophyll absorption). Another plant pigment of seasonal importance is anthocyanin - a red pigment produced by some vegetation during the autumnal senescence which gives rise to the vibrant red leaf colours associated with this season (Gates, 1970). The spectral response of vegetation

in the visible portion of the spectrum is thus determined by the relative levels of these various pigments (Jones, 1992).

In the near-infrared (NIR) portion of the spectrum, the interaction between the internal structure of the leaves and the incident solar radiation regulates the spectral response of the vegetation, with approximately half of the incident energy being reflected, half being transmitted and a minimal amount being absorbed (Gates et al., 1965). Compared to the spectral response of vegetation in the visible portion of the spectrum, the response in the NIR is characterized by a dramatic increase in reflectance at 0.7 μm and very high transmittance (Sinclair et al., 1974). With the presence of multiple leaf layers, the amount of reflectance from the vegetation increases through the phenomenon of additive reflectance which is facilitated by the increase in transmittance (Hoffer, 1978).

The spectral response of vegetation in the NIR portion of the spectrum is commonly more diverse than the response in the visible portion of the spectrum and this increased variation has been used to discriminate species types and to detect stress (Curran, 1985). Differences in the internal structure of the leaves between species result in distinct differences in reflectance in the NIR, although differences between species in the visible portion of the spectrum may be negligible (Gausman et al., 1978; Curran, 1980; Curtis, 1980). The NIR has also been shown to exhibit more structural variability for vegetation, particular in areas with dense vegetation (Chavez, 1992). Various forms of vegetation stress can alter the internal structure of the leaf and thereby alter its spectral response in the NIR (Jensen, 1983). The optimum spectral region for detecting biomass is in the NIR, from 0.74 μm to 0.90 μm (Tucker, 1978).

Finally, in the mid-infrared portion of the spectrum, the total moisture content of the vegetation regulates the spectral response and much of the incident energy is absorbed by the water in the leaf and the remainder is reflected (Knipling, 1970; Wooley, 1971; Curtis, 1980). The primary water absorption band is located at 2.7 μm (Hoffer, 1978). Other water absorption bands, each with successively less absorption, occur at 1.9 μm, 1.4 μm, 1.1 μm, and 0.96 μm. The reflectance peaks in the mid-infrared occur at 1.6 μm and 2.2 μm (Gates, 1970). Beyond 1.3 μm, the reflectance of a leaf in this portion of the electromagnetic spectrum is approximately inversely related to the degree of water absorption in the leaf (Thomas et al., 1987). Absorption in turn, is a function of the total amount of water in the leaf, which itself is a function of both the percentage moisture content of the leaf and the leaf thickness. If the moisture content of the leaf decreases, reflectance over all wavelengths increases. Decreasing amounts of moisture influence the internal structure in the leaf, which thereby impacts the spectral response in the NIR. Reflectance peaks in the mid-infrared occur at about 1.6 μm and 2.2 μm between the primary water absorption bands (Jensen, 1983). The optimum region for measuring moisture content in the MIR is between 1.55 μm and 1.75 μm (Tucker and Garrett, 1977; Gausman et al., 1978).

Plant pigments, leaf structure and moisture content control the spectral response of vegetation in the each of the visible, near-infrared and mid-infrared portions of the electromagnetic spectrum. Variations in the aforementioned spectral response characteristics of healthy green vegetation however, may result from differences in the geometry, morphology, and physiology of the leaf, plant and canopy, further

complicating the spectral response (Colwell, 1974; Hoffer, 1978; Thomas et al., 1987). Specifically, these additional factors include the amount, geometry and configuration of cover, the characteristics of the non-leaf components in the scene, the characteristics of the shadow components and other environmental variables such as atmospheric conditions. For example, Colwell (1974) determined that reflectance in the red wavelengths is dependent upon the percent cover, the look angle and the solar zenith angle. Similarly, reflectance in the NIR is positively correlated with crown closure, while reflectance from the visible portion of the spectrum has a negative correlation with crown closure (Colwell, 1974). A change in the orientation of vegetation from horizontal to vertical, caused by wilting or wind increases reflection in the visible and decreases reflection in the NIR (Suits, 1972). The spectral response of vegetation is therefore an extremely complex interaction of chemistry, geometry, environment and wavelength which is clearly difficult to characterize on the basis of spectral information alone.

2.2 Spectral and Spatial Properties of Landsat Thematic Mapper Data

The Thematic Mapper is the second generation multi-spectral scanning instrument in the Landsat program. Conceived as a civilian earth resources satellite program to provide data to resource managers and earth scientists, the first Landsat satellite, Landsat-1 was launched on July 23, 1972 (Earth Observation Satellite Company, 1994a; Scheffner, 1996). This satellite carried the predecessor to TM, the Multi-Spectral Scanner (MSS) instrument. Landsat-4, launched in 1982, was the first Landsat satellite to carry the TM sensor. Compared to the MSS, the TM sensor has an improved spatial

resolution, additional spectral bands and improved geometric precision (Avery and Berlin, 1992). Landsat-5, the sensor currently used operationally for routine TM data collection, was launched in 1985. Landsat-5 has a repetitive, circular, sun-synchronous, near polar orbit at an altitude of 705 kilometres (Lillesand and Kiefer, 1987). The satellite provides a sixteen day, 233-orbit cycle with a swath overlap that varies from seven percent at the equator to nearly 84% at 81° north or south latitude. The 14.5 orbits which are completed each day have an inclination angle of 98.2° with respect to the equator (Hord, 1982; Curran, 1985).

Radiometrically, the wavelength range for the TM sensor extends from the visible (blue), through the mid-infrared into the near-infrared portions of the electromagnetic spectrum. The widths of the reflective TM bands are provided in Table 2.1 and illustrated in Figure 2.1. Sixteen detectors for the visible, near-infrared and mid-infrared wavelength bands in the TM sensor provide 16 scan lines on each active scan and four detectors for the thermal-infrared provide four scan lines on each active scan (Lillesand and Kiefer, 1987). The TM scanner is bi-directional, collecting data in both the forward and backward sweeps of its scan mirror, which has scan angles of ± 7.7° from nadir (Earth Observation Satellite Company,1994a). The sensor has a spatial resolution of 30 metres for the visible, near-infrared and mid-infrared wavelengths and a spatial resolution of 120 metres for the thermal-infrared band. The analog to digital signal conversion is quantified over 256 data numbers, giving the TM data 8-bit radiometric precision (Lillesand and Kiefer, 1987).

Table 2.1 The Six Reflective TM Bands.

Band	Description	Sensitivity (μm)
1	blue (B)	0.45-0.52
2	green (G)	0.52-0.60
3	red (R)	0.63-0.69
4	near-infrared (NIR)	0.76-0.90
5	mid-infrared (MIR)	1.55-1.75
7	2nd mid-infrared (MIR2)	2.08-2.35

2.2.1 Vegetation Classification Using Thematic Mapper Data

The spectral and spatial properties of TM imagery are well suited to vegetation discrimination. The narrow green band of the TM sensor (band 2) covers the green reflectance peak from vegetation and the narrow red band (band 3) is spectrally located within the chlorophyll absorption region. Band 4 operates in the best spectral region for distinguishing vegetation types and conditions and band 5 is responsive to changes in leaf tissue water content (Lillesand and Kiefer, 1987; Avery and Berlin, 1992). The majority of these bands are located in the optimal spectral locations of the electromagnetic spectrum for detecting features relevant to vegetation discrimination, as outlined in Section 2.1.

The question of which bands or band combinations are most effective for vegetation discrimination is not readily determined from an examination of the various raw and synthetic bands and band combinations which have been used in previous research. It is generally acknowledged that one band from each of the three spectral zones (visible, near-infrared and mid-infrared) is useful for the general discrimination among forest vegetation (Benson and De Gloria, 1985; Ahern and Archibald, 1986). A combination of bands 3, 4 and 5 is most often used, with band 4 providing the most

information (Shen et al., 1985; Hepner, 1990; Moore and Bauer, 1990; Bolstad and Lillesand, 1992; Saxena et al., 1992; Conese and Maselli, 1993; Sader, 1995). Previous research has confirmed that the combination of bands 3,4,5 represents the majority of information contained in TM data collected in northern temperate forest areas and also provides a suitable equipoise between processing efficiency and classification accuracy (Nelson et al., 1984; Horler and Ahern, 1986; Moore and Bauer, 1990; Bolstad and Lillesand, 1992). Some studies have however reported using all six reflective TM bands (Belward et al., 1990; Blonda et al, 1991; Civco, 1993).

Correlation matrices derived from imagery taken over forest environments have indicated that there is a high degree of correlation amongst the visible bands (1,2 and 3) and also a high degree of correlation between the mid-infrared bands 5 and 7 (Beaubien, 1994). Band four is particularly informative in forested environments followed by band 5. In the visible region of the spectrum, band 3 is preferred because it is not as strongly influenced by variations in the atmosphere. The principle eigenvector matrix ensuing from a sampling of general forest cover, further supports the selection of bands 3, 4 and 5 (Beaubien, 1994) because bands 4 and 5 contribute the majority of information to the first and second principle components with negative input from band 4 in the second component. The third component receives most of its information from the visible portion of the spectrum, with negative input from band 5. Others have argued that the optimal set of TM bands to use for forest discrimination is dependent on the thematic content of the classification and the specific features present in the scene (Conese et al., 1993).

Landsat Thematic Mapper data has been widely used in forest management applications (Horler and Ahern, 1986; Peterson et al., 1986; Williams and Nelson, 1986; Fung, 1990; Moore and Bauer, 1990; Cohen, 1992; Congalton et al., 1993; Fiorella and Ripple, 1993a; Fiorella and Ripple, 1993b; White et al., 1995; Schriever and Congalton, 1995; Brondizio et al., 1996; Grigetti et al., 1997; Hyppanen, 1997; Jasinski, 1997; Langford, 1997; Cohen et al., 1998; Deppe, 1998; Hagner and Rigina, 1998). Numerous studies have used TM imagery to identify and/or measure a wide variety of forest attributes (Butera, 1986; Franklin, 1986; Matejek and Dubois, 1988; Peterson et al., 1988; Plummer, 1988; Smith, 1988; Spanner et al., 1989; Khorram et al., 1990; Spanner et al., 1990; Stenback and Congalton, 1990; Collins and Woodcock, 1994; Adams et al., 1995; Sant'Anna et al, 1995; Iiasaka et al., 1996; Collins and Woodcock, 1997; Franklin and Luther, 1997; Hall, 1997). The British Columbia Forest Service has used TM imagery for its forest inventory program for over fifteen years, mainly for the identification of forest disturbances from clear-cutting and fires (Heygi and Quenet, 1982; Pilon and Wiart, 1990). Despite its widespread application in forestry, it is the human interpretation of aerial photography and not the use of TM imagery, which continues to be the remote sensing methodology of choice for operational forest inventory - largely due to issues related to scale and the need for detailed attributes which cannot be resolved with the 30 metre resolution offered by the TM sensor (Franklin, 1994).

2.3 Classification Algorithms

The search for the optimal classification algorithm for vegetation cover has been the subject of extensive research in remote sensing for the past three decades. Originally, relatively simple mathematical theories and models were used to group image pixels with similar spectral properties into homogeneous units. However, advances in other fields such as computing science and artificial intelligence have resulted in the development of more complex and computationally intensive classification algorithms which may use both the spectral and spatial (texture and context) information in the image along with non-image data sources (ancillary data), multiple-source and multi-date imagery. Some researchers have focussed on the creation of new methods and algorithms, while others have attempted to develop new methodologies that may combine a series of methods to solve a particular classification problem.

Argialas and Harlow (1990) outlined three levels of approach to the classification of remotely sensed images. Operating on a per-pixel basis, the low level approach relies primarily on the spectral reflectance properties of land cover types. Medium level approaches on the other hand, utilize the properties of texture, context and shape in addition to spectral information. High level approaches employ domain specific knowledge of spectral, spatial and temporal properties. Civco (1993) describes the ideal classification approach as a region based, hierarchically structured and knowledge oriented method which is able to span spectral, temporal and spatial domains.

Although conventional classification techniques such as the maximum likelihood and the minimum distance to means are categorized as lower level techniques (Argialas

and Harlow, 1990) they should not be completely disregarded. The maximum likelihood in particular, continues to be used extensively for vegetation classification because of its computational simplicity and the relative ease with which it is applied. The maximum likelihood classification algorithm has been used extensively in vegetation classification applications (Gastellu-Etchegorry, 1990; Moore et al., 1990; Bolstad et al, 1992; Saxena et al., 1992; Fiorella and Ripple, 1993b; Franklin et al, 1994; Mathieu-Marni et al., 1995; Sader, 1995).

Recent trends have sought to elevate these conventional techniques to medium level approaches by incorporating ancillary data and textural properties (Niemann, 1993). Alternatives to the conventional classification techniques have developed in a constant effort to improve classification accuracies (Skidmore et al., 1987; Skidmore et al., 1988). These techniques include contextual classifiers, fuzzy logic classifiers and neural networks. Other methods which combine different methods have also been developed and include guided clustering, decision trees and hierarchical classifiers. An interest in the incorporation of ancillary data to improve classification accuracy has resulted in the use of evidential reasoning and possibility based approaches (Peddle and Franklin, 1991; Leiss et al., 1995).

Whether conventional or alternative, this discussion will focus primarily on supervised classification techniques. The goal of any supervised classification algorithm is to manufacture the best possible partition of the feature space. A supervised classification involves the transformation of the original image data through this partitioned feature space to create an output thematic map. A thematic map, in this

context, is therefore a reduction of the original remotely sensed image to a discrete set of regions of interest (Swain and Davis, 1978).

The objective of classification is to group together a set of observational units on the basis of their common attributes (Ken and Coker, 1992). The end product of a classification should be a set of groups derived from the units of observation where, typically, units within a group share more attributes with one another than with units in other groups. For vegetation classification, the unit of observation is typically the "stand", defined as a relatively homogenous area with respect to species composition, structure, and function.

2.3.1 Conventional Classification Methods

Although conventional classification techniques are assumed to be parametric, they include both parametric and non-parametric approaches. Conventional classification methods, for the purposes of this discussion, include the maximum likelihood (MLC) algorithm and most of the more simplistic methods developed and used prior to the MLC such as the Euclidean distance classifier and the unsupervised clustering approaches such as the K-means. A non-parametric approach would be one in which no assumptions are made concerning the underlying statistical distribution of the data. Cortijo and Perez de la Blanca (1997) identified three groups of spectral non-parametric classifiers and these include: (1) nearest neighbour estimation techniques which estimate the probability that a pixel belongs to a particular class; (2) nearest neighbour classification rules which directly estimate the a posteriori probability that a pixel belongs to a particular class; and (3) classification trees which recursively split the

feature space using binary questions. In contrast, parametric approaches rely on typifying the probability distribution of the image data by a representative function, which is commonly the normal or Gaussian distribution (Swain and Davis, 1978). The advantages of these parametric approaches are that since only the mean and standard deviations for each class are required to be stored as parameters (not the entire image), the method can theoretically extend to a feature space with unlimited dimensions (input bands) and classes, while still remaining computationally efficient. To summarize, parametric approaches depend on a assumed *model*, whereas non-parametric approaches depend on *data*.

A major trend in remote sensing literature has been the abnegation of conventional, parametric, pixel-based classifiers that have provided inadequate levels of accuracy in the discrimination of vegetation classes (Blonda et al., 1991; Bischof et al., 1992; Civco, 1993; Mouchot et al., 1994; Peddle et al., 1994; Sui, 1994; Miller et al., 1995). Particularly where there are highly overlapping class boundaries in the feature space, the use of non-parametric methods are preferable (Cortijo and Perez de la Blanca, 1997). Civco (1993, p. 173) characterized the status quo of conventional "classification science" as follows:

> "The traditional computer-assisted processing of digital multi-spectral remote sensing data for the purpose of the classification of land cover type has involved per pixel algorithmic analyses of spectral reflectance properties, often using a maximum likelihood decision rule, based on the assumption of a normal distribution."

Civco (1993) generalized several problems associated with the conventional statistical classifiers and these generalizations are summarized in Table 2.2.

Table 2. 2 Characteristic Weaknesses of Conventional Classifiers

Characteristic	Result
• classifier operates on a pixel by pixel basis	• fragmented, salt-and-pepper classification result
• use a computer-implemented algorithm	• methods are fixed, inflexible and do not easily permit variation in their execution • pixels which deviate from the training class mean will be misclassified or unclassified
• based on spectral properties for single data	• ancillary data is not easily incorporated into an algorithm-based computer program • multi-temporal imagery often not considered
• size, shape, texture and pattern are usually not considered • contextual attributes such association and proximity are not considered	• poor overall classification performance
• highly scene dependent	• classifiers perform sub-optimally in different location, time or cover types

It shall be demonstrated in the discussion of alternate techniques in Section 2.3.2 and further in Chapter 3, that many of characteristic weaknesses outlined in Table 2.2 may also be applicable to some of the newer alternative classification methods, particularly when these techniques are used in isolation. Successful classification, in the context provided by Civco (1993), involves a classification strategy, within which any number of classification tools (algorithms) may be configured and implemented to facilitate the extraction of the informational classes from the spectral information. Examples of such methods are outlined in Section 2.3.2.4.

2.3.1.1 Minimum Distance to Means

The minimum distance to means or Euclidean classifier is a non-parametric conventional technique which determines the mean spectral value for each image band and establishes a mean vector for each output class (Lillesand and Kiefer, 1987). This technique uses the measurement of the Euclidean distance within the multi-dimensional

feature space, from the mean digital number (DN) value to the location of the unknown pixel in the feature space (Figure 2.2).

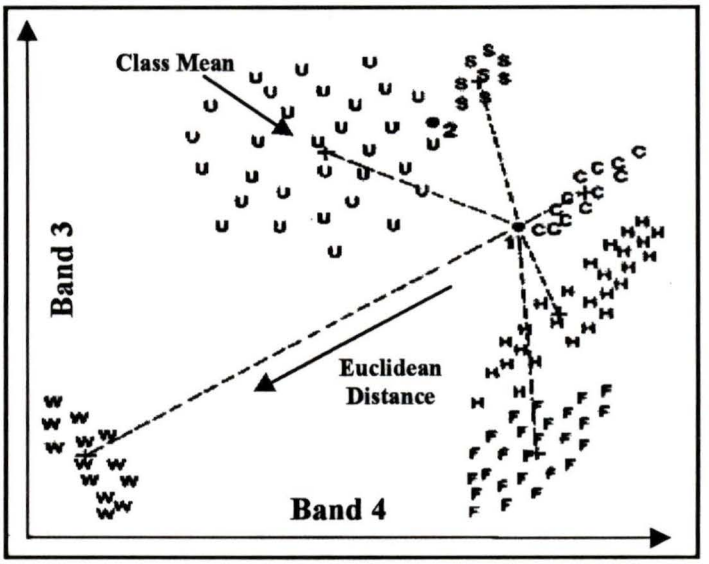


Figure 2.2 A two-dimensional representation of the minimum distance to means classifier
(Source: Adapted from Lillesand and Kiefer, 1987)

The unknown pixel is therefore assigned to the class for which the Euclidean distance is the shortest, unless the unknown pixel lies beyond an analyst-defined distance from the class mean, in which case the pixel would be classified as unknown (Thomas et al., 1987). Although this technique is simple and computationally efficient, it is insensitive to different degrees of variance in the spectral response data. This method is not widely used in applications where spectral classes are close to one another in the measurement space and have a high variance (Curran, 1985).

The Mahalanobis distance classifier is a parametric modification of the minimum distance to means method. While this method also evaluates the Euclidean distance between an unknown pixel and the class mean value, the Mahalanobis method

modifies the Euclidean distance by dividing it by the variance/covariance matrix for the class, in the appropriate n-dimensional direction (from the unknown pixel to the mean value for the class). The Mahalanobis is therefore measuring the remoteness of a pixel in terms of the statistical distributions of the class training data in the n-dimensional feature space (Haralick, 1983; Thomas et al., 1987).

2.3.1.2 Parallelepiped

The parallelepiped, like the minimum distance to means, is also a non-parametric classification technique. Referred to as the box classifier, the parallelepiped functions by calculating the minimum and maximum values associated with a particular class from the training data and then uses these values to construct a box around the class in feature space. Pixels located outside the bounds of this box are excluded from that particular class. The multi-dimensional representations of these rectangular areas are parallelepipeds (Lillesand and Kiefer, 1987). The parallelepiped is a fast and efficient classification method which is sensitive to the class variance because it considers the range of values for each class. Unfortunately, classes which are highly correlated or highly covariant are poorly described by the rectangular decision regions of the parallelepiped as shown in Figure 2.3, where the overlapping rectangular decision regions have resulted in classification confusion. Strong degrees of correlation and covariance are frequently the rule rather than the exception in the spectral response patterns of vegetation and consequently, stepped decision boundaries or bespoke boxes (Figure 2.4) have been suggested as a remedy for this problem (Curran, 1985).

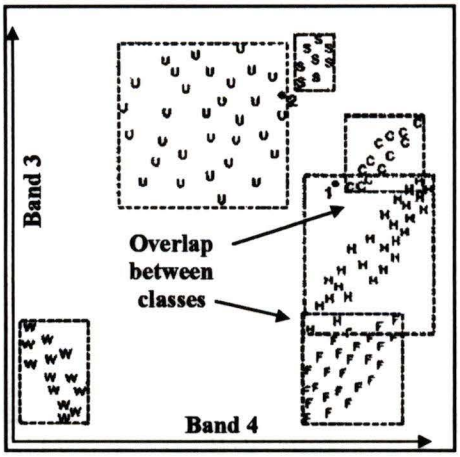


Figure 2.3 The Parallelepiped Partition of a Two Dimensional Feature Space.
(Source: Adapted from Lillesand and Kiefer, 1987)

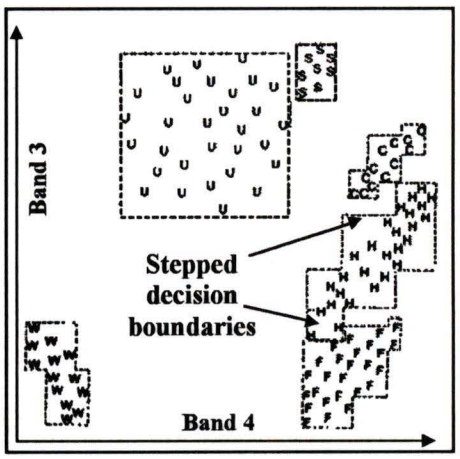


Figure 2.4 The Parallelepiped Partition of a Two Dimensional Feature Space with Bespoke Decision Boundaries.
(Source: Adapted from Lillesand and Kiefer, 1987)

2.3.1.3 ***Maximum Likelihood Classifier***

The maximum likelihood classifier has been the most extensively used classification algorithm in remote sensing - previously as the technique of choice and more recently as a benchmark against which the performance of new, alternative algorithms are compared (Yool, 1998). The MLC is based on the minimum distance algorithm, however the probability of a pixel belonging to a particular class is based on the distribution of points within the training data, with the assumption that the data is normally distributed (Lillesand and Kiefer, 1987). The variance and covariance of the training data is used to define the probability density function, which in turn, is used as the decision basis for class membership. (As was previously mentioned, assumptions of normality are made so that these higher order statistics can be used to represent the

distribution, increasing the speed and efficiency of MLC computations). The probability that a pixel belongs to a particular class is calculated and ellipsoidal decision boundaries are created. These ellipsoidal, equiprobability decision boundaries are used to represent the feature space as shown in Figure 2.5. Unknown pixels are classified into the class for which they have the highest probability or they are classified as unknown if the probability values are below the threshold set by the analyst.

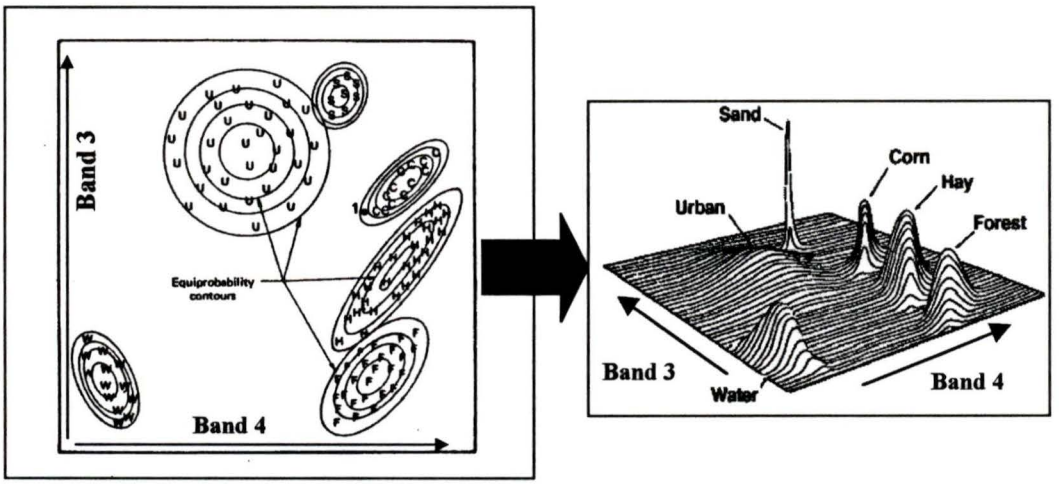


Figure 2.5 A two dimensional (left) and three dimensional (right) MLC partition of a feature space.
(Source: Adapted from Lillesand and Kiefer, 1987)

Despite the efficiencies gained from the assumption of normality, a large number of computations are still required for the MLC to classify each pixel, particularly if there are large numbers of output classes or input channels. Swain and Davis (1978) argued that the normal density function adequately models the probability processes for a large number of remote sensing applications and that classifiers adopting this assumption tend to be relatively robust, meaning that the overall classification accuracy is not sensitive to moderate violations of the normality assumption. Swain and Davis (1978) recommended that analysts exercise caution when using the Gaussian

distribution and offered two noteworthy caveats. First, adequate training samples must be provided to accurately estimate the mean and covariance measures for each class. The required number of samples is $n+1$ where n equals the number of image channels to be used as input for the classification. Otherwise, the variance and covariance matrices would be singular and it would not be possible to derive the probability density distribution function. In practice, the number of training samples is recommended to be 10 to 100 times the number of image channels. Secondly, where the assumptions of normality are clearly violated, for example where the distribution of a class is multi-modal rather than uni-modal, the classification must be approached with caution. Separability indices are useful tools for reducing the dimensionality of the data set, thereby lowering the number of required training samples. These indices are also useful for assessing the modality of the data (Thomas et al., 1987).

The basis of the MLC is the multi-dimensional normal probability density function. The probability density function and the likelihood function are not synonymous, with the likelihood function actually being *derived* from the probability density function. It is important to distinguish between a probability and a likelihood. A probability is an entity which exists before any measurement has been taken or any pixel has been classified. The quantity which exists after the classification has been completed is the likelihood, which is dependent on the number of repeat experiments conducted with the same set of probability density functions (Swain and Davis, 1978).

For N repetitive independent measurements of the same quantity D, the resultant probability P_a will be $P_a = (P(D))^N$. P_a is thus considered to be an a posteriori probability or likelihood. The likelihood value is derived therefore from the completion

of a series of N measurements, each of which was collected independent of the other measurements, with the same probability of obtaining an individual results $P(D)$. The resulting likelihood of each of N events yielding the same result is then the product of the individual probabilities of that event occurring as the result of a single independent measurement (Swain and Davis, 1978).

The Bayesian classifier is a variation of the MLC which uses two additional weighting factors applied to the probability estimate described above, for the purpose of minimizing the cost of misclassification. First a measure of the *a priori* probability or anticipated likelihood of occurrence for each class in a given scene is applied to the single event probability, $P(D)$, to reflect other influences such as spatial associations and seasonal influences (Thomas et al., 1987). Second, a weight associated with the cost of misclassification for each class in a given scene may also be used in conjunction with the Bayes optimal strategy of minimizing the average loss due to misclassification over all classes. The zero-one loss function is a variation of the Bayes optimal strategy and is a decision rule which produces results that have the minimum probability of error over the entire classified data set. The objective of the Bayes approach is to produce the most accurate classification possible, however the *a priori* probabilities are rarely known reliably in real world applications and the costs of misclassification are often obscure and difficult to define. It is therefore commonly assumed that the *a priori* probabilities for all of the classes are equal.

To summarize, the likelihood function is derived after a series of independent events, from the individual probability density functions. The likelihood function is evaluated for each pixel over the set of user specified classes and the maximum

likelihood value is determined and the pixel assigned to that class. Once MLC values are derived for each pixel, the multidimensional data channel space collapses to the single dimension likelihood space.

2.3.2 Alternate Classification Methods

Expectations of improved accuracy and efficiency have provided the impetus for the development of new classification methods. Image classification in remote sensing has been strongly influenced by developments in other disciplines, as manifested in the application of neural networks and fuzzy logic to remote sensing problems. Other techniques have also emerged which have sought to improve classification accuracy and efficiency by combining several different classification methods into one overall strategy, with each individual method being particularly effective at resolving specific classes of interest. Such techniques include decision trees and hierarchical analysis pathways. The number and variety of classification methods which have been developed is vast, and a comprehensive treatment of all possible methods is beyond the scope of this discussion. Some of the more commonly used methods and those most appropriate to vegetation and land cover discrimination are discussed in the following sections.

2.3.2.1 Contextual Classifiers

Remotely sensed imagery contains both spectral and spatial information. This spatial information is either textural or contextual. Texture is the spatial variability of tones within a contiguous group of pixels, while context is a description of the spatial relationships between pixels (Gurney and Townshend, 1983; Alonso and Soria, 1991; Rhyherd and Woodcock, 1996). Contextual classifiers are defined as those classification techniques which consider spectral and spatial characteristics simultaneously to achieve more accurate classification results (Toussaint, 1977; Alonso and Soria, 1991; Groom et al., 1996). The contextual information in an image is dependent on the following empirical relationships between pixel size and image object: (1) if the spatial resolution of the image is finer than the objects in the scene, then most of the measurements for a pixel will be highly correlated with neighbouring pixels; (2) if the spatial resolution is the same size as the objects in the scene, then the likelihood of neighbours being similar decreases; and (3) if the size of the pixel increases and many objects are found within a single pixel then the correlation between pixels increases (Woodcock and Strahler, 1987). Contextual information has been extensively applied to vegetation classification (Jensen, 1979; Wharton, 1982; Dutra and Marscanrechas, 1984; Franklin and Peddle, 1990; Gong and Howarth, 1992).

Kartikeyan et al. (1994) outlined three different approaches which have been taken to incorporate context into the classification process. In the first technique, pixels are grouped according to similar spectral responses and then the samples are classified en masse by comparison with class spectral signatures. The second method is the relaxation technique which is an iterative process that considers the joint distribution of

neighbouring pixels one at a time. The third technique attempts to decide on the label for one pixel based on the observation at all the other pixels in the image.

Contextual information has been used to improve classification accuracy, particularly where there is a high degree of spatial autocorrelation amongst the image pixels resulting from the interaction between the spatial resolution of the image and the landscape objects. Gong and Howarth (1992) developed an alternative method of contextual classification which used grey level vector reduction and grey level vector-occurrence frequencies to characterize and classify land use types. Swain et al. (1981) note that the final classification accuracy derived from the contextual method will depend largely on the assumptions underlying the classifier model. Gurney and Townshend (1983) reinforced this with the conclusion that although contextual information is a valuable tool in classification, the design and use of appropriate contextual rules is dependent upon the analyst's knowledge of the spatial relationships actually existing on the ground and on the relation between pixel size and object size in a scene.

2.3.2.2 Fuzzy Classifiers

Fuzzy classifiers incorporate fuzzy logic into the classification process. Fuzzy logic is essentially a simulation of the nebulousness and uncertitude which exists in nature. Fuzzy classifiers do not attempt to assign pixels into inflexible categories. Blonda et al. (1991) examined the use of fuzzy logic in a rule-based decision making process of image classification. This study concluded that the statistical descriptions provided by the maximum likelihood classification are only effective for sufficiently

homogenous classes and that fuzzy classifiers performed better where there was a large percentage of inhomogeneity. The fuzzy classifier used by Blonda et al. (1991) achieved an overall classification accuracy of 91% versus the 86% of the MLC. According to Wang (1990), the improvement offered by the fuzzy classifier is directly proportional to the percentage of inhomogeneity in the scene.

The unsupervised fuzzy clustering method developed by Gath and Geeva (1989) is one example of how fuzzy techniques are implemented for image classification. This method is referred to as the "unsupervised fuzzy partition-optimal number of classes", or UFP-ONC. With this method there are no *a priori* assumptions on the number of clusters in the data set and the validity of final clusters is based on the performance measures of hyper-volume and partition density. Gath and Geeva (1989) claim that the UFP-ONC algorithm performs well when there is large variability in cluster shapes, densities and number of points in each cluster. This algorithm was not designed specifically for image segmentation.

There are two phases to the UFP-ONC algorithm. In the first pass of the classifier, cluster seed points are identified using a k-means clustering algorithm. In the second pass of the classifier, these seed points are used by a modified maximum likelihood estimation algorithm to calculate the degree of class membership and to refine cluster shape. The performance measures (hyper-volume and partition density) are then calculated and the number of clusters are increased and the first three steps repeated until an optimum value for the performance measure is calculated or some minimum error is achieved. The UFP-ONC algorithm is designed to capitalize on the advantages of the fuzzy k-means and fuzzy maximum likelihood estimation, while minimizing the inherent weaknesses of these two

approaches: the fuzzy k-means algorithm performs poorly where there are variable cluster shapes and densities while the fuzzy maximum likelihood estimator requires reliable cluster centroids because the exponential component of this algorithm will converge to a local optimum in a rather narrow spectral region (Gath and Geeva, 1989).

Unsupervised fuzzy clustering algorithms are not included in the numerous scenarios outlined by Pal (1994) and Foody and Trodd (1996) for incorporating fuzzy set theory into image classification. More commonly used techniques include the flexible supervised fuzzy clustering approaches, as well as techniques for merging fuzzy theory with the output of conventional "hard" classifiers, such as the maximum likelihood (Foody et al., 1992; Foody et al., 1996). Combinations of techniques such as neural networks with fuzzy classifiers have also been suggested (Blonda et al., 1995; Foody, 1996).

The majority of the documented uses of fuzzy classifiers have utilized a supervised classification approach with the fuzzy c-means algorithm (Bezdek et al., 1984). The main advantage of the fuzzy c-means algorithm over the UFP-ONC algorithm is that the former can be adapted to operate under a supervised scenario, with user-defined training data providing the cluster seed points as in Key et al. (1989). The fuzzy c-means algorithm is a non-hierarchical clustering technique which begins by randomly assigning pixels to classes (or assigns them in accordance with user-defined seed points in a supervised approach) and then iteratively moves pixels to other classes with the aim of minimizing the generalized least-squared error.

In a fuzzy c-partition of a data set, membership functions characterize the membership of each pixel in all classes. Fuzzy membership values range from 0 to 1 and the memberships for each case sum to unity. Memberships close to one indicate a high

degree of similarity between a pixel and a class, whereas memberships close to zero indicate little similarity between a pixel and a class. The analyst can determine the degree of fuzziness of the algorithm by setting the value of the cluster weighting parameter. One of the key advantages of user-defined seed points is an obvious reduction in the amount of time which the algorithm will take to process the image.

Foody (1992a), found that the fuzzy c-means algorithm was more appropriate for modelling continuous phenomena, such as that found in heathland vegetation, when compared to more conventional classifiers. A supervised approach was used whereby an analyst provides the seed points for the class centroids used by the fuzzy c-means algorithm. Foody's (1992a) conclusions from this trial indicate that this algorithm can be used to discriminate between the end points of continua and that fuzzy class membership functions correspond with canopy composition.

Maselli et al., (1996) obtained fuzzy membership grades by using a statistical method based on a softened maximum likelihood classifier that was first proposed by Wang (1990). *A priori* probabilities were derived from the frequency histograms of training data sets and the MLC was altered so that the classifier attempted to mathematically maximize the a posteriori probability of correct attribution to a particular class for each unknown pixel. It was found that it was most efficient to apply this algorithm to areas of the image where pixels were mixed or difficult to attribute to one particular class. The complexity of producing and interpreting fuzzy memberships may reduce its utility to areas where the aforementioned condition exists. Palubinskas et al. (1995) also used the method where the strength of class membership is derived from a modified MLC classification, using a supervised approach with training data.

Blonda et al. (1991) developed a rule based classification and attempted to translate linguistic predicates such as "high", "medium" and "low" to a mathematical function which would assign class probabilities to pixels. This supervised approach requires the user to define the bounds for "high", "medium" and "low" classes and through this process a knowledge base is created which is then applied to the image. The output from the classifier are membership functions for fuzzy sets defined on the image spectral values. Similar methods have been used by Binaghi et al. (1997) to classify the glacier equilibrium line in the Italian Alps from multi-source imagery.

Only one of the papers examined employed an unsupervised methodology. Mouchot et al. (1994) used an unsupervised approach with the fuzzy c-means algorithm, although there was *a priori* knowledge of land use activities. A membership function was associated with each pre-selected category (so that pixels belonging totally to one fuzzy category could not totally belong to another category). Several attempts were made at interpreting the output from the fuzzy classifier and regrouping the clusters to create meaningful classes.

In summary, there are two primary methods whereby fuzzy set theory has been successfully applied to remotely sensed imagery. The first is through the fuzzy c-means algorithm which has the flexibility to operate in both a supervised and unsupervised manner – with the former being the most common manifestation of its application. The other method by which fuzzy set theory is incorporated into image classification is through the process of softening the output of conventional hard classifiers to produce fuzzy membership functions (Hadipriono, 1991; Foody, 1996).

2.3.2.3 Neural Networks

Neural networks are algorithms which imitate the computational structure of the brain. Pao (1989) referred to artificial neural networks as data transformers, because of the manner in which they transform a feature space into a class space. Artificial neurons are used as the basic unit of the network, just as a neuron is the fundamental unit of the nervous system. The artificial neurons form a network wherein information is passed from one neuron to the next, with incremental processing steps occurring along the connections between the neurons. Neural networks have been extensively applied to remote sensing image classification. A detailed analysis of neural networks is provided in Chapter 3.

2.3.2.4 Other Alternate Methods

The classification approaches considered in Section 2.3.2.4 are linked by virtue of the manner in which they combine different classification methods and algorithms or through the manner in which they integrate different sources of data, with the ultimate goal of producing a superior classification result. These methods include techniques that facilitate the incorporation of ancillary data, decision trees, hierarchical methods, evidential reasoning and guided clustering.

Some researchers have sought to improve classification accuracy by incorporating appropriate non-image ancillary thematic or spatial data before, during or upon the completion of the image classification (Hutchinson, 1982; Janssen et al., 1990; Niemann, 1991; Bauer et al., 1994; Joria and Jorgenson, 1996; Treitz et al., 1996, Hojsgaard et al., 1997). Some have argued that the use of Thematic Mapper spectral

information alone is insufficient for the discrimination of forest species (Benson and DeGloria, 1985; Niemann, 1993; Apan, 1997). Several studies have reported improvements to classification accuracy resulting from the incorporation of this ancillary information (Bolstad and Lillesand, 1992; Bruzzone et al., 1997; Carpenter et al., 1997; Medler and Yool, 1997) while some have found only minor improvements in overall accuracy (Kenk et al., 1988).

Prior to classification, ancillary information can be used to stratify the image into identifiable regions (Bauer et al., 1994) or the information can be used post-classification to identify sub-classes within the spectrally defined classes. During the classification, ancillary data may be incorporated either as an additional information channel or through the use of complex decision rules. Such a technique was implemented by Bolstad and Lillesand (1992), who compared a land cover classification obtained from the use of evidential and hierarchical decision rules to combine satellite data with soil and terrain information, to the classification obtained from a standard maximum likelihood method. The decision rule system developed by Bolstad and Lillesand (1992), is based on a framework of a supervised MLC with at-classification incorporation of additional spatial data defined by a set of decision rules. Niemann et al. (1989) highlighted the problem of using traditional cartographic databases as ancillary data because these databases are generalizations, and these generalizations can lead to increased confusion when combined with remotely sensed data.

Another alternative classification method is that of decision trees. A decision trees is defined as a classification procedure that recursively partitions a data set into

smaller subdivisions on the basis of a set of tests defined at each branch of the decision tree (Friedl and Brodley, 1997). Decision trees are non-parametric and are capable of handling non-linear relationships between features and classes and allow for missing values. Decision trees are also capable of handling both numeric and categorical inputs and have an explicit structure which is easily interpretable as shown in Figure 2.6 (Hansen et al, 1996). In recognition that no one algorithm will be optimal for all classes under consideration, hybrid decision trees allow different classification algorithms to be applied to different parts of the data set (Figure 2.7). The purported efficiency, flexibility and robustness of this technique makes it particularly useful for multi-temporal or multi-source data (Friedl and Brodley, 1997). Unfortunately, considerable effort is required to design a decision tree, although some researchers have suggested using a set of decision rules for automatic "tree growing" (Yoshikawa et al., 1995). Another disadvantage of decision trees is that they often attempt to optimize overall classification accuracy, to the detriment of smaller classes.

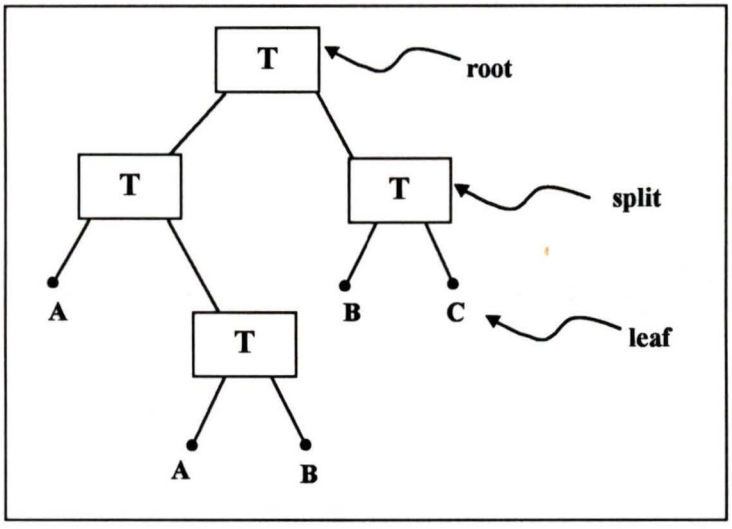


Figure 2.6 A decision tree classifier. Each box is a node at which tests (T) are applied to recursively split the data into successively smaller groups. The labels (A,B,C) at each leaf node refer to the class label assigned to each observation
(Source: Adapted from Freidl and Brodley, 1997)

A method very similar to that of a decision tree, although less systematic is the hierarchical classification method. An example of a hierarchical procedure developed by Lauver and Whistler (1993) for the identification of natural grassland areas from TM data is shown in Figure 2.8. In the example presented in Figure 2.8, both unsupervised and supervised techniques have been combined sequentially to eliminate areas of non-grassland and to produce a map of low and high quality grassland areas. Discriminant analysis was used to select the image channels best able to distinguish between the high and low quality grassland areas and then standard MLC was used to complete the classification process. Other researchers have attempted to implement a similar hierarchical approach with the MLC in order to improve accuracies (Ediriwickreme and Khorram, 1997).

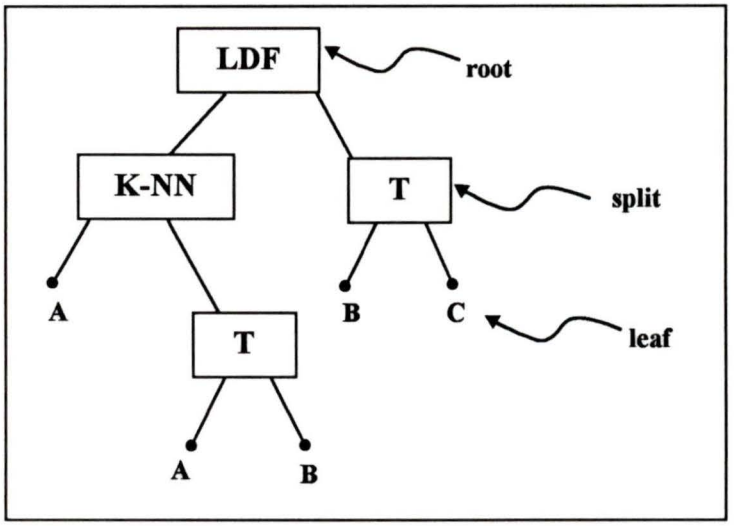


Figure 2.7 A hybrid decision tree classifier. In this example, the splitting rule at the root node is a linear discriminant function (LDF), whereas the left subtree is a *k* nearest-neighbour classifier (K-NN), and the right tree is a univariate decision tree.
(Source: Adapted from Friedl and Brodley, 1997)

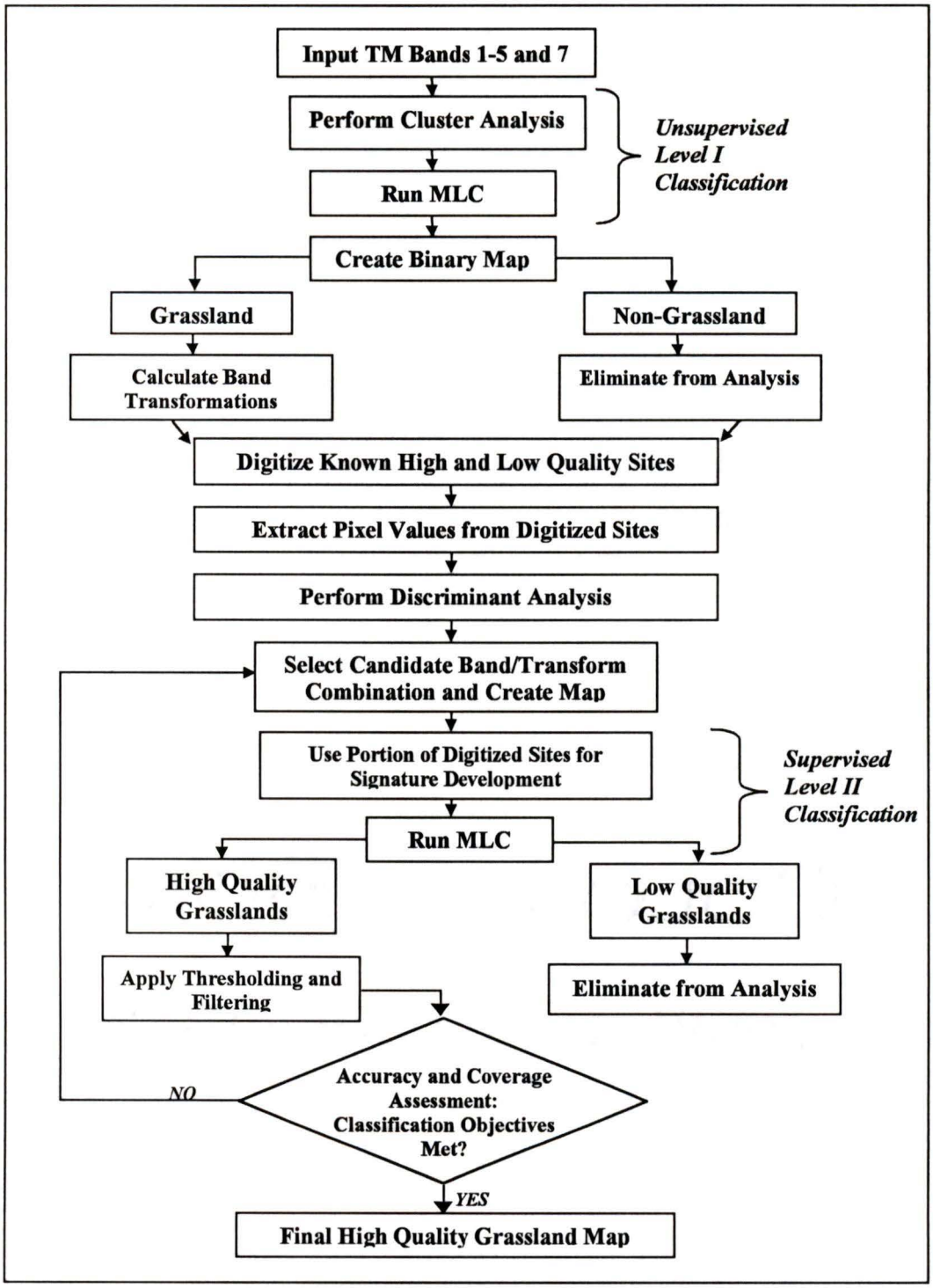


Figure 2.8 Schematic Diagram Of The Hierarchical Classification And Methods Used To Produce A Map Of Potential High Quality Grassland.
(Source: Adapted from Lauver et al., 1993)

Evidential reasoning is another alternative classification strategy which is free of statistical and data-type (i.e. nominal thematic data, ordinal data and directional information) assumptions and also has an explicit mechanism for handling data uncertainty (Peddle et al., 1994). Evidential reasoning involves reasoning in a hypothesis space (a set of vegetation classes) and with accumulating evidence regarding specific hypothesis (classes) within that hypothesis space (Wilkinson and Megier, 1990). The evidence is combined via a "belief" function which fuses belief onto the optimal hypothesis (classification result).

Wang and Civco (1994) used evidential reasoning to incorporate ancillary data with image spectral information. In this approach, a traditional MLC was first performed and a preliminary accuracy assessment was completed (Figure 2.9). From this, a set of possibly misclassified pixels (PMPs) was identified which were passed through an evidential reasoning model where the ancillary data was used in the classification of these pixels. The evidential reasoning model uses Dempster-Schaefer's theory of evidence and the rule of orthogonal sums to provide a quantitative basis for the evidential combination of ancillary data and an initial classification of Landsat TM data (Wang and Civco, 1994).

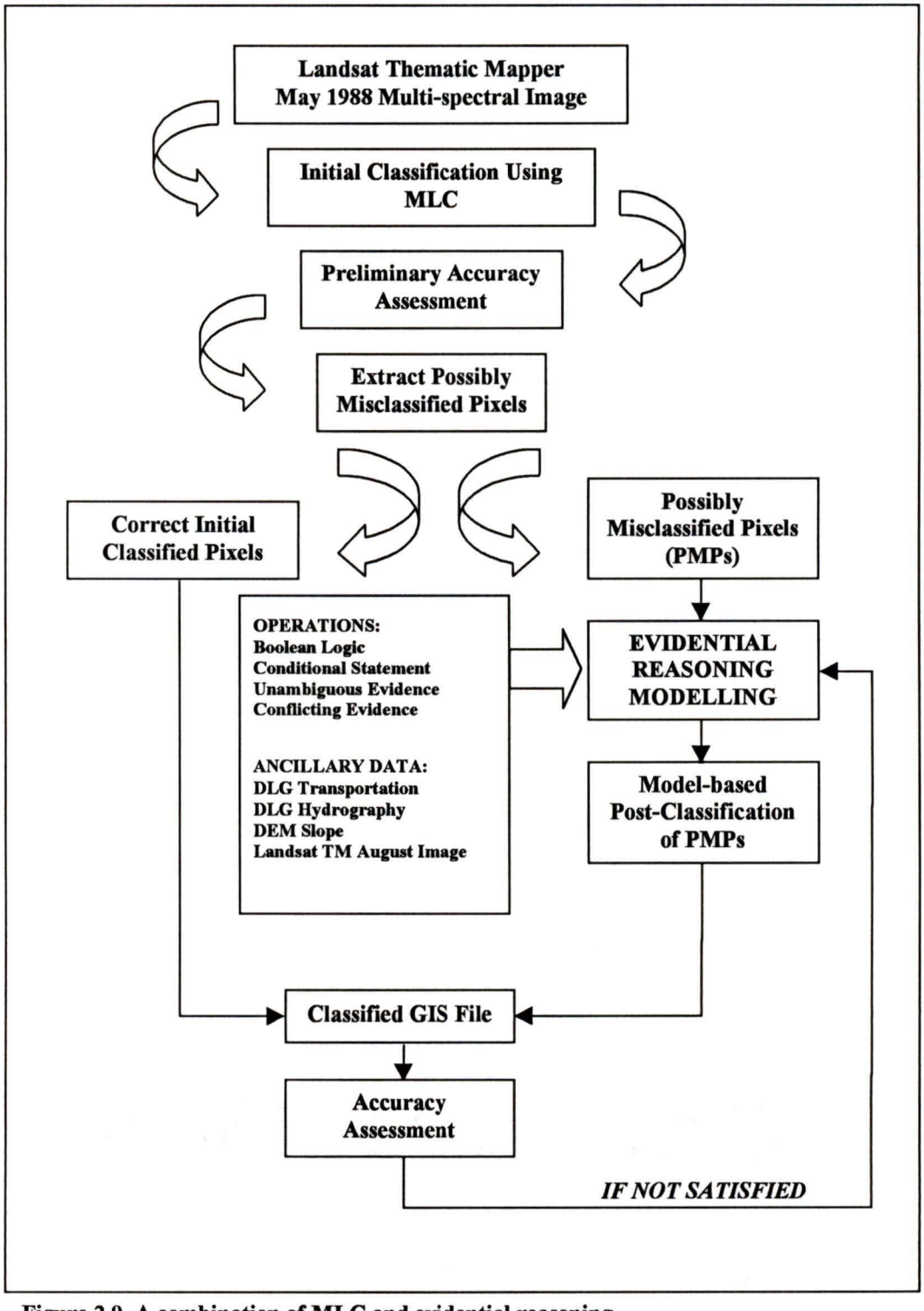


Figure 2.9 A combination of MLC and evidential reasoning.
(Source: Adapted from Wang and Civco, 1994)

Wilkinson and Meiger (1990) implemented evidential reasoning using the Gordon-Shortliffe algorithm which functions only in strict classification hierarchies where each hypothesis has no more than one direct ancestor above it. They used MLC and have output which includes both the maximum likelihood class for each pixel and a full set of computed likelihoods for each class. This information was then combined with an expert system with contextual rules and geographic information from a GIS (elevation, slope and aspect and physical zoning). For each pixel, a set of expert system rules apply and these rules produce further evidence which must be weighted by the algorithm in the derivation of the final class assignments. The results of this approach however are strongly dependent on the quality of the contextual rules and the geographic data which are used to provide the extra evidence.

Peddle (1993) developed a new implementation of the evidential reasoning technique which produced greater accuracies than MLC or linear discriminant analyses. Peddle et al. (1994) compared the classification results of MLC, two different artificial neural networks (ANNs) and evidential reasoning. They found that the neural networks achieved the highest level of accuracy, followed by evidential reasoning and MLC. Duguay and Peddle (1996) also found that evidential reasoning gave better results than the ANN without the high level of analyst interaction required by the neural network approach. Evidential reasoning was seen as a more objective approach than the neural network and was faster to implement.

Guided clustering is an alternate method which makes use of both supervised and unsupervised approaches and avoids many of the problems associated with each procedure individually. This method, outlined in Figure 2.10, uses analyst defined

training data for a set of target classes which is then clustered into spectrally separable sub-classes using unsupervised approaches. A maximum likelihood classification is then performed, using the resulting sub-classes from the unsupervised clustering process as training data. The classes resulting from the MLC are then collapsed back to the original number of target classes. Post-processing procedures are used to smooth the classification (Bauer et al, 1994). Others have successfully implemented a similar guided clustering approach (Fleming and Hoffer, 1977; Gaydos and Newland, 1978; Hoffer, 1979; Mayer and Fox, 1981; Hepner et al., 1990, Guo et al., 1994; Carbone et

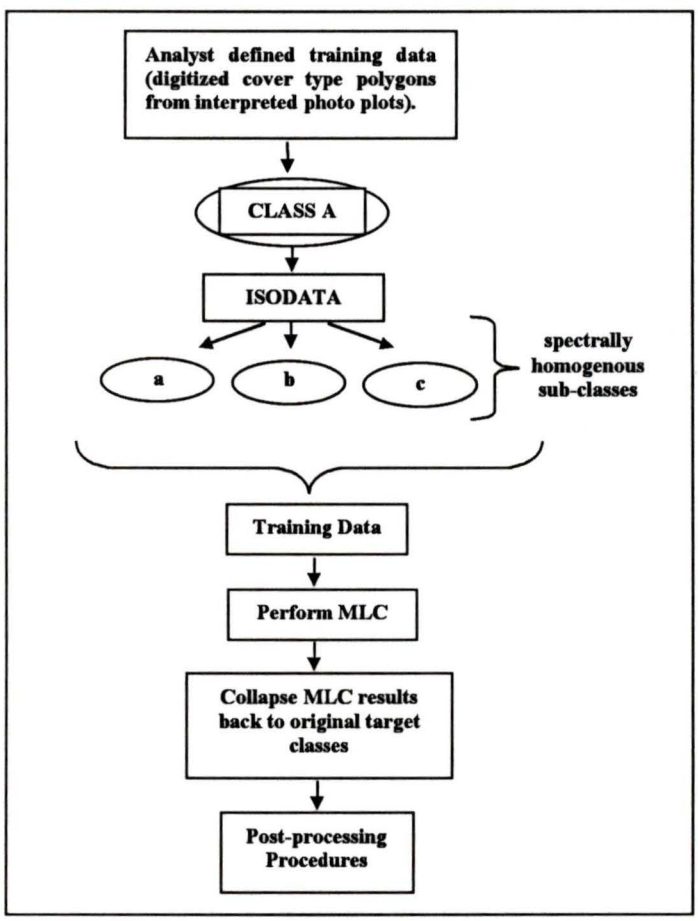


Figure 2.10 A Schematic Diagram Of Guided Clustering.
(Source: Adapted from Bauer et al., 1994)

al., 1996).

Similar to guided clustering is a hybrid classification system developed by Hardin (1994) which utilized both unsupervised nearest-neighbour classification techniques along with MLC. Because true unsupervised classification can produce unusable results, is time consuming and can limit analyst interaction, many select a semi-supervised or hybrid approach in preference to the completely unsupervised method of Campbell (1987). Figure 2.11 shows an example of this hybrid approach where a random sample of pixels is extracted from the image and passed to the clustering algorithm. Using this hybrid approach, descriptive statistics are calculated and for each resulting cluster and a classifier such as MLC is used to assign the image pixels to a spectral class. The advantages of this approach are speed, flexibility and more potential for analyst interaction. Hardin (1994) suggested using nearest neighbour in place of MLC for the pixel assuagement, especially when training data sets are large and parametric classification is low.

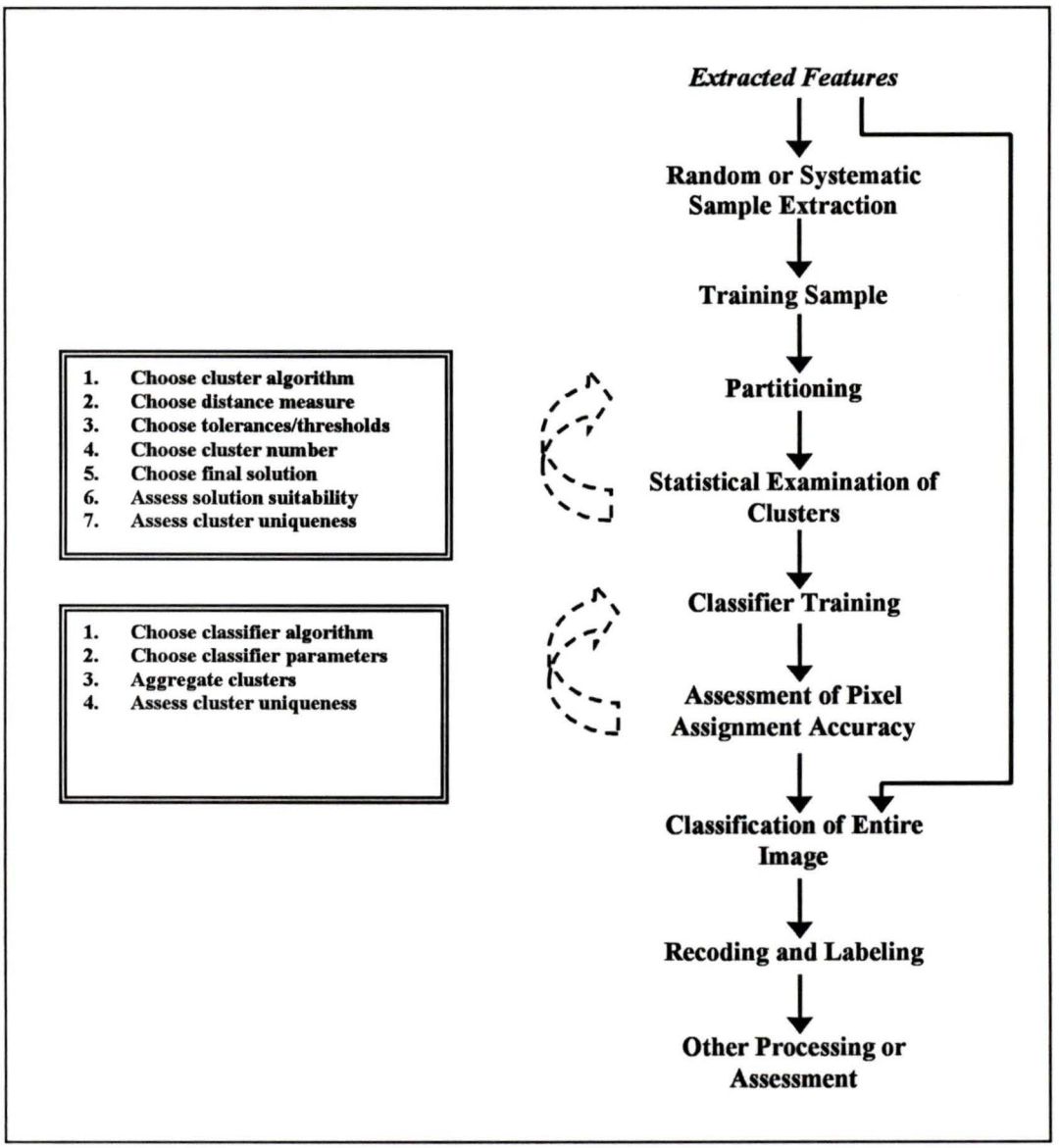


Figure 2.11 One Approach To Hybrid Classification. Iterative Experimentation In The Partitioning Phase Requires That The Training Sample Be Clustered Only During Each Iteration.
(Source: Adapted from Hardin, 1994)

Wilkinson et al. (1995) proposed a method of combining statistical and neural approaches for classification, acknowledging that each method will divide the feature space into different hyper-regions on the basis of very different mathematical models with different strengths and weaknesses and that these methods should not be seen as

competitors but as complementary tools for data analysis. Wilkinson et al. (1995) sought to develop a model which would capitalize on the relative strengths of these two different approaches. The hyper-regions generated by the two methods cannot be combined, however areas of joint agreement are found where these hyper-regions intersect, strongly indicating that the pixels should belong to that class. For those pixels which fall outside the area of joint agreement, a third classifier (a neural network) is used for classification. A diagram of this procedure is offered in Figure 2.12.

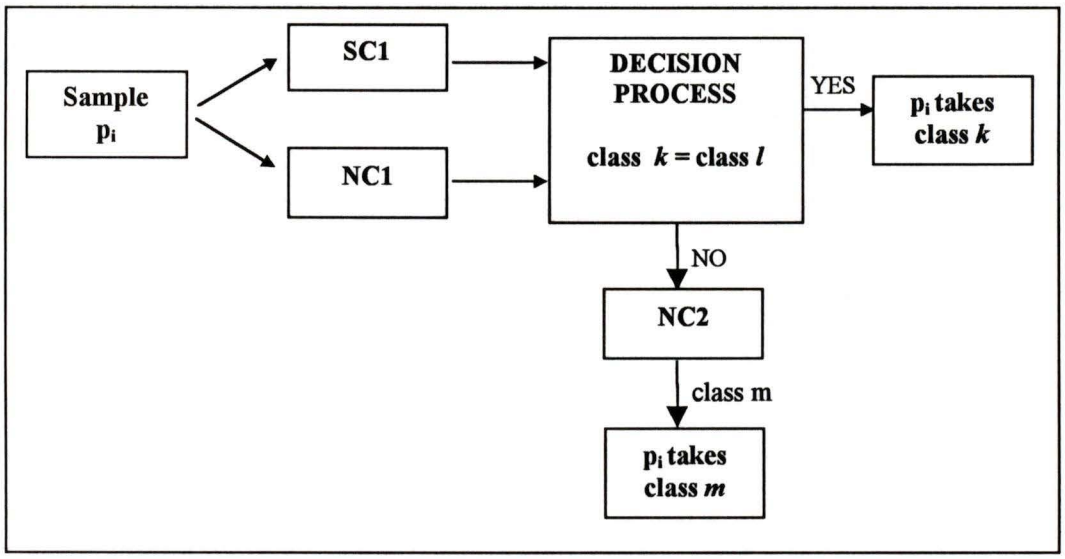


Figure 2.12 An Integrated Classification Procedure Using A Statistical Classifier And A Neural Network.
(Source: Adapted from Wilkinson et al., 1995)

2.4 Summary

Vegetation has complex spectral properties, resulting in challenging classification problems. TM data has been extensively used in vegetation classification applications, with varying degrees of success. Prior to the development of other sensors and platforms in recent years, TM has been the only reliable sensor of its kind to

combine the spectral and spatial properties suitable for discriminating vegetation at the landscape level. A wide variety of classification methods, both parametric and non-parametric, have been used for vegetation classification in previous research. It is often difficult to separate the success of the methods from either the target objects, the biophysical setting of those targets and the image pre-processing procedures. Comparative studies often evaluate techniques against the maximum likelihood classifier, which has emerged, perhaps erroneously, as the de facto classification method in remote sensing. The new alternative techniques hold promise for image classification, however their inherent flexibility is accompanied by a complexity which may limit their operational implementation. The combination of multiple classification "tools" into an overall classification strategy as per Civco (1993), may prove to be the most pragmatic approach - capitalizing on the benefits of both conventional and alternative classification methods.

Research into the classification of vegetation in remotely sensed images can be characterized as primarily empirical, "proof of concept" studies which provide classification results that are difficult to reproduce. Furthermore, the important procedures which occur before and after classification (image corrections and enhancements, spatial filtering) are rarely documented fully. This latter information is of paramount importance because the techniques involved are often not commutative and understanding which techniques were used and the sequence in which they were used is critical (Duggin and Robinove, 1990). These weaknesses in existing research suggest that there is a requirement for more systematic research into the use of these

various classification practices in different geographic locations, as well as a need for the improved documentation of image processing and classification procedures used.

CHAPTER THREE

NEURAL NETWORKS AND IMAGE CLASSIFICATION

The previous discussion has demonstrated that there are numerous alternate classification techniques and approaches used by researchers for image classification, many of which are particularly relevant to vegetation classification. This chapter focuses on the use of artificial neural networks for image classification and begins with a definition of artificial neural networks, followed by a description of the components which control the behaviour of the network with specific emphasis on the multi-layer perceptron – the type of network used in this research. Following this, the application of artificial neural networks to remote sensing image classification will be reviewed, with a particular emphasis on the advantages and disadvantages of neural networks for this purpose as well as an examination of some alternative methods by which neural networks have been implemented. This chapter should demonstrate the suitability of neural networks for image classification, particularly for complex targets such as vegetation. While there is a large body of research developing on the use of neural networks, their common operational use remains encumbered by the "black box" nature of their architecture and the complexity of their implementation.

3.1 Artificial Neural Networks: A Definition

Artificial neural networks (ANN) are defined as information processing systems which have certain performance characteristics in common with biological neural networks (Fausett, 1994). They can be seen as collections of mathematical models that

emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning (Lippmann, 1987). Both artificial and biological neural networks are made up of groups of connected neurons. Biological networks are composed of millions of cells (neurons) each of which is joined to other neurons by connective links (synapses) and these networks learn by altering the strengths of their synaptic connections (Rumelhart et al., 1986). Artificial neural networks are extremely simplified versions of their biological counterparts - their complexity is limited to a certain extent by the computer hardware used for their processing. Neural networks are useful for information processing because they are adaptable, robust and non-parametric (Lau, 1992). The latter has strong implications for remote sensing, because ANNs make no assumptions regarding the underlying distributions of the source data (Masters, 1994).

Neural networks have a wide range of applications, including storing and recalling data patterns, classifying patterns, performing mapping based on input patterns and finding solutions to constrained optimization problems (Lippmann, 1987). In remote sensing, artificial neural networks have been used primarily for image classification, although other applications include geometric rectification (Liu and Wilkinson, 1992), feature extraction (Excell and Studholme, 1990; Ryan et al., 1991), feature selection (Ryan et al., 1991; Kimes et al., 1996) data compression (Fang, 1992; Manikopoulos, 1992), change detection (Gopal and Woodcock, 1996), stereo matching (Loung and Tan, 1992), and pattern matching (Paola and Schowengerdt, 1995a).

The concept of neural networks was originally developed in the 1940's with the first formal treatment of ANNs by McCulloch and Pitts (1943). Later that decade, Hebb

(1949) developed the concept that the weight on an input neuron should be augmented to reflect the correlation between the input and the unit's output and this "Hebb rule" has since played a prominent role in the development of ANN learning schemes. In 1957, Rosenblatt invented the perceptron and in 1962, Rosenblatt's research findings were published in his book *Principles of Neurodynamics*. Another significant work was Minsky and Papert's 1969 monograph *Perceptrons*, demonstrated the limitations of perceptrons and highlighted the inability of multi-layer perceptrons to achieve optimal results. In their work, Minsky and Papert presented what would be a fundamental obstacle to ANN development: how much does each unit in a multi-layer network contribute to the error between the actual and desired output? At this point in the history of ANN development, interest waned substantially. However, in these so called "dark ages" of ANN, significant research was still being conducted (Fukushima, 1975; Kohohen, 1982; Grossberg, 1988), albeit in the shadow of more impressive developments in the fields of artificial intelligence and expert systems.

A renaissance in ANN research occurred in the 1980's as a result of two significant advances in the field. First, Hopfield (1982) demonstrated that a highly interconnected network of units, initiated in some random state could achieve a stable final state with the use of feedback loops in the connection paths between the units. The second and probably more significant breakthrough was the development of the backpropagation algorithm by Rumelhart, Hinton and Williams (1986) which addressed the problems with multi-layer perceptrons raised by Minsky and Papert (1969). These milestones in ANN research, combined with advances in computing technology and a

proliferation of ANN algorithms led to a resurgence in ANN research and in the development of applications for ANN in the early 1990's.

There are several assumptions which are common to all neural networks (Atkinson and Tatnall, 1997): (1) information processing occurs at numerous simple elements called neurons; (2) signals are passed between neurons over connection links; (3) each connection link has a weight associated with it, which commonly multiplies the signal transmitted; (4) each neuron applies an activation function to its net input (the sum of the weighted input signals). The activation function is usually non-linear and is used to determine the output signal of the neuron (Hassoun, 1997). An artificial neural network is therefore characterized by its architecture, its activation function and its learning algorithm (Fausett, 1994)

3.2 Artificial Neural Network Architecture

The architecture of the neural network refers to the structure of its individual components. The network consists of a series of processing nodes which are organized into layers. A simplified three layer ANN is shown in Figure 3.1. The first layer or input layer functions as a distribution structure for the data being presented to the network and no processing occurs at this layer. The input layer is commonly followed by one or more processing layers and then finally, the output layer. The layers between the input and output layers are referred to as hidden layers. The interconnections between each node (on two separate layers) have a weight associated with them. When a value passes along the connection between two nodes, it is multiplied by the weight.

It is the weight values which contain the distributed learning information of the network (Paola, 1994). There are no interconnections between nodes in the same layer.

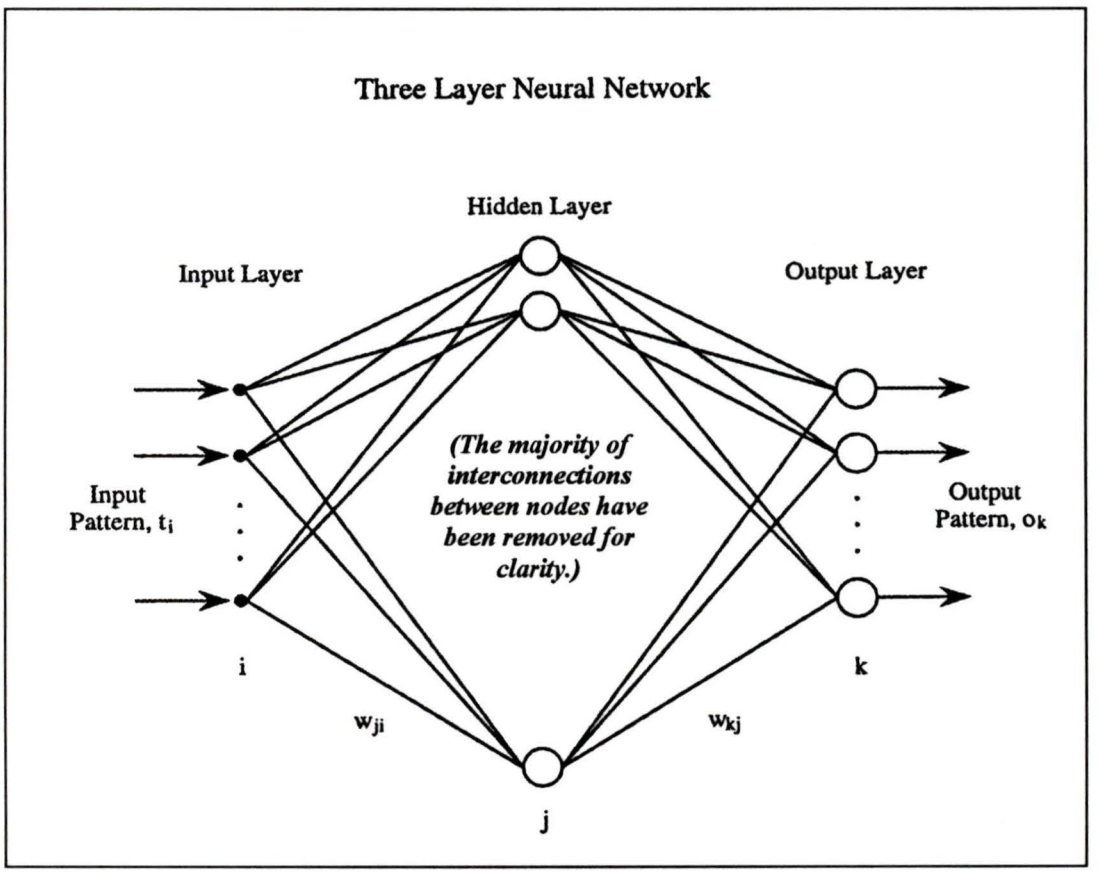


Figure 3.1 A Simplified, Three Layer Artificial Neural Network.
(Source: Adapted from Paola, 1994).

3.2.1 Processing Nodes and Weighting Factors

The neuron is the elemental processing unit of the neural network as shown in Figure 3.2. Also referred to as processing nodes, these neurons mimic biological neurons and perform two functions: (1) sum the value of their inputs; and (2) pass this sum through an activation function to produce the node's output value (Lau, 1992).

Processing nodes are organized into layers and every node in one layer of the network is directly linked to every node in the adjacent layer. There are no interconnections between nodes in the same layer. Neurons in the same layer will behave in a similar fashion, will have the same activation function and the same patterns of connection to other nodes. Each connective link has a weight associated with it. The weights are the information used by the neural network to solve the problems or in this context, assign pixels to classes (Masters, 1994). Through a process of supervised learning, a sequence of training patterns, each with an associated output vector, is presented to the network. The output vectors are adjusted according to the learning algorithm. All of the weights must be initialized to random values because it is theoretically impossible to obtain a set of unequal weights containing the distributed knowledge of the network if a set of equal weights is used in the initial configuration (Rumelhart et al., 1986).

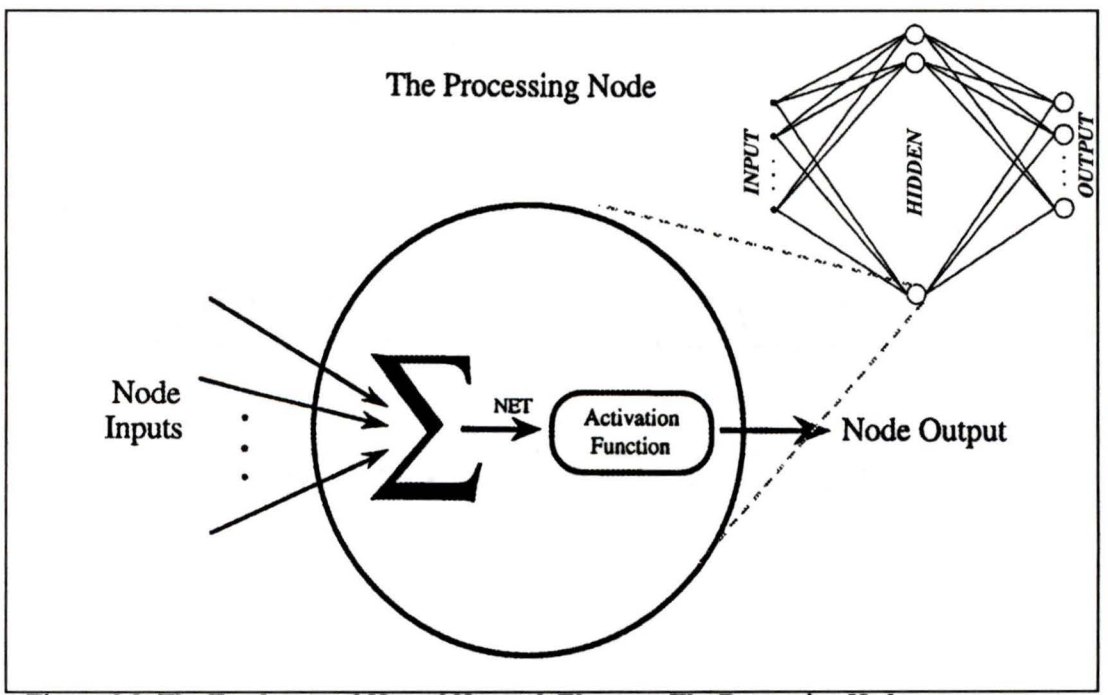


Figure 3.2 The Fundamental Neural Network Element: The Processing Node.
(Source: Adapted from Fausett, 1994 and Paola, 1994)

In Figure 3.3, neuron *B* is receiving input from neurons *A1*, *A2* and *A3*. The activations or output signals from these three neurons are *a1*, *a2* and *a3* and the weights are *w1*, *w2* and *w3*. The net input, *b_in*, is therefore the sum of the weighted signals from neurons *A1*, *A2* and *A3* as summarized in Equation 3.1 (Fausett, 1994):

$$b\_in = w1A1 + w2A2 + w3A3 \qquad \textit{[Equation 3.1]}$$

The activation of neuron *B*, *b*, is given by some function of its net input,

where $b = f(b\_in)$.

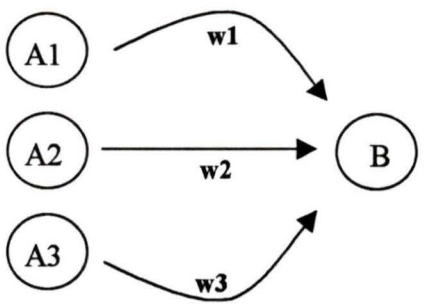


Figure 3.3 A Set Of Simplified Artificial Neurons And Their Weightings.

3.2.2 Sigmoid Activation Function

Any differentiable function can be used as the activation function however the sigmoid activation function (Figure 3.4) is the most common. The sigmoid activation function is nonlinear and has a range of 0 to 1. The non-linearity associated with the activation function contributes to the power of the multi-layer perceptron for solving complex problems.

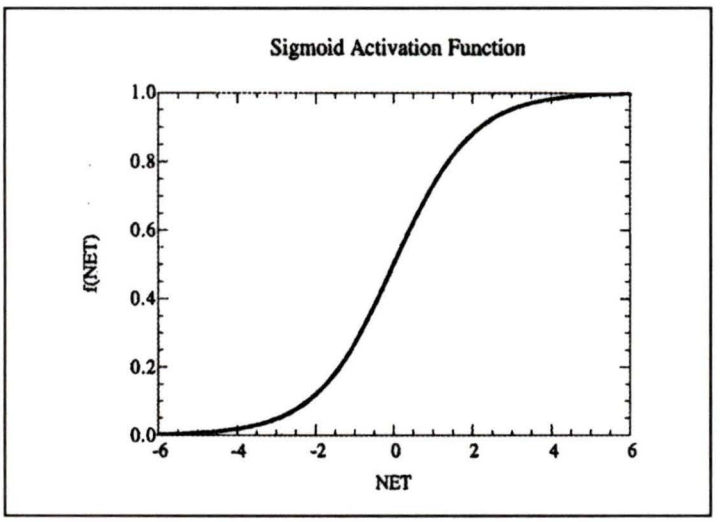


Figure 3.4 The Sigmoid Activation Function.

The equation to calculate the activation of neuron *B* is as shown in Equation 3.2 (Fausett, 1994):

$$f(x) = \frac{1}{1 + \exp(-x)}$$

[Equation 3.2]

In Figure 3.5, an additional layer has been added to the example previously illustrated in Figure 3.3. Once the activation function for neuron B has been calculated, its signal will be sent to both neurons *C1* and *C2* in the next network layer. The signal received at neurons *C1* and *C2* will not be identical because the signal will have been scaled by weights *w1* and *w2*. In a typical network the activations of *C1* and *C2* would depend on output from several neurons and not just one as portrayed in this simplified example. The basic operation of a neural network is to sum all of the weighted input signals and apply an activation function (Fausett, 1994).

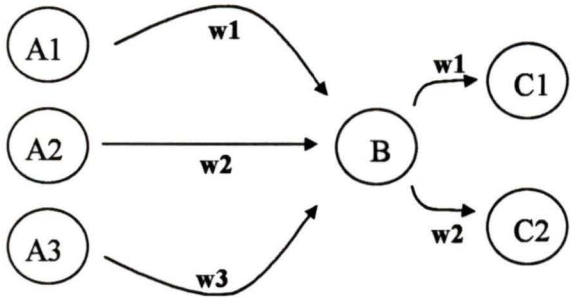


Figure 3.5 An Additional Layer Of Nodes Added To The Example From Figure 3.3.

Although the network structure presented in this example is very simple, the presence of a hidden unit and a non-linear activation function give it the ability to solve more complex problems than could be achieved with only input and output units. It is however more difficult to train networks with hidden layers (Lau, 1992).

3.2.3 Backpropagation Training Algorithm

Prior to the development of the backpropagation training algorithm by Rumelhart et al. (1986), there were no effective algorithms for adjusting the interconnecting weight values to minimize the overall training error in multi-layer networks (Lippman, 1987). Presently, backpropagation is the most commonly used training algorithm (Paola, 1994; Atkinson and Tatnall, 1997). The generalized delta rule or backpropagation is an iterative gradient descent training procedure, whereby the overall squared error between desired and actual outputs are minimized. The training data used by the backpropagation algorithm involves two data vectors. The input data vector is the pattern to be learned and this is analogous to training patterns used for the MLC. The desired output vector is the set of output values that should be produced by the network

upon recall of the input training patterns and is analogous to the class assignment values (Lippmann, 1987).

A detailed mathematical explanation of how backpropagation functions is given by Rumelhart et al. (1986) and a more general explanation is offered here. Essentially, backpropagation is a straightforward application of the chain rule of elementary calculus (Werbos, 1994). The error of one input training pattern is given by the following equation (Lippmann, 1987):

$$E = \tfrac{1}{2}\sum\left(d_k - o_k\right)^2$$

[Equation 3.3]

where d_k is the desired output vector and o_k is the actual output vector. The value of the actual output vector is determined by the evaluation of the activation function using the sum of the values produced at node k (corresponding to **NET** in Figure 3.2). The value of **NET** is itself a function of the weights between the hidden and the output layers and of the output from the hidden layer nodes. In order to minimize the error, the change in the weights from one iteration to another must be proportional to the derivative of the error with respect to the weights summed over all of the training patterns (Paola, 1994).

The learning rate determines what percentage of the step (in the gradient descent) is taken towards the minimum error in each iteration. In order to minimize the error at the output end of the network, the values of the weights must be adjusted accordingly and therefore (since there is no feedback loop) the error must be propagated back through the network from the output layer (hence the name "backpropagation"). Each set of

weights is updated by a function that is described in terms of the previous set of weights. As each training pattern is presented, the term at each node is summed and also, the total error between desired and actual output is also summed. If the error is still above some predetermined threshold when the training cycle is complete, the weights adjust as described above and the training continues (Atkinson and Tatnall, 1997).

The purpose of training is to achieve a balance between the ability of the network to respond correctly to the input patterns that are used for training and the ability to give reasonable responses to input that is similar, but not identical, to that used in training (e.g. to generalize to other data). The training of a backpropagation network involves three stages: (1) feed forward of the input training patterns; (2) calculation and backpropagation of the associated error; and (3) adjustment of the weights (Fausett, 1994). Once training is complete, the trained network is then applied to the complete image and functions much like a hard wired circuit, with only the feed-forward calculations being computed. Training is slow, but once trained, processing a set of data is fast relative to a maximum likelihood classification (Paola, 1994).

The neural network learns by changing the weights by an amount proportional to the difference between the desired output and the actual output. That is, the error at that unit times the output of the unit feeding into the weight. In the forward pass, the outputs are calculated and the error at the output units is calculated. In the backward pass, the output unit error is used to alter the weights on the output units and then, the error at the hidden nodes is calculated by backpropagating the error at the output units through the weights and the weights on the hidden nodes are then altered using these

values (Masters , 1994). These forward and backward passes continue until the error is minimized (Atkinson and Tatnall, 1997).

To obtain a true gradient descent of the overall training error, the weight change terms should be summed for all training patterns and the weights should not be adjusted until after the entire sum is complete and all the training patterns have been presented to the network. This is termed batch or epoch training (Lippmann, 1987). It is technically easier to implement sequential training however when adjustments occur after the presentation of each training pattern. Not only is this more time consuming, but it is not a true gradient descent method (Pao, 1989).

3.2.4 Learning and Momentum Rates

Other parameters which must be determined before the network is trained are the learning and momentum rates. As was mentioned previously, the learning rate establishes the step size for the error minimization in each weight adjustment (Lippmann, 1987). Normally, the learning rate is established through a time consuming process of trial and error, however, recently researchers have been implementing adaptive learning rates which increases the training speed because the rate of learning is automatically adjusted to the highest value which does not cause network instability (Paola, 1994; Heermann and Khazenie, 1992.)

Backpropagation is not guaranteed to find the global minimum error. During training, the network will take the steepest descent from the current position to one of lower error (Caudill, 1988). If the network encounters a valley (local minimum) in the error space, the network may stall and the error will not decrease to the global minimum

value. The ANN may also oscillate between two error values – indicating network instability. Generally, the higher the learning rate, the more likely it is that oscillation will occur (Pack et al., 1997). One way to avoid this problem is to add some fraction of the weight change calculated in the previous iteration to the weight update formula. The added push from this term can prevent the network from becoming stalled in a local minimum. This term is referred to as the momentum rate and an optimum value for this term must also be established through trial and error, although adaptive momentum rates can also be programmed into the network structure (Bhattacharya and Parui, 1997).

3.2.5 Training Convergence Criterion

Another network parameter which must be specified is the training convergence criterion. Since it is not feasible that the ANN will achieve an error of zero, some criterion must be used to establish an acceptable level of error and halt the training process (Rumelhart et al., 1986). This criterion must also be determined experimentally. It controls the degree of specialization versus generalization. If the network is over-trained it will become too specific to the training patterns, however if it is under-trained it will be unable to discriminate between the classes (Paola, 1994). The pragmatic solution to this problem is to save the network configuration at regular intervals throughout the training cycle so at the end there are multiple networks to chose from, all of which have different degrees of training. The convergence criterion also controls the training time.

3.3 The Multi-Layer Perceptron (MLP)

Perceptrons, developed by Rosenblatt in 1952, are single nodes which receive weighted inputs and threshold the result according to a predefined rule. The perceptron can thus classify linearly separable data but is unable to handle non-linear data (Atkinson and Tatnall, 1997). The perceptron was then adapted by Widrow and Hoff (1960) by the application of a one layer delta rule which facilitated supervised learning in the network via a least mean squares algorithm. The multiple layers of an MLP allow it to separate non-linear data and as the number of hidden layers increases, the network is able to learn more and more complex problems, but at the same time, loses its ability to generalize to other data (Lippmann, 1987). The MLP is the most widely known neural network architecture and is more powerful than a single layer network, as it has been proven that any pattern recognition problem can be learned given two hidden layers of units (Rosenblatt et al, 1986). The MLP consists of an input layer, one or more hidden layers and an output layer. Information flows forward from each layer to the next with no feedback and each layer is fully connected to the next. The nodes in the hidden layers operate on the sum of the inputs with sigmoid function.

The MLP functions as follows. First, the elements of the input vector are applied, one each, to the input nodes. The input nodes do not alter the input vector elements, except perhaps by linear scaling (Fausett, 1994). Then, the response of each of the nodes in the first hidden layer is evaluated by summing the products of the values on the input nodes with the connecting weights and then evaluating a sigmoid function with this sum as the argument. The next hidden layer (if it existed) would be evaluated

in the same manner and finally the output layer is also evaluated as previous layers but without the sigmoid activation function (Foody and Arora, 1997).

The MLP is popular in remote sensing applications but is not without its limitations. The fully connected nature of the MLP result in an evaluation time which scales approximately as the third power of the number of nodes (Heermann and Khazenie, 1992). The number of weights used directly influences the success of the output. As the number of weights increases the processing time of the MLP will also increase as all of the weights must be used in every evaluation. Two important criticisms of MLP stem from this problem: (1) it should not be necessary for every training vector to use the same weights unless the input vectors are correlated in some fashion; (2) for any given input, the number of weights relevant to the solution of the problem should be rather few in principle (Lippmann, 1987). A method needs to be identified that will facilitate the selection of the relevant weights for the application (which would thereby speed up the training process). Most implementations of MLP initiate the weights to random values (Paola, 1994).

3.4 Neural Networks Applications for Remote Sensing Image Classification

A wide variety of data sources have been analyzed using ANN, including Landsat and airborne multi-spectral scanner (MSS) (Ersoy and Hong, 1990; Benediktsson et al., 1993), Landsat TM (McClelland et al., 1989; Heermann and Khazenie, 1992; Moody et al, 1994; Baraldi and Parmiggiani, 1995; Wilkinson et al., 1995; Duguay and Leverington, 1996), synthetic aperture radar (SAR) (Decatur, 1989a; Decatur, 1989b; Dawson et al., 1994; Pierce et al., 1994; Grossberg et al., 1995),

passive microwave (Maslanik et al., 1990; Chen et al., 1995), Indian Remote Sensing Satellite (IRS) (Bhattacharya and Parui, 1997), Satellite pour l'Observation de la Terre (SPOT) (Dreyer, 1993; Tzeng et al., 1994; Bhattacharya and Parui, 1997) and Advanced Very High Resolution Radiometer (AVHRR) (Gopal and Woodcock, 1994). Many authors have declared that ANNs provide a viable alternative to conventional parametric classification methods (Heermann and Khazenie, 1990; Bruzzone et al., 1995). Civco (1993) boldly declared:

> "Given continued and expanded research in this area, it is expected that artificial intelligence approaches to image processing and pattern recognition will eventually surpass traditional techniques such as per-pixel maximum likelihood classification and perhaps some day rival human photointerpreters" (Civco, 1993, p.184).

Others have been less optimistic and more critical regarding the prospects of ANN for remote sensing image classification (Mulder et al., 1991).

3.4.1 Advantages and Disadvantages of Neural Networks for Image Classification

There are several advantages associated with the use of neural networks for image classification. One of the primary advantages of neural networks is their ability to learn by example and to generalize this learning to new data independent of statistical distribution assumptions (Foody and Arora, 1997). Because neural networks are non-parametric and involve no assumptions regarding the statistical distribution of the underlying image data, the feature space partitions are adjusted to conform to the natural distribution of the data, whether it be Gaussian, multi-modal or disjointed (Paola, 1994; Solaiman and Mouchot, 1994; Bruzzone et al., 1995). As was noted in previous sections, parametric methods assume that each class in the feature space has an

N-dimensional multivariate Gaussian distribution, where *N* is the number of input image bands (Atkinson and Tatnall, 1997). ANNs perform more effectively than their parametric counterparts when the feature space is complex and the distribution of the data deviates from the Gaussian norm (Liu and Xiao, 1991; Schalkoff, 1992; Benediktsson and Sveinsson, 1997). Thus, while parametric approaches adhere to an assumed model, neural networks rely on the image data directly, making them a more flexible tool for image classification (Atkinson and Tatnall, 1997). Parametric methods such as the MLC will however have the advantage over ANNs when the assumption of normality is not violated (Benediktsson et al., 1990; Peddle et al., 1994). Researchers have linked differences in classification accuracy between ANN and MLC to the distribution assumptions made by the MLC (Ritter et al., 1990; Fierens et al., 1994).

Another advantage of neural networks is the ease with which they can incorporate multi-spectral, multi-temporal and multi-source data (Civco and Wang, 1994), thereby facilitating synergistic studies (Benediktsson and Sveinsson, 1997). ANNs avoid the problem common in statistical multi-source data analysis of having to specify the degree of influence attributable to each data source in the classification. This additional information provides *a priori* knowledge and realistic physical constraints to the analysis, thereby improving classification accuracies (Foody, 1995a; Foody, 1995b). ANNs are able to effectively integrate ancillary data because they make no assumptions regarding the level of measurement precision or noise in the data (Hepner et al., 1990; Civco and Wang, 1994; Foody and Arora, 1997). It must be acknowledged that the additional valuable information supplied by the ancillary data may also be accompanied by noise, redundancy and confusion (Wang and Civco, 1996). The MLC is also capable

of incorporating ancillary data, providing the data is normally distributed. Ancillary data is integrated into a neural network classification by the addition of supplementary input nodes to the network, as shown in Figure 3.6. ANN as a classification tool is perhaps most appropriate when the data are of multiple types and cannot be modeled by a convenient multivariate statistical model (Benediktsson et al., 1993; Paola and Schowengerdt, 1994).

The ability of neural networks to function with the minimal training sets is often cited in the literature as a significant advantage of neural networks for image classification. There is however much debate over the importance of the size and quality of the training data that is required by the ANN. Benediktsson et al. (1990) concluded that the success of neural networks depends largely on the quality of the data set used to train the network.

Although an ANN will function with a smaller training set, the training data must still be representative of the image data, as is the requirement for most other supervised classification methods. Paola and Schowengerdt (1995b) found that the MLC was much more sensitive to the purity of the training sample than the ANN and other researchers have similarly found that ANNs can accept lower quality training data than the MLC (Yoshida and Omatu, 1994). Hepner et al. (1990) claimed that neural networks could transcend spectral variance and perform well even when presented with the most pure and minimal of training sets. Some studies however have concluded that ANNs are more sensitive to the representativeness of the training data (Benediktsson et al., 1993; Kim et al., 1995). Blonda et al. (1995) found that the number of iterations required for network convergence increased with the complexity of the training data and

moreover, that the computational costs for poor training data were faulty overlapping decision boundaries in the feature space.

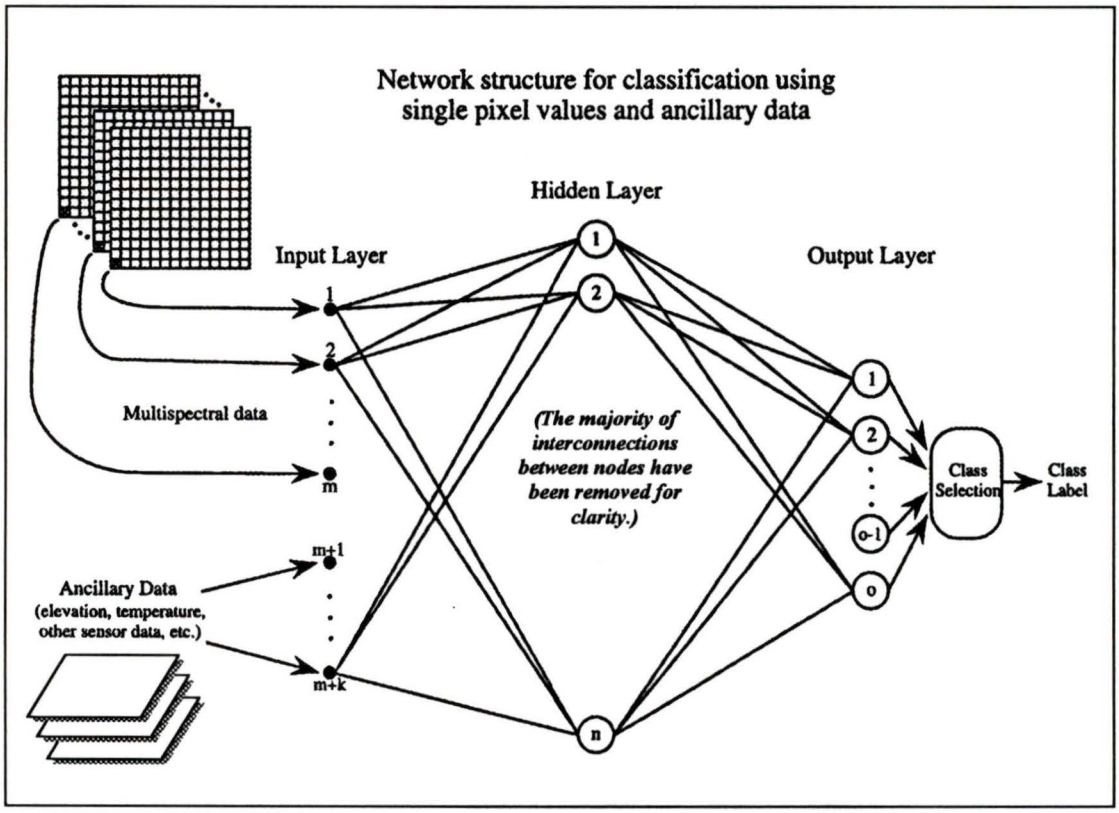


Figure 3.6 The Incorporation Of Ancillary Data Into An Artificial Neural Network.
(Source: Adapted from Paola, 1995)

Some researchers have linked the size of the training data set to the level of classification accuracy attained by an ANN (Hepner et al., 1990; Heermann and Khazenie, 1992; Benediktsson et al., 1993; Foody et al., 1995b; Kim, 1995; Blamire, 1996). Zhuang et al. (1994) found that an increase in the size of the training data set resulted in a small but statistically insignificant increase in accuracy, however with a correspondingly large increase in the network's training time. In their study, they

concluded that approximately 5% to 10% of the image data was needed to train a neural network classifier adequately to obtain satisfactory performance. Similarly, Gong and Chen (1996) found that using 10% of the image data was suitable for training and that the selection of random training samples gave better accuracy than contiguous samples. For the MLC it is desirable to select pure pixel areas for training, whereas the neural network requires training information from along the boundary between various categories and thus, atypical examples are more appropriate for neural network training (Bischof et al., 1992).

The status quo in remote sensing applications of ANN has been the presentation of equally sized training samples for each class under consideration, in order to avoid bias in the training process. Some researchers have used variable class sample sizes weighted deliberately to bias the training data. In this manner, the biased training data is similar to the *a priori* information used in conventional parametric methods (Foody et al., 1995a). Blamire (1996) hypothesized that broadly defined classes with a high degree of intra-class variability should be trained on larger samples than more narrowly defined classes. Furthermore, Blamire (1996) suggested that identical sample sizes may not be appropriate for avoiding bias in class learning and hypothesized that class definitions or the degree of intra-class variability should be the principal determinant of class size, rationalizing that neural networks need to spend longer periods of time training on complex classes in order to achieve an accurate representation of class structure.

Foody (1995a) found that higher classification accuracies could be achieved by incorporating *a priori* knowledge into the network. One method to incorporate this *a*

priori information when there is a large sample of training data available is to adopt Blamire's (1996) approach of varying the number of training pixels for each class according to the distribution of the classes in the image (Foody et al., 1995b). An alternative method used by Foody (1995a) and one that is more successful when there is a minimal training set available, is using *a priori* probabilities directly to modulate the activation level of the neural network. For this method, the output activation level of each output unit is multiplied by the prior probability of the class associated with it and therefore, class allocations are made on the basis of these modified activation levels.

While the number of training samples required by the MLC is quadratic with respect to the number of bands, the number of parameters required for an ANN is dependent on the hidden layer size and not necessarily on the number of bands (Paola and Schowengerdt, 1995b). ANNs therefore have the capacity to handle hyper-spectral data such as AVIRIS with 224 image channels that would require a minimum of 225 training samples per class for the MLC, or preferably between 2250 and 22500 training samples. A neural network training sample size as small as 30 pixels have been used to successfully classify AVIRIS data (Paola, 1994). An ANN is capable of outperforming the MLC in situations where the data is normally distributed, but the number of training samples is limited. As the complexity of the training data increases, the number of iterations required for network convergence also increases (Blonda et al., 1995).

Paola (1994) presented one of the most interesting advantages offered by neural networks: their mutually exclusive training property. He argued that ANNs produced a fundamentally different type of classification because the ANN uses the available null data during the training process to adjust the decision region boundaries. The training

data is therefore used to both enhance and suppress class membership allocations. It is this mutually exclusive training property, coupled with multiple layers of neurons, that allows the network to form non-linear decision boundaries in the feature space (Hepner et al., 1989; Hepner and Ritter, 1989; Solaiman and Mouchot, 1994).

As instruments of image classification, artificial neural networks have a significant number of disadvantages associated with them as well. Perhaps the greatest of these is the black box nature of ANNs. Neural networks are often referred to as a black box tool where the contribution of individual data sources and information about intermediate network states is neither observed nor fully discovered (Civco and Wang, 1994). Serpico and Roli (1995) referred to this disadvantage as "the opacity problem". In addition, the distribution-free nature of ANNs make statistical modeling and interpretation difficult (Key et al., 1989; Kanellopoulos et al., 1992). The manifestation of this is that while neural networks may perform more accurately than the MLC, it is often difficult to determine exactly why the ANN provided superior results (Kanellopoulos et al., 1991). One of the frustrations stemming from this black box approach is the trial and error method by which the neural network architecture is established. Dreyer et al. (1993) found that the optimal configuration of the network, although difficult to discern, was important for its effective performance. Compounding this problem is the length of time required to train a neural network for a particular application. Paola and Schowengerdt (1995b) concluded that the long training time of ANNs was the single most important drawback to their routine operational use. The time required to train a network is directly related to the size of the network and the amount of data used in the training process. Heermann and Khazenie

(1992) found that as the quantity of input to the neural network increases, so does the computation time by a factor of N^2, where N is the number of input features.

The "Hughes Phenomena" as described by Foody and Arora (1997) is the process by which class separability and ultimately classification accuracy increase initially with an increase in the number of input channels, beyond which additional channels have either no significant effect or lead to a reduction in overall accuracy. Therefore, it is wise to exclude bands which have low information content or which are highly inter-correlated and redundant. Benediktsson and Sveinsson (1997) examined feature selection strategies which would be beneficial for reducing the dimensionality of a data set and thereby the number of input nodes and overall complexity of the network. This increases efficiency of training while still maintaining a high level of accuracy. Benediktsson and Sveinsson (1997) used PCA and discriminant analysis but found that these methods do not take full advantage of the way neural networks define complex decision boundaries. They found that decision boundary feature selection was the best method to use with neural networks because it makes no assumptions regarding the distribution of the underlying data and fits the ANN strategy very well. They concluded that feature extraction methods are very important for ANN and can be used to find the best representation of the input data to the network. Chae et al. (1997) successfully used PCA to reduce the complexity of the ANN and decrease training time. Sergi et al. (1995) found that ANN classifications using PCA produced superior results to classifications using the normal spectral channels of TM.

Artificial neural networks are computationally intensive and complex (Clark and Canas, 1995). Their lack of transparency and the time required to train the network are

related issues because the architecture of the neural network will ultimately affect the efficiency of the network. One of the neural network parameters which must be supplied by the analyst are the initial values of the weights which lay on the interconnections between the nodes. These weights contain the distributed learning information of the network and therefore are critical components that are initialized by most ANN software to random values. Paola and Schowengerdt (1994) cited the need for user supplied initialization variables as one of the main weaknesses of neural networks. The results of an ANN are subject to random variation, since these weights are randomized prior to training although experimentation has shown that this is not a significant problem. Paola and Schowengerdt (1997) found that this initial weight randomization was a primary factor for network performance and hidden layer size was a secondary factor. This initialization effect decreases as more hidden layer nodes are added to the network.

Determining the appropriate number of layers and nodes can also be problematic. While recent work has attempted to provide heuristics for network configuration (see section 4.8.1.1) the selection of an appropriate architecture can be an extremely daunting task for the novice. Foody and Arora (1997) found that classification accuracy was significantly influenced by the size of the training set, discriminating variables in the network architecture and the nature of the testing set used. The architecture of the network has an effect – such that large and complex networks may have higher accuracies but these are not statistically significant. A similar effect was found by Leverington and Duguay (1996) who concluded that neural network behaviour with changing network performance is not necessarily consistent, as

is apparent when increasing the number of iterations does not lead to increases in classification accuracy and similarly, altering the number of hidden nodes also has no significant effect on accuracy. Paola and Schowengerdt (1997) therefore concluded that hidden layer size made little difference to the final classification accuracy and that the optimization of the number of hidden layer nodes was unnecessary.

New methods are being developed which would determine the optimal size of a network for a given set of input data (Dreyer et al, 1993; Dawson et al, 1994). Mozer and Smolensky (1989) optimized an ANN by calculating the relevance of each neuron in the input and hidden layer and then subsequently removed the least important neurons. Le Cun et al. (1990) optimized by removing unnecessary interconnecting weights. Dawson et al. (1994) found that by reducing the number of redundant and/or useless units through a network pruning technique, the network training time is reduced. The underlying philosophy behind these network pruning strategy is based on "Occam's Razor", an empirical rule based on the observation that the neural network with the fewest number of neurons gives the best generalization (Dreyer et al., 1993). Foody and Arora (1997) found that a larger, more complex network is better equipped to correctly characterize the training data but may be less able to generalize to data on which it has not been trained. Mayer et al. (1996) attempted to address the problem of finding an appropriate ANN topology by using a genetic algorithm to adapt topology to their specific application. Various methods for topological specialization exist including network growing, network pruning and evolutionary design. Zhou and Civco (1996) similarly found that an ANN with a genetic algorithm as a learning mechanism

can avoid the over-fitting problem by using the fewest number of hidden nodes possible and thereby guaranteeing the most conservative generalization.

Benediktsson et al. (1993) cited the ease with which a network can become over-trained and therefore over-specific to the training data and less able to generalize to the remaining image data as another disadvantage. This problem can be overcome by using a cross-validation technique, where a validation test set is used to evaluate the learning process in addition to the training samples (Fischer and Gopal, 1993). Bhattacharya and Parui (1997) also used this cross-validation technique. The errors are computed on the training and test set at regular intervals for five consecutive sweeps and initially errors on both the test set and the training set will decrease and then the error on the test set will start to increase even though the error on the training set is still decreasing. At the point in time when the error on the test set increases for at least three consecutive sweeps, training is stopped.

It is important to note however, that once the neural network has been trained and has converged, it becomes a very efficient feed-forward circuit which may then be used on different image dates of the same scene, on a completely different scene, or as some have suggested, on a scene collected from a different sensor (Côté and Tatnall, 1995; Bruzzone et al., 1997). Paola and Schowengerdt (1995b) also found that an ANN, once trained, performed classifications faster than the MLC, however the ability to amortize the long training period of the ANN over several applications of the trained network to multiple or multi-temporal images may not be advantageous if the images need to be normalized for atmospheric and sun angle differences.

Sui (1994) identified several additional problems of ANNs. These include: (1) the local minima problem (there is no way to tell if the global minimum error has been achieved by the network); (2) the network paralysis problem (which can be minimized by the use of small momentum terms and low learning rates); and finally (3) the convergence rate problem (there is no proof that backpropagation learning will ever converge within a finite number of learning iterations – most ANN will converge but training process is unpredictable and lengthy).

To summarize, the main advantages of neural networks as tools for remote sensing image classification are as follows:

1. ANNs are non-parametric. They do not depend on assumed *models* but rather on the image *data* itself and conform to the natural distribution of the data.
2. ANNs have the ability to learn by example and to *generalize* this learnt information.
3. ANNs are able incorporate ancillary data with ease.
4. ANNs have the ability to function with a minimal training set.
5. ANNs possess a mutually exclusive training property. Training data is therefore used to both enhance and suppress class memberships.
6. Once trained, ANNs are very fast and portable.

Conversely, the main disadvantages of neural networks are as follows:

1. ANNs have a black box nature, commonly referred to as "the opacity problem". The inner-workings of the ANN remain somewhat mysterious.
2. ANNs have extremely lengthy *training* times.
3. The Hughes Phenomena describes the diminishing returns of adding more spectral input bands.
4. ANNs are computationally intensive and complex.

5. ANNs require user supplied initialization variables. Most algorithms implement random initial values for weights and this leads to the randomization problem.
6. The determination of ANN architecture through trial and error.
7. There is the potential to over-train the neural network.
8. The local minima problem may cause the ANN to stall.

Based on this review of the advantages and disadvantages of neural networks, it is apparent that neural networks can be expected to be most successful in situations where the feature space is complex, the distribution of the image data deviates from the Gaussian norm and there is limited training data available. There are however several disadvantages which will clearly hamper the operational use of neural networks for image classification, in particular, the manner in which the network architecture must be determined for each application. Clearly, there needs to be a stronger theoretical understanding of how the network architecture is reflected in the results of the classification.

3.4.2 Alternative Applications of Neural Networks

There are three main trends in the alternative applications of neural networks. These include: (1) the use of alternative neural networks architectures other than the MLP and the combination of neural networks with other classification methods; (2) the combination of many networks in modular or hierarchical neural network; and (3) the incorporation of ancillary data or textural information. There are over thirty different types of neural networks currently being used by researchers in a wide variety of disciplines (Sui, 1994). Many researchers in remote sensing have attempted to use

neural networks other than the multi-layer perceptron for their applications and some have argued that the MLP will eventually be replaced by a more efficient neural network (Salu and Tilton, 1993). Another focus in neural network research in remote sensing has been the combination of neural network models with other techniques that serve to enhance network performance. For example, Benediktsson et al. (1991) combined statistical consensus theory with a neural network to develop a consensual neural network, Clark and Canas (1995) suggested an operational system which incorporated neural networks and genetic algorithms in a complementary approach and Wilkinson et al. (1992) combined neural networks and expert systems.

Tzeng et al. (1994) modified the standard backpropagation algorithm used by the MLP with a Kalman filtering technique which increased the convergence rate in the learning stage and enhanced the separability of classes for highly non-linear boundary problems. Moody et al. (1996) suggested that neural networks may be ideal for sub-pixel classification in coarse resolution data because of an observed correspondence between the magnitude of the neural network outputs and the relative proportions of the sub-pixel classes and the increase in network uncertainty as pixels become increasingly mixed. Murai and Omatu (1997) combined a neural network with the k-means clustering algorithm and knowledge based processing in a hierarchical approach, with the latter technique being used for fixing any misclassifications of pixels resulting from the ANN. Liu and Xiao (1991) found that the blocked backpropagation algorithm performed more efficiently than a standard backpropagation algorithm when data sets are very large in size and there are spectral overlaps among classes. Fogelmann Soulie (1991) in examining the history of ANNs, highlighted the evolution from simple

networks to hybrid and complex architectures. They concluded that the way of the future for ANN research would be an integration of neural networks and conventional techniques within hybrid systems.

The other main focus of research has been the development of modular or hierarchical networks. Blonda et al. (1995) applied a modular neural network to the classification of poor, low dimensional remotely sensed data. The modular nature of the network reduced the computational complexity of the training stage, thereby increasing the efficiency and generalization capability of the ANN, with no loss in accuracy. In their research, they determined that by incorporating fuzzy logic into the neural network, it could more readily model uncertain and ambiguous data which is often encountered in real applications as did Zhang and Hoshi (1994). Earlier work by Blonda et al. (1994) also supports this need for modularization and successive learning for high dimensional data. Ito and Omatu (1997) also used a modular approach which incorporated a k-nearest neighbour technique to classify unknown pixels. San Miguel-Ayanz and Biging (1997) used a modular approach for forest classification from TM and SPOT data and incorporated a band selection process for each network. Many others have also used this modular approach (Benediktsson et al., 1991; Kiang, 1992; Serpico and Roli, 1995).

Texture information, generally in the form of a 3x3 window of image data improved classification accuracy in several studies (Hepner et al., 1990; Key et al., 1990; Bischof et al., 1992; Kushwaha, 1994; Miller et al., 1995; Arai, 1997; Bruzzone, 1997). Contrary to these results, Gong et al. (1996) found that texture measures did not improve classification accuracies. Both Moody et al. (1994) and Medler and Yool

(1997) were able to successfully improve classification accuracies by incorporating additional non-image information.

3.5 Summary

Artificial neural networks, as tools for remote sensing image classification, have been extensively researched in the past decade. While researchers continue to experiment with different applications of neural networks and different configurations of the neural network parameters, there appears to be a need for a stronger theoretical understanding of how the ANN functions and how the various ANN parameters may influence the classification output. Empirical research into the application of ANNs for image classifications has resulted in the development of several heuristics for the successful configuration of an ANN.

There are several advantages and disadvantages associated with the use of neural networks for image classification. The fact that ANNs are non-parametric and are able to generalize their learned information to other data are their primary advantages, while the main disadvantages of ANNs are undoubtedly their opacity and the length of time required for their construction and training. It is clear from reviewing the existing literature that ANNs are viable classification methods which have been successfully implemented for vegetation classification on a wide range of remotely sensed data sources. The use of ANNs in remote sensing will continue to be influenced by the developments in ANN research in other disciplines and new neural network configurations and theoretical insights may lead to the increased use of ANNs for remote sensing applications.

CHAPTER FOUR

METHODOLOGY

4.1 Characterization of the Project Study Area

The project study area is approximately 65000 hectares in size and extends from 49°00′00″N to 49°12′00″N and from 121°24′00″W to 121°48′00″W as shown in Figure 4.1 and Figure 4.2 (a) and Figure 4.2 (b). The area encompasses much of the Chilliwack River Valley, which is oriented west to east, and which corresponds to the region covered by four contiguous 1:20000 TRIM mapsheets: 92H.002, 92H.003, 92H.012 and 92H.013. The study area is dominated by the presence of Chilliwack Lake at the eastern end of the valley.

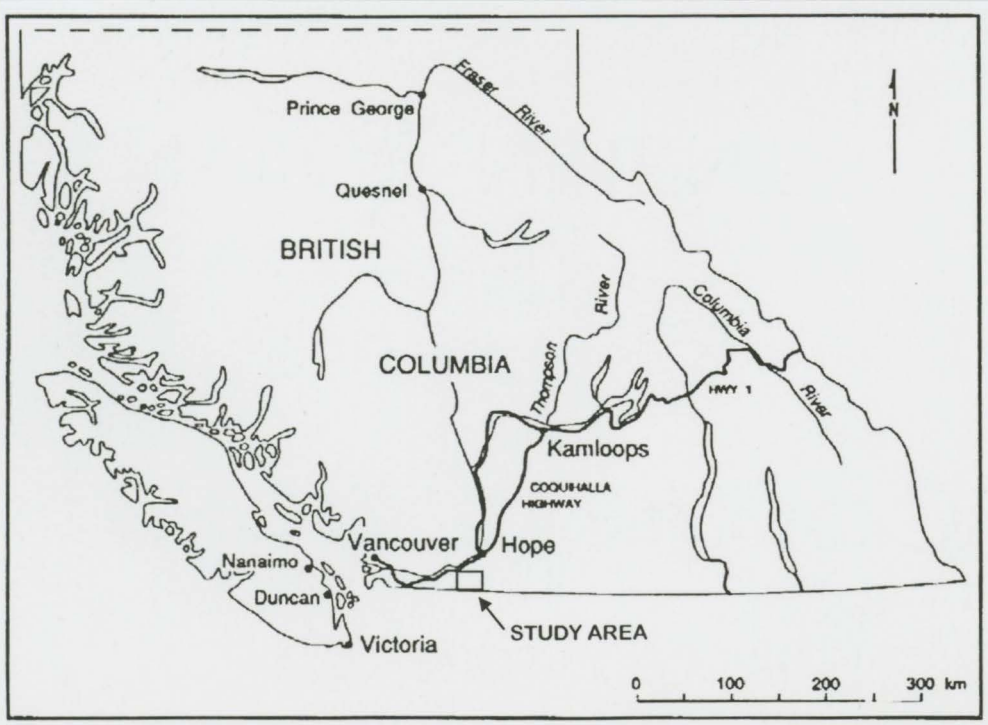


Figure 4.1. The Project Study Area.

Figure 4.2 (a). Project Image (July, 1994), Thematic Mapper Bands 1, 2, and 3.

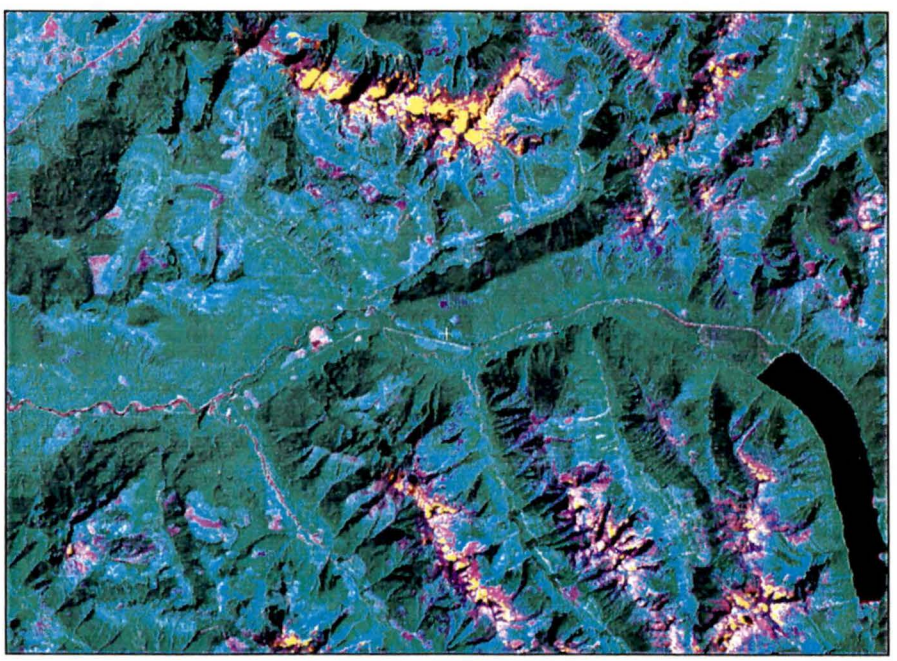


Figure 4.2 (b). Project Image (July, 1994), Thematic Mapper Bands 3, 4 and 5.

The study area is dominated by three main biogeoclimatic zones (Green and Klinka, 1994): 58% Coastal Western Hemlock (CWH), 39% Mountain Hemlock (MH) and 3% Alpine Tundra (AT) (Figure 4.3) and five biogeoclimatic sub-zones: dry maritime (dm) (15%), dry sub-maritime (ds) (6%), moist sub-maritime (ms) (31%), moist maritime (mm) (38%) and very wet maritime (vm) (7%) (Figure 4.4). The predominant biogeoclimatic unit in the area is the Southern Moist Sub-maritime Coastal Western Hemlock Variant (CWHms1) (Green and Klinka, 1994). The climate of this variant is described as transitional between the coast and the interior regions of the province and is characterized by moist, cool winters with heavy snowfall, particularly at higher elevations, and cool, relatively dry summers. Western hemlock (*Tsuga heterophylla)*, Douglas-fir (*Pseudotsuga menziesii)*, western red cedar (*Thuja plicata)* and amabilis fir (*Abies amabilis)* are considered the dominant species in this variant (Green and Klinka, 1994). Several other sub-zones and variants are found within the study area and these are summarized in Figure 4.5 and Table 4.1.

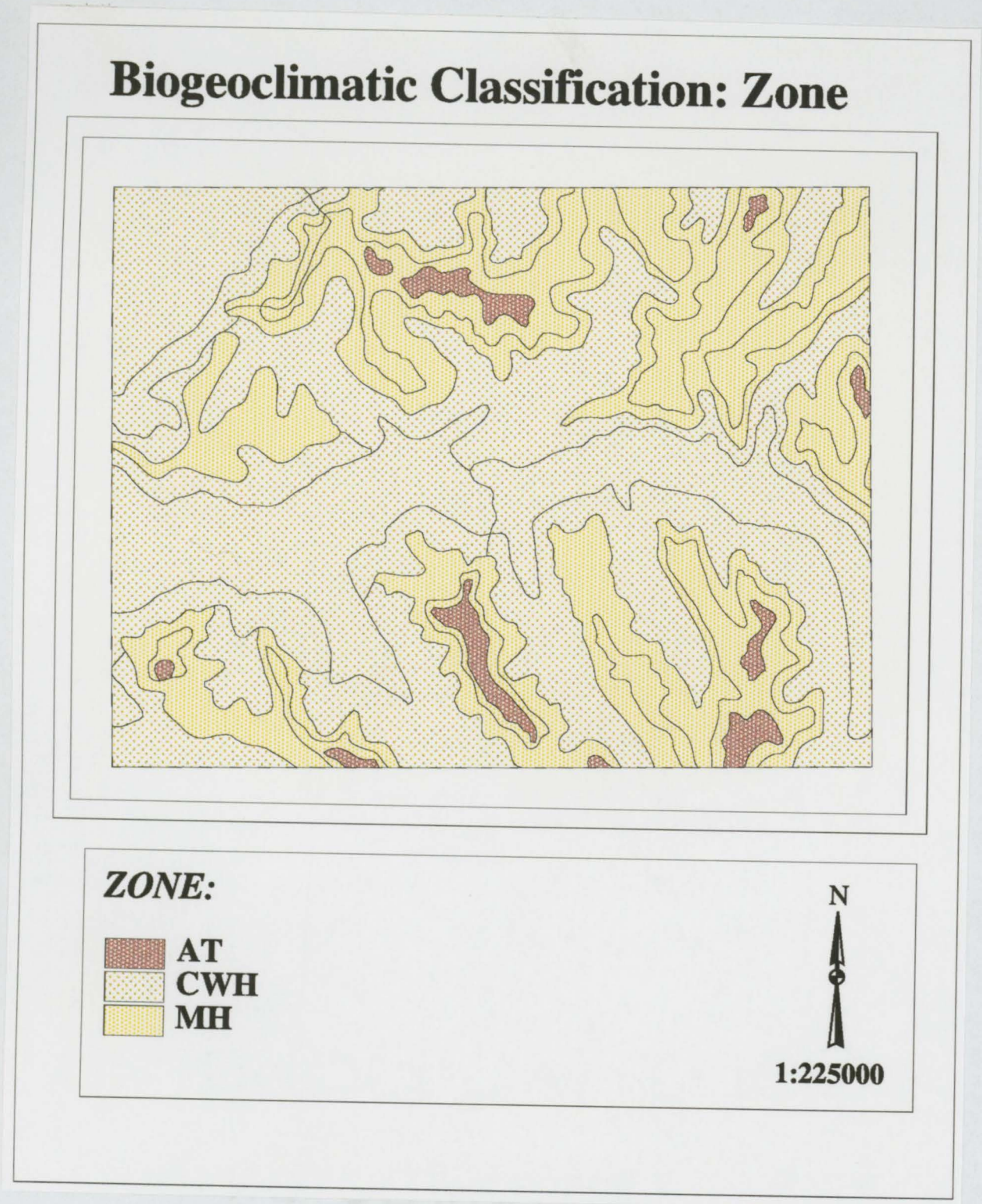


Figure 4.3. Biogeoclimatic Zones Found Within The Study Area.
(Source Data: Ministry of Forests, Research Branch, 1996)

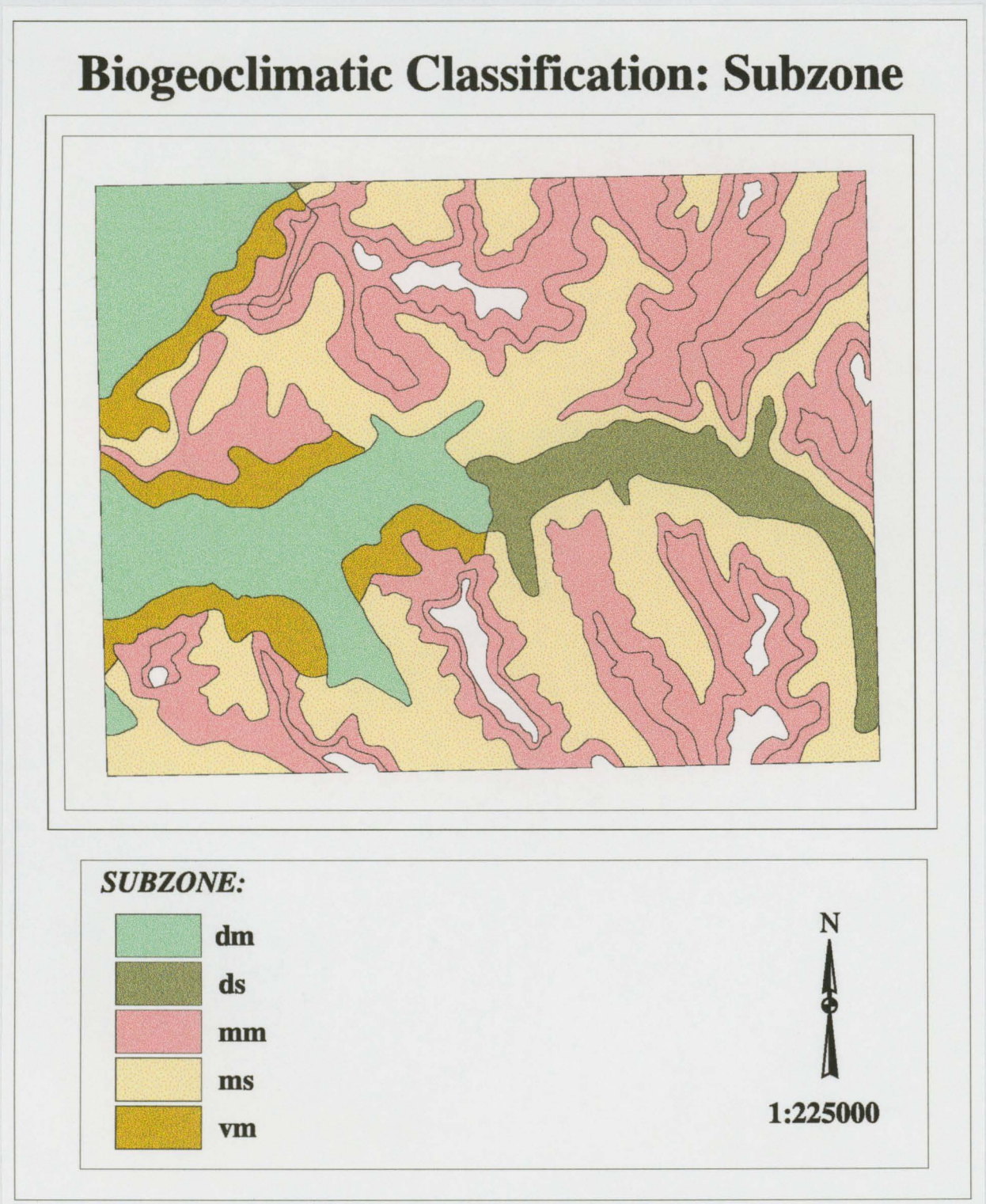


Figure 4.4. Biogeoclimatic Sub-Zones Found Within The Study Area.
(Source Data: Ministry of Forests, Research Branch, 1996)

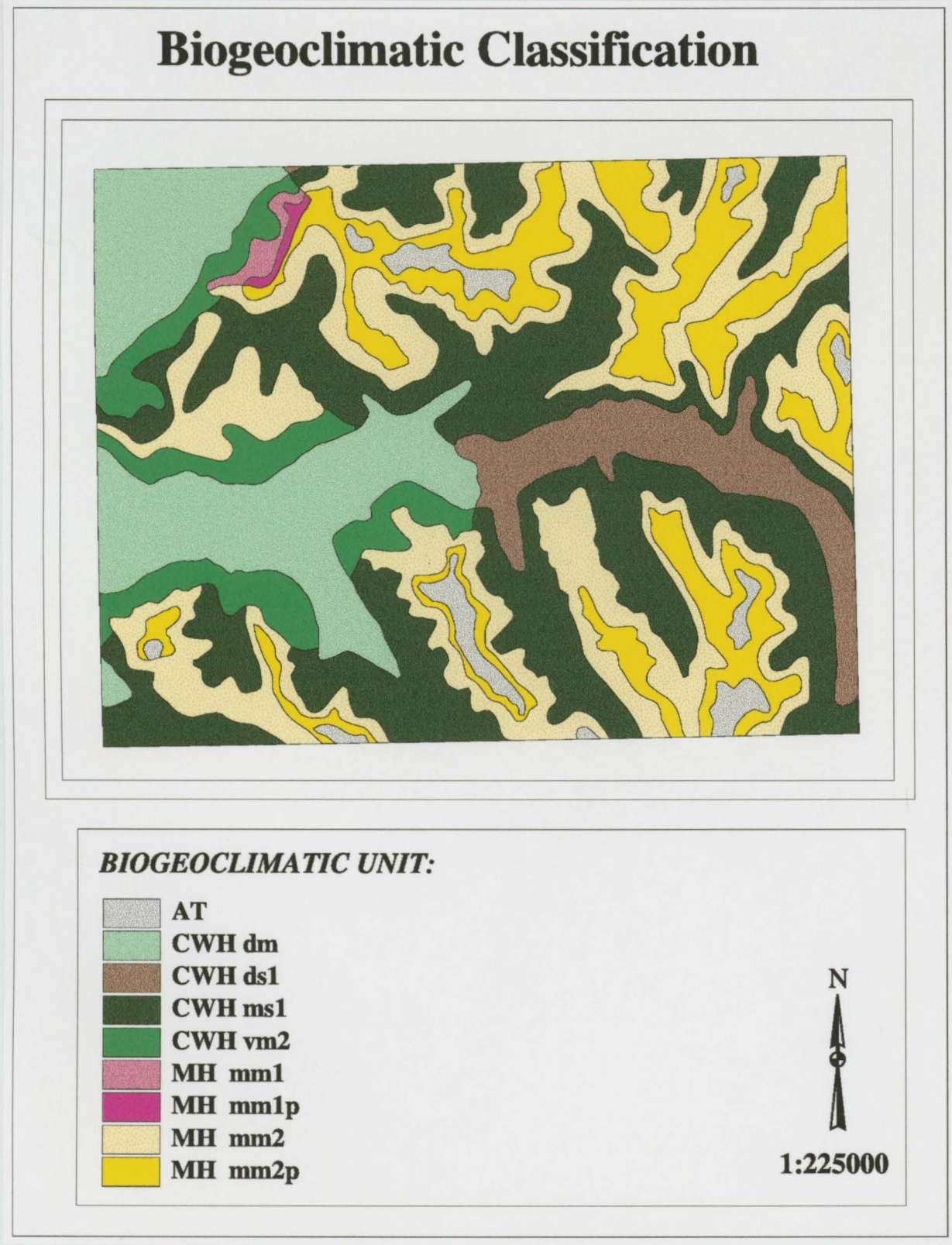


Figure 4.5. Complete Biogeoclimatic Classification Within The Study Area.
(Source Data: Ministry of Forests, Research Branch, 1996)

Table 4.1 BEC Distribution Within The Study Area.
(Source Data: Ministry of Forests, Research Branch, 1996)

BEC CLASS	HECTARES	PERCENT
AT	1768.55	2.72
CWHdm	9541.59	14.69
CWHds1	4116.66	6.34
CWHms1	19881.45	30.62
CWHvm2	4354.23	6.71
MHmm1	327.53	0.52
MHmm1p	158.99	0.24
MHmm2	15647.59	24.09
MHmm2p	9138.32	14.07
TOTAL	***64934.95***	***100.00***

The leading forest cover species in the region are shown in Figure 4.6. Forest inventory information verified that the dominant leading species in the region were Douglas-fir (*Pseudotsuga menziesii)* (32.99%), amabilis fir (*Abies amabilis)* (29.52%) and western hemlock (*Tsuga heterophylla)* (23.71%). Broadleaf species accounted for only 5.65% of the leading species in the area. A complete summary of the distribution for all of the forest cover species is included in Table 4.2. The non-productive forest components of the landscape in the study region are primarily alpine (64%) and alpine forest (12%), as summarized in Figure 4.7 and Table 4.3.

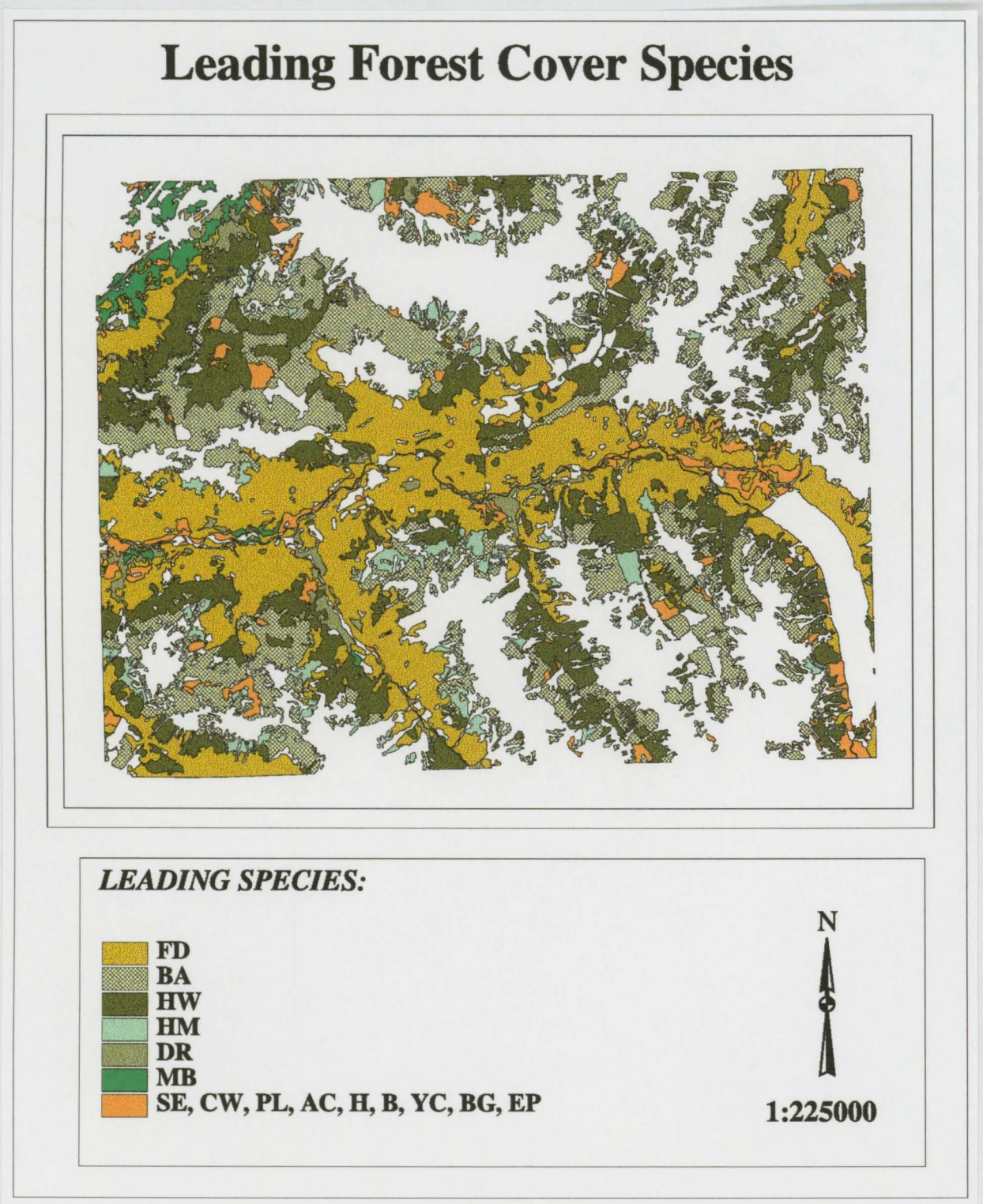


Figure 4.6. Distribution Of Leading Forest Cover Species.
(Source Data: Ministry of Forests, Resources Inventory Branch, 1996)

Table 4.2 Leading Forest Cover Species.
(Source Data: Ministry of Forests, Resources Inventory Branch, 1996)

SPECIES	CODE	AREA (ha)	PERCENT
Balsam Poplar	AC	348.58	0.86
Fir	B	96.48	0.24
Amabilis Fir	BA	11954.26	29.52
Grand Fir	BG	75.94	0.19
Western Red Cedar	CW	388.29	0.96
Red Alder	DR	1135.42	2.80
Common Paper Birch	EP	58.87	0.15
Douglas Fir	FD	13357.08	32.99
Hemlock	H	331.15	0.82
Mountain Hemlock	HM	1512.59	3.74
Western Hemlock	HW	9600.17	23.71
Broadleaf Maple	MB	745.83	1.84
Lodgepole Pine	PL	385.96	0.95
Engelmann Spruce	SE	399.82	0.99
Yellow Cedar	YC	98.54	0.24
TOTAL		***40488.98***	***100.00***

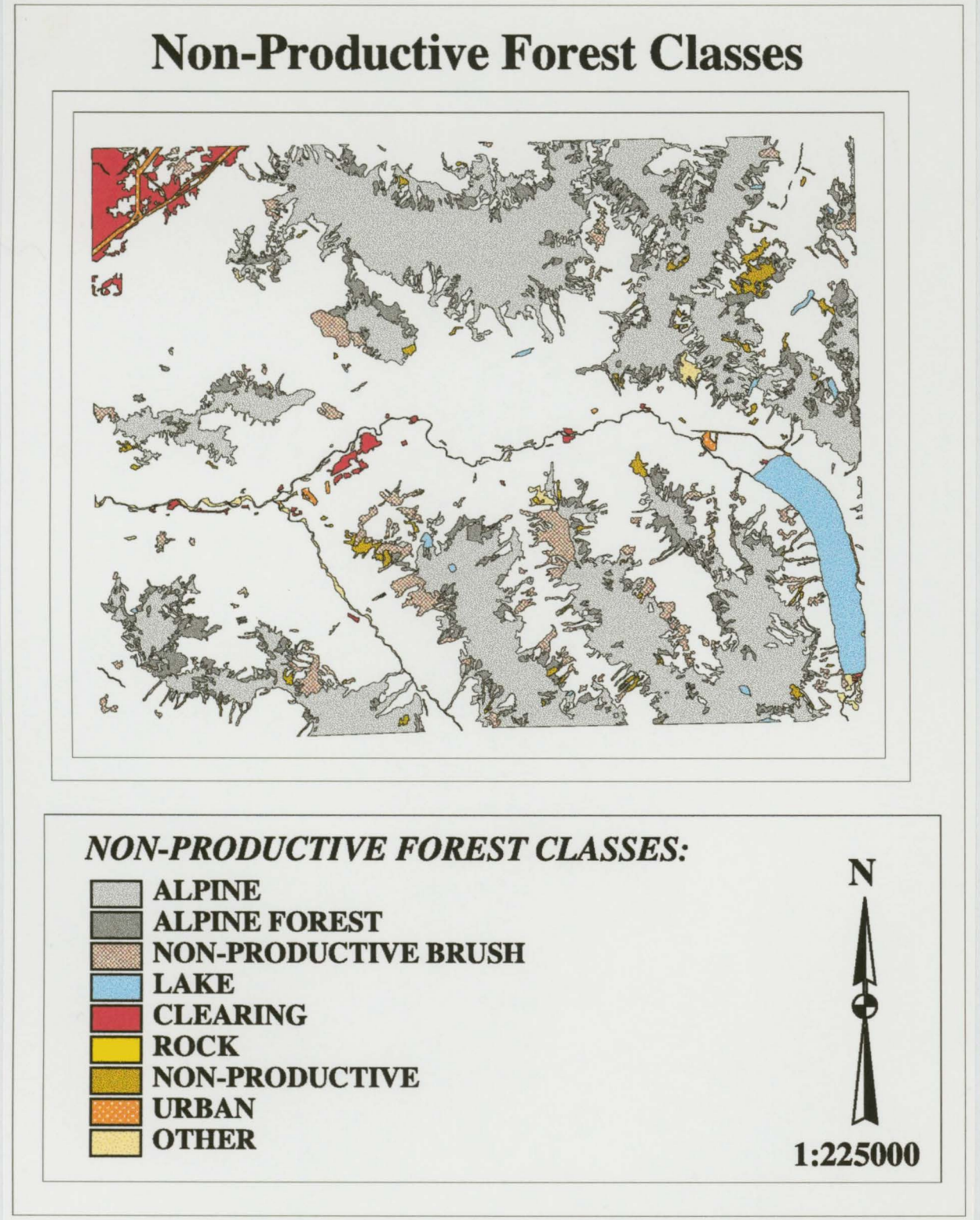


Figure 4.7. Distribution Of Non-Productive Forest Classes.
(Source Data: Ministry of Forests, Resources Inventory Branch, 1996).

Table 4.3 Non-Productive Forest Classes.
(Source Data: Ministry of Forests, Resources Inventory Branch, 1996)

NON-PRODUCTIVE DESCRIPTOR	CODE	AREA (HA)	PERCENT OF TOTAL AREA
Alpine	A	17456.06	64.55
Alpine Forest	AF	3273.88	12.11
Clearing	C	967.69	3.58
Gravel Pit	G	16.82	0.06
Lake	L	1280.12	4.73
Non-Productive	NP	623.92	2.31
Non-Productive Brush	NPBR	1936.61	7.16
Non-Productive Burn	NPBU	101.09	0.37
Rock	R	830.46	3.07
River	RIV	229.47	0.85
Swamp	SWAMP	51.78	0.19
Urban	U	275.88	1.02
TOTAL		***27043.79***	***100.00***

The topography in the region is highly variable, with an elevation range of 0 to 2400 metres and a mean elevation of approximately 1300 metres and a standard deviation of 500 metres. The variation in the topography is represented in Figure 4.8.

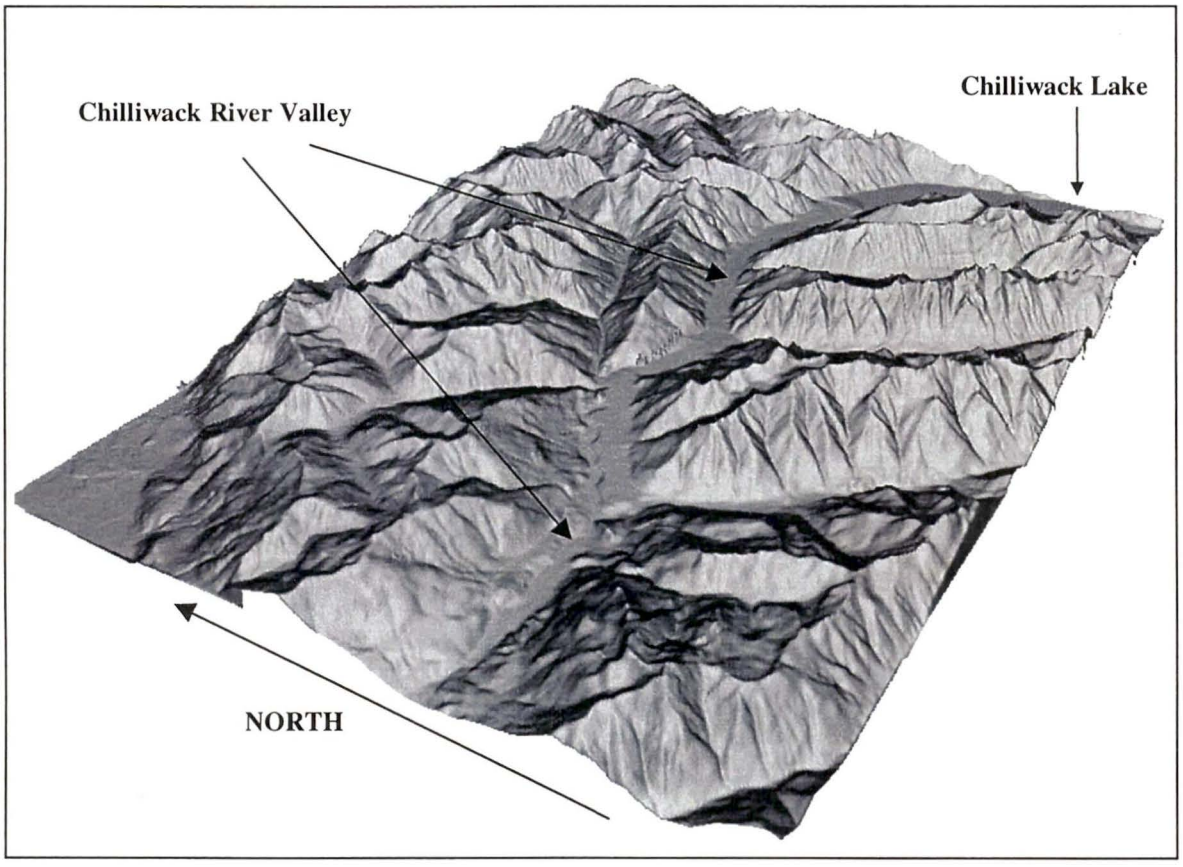


Figure 4.8. Three Dimensional Rendering Of Study Area Topography Derived from TRIM Elevation Model.

4.2 Project Imagery

The imagery used for the vegetation classification trials included sub-scenes (987 pixels by 723 lines) from four Landsat 5 Thematic Mapper quadrant scenes, collected on the following dates: August 29, 1993; September 30, 1993; July 31, 1994; and October 3, 1994. All of these images cover the same geographical area: path 046, row 026, quadrant 1. The image meta data are summarized in Table 4.4.

Table 4.4. Landsat TM Image Meta Data.

Image Date	Time	*Scene Center*		*Corrections*	
		Latitude	Longitude	Geometric	Atmospheric
29/08/93	18:17:31	N 49:18:13	W 121:43:46	Precision	Calibrated
30/09/93	18:17:25	N 49:18:24	W 121:43:10	Systematic	Calibrated
31/07/94	18:12:50	N 49:18:21	W 121:47:15	Systematic	Calibrated
03/10/94	18:11:10	N 49:18:24	W 121:48:23	Systematic	Calibrated

The August image date was precision georeferenced, while the other three dates were systematic georeferenced. The raw data that the Landsat satellites collect contains geometric distortions that must be compensated for before it can be used. The image vendor preprocesses the imagery to compensate for distortions caused by: the motion of the Earth; the motion of the spacecraft; the motion of the sensor; detector placement on the prime and cold focal planes; and the curvature of the Earth (Earth Observation Satellite Company, 1994a). Systematic georeferencing orients the image with "north up" and corrects the image to a user-specified map projection (UTM NAD 27). Precision georeferencing provides greater positional accuracy than systematic processing and for this process, ground control points (GCPs) as well as a map projection are used to spatially align the image. The difference between the two is reflected in the measured geometric error reported in the image work order. The

absolute RMS is 62.9 metres for the precision georeferenced image compared to 1771.6 metres for the systematic processed images.

Radiometric adjustments performed by the image vendor calibrate the sensor to account for the independent behavior of each of the on board detectors (for TM there are 16 detectors for each non-thermal band) so that all of the detectors use the same data number for the same input radiance. The purpose of the radiometric correction is to produce an output digital number (DN), which normalizes the response of each individual detector to some common radiance range. The internal radiometric calibration system for the TM reflective bands consists of three calibration lamps and a shutter, all of which are seen by the detectors at the end of each scan. The raw calibration data consists of up to 1,024 bytes of calibration data starting at the end of a TM image scan (Earth Observation Satellite Company, 1994b).

Initial assessments of the imagery indicated that there was no excessive cloud cover, haze or noise in the image. Some cloud cover was present in the high elevation areas in the northwest corner of the July 1994 and September 1993 quadrant scenes, however these were beyond the bounds of the study area. Summaries of the image statistics are included in Appendix A.

4.3 British Columbia Land Cover Classification System

One of the assumptions implicit in any remote sensing image classification application is that the desired informational classes have a strong correlation with spectral classes and that this correlation is invariant across the entire image (Duggin and Robinove, 1990). For the research reported in this thesis, the informational classes were pre-defined by the structure of the British Columbia Land Cover Classification System (BCLCCS), which combines the physiognomic and floristic properties of existing vegetation. A selected portion of the BCLCCS, for which it was thought plausible training data could be derived from existing forest inventory information, was selected for investigation. To determine the spectral separability of the classes at the various levels of the BCLCCS, histograms of DN values in Landsat Thematic Mapper band 4 for each of the relevant classes were created and are included in Appendix B.

The BCLCCS was designed by the Vegetation Inventory Working Group (a component of the provincial Resources Inventory Committee) to provide the framework for a more objective and value neutral vegetation inventory than the existing provincial forest inventory which is strongly oriented towards timber management. The broader scope of the new Vegetation Resources Inventory (VRI) was designed to facilitate the exchange of information for the purposes of integrated resource management (Bancroft, 1995). The VRI was proposed to fulfill three main objectives: to determine (1) the amount and (2) the location of vegetation cover in the province; and (3) to monitor changes in the amount and location of vegetation resources over time. The new inventory is to be implemented in a three phase process: (1) air photo interpretation will be used to establish initial attribute estimates; (2) ground sample measurements will be

used to verify these attribute estimates; and (3) statistical analysis will be used in the adjustment of the initial estimates from the photointerpretation, based on the results of the ground sampling.

The BCLCCS is a hierarchical classification system which is organized into a series of five levels ranging from broad, generalized classes at level one to more specific, detailed classes at level five. The structure of the five levels differs for the vegetated and non-vegetated polygons, as indicated in the generalized schema presented in Figures 4.9 and 4.10.

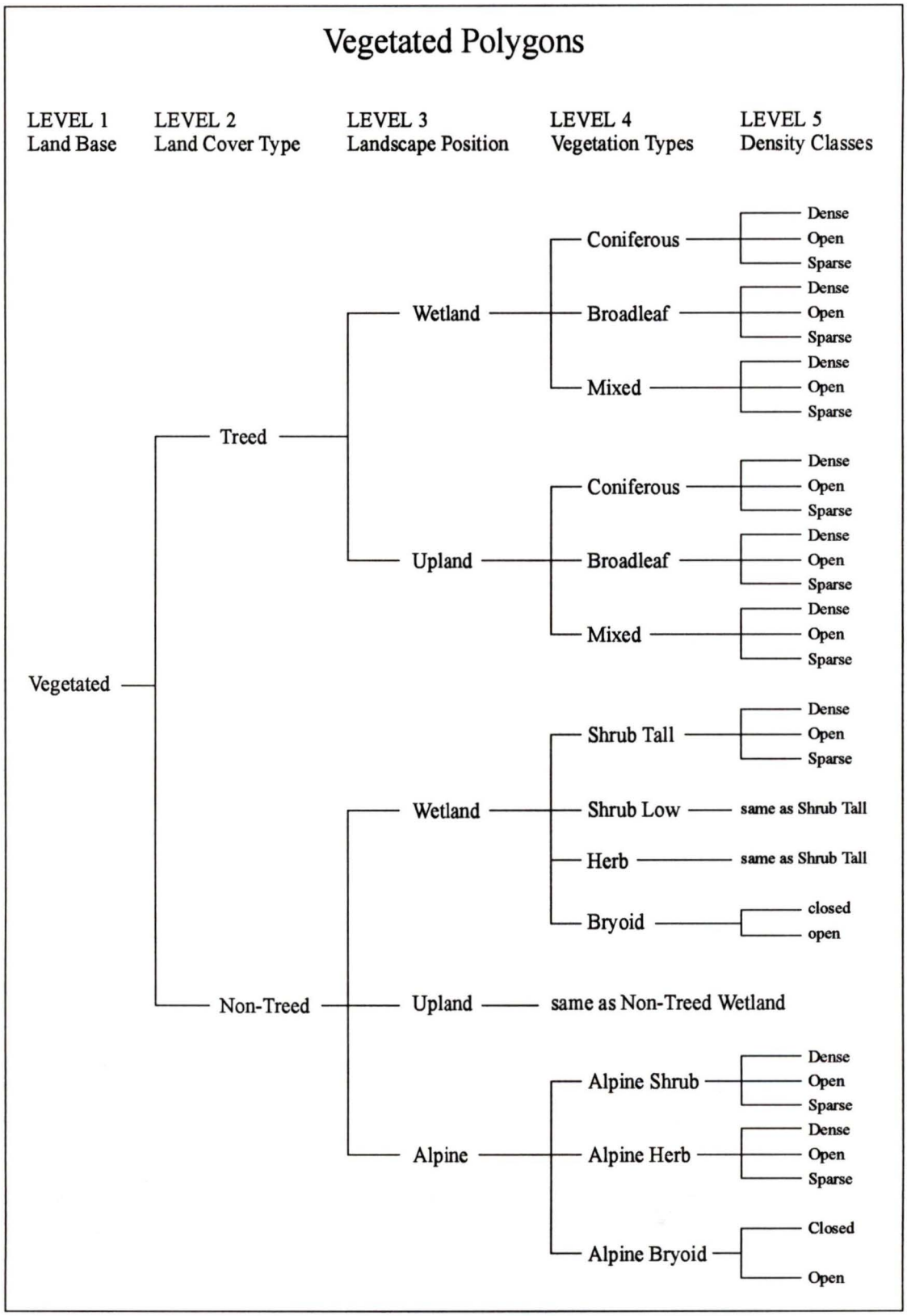


Figure 4.9 Simplified Land Classification Scheme (Vegetated Polygons).

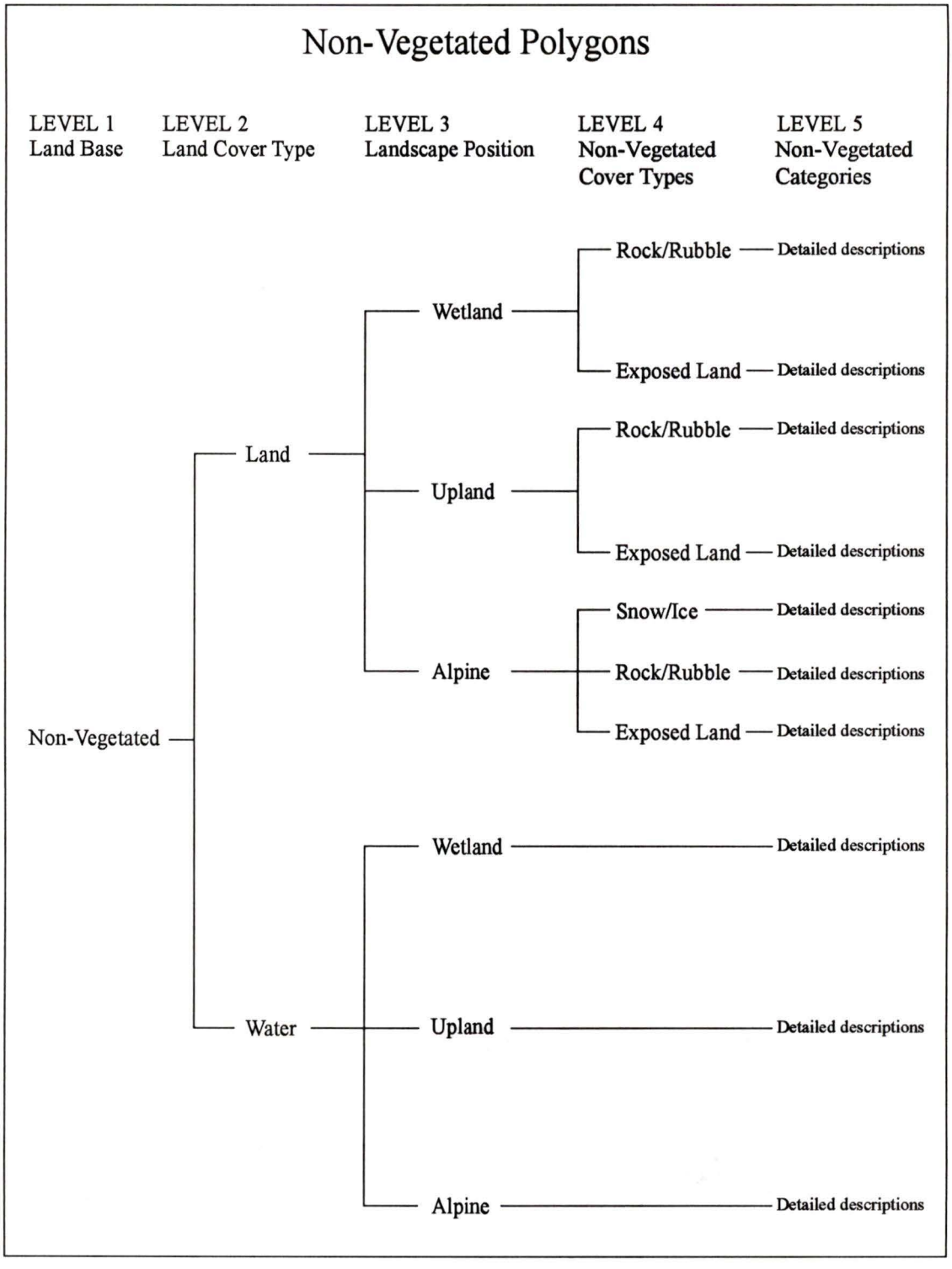


Figure 4.10 Simplified Land Classification Scheme (Non-Vegetated Polygons).

The BCLCCS is based on the existing vegetation cover as opposed to climax or potential vegetation and satisfies the twelve criteria outlined by Bones (1993) for standard international forest definitions and classifications. Cover can be vegetated or non-vegetated. Vegetated cover is either treed or non-treed; non-vegetated cover is either land or water. In most cases, homogeneous areas (polygons) are delineated on mid-scale aerial photographs (1:10,000 to 1:20,000). Polygon delineation is driven by the following primary attributes: the BCLCCS criteria; vegetation attributes; mensurational attributes; and ecological attributes (Resources Inventory Committee, 1996a). The minimum polygon sizes suggested in the VRI guidelines are two hectares for areas with distinct boundaries and five hectares for areas with indistinct boundaries. Allowances are made for flexibility in these minimum sizes depending on the local context (Resources Inventory Committee, 1996a).

The purpose of the BCLCCS as outlined in the VRI Photo Interpretation Procedures Manual (Resources Inventory Committee, 1996a) is as follows:

> "The purpose of the B.C. Land Classification Scheme is twofold. First, the land classification can be derived for each polygon (or portion thereof), based on the photo interpreter's attribute estimates. The land classification of each polygon is summarized as a seven letter code to facilitate broad land classification reporting, and also to provide a link for comparison of land classification accuracy with Phase II field sampling data. Second, the B.C. Land Classification Scheme provides the criteria for distinguishing cover types within the polygon. These criteria are critical for the assessment of specific tree, shrub, herbaceous, bryoid and non-vegetated communities within polygon boundaries (referred to as land cover components)." (Resources Inventory Committee, 1996a, p. 7)

Under current procedures, the photo interpreter does not derive the seven letter land classification code, rather the photo interpreter defines a series of attributes and the BCLCCS is derived from the attributes identified by the photo interpreter on the mid-

scale photography. Other derived attributes include the dominant polygon soil moisture regime, diameter at breast height, volume and site index.

4.3.1 Training Data Generation

The existing forest inventory information was used as a proxy for the BCLCCS and classes were constructed using BCLCCS membership rules. The Ministry of Forests has adopted a similar procedure to roll-over the existing inventory to the new VRI specifications (Salkeld, 1998). The generation of training data for each level of the BCLCCS was restricted to the area encompassed by the previous level of the classification hierarchy, thereby ensuring that all polygon attributes are mutually exclusive for each classification level.

4.3.2 Level 1: Land Base

The first level of the BCLCCS determines the presence or absence of vegetation within the polygon. Cover is divided into two main classes: vegetated (V) and non-vegetated (N). A polygon in which five percent of the total surface area is covered by vegetation (trees, shrubs, herbs and cryptograms exclusive of crustose lichens) is considered to be vegetated. Conversely, those polygons with less than five percent vegetation cover are classified as non-vegetated.

For the training data, the percentage crown closure from the forest inventory was used to determine membership in this class. In the existing forest inventory, crown closure is assessed from aerial photographs and is based on the amount of ground area covered by tree crowns (Resources Inventory Branch, 1996). Under the new VRI

guidelines, photointerpreters will be required to assess crown closure for all vegetation types, not solely trees, and therefore in future operational VRIs, the distinction between V and N will be derived from the sum of *all* of the vegetation crown closures in the polygon (Resources Inventory Committee, 1996a).

4.3.3 Level 2: Land Cover Type

The second level of the BCLCCS classifies the polygon according to land cover type: treed or non-treed; land or water. Polygons which are classified as vegetated at level one are further divided into treed (T) or non-treed (N) classes at level two of the BCLCCS. Treed polygons must have at least ten percent crown closure of any sized tree species. Proxy data for this level was generated by selecting those vegetated polygons from the forest inventory which were treed (where tree species attributes were populated) and which had a crown closure that was greater than ten percent. Conversely, vegetated polygons not matching these criteria were classed as non-treed. VRI guidelines indicate that the estimate of tree crown closure will be used operationally to distinguish treed and non-treed polygons (Resources Inventory Committee, 1997).

The non-vegetated polygons identified at level one are further divided into the land cover types of land (L) or water (W). The latter includes lakes, reservoirs and rivers/streams. The non-productive forest code from the forest inventory was used to identify lakes and rivers as water. The remaining areas were classified as land.

4.3.4 Level 3: Landscape Position

The third level of the BCLCCS describes the location of the polygon relative to elevation and drainage and is described as either alpine, wetland or upland. The alpine class (A) applies to non-treed areas above the tree line. It is defined as treeless (less than one percent tree cover), dominated by shrubs, herbs, bryophytes, and lichens and commonly rock, ice and snow. The alpine class excludes those areas which are identified as alpine tundra parkland sub-zones in the biogeoclimatic classification system (Resources Inventory Committee, 1997).

For the training data, alpine areas were identified by using the non-productive forest descriptor from the forest inventory. As indicated in Table 4.3, alpine areas accounted for 65% of the non-productive forest portion of the study area. Alpine forest was not included in this class because all alpine forest areas had crown closures of greater than one percent. The BEC data was used to exclude all areas of alpine tundra parkland sub-zones which intersected with the areas identified as alpine from the forest inventory.

The wetland class (W) is defined as land having the water table near, at or above the soil surface or which remains saturated for a period of time long enough to promote wetland or aquatic processes. The latter processes are indicated by poorly drained soils, specialized vegetation, and various kinds of biological activity which are specially adapted to the wet environment. Wetland information was taken from the forest inventory non-forest descriptor and from the TRIM data. The broad class which therefore encompasses all non-wetland ecosystems and non-alpine areas was upland (U).

The VRI guidelines indicate that alpine polygons will be considered those which fall in the alpine landscape regions, as defined above. Wetland and upland areas will be identified on the basis of the soil moisture regime in the polygon, as estimated by the air photo interpreter.

4.3.5 Level 4: Vegetation Type

The fourth level of the BCLCCS classifies the vegetation types and non-vegetated cover types. Vegetated treed polygons identified in levels 1 to 3 are further divided into three distinct groups: coniferous, broadleaf and mixed. The treed polygon is classified as coniferous (TC) when the total crown cover of coniferous trees (of the order *Coniferae*) is 75% or more of the total tree crown cover and trees cover a minimum of 10%, by crown cover, of the total area. The polygon is classified as broadleaf (TB) when the total crown cover of broadleaf trees (of the order *Angiospermae* in the subclass *Dicotyledoneae*) is 75% or more of the total tree crown cover and trees cover a minimum of 10% by crown cover, of the total area. The mixed class (TM) occurs when neither coniferous or broadleaf covers more than 75% of the total tree crown cover and trees cover a minimum of 10%, by crown cover, of the total area. The average leading species percentage per polygon was 68% and therefore, there were a number of polygons which were classified as mixed.

Training data for these areas were identified by selecting polygons which had layer and ordering attributes equal to one and greater than 75% crown closure followed by a leading species (all possible coniferous or broadleaf species being considered). For

the mixed class, similar rules applied, except the percentage of coniferous or broadleaf species must be less than or equal to 75%.

Non-treed polygons are grouped into shrubs, herbs and bryoids at this level, however there was no information in the forest inventory to simulate this sufficiently so no non-treed classes were considered at levels four or five of the BCLCCS hierarchy. VRI guidelines indicate that the photo interpreter must describe the type of vegetated cover observed and therefore will provide information for level four directly from Phase I.

4.3.6 Level 5: Density Classes

The vegetated polygons delineated and described in levels 1 to 4 of the BCLCCS are further classified into density classes at level five. An interpretation of the coverages and density of vegetation with the polygon is made. Treed vegetation types have three density classes: dense (DE), open (OP) and sparse (SP). Dense tree cover is defined as crown closure between 61% and 100%. Open tree cover has between 26% and 60% crown closure and sparse tree cover has between 10% and 25% crown closure. Training data were created for this level using estimates of crown closure from the forest inventory and in future VRIs, vegetation crown closure will be used to determine density classes for trees as well as shrubs, herbs and bryoids.

4.4 Image Pre-Processing

Preprocessing operations included atmospheric and geometric corrections in addition to procedures performed by the image vendor prior to data delivery. The

conversion of DNs to apparent reflectance and the registration of the images to a common map projection were seen as necessary prerequisites to further image analysis.

4.4.1 Atmospheric Correction

There are various methods to radiometrically calibrate a digital image, depending on the information available about the sensor, atmosphere and ground truth for the data of image acquisition (Barker, 1983; Slater, 1987; Price, 1987; Isaacs and Vogelmann and Rock, 1988; Chavez, 1989; Teillet et al., 1990; Hill, 1991; Wrigley et al., 1992; Chavez et al., 1994; Pons and Sole-Sugranes, 1994; Moran et al., 1995; Williams et al., 1996; Chavez, 1997). Atmospheric correction methods which use atmospheric models independent of the remotely sensed image itself, function by the selection of a standard atmosphere and set of atmospheric conditions and information regarding ground visibility at the time of image acquisition. H5S is a computer code developed by Tanre et al. (1986), which estimates the apparent reflectance of a pixel by accounting for the effects of gaseous absorption, scattering by molecules and aerosols and inhomogeneity in the ground reflectance. H5S also provides estimates for gaseous transmittance and surface irradiance (Tanre et al., 1990).

H5S works in terms of reflectance and therefore image data numbers must first be converted to radiance values and then subsequently converted to apparent (Lambertian) reflectance before the atmospheric corrections may be applied. The correction equations utilized in the method adopted by this study allow the final atmospheric corrections to be applied to the image data numbers (DN) directly, in order to avoid information loss in the conversion (Piekutowski, 1996). Conversion of the

mean DN values for each image channel to absolute reflectance is still necessary however, because the image reflectance values are used in the process of determining the parameters of the final correction equation.

Conversion to radiance is achieved by one of two methods. The first method is more generic and nature and involves the application of the following equation to the image channels (Markham and Barker, 1986; Lillesand and Kiefer, 1987):

$$L = ((LMAX-LMIN) / 255) * DN + LMIN$$
$$L = LMIN + \beta Qcal \qquad \textit{[Equation 4.1]}$$

where:

L	=	spectral radiance for the band ($mW\ cm^{-2}\ \mu m^{-1}\ sr^{-1}$)
$LMIN$	=	spectral radiance at Qcal = 0 ($mW\ cm^{-2}\ \mu m^{-1}\ sr^{-1}$)
$LMAX$	=	spectral radiance at Qcal = Qcalmax ($mW\ cm^{-2}\ \mu m^{-1}\ sr^{-1}$)
$Qcal$	=	digital signal level
$Qcalmax$	=	maximum digital signal level (for 8 bit data = 255)
β	=	(LMAX-LMIN) / Qcalmax

Alternatively, the second method is sensor-specific and uses standardized gains and offsets. Using this method, DNs can be converted to absolute radiance values using the following equation:

$$L^* = (DSL - DSLo) / G \qquad \textit{[Equation 4.2]}$$

where:

L^*	=	apparent radiance at the sensor ($Wm^{-2}\ sr^{-1}\ \mu m^{-1}$)
DSL	=	image digital signal level (counts)
$DSLo$	=	calibration offset coefficient (counts)
G	=	calibration gain coefficient (counts/radiance)

The offsets used for the conversion can either be obtained from the imagery documentation or published standard offsets can be used (Thome et al., 1994; Teillet

and Fedosejevs, 1995; Teillet, 1996). Use of the standard gain coefficients to convert digital counts to radiance and subsequently to reflectance can be problematic. Investigators are finding that the top-of-the-atmosphere reflectances are becoming too low over time to allow for a proper atmospheric correction used (Thome et al., 1985; Teillet and Fedosejevs, 1995; Teillet, 1996). Table 4.5 (Teillet, 1996) provides the equations which are used to calculate the gain coefficients.

Table 4.5. Equations Used For The Calculation Of The Gain Coefficients.
(Source: Teillet, 1996).

TM Band	Calibration Gain Coefficient (counts/($Wm^{-2}sr^{-1}\mu m^{-1}$))	Characteristic Wavelength (μm)	Solar Irradiance ($Wm^{-2}\mu m^{-1}$)
1	G=(-3.58E-05)*D+1.376	0.4863	1959.2
2	G=(-2.10E-05)*D+0.737	0.5706	1827.4
3	G=(-1.04E-05)*D+0.932	0.6607	1550
4	G=(-3.20E-06)*D+1.075	0.8382	1040.8
5	G=(-2.64E-05)*D+7.329	1.677	220.75
7	G=(-3.81E-04)*D+16.02	2.223	74.96
where D = days since launch (March 1984)			

Table 4.6. Offset and Gain Coefficients for Study Imagery.
(Calculated Using Equations in Table 4.5)

Image Date	TM Band	Offset	Gain
31-Jul-94	1	2.523	1.2398526
31-Jul-94	2	2.417	0.657137
31-Jul-94	3	1.452	0.892448
31-Jul-94	4	1.854	0.953304
31-Jul-94	5	3.423	7.2286008
31-Jul-94	7	2.633	15.8751057
3-Oct-94	1	2.523	1.2375614
3-Oct-94	2	2.417	0.655783
3-Oct-94	3	1.452	0.891783
3-Oct-94	4	1.854	0.951256
3-Oct-94	5	3.423	7.2269112
3-Oct-94	7	2.633	15.8726673
29-Aug-93	1	2.523	1.2628362
29-Aug-93	2	2.417	0.670619
29-Aug-93	3	1.452	0.899125
29-Aug-93	4	1.854	0.973848
29-Aug-93	5	3.423	7.2455496
29-Aug-93	7	2.633	15.8995659
30-Sep-93	1	2.523	1.2616906
30-Sep-93	2	2.417	0.669947
30-Sep-93	3	1.452	0.898792
30-Sep-93	4	1.854	0.972824
30-Sep-93	5	3.423	7.2447048
30-Sep-93	7	2.633	15.8983467

Once the DNs are converted to radiance values they can then be converted to reflectance using the following equation (Markham and Barker, 1985):

$$\rho^*_{0\text{-}1} = (\pi L d^2 / (Esun_\lambda \cos \phi) \qquad \textit{[Equation 4.3]}$$

where:

$p^*_{0\text{-}1}$ = apparent reflectance scaled from 0 to 1
d = earth-sun distance in astronomical units
$Esun_\lambda$ = mean solar exoatmospheric irradiance (mW cm^{-2} μm^{-1})
ϕ = solar zenith angle

A series of steps are used for generating the atmospheric correction and applying it to the image using the H5S atmospheric code. The initial step is to collect

the appropriate atmospheric information required by H5S and determine a standard atmospheric model and an appropriate aerosol condition. Information on horizontal visibility at the time of data acquisition must also be obtained. The National Climate Data Center in the United States maintains data for approximately 8000 active global stations for twelve parameters, including ground visibility. Three weather stations were selected to provide the necessary atmospheric information and these are included in Table 4.7.

Table 4.7 Selected Weather Stations.

Station ID	Code	Description	Latitude/Longitude
718920	CYVR	VANCOUVER INTL ARPT	49°11'N 123°10'W
711080	CYXX	ABBOTSFORD AIRPORT	49°02'N 122°22'W
711130	CWZA	AGASSIZ (AUTO)	49°15'N 121°46'W

Once all of the appropriate information required by H5S has been collected, the second step is to examine the histograms of the raw image bands and convert the mean DN values to apparent reflectance using equations 4.1 or 4.22 to calibrate DNs to radiance values and then equation 4.3 to convert to apparent reflectance. These apparent reflectance values are subsequently used as a constant value for background reflectance over the entire image (apparent reflectances for the homogenous surface in the H5S model).

The third step in determining the atmospheric correction equation involves using the input data from the first step and the mean apparent reflectances from the second step and running H5S in "reverse computation mode" for each spectral channel of interest. The result of this step is an estimate of the background reflectances for the environment. The fourth step in the process is to run H5S in the "adjacency option

mode" where the values generated from step 3 are used as background reflectance and a target reflectance is selected based on the dominating cover type in the image (forest). The size of the target area also must be specified and determining an appropriate target size may require experimentation depending on the objects comprising the image (e.g. large tracts of homogenous forest land versus small heterogeneous fields.) Past research has indicated a target size of 0.5 km is appropriate for use with forest scenes similar to the project imagery used in this study, and on trial runs this value provided good results (Teillet, 1996). The purpose of this step is to experiment with target sizes to select an optimal size.

The fifth step involves re-running H5S in the "adjacency option mode" once an optimal target size has been selected with target reflectances of 0 % and 100 % for each channel. The environmental reflectances are set to the background reflectances derived from step 2. The output from this fifth step will be the apparent reflectance values (ρ_0 and ρ_1) values which are then used in calculating the coefficients used in the final atmospheric correction equation.

Using the ρ_0 and ρ_1 values from step 5, the atmospheric correction equation (Piekutowski, 1996) takes the following form:

$$\rho_{0\text{-}1} = \gamma \rho^*_{0\text{-}1} - \delta$$
$$= \gamma \alpha L_\lambda - \delta \qquad \textit{[Equation 4.4]}$$

where:

$\rho_{0\text{-}1}$	=	true surface reflectance scaled from 0 to 1
γ	=	$1 / (\rho_1^* - \rho_0^*)$
δ	=	$\rho_0^* / (\rho_1^* - \rho_0^*)$
ρ_0^*	=	apparent reflectance of a surface with a true reflectance = 0
ρ_1^*	=	apparent reflectance of a surface with a true reflectance = 1

In order to utilize equation 4.4, the image values must first be converted to reflectance. Equation 4.5 (Piekutowski, 1996) allows the image DNs to be directly converted to corrected reflectance:

$$
\begin{aligned}
\rho_{0\text{-}255} &= 255\,\rho_{0\text{-}1} \\
&= 255\,[\gamma\,\alpha\,L_{\lambda} - \delta] \\
&= 255\,[\gamma\,\alpha\,L_{min\lambda} + \gamma\,\alpha\,\beta\,Q_{cal}\,\delta] \\
&= 255\,\gamma\,\alpha\,\beta\,Q_{cal} + 255\,(\gamma\,\alpha\,L_{min\lambda} - \delta) \\
&= D\,Q_{cal} + C
\end{aligned}
$$

[Equation 4.5]

where:

$\rho_{0\text{-}255}$ = true surface reflectance scaled from 0 to 255
D = $255\,\gamma\,\alpha\,\beta$
C = $(\gamma\,\alpha\,L_{min\lambda} - \delta)$

In order to avoid information loss through data compression, the conversion from digital signal levels to atmospherically corrected surface reflectances was completed in one step using Equation 4.5. Information loss will be greater if DNs are first converted to apparent reflectances (Piekutowski, 1996). It is possible to scale reflectance values over the 8-bit range if the image contains a limited range of reflectance values, which was not the case for this study.

4.4.2 Geometric Correction

Landsat TM data is considered to have "a great deal of inherent geometric fidelity and minimal residual distortion" (Schowengerdt, 1983), upon completion of preprocessing. The purpose of the geometric correction was to ensure that all of the data sources for the project had a common projection (UTM NAD83), thereby facilitating the incorporation of georeferenced ancillary information into the

classification process. A nearest neighbour third order polynomial was used for the correction procedure. Nearest neighbour resampling was selected in order to avoid alteration of the original grey values of the image (Lillesand and Kiefer, 1987). A third order polynomial was selected as being most appropriate for these images because it was the transform which provided the lowest residual mean square error (RMS). An image to vector correction was performed using PCI™ GCPWORKS™ software. The Terrain Resource Information Management TRIM planimetric coverages were imported into PCI™ from ARC/INFO™ and used as the vector source. Approximately twenty ground control points (GCP) were derived from road and river intersections for each of the images processed. The RMS errors for each of the images is summarized in Table 4.8.

Table 4.8 Image to Vector Correction RMS Errors.

Image	RMS Error (pixel)
290893sub.pix	0.76
300993sub.pix	0.81
310794sub.pix	0.80
031094sub.pix	0.97

4.4.3 Topographic Influences

The topography of the study area is highly variable and extensive consideration was given to the potential influence this varying topography would have on the final classification results. It is important here to distinguish between the two different problems which topography can introduce into a remote sensing image classification project: the topographic effect and topographic displacement. The topographic effect refers to the differential illumination of ground surfaces due to slope angle and aspect

variations, resulting in surface cover types having a wide range in radiance values (Holben and Justice, 1980). To further complicate this effect, it has been shown in past research that different types of vegetation respond differently to direction and illumination effects (Holben and Justice, 1980; Leprieur et al., 1988; Thomson and Jones, 1990) and that slope-aspect corrections must be class specific, with no single formula being sufficiently general to accommodate various forest types (Teillet et al., 1982).

Topographic normalization refers to the methods used to compensate for the topographic effect. There are several different normalization procedures which have been developed (Leprieur et al., 1988; Civco, 1989; Colby, 1991; Justice et al., 1991; Conese et al., 1993; Ekstrand, 1996), the most common being the Lambertian correction function which assumes a Lambertian surface (a perfectly diffuse reflector, appearing equally bright from all viewing directions). For this function, radiance is assumed to be proportional to the cosine of the incident angle of the sun (Holben and Justice, 1980). Previous research has shown that this assumption is clearly not valid for the non-Lambertian behaviour exhibited by most vegetation (Ekstrand, 1996) and hence, more complex methods originally developed by Minnaert (1941) were adapted for remote sensing applications by Smith et al. (1980). Given the complexity of adequately compensating for topographic influences on reflectance, a decision was made not to perform topographic normalization procedures on the project imagery. This afforded the opportunity to examine how the neural network and MLC methods would deal with the problem of topographic effect and how the classification results would change as

different image dates with different sun angles and relatively more or less shadow were used.

Topographic displacement is the other problem introduced by topography. This displacement occurs when images of mountains and other topographic features are displaced from their desired orthographic position in the image (Avery and Berlin, 1992). In an optical image the displacement is away from the sensor, and is relatively small when the sensor is nearly overhead (as is the case in the Landsat 5 satellite). The displacement increases with increasing incidence angle. Orthorectification is the process used to compensate for this topographic displacement.

A comprehensive effort was made to use the orthorectification software developed by PCI™ to correct the project imagery for topographic displacement with unsatisfactory results. The spatial configuration of the orthorectified image had a poor spatial alignment with the forest inventory and TRIM data, relative to the images which had only been geometrically corrected and a decision was therefore made not to use the orthorectified satellite images for this project. The poor results of the orthorectification are largely caused by the requirement of the PCI™ SATORTHO software that the entire image (full scene) be corrected and in this instance, accurate ground control points and a digital elevation model were available only for the study area sub-scene. A superior orthorectification result may have been achieved with a more accurate set of ground control points *for the entire image*, as other studies using the same software have used ground control points collected with the aid of a Global Positioning System (Baker et al., 1995; Cheng and Toutin, 1995).

4.5 Image Enhancements

4.5.1 Principal Component Analysis

Principal component analysis (PCA) is a mathematical transformation which rotates the axes of multi-dimensional image space in the direction of maximum data variance (Lillesand and Kiefer, 1987). The new synthetic bands generated by this transformation are referred to as components and are linear combinations of the original set of bands which are orthogonal to one another, thereby eliminating any mathematical correlation between them (Chavez and Kwarteng, 1989). The axes are arranged so that the first component contains the greatest amount of variance with each subsequent component containing less variance than the one preceding it. The sum of the variances for all of the principal components is equal to the total variance in the original input bands (Schowengerdt, 1983). PCA is used most often to diminish the dimensionality of the data by reducing the redundancy in spectral information between highly correlated spectral bands. The common method associated with TM data is to use all six reflective bands as input to the PCA and then use the first three principal components for analysis. This standard method of PCA was used for this project.

Sergi et al. (1995) used principal components with neural networks and TM imagery, comparing the classification results obtained using the principal components to those obtained with the original spectral bands. Differences in classification results were not significant, although the classification results generated with the principal components were considered superior. The noise and redundancy reduction inherent in the PCA was seen as providing better training data to the neural network, which was

thus able to generalize more effectively, resulting in more homogenous classified regions.

4.5.2 Band Ratios

Ratio images are enhancements resulting from the division of DN values in one spectral band by the corresponding values in another band. The ratio compensates for variations in brightness or albedo caused by topography (Avery and Berlin, 1992). Noise removal is an important prelude to the preparation of ratio images since ratioing enhances noise patterns that are uncorrelated in the component images. Ratios only compensate for multiplicative illumination effects and not those that are additive, such as haze (Lillesand and Kiefer, 1987).

There are thirty possible ratio combinations for the six reflective TM bands. Three band ratios determined to be useful for intensifying forest canopy characteristics (Bauer et al., 1994) were selected for this study: TM4/TM3, which provides information with respect to canopy condition (Fiorella and Ripple, 1993b); TM4/TM2 which provides information on wetlands (Jensen, 1983) and TM5/TM4 is useful for identifying conifer canopy structure (Peterson et al., 1986).

4.5.3 Normalized Difference Vegetation Index

The Normalized Difference Vegetation Index (NDVI) is a simple ratio based index which ranges from -1 to 1. The NDVI was first developed by Kriegler et al. (1969) and later refined by Rouse et al. (1973). It is calculated by subtracting the red band (R) from the near-infrared band (NIR) and dividing this by the sum of the near-

infrared and red bands: *(NIR - R) / (NIR + R)*. The NDVI is moderately sensitive to soil background and the atmosphere, except in situations of low plant cover. The summed denominator largely compensates for changing illumination conditions, surface slope and viewing aspects. Although numerous other vegetation indices exist, the simplicity, dynamic range and sensitivity of the NDVI to changes in vegetation cover make it the preferred vegetation index for broad band sensors such as Thematic Mapper (Ray, 1995).

4.5.4 Tasseled Cap Transformation

The Tasseled Cap Transformation was originally developed by Kauth and Thomas (1976) for crop monitoring applications using Landsat Multi-spectral Scanner (MSS) data. The Tasseled Cap is a linear transformation which involves rotation of the image data in such a manner as to capture most of the variation in a two dimensional view. Kauth and Thomas (1976) directly correlated physical characteristics of agricultural fields to the features discriminated in this two dimensional view. The two components, which accounted for 95% of the variation in an MSS scene, were identified as brightness, a weighted sum of all bands defined in the direction of maximum variation in soil reflectance, and greenness, a contrast between near-infrared and visible bands, oriented approximately orthogonal to the brightness component (Lillesand and Kiefer, 1987).

Crist and Cicone (1984) extended the concept of the tasseled cap to simulated Landsat Thematic Mapper data, finding that the six reflective TM bands translated into a three dimensional view: brightness, greenness and wetness. These three components

related to planes of soil, vegetation and a transition zone between soil and vegetation (Lillesand and Kiefer, 1987). The use of Tasseled Cap has been extended to forestry applications (Bauer et al., 1994; Fiorella and Ripple, 1993a). The wetness component relates to canopy and soil wetness and has been found useful for wetland delineation and in the discrimination between upland and lowland forest types. Greenness and to a lesser extent brightness, is used as an indicator of vegetation cover (Bauer et al., 1994). The Tasseled Cap Transformation is performed in PCI™ using TASSEL. The empirically derived TASSEL coefficients used in the transformation are included in Table 4.9.

Table 4.9 Tassel Cap Coefficients.
(Source: Crist and Cicone, 1984)

	A1	A2	A3	A4	A5	A7
Brightness	0.3037	0.2793	0.4743	0.5585	0.5082	0.1863
Greenness	-0.2848	-0.2453	-0.5436	0.7243	0.0840	-0.1800
Wetness	0.1509	0.1973	0.3279	0.3406	-0.7112	-0.4572

The Tasseled Cap Transformation takes the following form (for each of the components) (Crist and Cicone, 1984):

$$TC = A1(TM1) + A2(TM2) + A3(TM3) + A4(TM4) + A5(TM5) + A7(TM7) \quad \textit{[Equation 4.6]}$$

4.6 Feature Selection

The discussion in Chapters 2 and 3 indicated that feature or band selection is an important process for both the MLC and the ANN classifiers. The selection of an appropriate set of bands which presents the greatest amount of information helps to reduce the number of training samples required for the MLC and reduces the training

time for the ANN. A variety of bands and synthetic bands were created and submitted to a feature selection process.

The bands used as input to the classifications were selected through a combination of a branch and bound algorithm included in the PCI™ software and the examination of spectral plots for each of the initial thirty-three target classes. Appendix C contains these spectral plots of the mean reflectance values with the standard deviation for each of the image bands, both synthetic and original, for each of the informational classes. The divergence analysis was used to select an optimum set of input bands which have a high degree of separability for the features being classified. While this type of band set selection is beneficial for enhancing the results of a statistical based classifier, such as the maximum likelihood, it was less clear how it would enhance the results of the neural network classification. A detailed exposition of feature selection is given by Narendra and Fukunaga (1977), Swain and Davis (1978), and Mausel et al., (1990). The branch and bound algorithm determines the optimum channel set for class discrimination given a set of image channels and a specific target class or set of target class signature segments. This optimum channel set is determined using the divergence or dissimilarity between signature segments. The dissimilarity is calculated from the class sample means and covariance matrices and a decision rule is used to extend this dissimilarity to all of the segment class pairs. The minimum pairwise divergence decision rule was utilized as it maximized the classification accuracy for those classes with the greatest degree of confusion between them (Narendra and Fukunaga, 1977).

The motivation for selecting an optimum band set for each level of the BCLCCS stems from the number of parameters necessary to evaluate the quadratic during classification (Swain and Davis, 1978; Paola, 1994). The number of training samples required to provide accurate estimates of the class parameters is estimated at between ten and one hundred times the number of bands used as input for the maximum likelihood classification (Swain and Davis, 1978). The number of training samples required can be more specifically estimated using the following equation (Paola, 1994):

$$N_{ML} = 2 + \text{classes} * \text{bands} + \tfrac{1}{2} * \text{classes} * (\text{bands}^2 + \text{bands}) \qquad \textit{[Equation 4.7]}$$

The number of parameters for the maximum likelihood method is quadratic with respect to the number of bands. If Equation 4.7 is evaluated for BCLCCS level one using the complete set of sixteen bands and band enhancements, 306 training samples would be required for each class. However, if a subset of six bands is selected, only 56 samples per class would be required and this value is commensurate with the number of training samples recommended by Congalton (1991) for statistically valid correspondence assessments.

Feature selection was used in this study to reduce the dimensionality of the data set and to ensure that the computational requirements of the maximum likelihood classifier were satisfied. To maintain valid comparisons, the same bands which were selected at each level of the classification were also used as input to the neural network. Additional neural network trials which experimented with the use of the six spectral bands as input were also attempted however because previous researchers have indicated that a mere reconfiguration of the spectral information may provide little

improvement in the neural network's discrimination of target classes (Sergi et al., 1995).

The selected channels are detailed in Table 4.10. Note that the raw bands were processed through the branch and bound algorithm separately from the band enhancements. For example, the channels selected for "aug1a" are the optimal set of three channels from the six reflective raw bands for the August image date and level one of the classification (vegetated versus non-vegetated). Conversely, the channels selected for "aug1b" are the optimal set of three channels from the ten band enhancements for the August image date and level one of the classification. Bands have similarly been selected for the July, September and October image dates.

Table 4.10 Selected Image Channels.

DATE	BAND_1	BAND_2	BAND_3
aug1a	TM3	TM5	TM7
aug1b	PC3	BR4/2	TCW
aug2a	TM1	TM3	TM5
aug2b	PC1	PC2	TCG
aug3a	TM3	TM5	TM7
aug3b	PC3	BR4/3	TCW
aug4a	TM2	TM4	TM5
aug4b	PC2	NDVI	TCG
aug5a	TM3	TM4	TM7
aug5b	BR4/3	NDVI	TCG
sept1a	TM3	TM5	TM7
sept1b	PC1	PC2	TCG
sept2a	TM1	TM3	TM5
sept2b	PC1	PC2	TCG
sept3a	TM3	TM4	TM5
sept3b	BR4/2	BR5/4	NDVI
sept4a	TM1	TM2	TM5
sept4b	PC2	TCB	TCG
sept5a	TM3	TM5	TM7
sept5b	PC2	BR4/3	NDVI
july1a	TM3	TM4	TM5
july1b	PC1	TCB	TCG
july2a	TM1	TM3	TM5
july2b	PC1	PC2	TCG
july3a	TM3	TM4	TM7
july3b	PC3	BR4/3	NDVI
july4a	TM2	TM4	TM5
july4b	BR4/2	NDVI	TCB
july5a	TM3	TM5	TM7
july5b	PC3	TCG	TCW
oct1a	TM3	TM5	TM7
oct1b	PC1	PC2	TCG
oct2a	TM1	TM3	TM5
oct2b	PC1	PC2	TCG
oct3a	TM3	TM4	TM5
oct3b	BR4/3	NDVI	TCB
oct4a	TM2	TM5	TM7
oct4b	PC2	NDVI	TCG
oct5a	TM3	TM5	TM7
oct5b	BR4/3	BR4/2	TCW

TM1–TM7 = raw bands; NDVI = normalized difference vegetation index; BR4/2, BR4/3, BR5/4 = band ratios; TCB, TCG, TCW = tasseled cap brightness, greenness and wetness; PC1, PC2, PC3 = principal components.

4.7 Image Masks

Image masks were used to facilitate the classification process in three distinct ways. First, masks of training areas were created from the forest inventory data in ARC/INFO™ and imported into PCI™. Since all the data sources were registered to a common base, the training masks were readily overlaid onto the TM image. The purpose of these masks was twofold: (1) to constrain the classification process at subsequent levels of the BCLCCS hierarchy; and (2) to provide training samples for the classification algorithms. For example, discrimination of treed and non-treed areas at level two was constrained to those areas identified as vegetated (level one) from the forest inventory. Training samples for treed and non-treed classes were selected from the treed and non-treed masks created from the forest inventory.

The second important use of image masks was for reducing classification confusion caused by water features. Masks of lakes and double line rivers were created and subsequently subtracted from the aforementioned masks created from the forest inventory. Finally, masks were used to exclude areas of agricultural and urban development, since the focus of the classification is on vegetation cover as represented in a specific segment of the BCLCCS hierarchy. There was a substantial area of agricultural development in the northwest corner of the image which was not masked out of the original maximum likelihood trials.

4.8 Classification Methods

4.8.1 Neural Network Architecture

The PCI™ neural network classification software was used to create and train the neural networks used in this study and ultimately to classify the image through the created network. The PCI™ software uses a standard backpropagation neural network with the generalized delta rule for supervised learning, adapted from Pao (1989). Input data are linearly scaled between 0 and 1 using channel minimums and maximums before input to the network, or conversely, the analyst has the option of scaling the data independently and storing to a 32 bit channel prior to inputting them into the network. To calculate the output from a hidden layer unit or output layer unit, the net input to that unit is first calculated using the following (Pao, 1989):

net_i = sum of (weight_i * output_i) ***[Equation 4.8]***

where ***weight_i*** is the weight for the link from a unit in the preceding layer and ***output_i*** is the output value of the same unit in the preceding layer. Once this net input is summed, is passed through the following sigmoid activation function to produce the unit's output (Pao, 1989):

output_j = 1.0 / (1.0 + exp (-(net_i))) ***[Equation 4.9]***

where ***exp*** is the exponential function of base e. The sigmoid activation function will have a range of 1, when ***net_i*** goes to infinity and 0, when ***net_i*** goes to negative infinity. This is what occurs on a *forward pass* through the network.

For the backpropagation algorithm, the second phase of training is a *backward pass* through the network which is used to reduce the error between the actual and the

desired output by calculating the errors and adjusting the weights to lower the errors. Two errors are reported by the PCI™ algorithm: individual error and normalized total error. The individual error is the sum of the errors of the output layer units (Pao, 1989):

individual error = sum of (abs (target – output))) ***[Equation 4.10]***

where ***target*** is the target value for the output layer unit and ***output*** is the activation value produced by the neural network for that output layer unit. The normalized total error is calculated as follows (Pao, 1989):

normalized total error = 0.5 * (sum of (individual error * * 2))/NSAMPLE

[Equation 4.11]

where the ***(sum of (individual error * * 2))*** is the sum of the squares for the individual error and ***NSAMPLE*** is the number of samples used for training the neural network.

There are several variables which must be set by the analyst in the initial creation and subsequent training of the neural network. In the creation of the network the following parameters must be defined: the input channels, the number of input and hidden units, input training bitmaps, the value to be assigned to each class. These variables establish the network architecture. The weights associated with the connections between units, which are adjusted through the backpropagation of error in the training of the network, are initialized to random values between –0.5 and 0.5.

For the training of the neural network, the following parameters must be specified: momentum and learning rate, followed by the maximum allowable total error, the maximum allowable individual error and the maximum allowable number of learning iterations. The training of the neural network ceases upon the satisfaction of one of three conditions: (1) the maximum total error is less than the maximum

allowable total error; (2) the maximum individual error is less than the maximum allowable individual error; or (3) the number of iterations completed exceeds the maximum allowable number of iterations. If training ceases due to the latter condition the network will not have converged and the network will either have to be processed for additional learning iterations or reconfigured with smaller momentum and learning rates.

4.8.1.1 ***Neural Network Heuristics***

One of the main disadvantages associated with neural networks is the number of network parameters which must be configured through trial and error. Some researchers have attempted to establish guidelines which would facilitate the configuration of the parameters for remote sensing applications. The number of hidden layer nodes is the most difficult parameter to set. The heuristics of Kannellopoulos and Wilkinson (1997) and Paola (1994) for hidden layer size, which are based on extensive practical experience with neural network applications in image classification, were used in establishing the network architectures for the classification trials conducted in this study. These heuristics are summarized in Table 4.11.

Table 4.11 Heuristics for the Number of Hidden Layer Nodes.

Neural Network Parameter	*Kanellopoulos and Wilkinson (1997)*	*Paola (1994)*
first hidden layer in a network with 2 hidden layers	number of hidden nodes should be equal or double the number of inputs and perhaps four times as many	number of hidden layer nodes should equal the number of output classes and be independent of the number and format of the input classes
second hidden layer in a network with 2 hidden layers	number of hidden nodes should be 2-3 times the number of output classes	
if only one hidden layer	number of hidden nodes should be equal to the higher of the two numbers derived above	

The number of hidden units included in a hidden layer influences the computational capacity and efficiency of a neural network. Increasing the number of hidden units may facilitate more complex classification problems, however it renders the network more difficult to train, requiring larger training datasets and an increased number of training iterations (Fausett, 1994). Kanellopoulos and Wilkinson (1997) indicate that the number of hidden units in a network with only one hidden layer should be equal to or double the number of input units. Conversely, other heuristics for neural network configuration suggest that the number of hidden layer nodes should equal the number of output classes and be independent of the number and format of the input classes (Paola, 1994). As with other components of the neural network, the optimum number of hidden layer nodes is often determined through trial and error.

The number of hidden layers in a neural network influences the network's computational capacity and efficiency in a manner similar to that of the number of

hidden nodes per hidden layers. Single hidden layers are considered suitable for most classification problems involving fewer than twenty classes, beyond which the flexibility of two or more hidden layers may be required (Kannellopoulos and Wilkinson, 1997). Additional hidden layers are also useful if the distribution of the spectral values for a given class are considered discontinuous (Paola, 1994).

4.8.2 *Maximum Likelihood Classification*

The full maximum likelihood classifier in PCI™ utilizes the Gaussian threshold stored in each class signature to determine whether or not a pixel falls within a certain class. The Gaussian threshold is the radius (in standard deviation units) of a hyper-ellipse surrounding the mean of the class in feature space. The bias is used to resolve overlap between classes and weights one class in favour of the other. The null class is used for those pixels which do not fall within the hyper-ellipse of any class. One of the major assumptions is that the training data are well selected and the target classes have a Gaussian distribution.

The equation implemented in the PCI™ software is the Mahalanobis minimum distance classifier defined by the following equation (PCI, 1995):

$$Gi(X) = -\tfrac{1}{2}(X - Ui)^t \, Ci^{-1} \, (X - Ui) - (d/2)\log(2TT) - (\tfrac{1}{2})\log(|Ci|) + \log(Pi)$$

[Equation 4.12]

where:

Gi (X)	is the result for class i on pixel X
d	is the number of channels in the classification
X = (x1,...,xd)	is the (d by 1) pixel vector of grey levels
Ui	is the (d by 1) mean vector for class i
Ci	is the (d by d) covariance matrix for class i
TT	is pi = 3.1415...
|Ci|	is the determinant of the covariance matrix
Pi	Bi/SUM(Bi) is the *a priori* probability for class i
Bi	is the bias for class i
SUM(Bi)	is the sum of biases for all classes used
t	as a superscript denotes transpose
-1	as a superscript denotes inverse
Ti	is the threshold value for class i

The values for ***d***, ***Ui***, ***Bi***, ***Ti***, Ci^{-1} and ***|Ci|*** are obtained from the class signature segment. The shape and orientation of the hyper-ellipsoid is defined by the matrix Ci^{-1}, while the position is determined by ***Ui*** and the size by ***Ti***.

4.9 Summary

This chapter has described both the project area and project imagery. The informational classes for the image classification are based on a selection of classes from the British Columbia Land Cover Classification Scheme (BCLCCS) and descriptions of these classes and their defining criteria were provided. The image pre-processing procedures included both atmospheric and geometric corrections and the

methods used for both are fully explained as are the band enhancements which were created. A feature selection process based on divergence analysis was implemented to select the optimal band set for each of the five levels of the BCLCCS under consideration. Image masks were used to eliminate irrelevant cover types and reduce classification confusion. Finally, this chapter provide descriptions of the specific algorithms used for the image analysis, as implemented in the selected software package.

CHAPTER FIVE

RESULTS AND DISCUSSION

5.1 Correspondence Assessment Rationale

An accuracy assessment is customarily conducted by selecting a sample of reference locations, and comparing the attributes at these reference locations to the classification output generated from the remotely sensed image. The reference sample should be independent of the data initially used to train the classification algorithm (Stehman, 1997). Errors in the allocation of pixels to classes are tabulated and investigated empirically in the form of an error matrix or contingency table (Congalton et al., 1983; Rosenfield, 1986; Lark, 1995). The contingency table is a square array of numbers set out in rows and columns which express the number of pixels assigned to a particular class (rows), relative to the actual class as verified against some form of reference data (column), and is the common method for portraying classification accuracy (Card, 1982; Colwell, 1983; Congalton et al., 1983). One of the primary drawbacks of the error matrix is its inability to provide information on the spatial structure of the uncertainty in a classified scene (Congalton and Green, 1993; Canters, 1997).

Accuracy assessments often seek to verify the performance of the classifier as opposed to determining the accuracy of the final thematic map product – two different objectives which are frequently confused (Richards, 1996). In order to determine the accuracy of the final thematic map, the probabilities of the classes under consideration must be known *a priori*. Certain summary measures of overall accuracy, such as the kappa coefficient (Stehman, 1997), involve similar assumptions about fixed class

probabilities and the issue of whether the row and column marginals of an error matrix are fixed or free to vary is the basis of much debate in the remote sensing literature.

By convention, the term accuracy assessment implies that the classification results are compared to ground truth data which has been specifically collected for the purposes of the assessment (Hord and Brooner, 1976; Rosenfield, 1982; Rosenfield et al., 1982; Congalton and Biging, 1992). In this study, the collection of ground truth data through field investigation was not practical and the results of the classification trials have therefore been compared to the attributes associated with the forest inventory ground and air calls (see section 5.2). It is important to acknowledge that there are uncertainties associated with using the forest inventory information as a surrogate for ground truth data. Firstly, the ground and air calls are point measures to which generalized polygonal attributes from the forest inventory are being associated. This should not be problematic, considering the ground and air calls were originally used as point estimates for the verification and adjustment of the inventory attributes initially assigned to the polygons by photointerpreters.

Secondly, the proxy vegetation inventory data which was generated from the forest inventory and which was used for both the training and the correspondence assessment is merely a simulation, and although the BCLCCS guidelines were employed in the construction of the vegetation classes from the forest inventory, the forest inventory attributes themselves are interpretations and generalizations of reality which also have a certain degree of error associated with them. Specifically, the Ministry of Forests requires inventories to conform to an 85% attribute accuracy standard. Adherence to this standard has been highly variable. The results of inventory audits conducted by the

British Columbia Ministry of Forests indicate that the correct identification of leading species in mature stands has ranged from a low of 32% to a high of 84% and a median of 51% (Gilbert, 1998).

In addition to sources of error associated with the forest inventory attributes, there are three sources of spatial or positional error which must be considered as well. The Thematic Mapper images were registered to the TRIM base, which has an accuracy of plus or minus ten metres (Geographic Data B.C., 1996), while the forest inventory has been collected with a variety of methods from both aerial photography and satellite imagery which may have varying levels of accuracy (Resources Inventory Branch, 1996). Secondly, there is a positional error introduced by the registration of the images which was, on average plus or minus 0.835 of a pixel for all of the image dates, however for Thematic Mapper data that translates into a positional error of approximately plus or minus twenty-five metres (assuming a 30 metre pixel). Finally, there are positional errors resulting from the topography of the study area which has an elevation range of 2400 metres. The nature of these topographic errors were discussed in Section 4.4.3.

Having acknowledged many of the potential sources of error, the correspondence assessment between the classification results and the proxy vegetation inventory data provides an adequate level of information for the objectives of this research. No claims however, may be made concerning the degree to which these classification methods accurately represent vegetation classes as they exist on the ground.

5.2 Correspondence Assessment Reference Data: Forest Inventory Ground and Air Calls

A ground call is defined by the British Columbia Ministry of Forests (Resources Inventory Branch, 1996) as a forest type description supported by a minimum of three representative sample trees measured from the leading or dominant species in the polygon. Similarly, an air call is a forest type description made from a low flying aircraft or helicopter. The purpose of these calls is to provide an indication of the reliability of the polygon attributes as initially defined by the photointerpreters (Resources Inventory Branch, 1996). Both ground and air calls are selected using a stratified random sampling methodology. The precise location and attributes of these ground and air calls are recorded in the forest inventory database and they thus provide a useful source of reference information. There are a total of 850 ground and air calls available in the study area. Their distribution by class at each level is represented in Table 5.1 and Figure 5.1. These points are neither distributed evenly throughout the study area nor are they distributed evenly amongst the target classes and these unequal distributions will influence the reliability of the correspondence assessment results.

In recognition of the importance of the reference sample being independent of the data used for training the classifier, the reference points were subtracted from the training bitmaps. The proxy vegetation inventory attributes for the reference points were arrived at by a point-in-polygon overlay analysis which intersected the reference points with the generalized forest inventory attributes. This procedure allowed each test point to be assigned attributes for each of the five levels of the classification hierarchy which were being considered. The attributes for the test points are mutually exclusive at each level of

the classification system. Each classification was constrained to the area defined by a mask from the preceding level in the BCLCCS hierarchy.

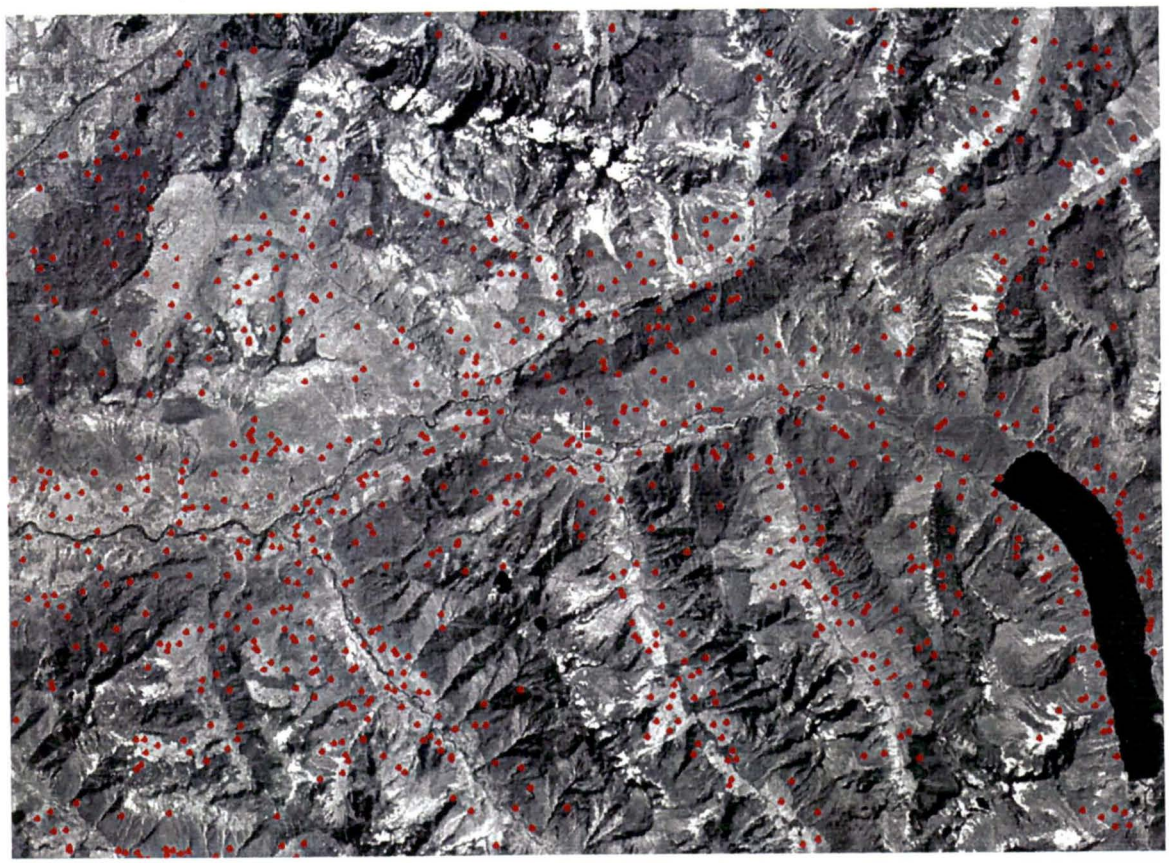


Figure 5.1 The Spatial Distribution of the Ground and Air Calls.

Table 5.1 Distribution of Test Points (Ground and Air Calls) by BCLCCS Level.

LEVEL	CLASS*	GROUND/AIR CALLS
1 – Land Base	V	758
	NV	92
2 – Land Cover Type	VN	39
	VT	719
3 – Landscape Position	NVLW	12
	NVLU	80
	NVLA	39
	VTU	706
	VTW	4
4 – Vegetation Type	VTUBR	6
	VTUCO	186
	VTUM	507
	VTWBR	1
	VTWM	2
5 – Density Classes	VTUBRDE	6
	VTUCODE	70
	VTUCOOP	103
	VTUCOSP	13
	VTUMDE	225
	VTUMOP	204
	VTUMSP	78
	VTWBRDE	1
	VTWCODE	1
	VTWMDE	1
	VTWMOP	1
* These classes are defined in Section 4.3		

Based on the number of available reference points presented in this table, only a subset of the classifications were considered suitable for correspondence assessment because they contained a sufficient number of test points. This evaluation was based on recommendations of Congalton (1991), who suggested the use of approximately fifty samples per class for error matrix development and this number was further verified by Hay (1979). The manner in which the samples are collected is also significant, as many of the statistical measures of correspondence assume that a simple random or stratified random sampling method has been used. The methods used to collect the ground and air calls meet this requirement (Resources Inventory Branch, 1996).

The following levels of the BCLCCS were selected as having sufficient reference samples to warrant a correspondence assessment to the forest inventory reference points: level 1, level 2, level 3 vegetated treed (3VT), level 3 non-vegetated (3NV), level 4 vegetated treed upland (4VTU), level 5 vegetated treed upland coniferous (5VTUCO), level 5 vegetated treed upland mixed (5VTUM). The results of the correspondence assessment for each of the trials were summarized into contingency tables, from which various measures of correspondence were computed.

5.3 Selected Measures of Correspondence

A variety of measures have been employed for describing the accuracy of image classifications. Stehman (1997) acknowledged that there is no general agreement in the literature as to which measures are appropriate for a given accuracy assessment objective, although the kappa coefficient appears to have been more widely employed than other measures (Rosenfield and Fitzpatrick-Lins, 1986; Fung and Ledrew, 1988; Foody, 1992b; Fitzgerald and Lees, 1994).

Stehman (1997, p.77) recommended that a measure of correspondence should be selected in accordance with the assessment objectives of the project:

> "Accuracy measures that are directly interpretable as probabilities of encountering certain types of misclassification errors or correct classifications should be selected in preference to measures not interpretable as such. User's and producer's accuracy and the overall proportion of area correctly classified are examples of accuracy measures possessing the desired probabilistic interpretation".

All summary measures of correspondence can obscure the detail found within the original error matrix and Congalton (1991) advocated the use of several measures for each assessment. The use of several different measures allows the analyst to compare the

relative rankings of the error matrices for each of these measures and because each of the measures incorporates different information from the error matrix, the relative rankings may differ (Congalton et. al., 1983). The technique of using multiple correspondence measures permits all possible information to be extracted from the error matrix and provides a more complete evaluation of classifier performance than that provided by a solitary measure (Aronoff, 1982; Story and Congalton, 1986). Three summary measures of correspondence will therefore be utilized: the overall proportion, of pixels correctly classified (P_c) along with the user's (P_U) and producer's (P_P) measures for individual classes; the kappa coefficient or khat statistic (k); and the modified kappa (k_n). The P_C and the modified kappa provide similar information, while the kappa provides unique information, particularly relevant to the visual appearance of the classification.

5.3.1 Proportional Measures of Correspondence

The proportional measures of correspondence have been selected primarily to provide a descriptive tool for the measure of classification error and confusion (Congalton, 1991). These measures include the overall percentage of correctly classified pixels (P_c), the producer's accuracy or error of omission (P_P), and the consumer's or user's measure of accuracy or error of commission (P_U) (Congalton et al., 1983; Story and Congalton, 1986). Confidence intervals were calculated for these proportional measures with the assumption that the test data used in the correspondence assessment were selected using stratified random sampling, acknowledging that the calculation of confidence intervals for overall probability measures is strongly dependent on the manner in which the samples were collected (Green et al., 1993; Janssen and van der Wel, 1994).

The overall proportion of area correctly classified is defined in Equation 5.1 (Stehman, 1997):

$$P_c = \sum_{k=1}^{q} p_{kk.}$$

[Equation 5.1]

where:

P_c	=	overall percentage of correctly classified pixels
p_{kk}	=	diagonal elements of the error matrix
q	=	number of land cover categories

The producer's accuracy for a cover type *j*, is the conditional probability that an area classified as category *j* by the reference data is classified as category *j* by the map and is given in Equation 5.2 (Stehman, 1997):

$$P_{Pj} = p_{jj} \; / \; p_{+j.}$$

[Equation 5.2]

where:

P_{Pj}	=	producer's accuracy for cover type *j*
p_{jj}	=	number of pixels classified in category *j*
p_{+j}	=	column total for category *j*

The producer's accuracy provides an indication of the ability of the classification algorithm to discriminate a particular class successfully while the user's accuracy is indicative of the probability that a pixel classified on the image actually represents that class on the ground (Story and Congalton, 1986; Congalton, 1991). The consumer's or user's accuracy for cover type *i*, is the conditional probability that an area classified as category *i* by the map is classified as category *i* by the reference data as defined in Equation 5.3 (Stehman, 1997):

$$P_{Ui} = p_{ii} \; / \; p_{+i.}$$

[Equation 5.3]

where:

P_{Ui}	=	user's accuracy for cover type *i*
p_{ii}	=	number of pixels classified in category *i*
p_{+i}	=	column total for category *i*

Snedecor and Cochran (1980) developed a standard test for comparing two population proportions and this test has been used in this study for the pairwise comparisons of P_c. This method will be utilized to assess the statistical significance of the difference between two given classification trials, at the 95% confidence level. The method for calculating the test statistic is provided in Equation 5.4:

$$z = \frac{\hat{P}_{c1} - \hat{P}_{c2}}{\sqrt{\hat{P}_{c1}(1-\hat{P}_{c1})/n_1 + \hat{P}_{c2}(1-\hat{P}_{c2})/n_2}}$$

[Equation 5.4]

where:

P_{c1} *and* P_{c2}	=	producer's measures of accuracy
n_1 *and* n_2	=	number of samples used for assessment

One of the assumptions associated with the pairwise comparison of matrices is that the reference samples used are independent. This assumption is violated in this instance since the same reference sample has been used in the creation of each error matrix at each level. The routine violation of this assumption results in the test statistic being only an approximation to the correct statistical test and will be regarded as such in

discussion of the results. Nonetheless, the pairwise comparison of P_c provides an interesting contrast to the pairwise comparison of kappa. P_c is an important measure because it is the base from which other summary measures are derived. Stehman (1997, p.79) argues that "users may chose to represent accuracy by another parameter, but it is not relevant to claim that P_c is incorrect or that it provides a biased measure of accuracy."

5.3.2 The Kappa Coefficient

The kappa coefficient was originally introduced to the remote sensing community by Congalton et al. (1983) and is defined as "a measure of the overall agreement computed for each matrix based on the difference between the actual agreement of the classification and the chance agreement" (Congalton et al., 1983, p.276). The kappa measure of agreement was developed by Cohen (1960) and has been widely used and promoted in the remote sensing literature. Rosenfield and Fitzpatrick-Lins (1986, p.226) recommended that "kappa be adopted by the remote sensing community as a measure of accuracy for thematic classification as a whole and for individual categories" and Fitzgerald and Lees (1994, p.368) argued "that the kappa test statistic be used in preference to the overall accuracy as a means of testing classification accuracy based on error matrices".

In the computation of kappa (khat), all of the cells in the error matrix are used (Bishop et. al., 1975). The correct formulation of kappa is provided by Bishop et al., (1975) and Hudson and Ramm (1987) in Equation 5.5:

$$\hat{k} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} x_{i+}x_{+i}}{N^2 - \sum_{i=1}^{r} x_{i+}x_{+i}}$$

[Equation 5.5]

where:

r	=	the number of rows in the matrix
x_{ii}	=	the number of observations in row *i* and column *i*
x_{i+} **and** x_{+i}	=	the marginal totals of row *i* and column *i* respectfully
N	=	the total number of observations

The approximate large sample variance for kappa as presented by Fleiss et al. (1969) and Bishop et al. (1975) is provided in Equation 5.6:

$$\hat{\sigma}^2[\hat{K}] = \frac{1}{N}[\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2{}^2)}{(1-\theta_2)^4}]$$

[Equation 5.6]

where:

$$\theta_1 = \sum_{i=1}^{r} x_{ii}/N$$

$$\theta_2 = \sum_{i=1}^{r} x_{i+}x_{+i}/N^2$$

$$\theta_3 = \sum_{i=1}^{r} x_{ii}(x_{i+} + x_{+i})/N^2$$

$$\theta_4 = \sum_{i=1}^{r} x_{ij}(x_{i+} + x_{+i})^2/N^3$$

The approximate large sample variance described above is used in the construction of confidence intervals for the kappa statistic. The kappa can be used to test the significance of each error matrix alone, to "determine whether the results presented in

the error matrix are significantly better than a random result, khat = 0" (Congalton, 1991, p. 41) as in Equation 5.7:

$$z \approx \frac{\hat{K}_1 - 0}{\sqrt{\hat{\sigma}_1^{\,2} + 0}}$$

[Equation 5.7]

where:

K_1 = kappa statistic
σ^2 = kappa variance

A pairwise comparison is considered significant at the 95% probability level if the *z* statistic exceeds 1.96 (Rosenfield and Fitzpatrick-Lins, 1986). The pairwise comparison is of the form:

$$z \approx \frac{\hat{K}_1 - \hat{K}_2}{\sqrt{\hat{\sigma}_1^{\,2} + \hat{\sigma}_2^{\,2}}}$$

[Equation 5.8]

where:

K_1 and K_2 = kappa statistics
σ_1^2 and σ_2^2 = kappa variance

Values of kappa typically lie on a scale from 0 to 1 with a kappa value of 0 indicating that the obtained level of agreement is equal to the level of chance agreement (i.e. the results are no better than a random result). Where kappa is positive (greater than zero), then the obtained agreement is greater than chance agreement, and conversely when kappa is negative (less than zero) then obtained agreement is less than that obtained by chance. A kappa value of 1 indicates a perfect level of agreement with the reference data (Rosenfield and Fitzpatrick-Lins, 1986).

Controversy surrounding the use of kappa was highlighted by Agresti (1996) who cites the dependence of the kappa definition of chance agreement on the marginal proportions: the stronger the agreement for a given pair of marginal distributions, the higher the value of kappa. As Stehman (1997, p. 80) argued:

> "The kappa adjustment for chance agreement is tantamount to assuming fixed map marginal proportions. But this assumption imposes a circularity in reasoning because the map (row) marginals are the result of the classification, not a fixed set of marginal proportions the map construction process was required to match."

Modified measures of agreement with different interpretations of chance agreement have been put forward by researchers who claim that kappa underestimates classification accuracy (Brennan and Prediger, 1981; Foody, 1992b; Ma and Redmond, 1995).

5.3.3 The Modified Kappa Coefficient

Foody (1992b) argued that the calculation of the proportion of chance agreement which is used for the calculation of the kappa coefficient is overestimated, resulting in kappa values which under represent the true level of classification accuracy. The calculation of the kappa coefficient is represented by Foody (1992b) as follows:

$$\hat{k} = \frac{P_o - P_e}{1 - P_e}$$

[Equation 5.9]

where P_o is the observed proportion of agreement and P_e is the proportion of chance agreement (Cohen, 1960; Rosenfield and Fitzpatrick-Lins, 1986). The proportion of chance agreement is calculated from the row and column marginals as follows (Foody, 1992b):

$$\sum_{i=1}^{n} P_r(i)P_c(i)$$

[Equation 5.10]

where n = the number of classes.

The proportion of chance agreement represented in Equation 5.10, being derived from observed row and column marginals, includes actual agreement as well as chance agreement, thus the calculation of kappa in Equation 5.9 does not represent the proportion of observed agreement less the proportion of chance agreement alone. By including actual agreement along with chance agreement in P_e, the magnitude of kappa is reduced, as is the level of apparent classification accuracy.

In the construction of error matrices for image classification, the row and marginal proportions are free to vary and are not fixed *a priori*, as is assumed in the calculation of kappa. Without any advanced knowledge regarding the land cover classes of the study region, there is an equal probability that the area would be classified into any one of n classes (Stehman, 1997). Therefore, the marginal proportion for each class are considered to be $1/n$ and the probability of chance agreement can thus be shown to be $1/n$ as well. The method for calculating the modified kappa index using this form of the probability of chance agreement is found in Equation 5.11 (Brennan and Prediger, 1981):

$$k_n = \frac{P_o - 1/n}{1 - 1/n}$$

[Equation 5.11]

The modified kappa therefore provides a measure of classification accuracy, from which has been removed only the proportion of chance agreement.

5.4 Classification Trials

The classification trials were designed with four main objectives: to determine the optimum configuration of the various neural network parameters for the specific project imagery and target classes, based on heuristics developed by Kannellopoulos and Wilkinson (1997) and Paola (1994); to compare the classification results of the neural network to those of a standard maximum likelihood classification; to determine if the geometric and atmospheric corrections performed on the project imagery altered the resulting classifications; and to assess the interpretive depth of the BCLCCS that can be consistently achieved with Landsat Thematic Mapper data. A summary of the classification trials conducted at each level is included in Appendix D, while Appendix E includes maps of the classification output from each of the trials.

5.5 Level 1: Land Base (Vegetated and Non-Vegetated)

Twenty-eight classification trials were processed for level one as outlined in Appendix D. Trials 1 through 4 and trial 26 were processed over the entire image, while the remaining trials were processed with areas of agricultural and urban development masked out (see section 4.8). A summary of the descriptive statistics for the overall measures of correspondence is included in Table 5.2. Values for P_c ranged from 50.84% to 87.17% with a mean overall level of correspondence of 73.36% and standard deviation of 7.89%.

Table 5.2 Level 1 Descriptive Statistics for All Trials (P_C).

DESCRIPTIVE STATISTICS			
Mean	73.357	Range	36.331
Standard Error	1.491	Minimum	50.839
Median	73.755	Maximum	87.170
Mode	71.822	Sum	2054.013
Standard Deviation	7.890	Count	28
Sample Variance	62.257	Confidence Level (95.0%)	3.059

A summary of the descriptive statistics for the producers' and consumers' measures of correspondence for both the vegetated and non-vegetated classes is presented in Table 5.3. An examination of these descriptive statistics indicates that there is a large discrepancy in the measured correspondence between the two target classes. The mean value for the producer's measure of correspondence (P_P) is 15.96% for the non-vegetated class compared to 90.65% for the vegetated class. The average P_U for vegetated class is a more moderate 78.51% compared to 32.58% for the non-vegetated class, which has a standard deviation of 13.18% compared to 10.26% for the vegetated class.

The discrepancy in the success with which the classification algorithm can identify the vegetated class but not the non-vegetated class may be explained by two factors. Firstly, the non-vegetated class represents a smaller portion of the study area and has fewer reference samples available for the correspondence assessment. Secondly, the reference samples which are available within the non-vegetated area have a poor spatial distribution, being located predominantly on the periphery of non-vegetated areas where there is greater potential for classification confusion. The dominance of the vegetated class in the reference data set and the strong levels of correspondence for the vegetated class have resulted in the inflated levels of overall correspondence for the level one classification trials (Fleiss, 1981).

Table 5.3. Level 1 Descriptive Statistics, Producers' and Consumers' Correspondence Measures for All Trials.

Descriptive Statistics	Producer (NV)	Producer (V)	Consumer (NV)	Consumer (V)
Mean	15.846	90.650	32.575	78.508
Standard Error	0.345	0.114	2.490	1.939
Median	15.530	90.680	34.442	78.260
Mode	14.800	90.200	35.560	83.190
Standard Deviation	1.824	0.603	13.175	10.259
Sample Variance	3.327	0.363	173.585	105.254
Range	8.190	2.230	55.550	47.450
Minimum	12.960	89.560	6.670	49.460
Maximum	21.150	91.790	62.220	96.910
Sum	443.674	2538.187	912.094	2198.226
Count	28.000	28.000	28.000	28.000
Confidence Level (95.0%)	0.707	0.234	5.109	3.978

The bias in the marginal proportions of the correspondence matrices are reflected in the kappa statistics which are on average, an order of magnitude less than P_c. This difference is caused by the imbalance in the distribution of the reference data between the vegetated and non-vegetated classes and the influence this imbalance exerts over the calculation of the kappa statistic. The modified kappa statistic provides a more cautious representation of the overall correspondence without being influenced by the imbalance in the error matrix.

The top ten ranked trials for each of the measures is included in Table 5.4. The rankings for the modified kappa are equal to those of the overall measure, while the khat is significantly different. Only trials 15, 16, 17, 10 and 9 are ranked consistently in the top 10 trials for all three measures. Table 5.4 indicates that there are indeed differences in the level of correspondence as measured by the proportional measures and the kappa value. Reliance on either one of these methods alone to report the correspondence of the classification trials would clearly be misleading and the use of all three measures provides a clearer picture of the perfomance of the ANN and the MLC.

Table 5.4 Relative Ranking of Classification Trials.

RANK	OVERALL	KAPPA	MODIFIED KAPPA
1	20	5	20
2	17	1	17
3	8	12	8
4	26	15	26
5	28	14	28
6	6	13	6
7	16	16	16
8	15	17	15
9	10	10	10
10	9	9	9

5.5.1 Number of Hidden Units in a Hidden Layer

Pairwise comparisons were conducted for a three layer neural network developed with a 6-2-2 structure (6 input units, 2 hidden units and 2 output units/classes) (Paola, 1994) and with a 6-6-2 structure (6 input units, 6 hidden units and 2 output units/classes) (Kanellopoulos and Wilkinson, 1997) as outlined in Section 4.8.1.1. These trial pairs differ only in the number of iterations for which they have been processed: 1000, 5000 and 10000 for comparison A, B and C respectfully. The results of these pairwise comparisons are presented in Table 5.5.

The greater overall correspondence achieved by the 6-2-2 network structure is significant at a 95% confidence level in both comparisons A and B. Conversely, in comparison C (10000 iterations), it is the 6-6-2 network structure (trial 14) which has a significantly higher level of overall correspondence than the 6-2-2 network structure (trial 11). The kappa values are higher for all of the 6-6-2 trials, including trial 12 and trial 13 which had lower overall accuracies than the 6-2-2 trials in comparisons A and B (although these differences in kappa are not significant at the 95% confidence level). The

modified kappa values, while indicating that the actual level of agreement is substantially lower than P_C, reflect the same trend as the latter, however they range from only 35% to 54% compared to 68% to 77% for P_C.

Table 5.5 Comparative Statistics: Number of Hidden Units in a Hidden Layer.

	Comparison A		***Comparison B***		***Comparison C***	
Structure	6-2-2	6-6-2	6-2-2	6-6-2	6-2-2	6-6-2
Trial	9	12	10	13	11	14
overall	0.7638	0.7182	0.7709	0.7182	0.6751	0.7362
comparative z statistic P_C	**3.0487**		**3.5665**		**-3.7177**	
kappa	0.07562	.09673	0.07659	0.09125	0.07043	0.0914
individual z statistic K	1.78827	2.12212	1.82602	2.00602	1.47114	2.05673
comparative z statistic K	**-0.3396**		**-0.2370**		**-0.3210**	
modified kappa	0.5276	0.4364	0.5420	0.4364	0.3501	0.4724

On the basis of a visual comparison, trials 12, 13 and 14 (6-6-2 structure) present classifications which correspond more closely to the overall distribution of the classes as derived from the forest inventory data. The individual z scores for trials 12, 13 and 14 (based on the kappa values) indicate that all three of the error matrices for these trials are significant at the 95% confidence level (greater than 1.96), while none of the 6-2-2 structure trials have significant individual *z* statistics. In comparison A, it is clear that trial 12 has defined more homogenous areas than trial 9. The lack of NV identified in trial 9 accounts for its higher P_C and the P_U values for trial 12 are 40% compared to 28% for trial 9. One of the striking visual differences between trial 9 and trial 12 is the success of the latter in identifying linear features such as roads and rivers. The trials in comparison B do not appear that different from those in comparison A. In comparison C at 10000 iterations, the appearance of the classifications change and the accuracy of the 6-6-2 network exceeds that of the 6-2-2 network. The latter (trial 11) is finally capable of

discriminating linear features, although not nearly as clearly as in trial 14 and appears very speckled or heterogeneous.

Histograms of the V and NV classes are included in Appendix B. The V and NV classes are very broadly defined classes with large intra-class variance and very similar mean reflectance values. The NV class however has a multi-modal distribution and this may contribute to the poorer performance of the MLC in discriminating this class. The increased number of hidden nodes in the 6-6-2 network trials appears necessary for effectively representing the complexity of the distributions for these two classes. This analysis suggests that the use of a greater number of hidden units may not result in higher levels of overall correspondence, however, it may result in classifications which are more realistic, relative to the proxy vegetation data (as reflected in the kappa values), particularly if a greater number of training iterations are employed. The increased complexity and training difficulty associated with increasing the number of hidden nodes is reflected in the results of comparison C, where it takes a greater number of iterations for a network with more hidden nodes to achieve a comparable level of accuracy to a network with fewer hidden nodes. The decrease in accuracy for the 6-2-2 network after 10000 iterations (trial 11) may indicate that the network has become overtrained and is therefore less able to generalize to the image data.

5.5.2 Number of Hidden Layers

A comparison was made between a four layer network with a 6-24-6-2 architecture (trial 21) and three different three layer networks (trials 7, 9 and 12) with a 6-6-2 architecture. These results, included in Table 5.6, indicate that the higher levels of

overall correspondence achieved by the three layer networks (trials 7, 9 and 12) are significant at the 95% confidence level for comparisons A, B and C. The three layer networks also have higher kappa and modified kappa values, although the comparisons based on kappa are not significant. Trial 12 (6-6-2) is the only matrix of the four which is significant.

Table 5.6 Comparative Statistics: Number of Hidden Layers.

	Comparison A		*Comparison B*		*Comparison C*	
Structure	6-24-6-2	6-2-2	6-24-6-2	6-6-2	6-24-6-2	6-24-2
Trial	21	9	21	12	21	7
overall	0.6331	0.7638	0.6331	0.7182	0.6331	0.7482
comparative z statistic P_C	**-7.7319**		**-5.0278**		**-6.8053**	
kappa	0.05092	0.07562	0.05092	0.09673	0.05092	0.07263
Individual z statistic K	0.9763	1.78827	0.9763	2.12212	0.9763	1.68484
comparative z statistic K	**-0.3678**		**-0.6613**		**-0.3208**	
Modified kappa	0.2662	0.5276	0.2662	0.4364	0.2662	0.4964

Visually, trial 21 (6-24-6-2) has a very poor correspondence with the test data in spite of its overall correspondence measure of 63%. This may be explained by the amount of NV identified in this classification (P_U = 52%) compared to trial 12 (P_U = 40%). It would appear that some areas of shadow have been included in class definition of NV in trial 21 and have contributed to the very heterogeneous and inferior appearance of the classification. These results would suggest that four layer neural networks are not necessary for achieving increased levels of correspondence and that they may actually hamper the networks' ability to define a useful classification in this instance where there are only two target classes.

5.5.3 Types of Input Features

Trials were executed which used the bands selected from the divergence analysis (see section 4.6) and also the original six spectral TM bands. The results of the input feature comparisons are included in Table 5.7. The trials used in the pairwise comparisons are identical in every respect except for the input units used. In comparison A and C, the trials which used the original TM bands as input have significantly higher levels of overall correspondence than the trials generated using the selected band set. In addition, the trials generated from the original TM bands seemed to have produced more consistent classification results, having a range of only 6% compared to the widely varying results of the selected band set with a range of 36%.

Table 5.7 Comparative Statistics: Types of Input Features.

	Comparison A		*Comparison B*		*Comparison C*	
Bands	3,5,7,9,11,16	1,2,3,4,5,6	3,5,7,9,11,16	1,2,3,4,5,6	3,5,7,9,11,16	1,2,3,4,5,6
Trial	19	22	20	23	21	24
overall	0.5084	0.7266	0.8717	0.6775	0.6331	0.7240
comparative z statistic P_C	**-12.4352**		**16.4033**		**-5.3702**	
kappa	0.04376	0.08214	0.05092	0.09194	0.08134	0.07408
individual z statistic K	0.7714	1.83405	0.97635	1.92223	1.61909	1.65618
comparative z statistic K	**-0.5309**		**-0.5797**		**-0.0118**	
modified kappa	0.0168	0.4532	0.7434	0.3549	0.2662	0.4484

In comparison B, trial 20 (selected band set) has the higher P_C and a lower kappa. Although trial 20 has the highest level of correspondence among the twenty-eight trials generated, this trial had one of the poorest classifications (in comparison to the forest inventory data) upon visual inspection. The trial 20 classification has identified very few areas of non-vegetated land, resulting in extremely high levels of correspondence for the vegetated class, and thereby resulting in the misleading P_c value for this trial. The kappa

statistic portrays a more realistic measure of the correspondence achieved by trial 20 relative to trial 23 in comparison B. The behaviour of trials 19, 20 and 21 is very erratic, suggesting that a four layer network may be unnecessary and unsuitable for this two class classification.

Overall, the original TM bands appear to provide a superior classification result from the neural networks for this level of the BCLCCS. This raises the question of the necessity of performing band manipulations and divergence analysis for bandset selection when using neural networks. The problem may be in the combination of synthetic and raw bands and it may have been more appropriate to use one or the other but not both together.

5.5.4 Number of Training Samples

The use of the forest inventory as training data provided a large resource for the selection of training pixels. Generally however, the number of pixels was not evenly distributed amongst classes, as indicated in section 5.2. The trials executed at level one had either 3.3 % (23500 pixels per class) or 16.5 % (118000 pixels per class) of the *total* image data. These trials which were compared had identical architectures and numbers of learning iterations. The results of the trial comparisons is included in Table 5.8.

Table 5.8 Comparative Statistics: Number of Training Samples.

	Comparison A		*Comparison B*	
Sample Size	23500	118000	23500	118000
Trial	5	6	7	8
overall	0.7542	0.8153	0.7482	0.8261
comparative z statistic P_C	**-4.0491**		**-5.1263**	
kappa	0.10402	0.05023	0.07263	0.06504
individual z statistic K	2.37392	1.27695	1.68484	1.58835
comparative z statistic K	**0.9135**		**0.1277**	
modified kappa	0.5084	0.6306	0.4964	0.6523

Trials 6 and 8, which had the largest sample sizes (118000 pixels per class) achieved significantly higher levels of overall correspondence, however they had lower kappa values and visually, were inferior in comparison to trials generated with much smaller sample sizes (trials 5 and 7). Furthermore, the large sample matrices were not significant, however the error matrix for trial 5 was significant. Visually, trial 6 has delineated only small amounts of NV, contributing to its higher overall accuracy. Trial 5 has a better representation of NV but is very speckled and poorly discriminates linear features. Trial 7 is a great improvement over trial 5 with a better discrimination of linear features and more homogenous regions – it has a lower accuracy but is still better than trial 8.

The inflated levels of correspondence for trial 6 and 8 result from a classification with a disproportional amount of area classified as vegetated. An explanation for the poor results of trial 6 and trial 8 may be that a large number of training samples renders the network too specific to the training data and thereby less able to generalize to the rest of the image data when the final classification is processed. Other studies have concluded that only 5-10% of image data is necessary for training.

5.5.5 Number of Learning Iterations

Fifteen comparisons were made between different trials having a range of 1000 to 50000 iterations. The results of these trials are summarized graphically in Figure 5.1 and in Table 5.9. In Figure 5.2, the two 6-6-2 networks differed in the values for their learning and momentum rates and the 6-24-6-2 structures differed in the image channels used as input. It is clear from these results that increasing the number of training iterations does not necessarily lead to increases in classification correspondence and hence, there are no obvious patterns which can be identified. In some trials, increasing the number of iterations has led to a greater level of overall correspondence as in trial 14, however such increases in P_c did not occur in trial 11 and trial 21. It would appear that the number of iterations alone may not be significant to overall network performance, and that the number of iterations has greater import when considered in conjunction with other neural network parameters. The optimum number of iterations therefore appears to be connected the configuration of the other structural elements in the neural network and the features being classified.

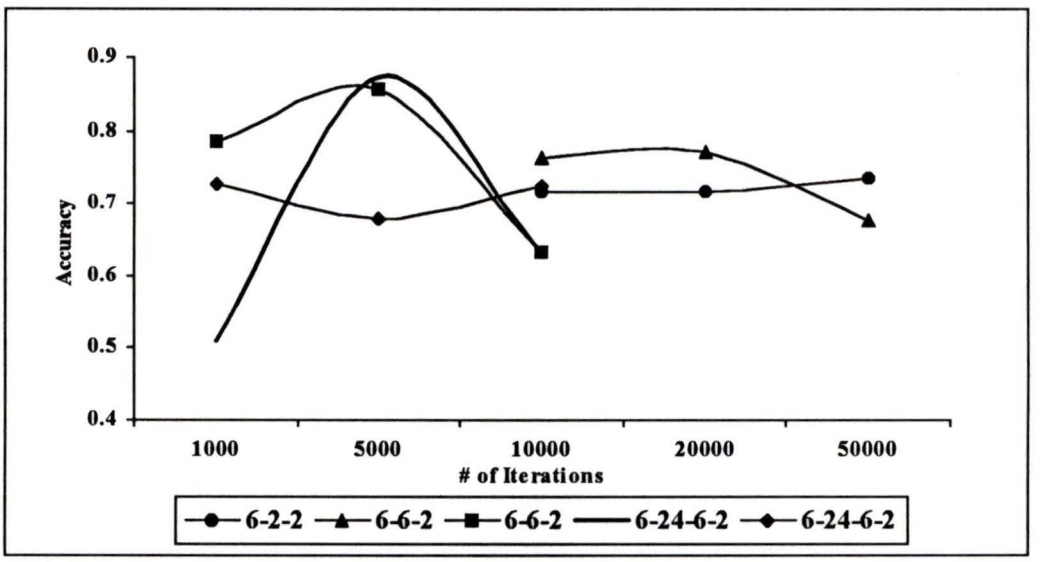


Figure 5.2 The Number Of Learning Iterations And Their Influence On Overall Accuracy.

Table 5.9 Comparative Statistics: Number of Learning Iterations.

	Comparison A		*Comparison B*		*Comparison C*	
Iterations	10000	20000	50000	20000	10000	50000
Trial	12	13	14	13	12	14
overall	0.7182	0.7182	0.7362	0.7182	0.7182	0.7362
comparative z statistic P_C	**0.0000**		**-1.137854**		**-1.1378**	
kappa	0.09673	0.09125	0.09140	0.09125	0.09673	0.09140
individual z statistic K	2.12212	2.00602	2.05673	2.00602	2.12212	2.05673
comparative z statistic K	**0.0851**		**-0.0023**		**0.0838**	
modified kappa	0.4364	0.4364	0.4724	0.4364	0.4364	0.4724

	Comparison D		*Comparison E*		*Comparison F*	
Iterations	10000	20000	20000	50000	10000	50000
Trial	9	10	10	11	9	11
overall	0.7638	0.7709	0.7709	0.6751	0.7638	0.6751
comparative z statistic P_C	**-0.4813**		**6.4758**		**5.9282**	
kappa	0.07562	0.07659	0.07659	0.07043	0.07562	0.07043
individual z statistic K	1.78827	1.82602	1.82602	1.47114	1.78827	1.47114
comparative z statistic K	**-0.0163**		**0.0968**		**0.0812**	
modified kappa	0.5276	0.5420	0.5420	0.3501	0.5276	0.3501

	Comparison G		*Comparison H*		*Comparison I*	
Iterations	1000	5000	5000	10000	1000	10000
Trial	16	17	17	18	16	18
overall	0.7865	0.8561	0.8561	0.6331	0.7865	0.6331
comparative z statistic P_C	**-4.8516**		**17.9406**		**10.6095**	
kappa	0.08755	0.08241	0.08241	0.06371	0.08755	0.06371
individual z statistic K	2.09330	1.75999	1.75999	1.26471	2.09330	1.26471
comparative z statistic K	**0.0819**		**0.2719**		**0.3642**	
modified kappa	0.5731	0.7122	0.7122	0.2662	0.5731	0.2662

	Comparison J		*Comparison K*		*Comparison L*	
Iterations	1000	5000	5000	10000	1000	10000
Trial	19	20	20	21	19	21
overall	0.5084	0.8717	0.8717	0.6331	0.5084	0.6331
comparative z statistic P_C	**-20.8257**		**20.1212**		**-7.0893**	
kappa	0.04376	0.05092	0.05092	0.08134	0.04376	0.08134
individual z statistic K	0.77114	0.97635	0.97635	1.61909	0.77114	1.61909
comparative z statistic K	**-0.0929**		**-0.4200**		**-0.4958**	
modified kappa	0.0168	0.7434	0.7434	0.2662	0.0168	0.2662

	Comparison M		*Comparison N*		*Comparison O*	
Iterations	1000	5000	5000	10000	1000	10000
Trial	22	23	23	24	22	24
overall	0.7266	0.6775	0.6775	0.7240	0.7266	0.7240
comparative z statistic P_C	**3.1328**		**-2.833898**		**0.1671**	
kappa	0.08214	0.09194	0.09194	0.07408	0.08214	0.07408
individual z statistic K	1.83405	1.92223	1.92223	1.65618	1.83405	1.65618
comparative z statistic K	**-0.1497**		**0.2727**		**0.1272**	
modified kappa	0.4532	0.3549	0.3549	0.4484	0.4532	0.3549

5.5.6 Momentum and Learning Rates

Three different momentum and learning rate combinations were selected based on those used successfully in other studies (see section 4.8.1.1). The results of these comparisons are included in Table 5.10.

Table 5.10 Comparative Statistics: Momentum and Learning Rates.

	Comparison A		*Comparison B*		*Comparison C*	
Momentum Rate	0.001	0.1	0.1	0.9	0.001	0.9
Learning Rate	0.010	0.025	0.025	0.1	0.010	0.1
Trial	9	15	15	18	9	18
overall	0.7638	0.7733	0.7733	0.6331	0.7638	0.6331
comparative z statistic P_C	**-0.6421**		**9.4951**		**8.7219**	
kappa	0.07562	0.09313	0.09313	0.06371	0.07562	0.06371
individual z statistic K	1.78827	2.18835	2.18835	1.26471	1.78827	1.26471
comparative z statistic K	**-0.2920**		**0.4463**		**0.1811**	
modified kappa	0.5276	0.5468	0.5468	0.2662	0.5276	0.2662

While there are no significant differences in P_c between trials 9 and 15 in comparison A, there are significant differences between the trials in comparisons B and C, with the more moderate learning and momentum rates of trial 15 and trial 9 having higher overall accuracies and kappa values. Visually, it would appear that the moderate momentum and learning rate used in trial 15 has provided the best overall classification results of the three trials and is the matrix with the highest kappa value and the only matrix with a significant individual z statistic. The learning and momentum rates must be determined through trial and error for each particular application and these results indicate that a tradeoff between network speed (learning rate) and network convergence (momentum rate) have produced the best results.

5.5.7 Different Image Acquisition Dates

The objective of the following comparisons was to examine the consistency of the neural network configuration when applied to different image dates for the same scene. There were four image dates available: July, August, September, October. The August image was of a better quality than the July image and hence it was used for the majority of experimental trials generated. An identical neural network was constructed on all four image dates and comparisons were made between the August image and each of the other image dates. The results of these comparisons are summarized in Table 5.11.

These results indicate that there is not a significant difference between the overall accuracies for the July and the August images (comparison B), however while the kappa values for these two trials are not significantly different, they are suprising since the July image date, which has the highest P_C at 74.97%, has a lower kappa than the August image date (and indeed, lower than the September image date). The August image (trial 1) is the only image date with an individual z statistic that is significant at the 95% confidence level. Since both July and August images were acquired within a period of full vegetation canopy and relatively high solar angles, it was anticipated that they would have similar levels of accuracy. Visually, the July image date has a superior appearance: linear features are more clearly distinguished and the classes are more homogenous.

For the trials of comparisons A and C, there is a significant difference in the overall accuracies at the 95% confidence level, with the September and October image dates having lower values for P_C. Visually, it is difficult to distinguish between the classifications of the August, September and October images, except that the latter two have discriminated more non-vegetated areas.

Table 5.11 Comparative Statistics: Different Image Acquisition Dates.

	Comparison A		*Comparison B*		*Comparison C*	
Date	August	September	August	July	August	October
Trial	1	2	1	3	1	4
overall	0.7389	0.7054	0.7389	0.7497	0.7389	0.6994
comparative z statistic P_C	**2.167**		**-0.6997**		**2.5547**	
kappa	0.09984	0.06312	0.09984	0.06096	0.09984	0.04693
individual z statistic K	2.24437	1.37979	2.24437	1.43382	2.24437	1.02447
comparative z statistic K	**0.5754**		**0.6318**		**0.8286**	
modified kappa	0.4778	0.4108	0.4778	0.4994	0.4778	0.3988

The topography within the study area results in shadows which influence the results of images acquired at a time of lower solar angles (September and October). In addition, the elevation of the study area means that summer is short and frost and leaf-off occur by mid to late September. Despite these factors, the classification results for all four dates are relatively consistent, although a July or August image date would be preferable.

5.5.8 Corrected Versus Uncorrected Imagery

Additional comparisons were made between the neural network classification results for the geometrically and atmospherically corrected imagery and the uncorrected imagery. The objective of this comparison was to determine whether or not the correction methods used were significantly detracting from the information content in the image. The results of these comparisons are included in Table 5.12. The results indicate that the greater overall correspondence found for the corrected image (trial 1) is significantly different from that of the uncorrected image (trial 25) at the 95% confidence level. The individual z statistic for the matrix of trial 1 is significant and its kappa value is higher than trial 25. This comparison demonstrates that geometric and atmospheric

corrections have not resulted in an inferior classification relative to the uncorrected image.

Table 5.12 Comparative Statistics: Corrected Versus Uncorrected.

	Comparison A	
Type	Corrected	Uncorrected
Trial	1	25
overall	0.7389	0.6778
comparative z statistic P_C	**3.949**	
kappa	0.09984	0.07251
individual z statistic	2.24437	1.51990
comparative z statistic K	**0.4190**	
modified kappa	0.4778	0.3556

5.5.9 Neural Network Versus Maximum Likelihood

The maximum likelihood classification is traditionally used in classification research as a benchmark of comparison for alternate classification methods (see section 2.3.1.3). A similar approach was adopted in this research and a comparison has been made between the classification results of the neural network and the classification results of the maximum likelihood classifier for every level of the BCLCCS considered. The results of the comparisons made at level one are included in Table 5.13. An initial comparison was made between the original neural network and a maximum likelihood classifier with no masks being applied to the image data (comparison A). The difference in overall correspondence between these trials (trials 1 and 26) is significant at the 95% confidence level, with the MLC achieving a higher value for P_c. These two classifications appear fundamentally different when examined visually. The MLC

discriminates the non-vegetated class poorly and therefore has a falsely inflated level of overall accuracy. This is reflected in the very low kappa values for trial 26 compared to trial 1.

Table 5.13 Comparative Statistics: Neural Network vs. Maximum Likelihood.

	Comparison A		*Comparison B*	
Type	Neural Network	Maximum Likelihood	Neural Network	Maximum Likelihood
Trial	1	26	1	28
overall	0.7389	0.8179	0.7389	0.8177
comparative z statistic P_c	**-5.1365**		**-5.123424**	
kappa	0.09984	0.04421	0.09984	0.00441
individual z statistic K	2.24437	1.13479	2.24437	1.09942
comparative z statistic K	**0.9409**		**0.9692**	
modified kappa	0.4778	0.6359	0.4778	0.6355

In an effort to improve on the results of trial 26, an additional MLC trial was created with the image masks referred to in Section 4.7 in place. The masks failed to improve the MLC's ability to discriminate the non-vegetated class. In comparison B, the difference between the new MLC and the neural network classification is significant at the 95% confidence level. Neither of the MLC matrices have significant individual z statistics as trial 1 does, as the latter has a higher kappa than both the MLC trials. The results of comparisons A and B indicate that the neural network is more successful at discriminating the vegetated and non-vegetated classes than the MLC, which appears to be particularly hampered by the multi-modal distribution of these two classes.

5.5.10 *Summary*

A number of conclusions may be made on the basis of the classification results generated for the first level of the BCLCCS:

- The mean P_C was 73.36% and the standard deviation was 7.89%.
- The trials with significant individual z-statistics were trials 1, 5 and 12 through 16.
- The number of hidden units should equal the number of input units (Kannellopoulos and Wilkinson, 1997).
- The use of a 3 layer network is preferable to a 4 layer network, particularly since the number of classes under consideration are less than twenty.
- The use of an optimal bandset (comprised of original TM bands and band enhancements) selected from the divergence analysis does not necessarily provide superior classification results from a neural network, although the enhanced statistical separability of the classes provided by this optimal bandset is preferred for the MLC.
- It is not necessary to use large numbers of training pixels, and in the interest of network performance, it is preferable to keep the number of training samples reasonable.
- The number of iterations required to train the network is difficult to gauge and the results of the comparisons made in this study provide no indication of the optimal number of training iterations to be used. The optimal number of training iterations for any given application appears to be inextricably linked to the configuration of other neural network parameters.
- The classification results obtained with a moderate momentum rate of 0.1 and a moderate learning rate of 0.025 were preferable to those obtained with other

combinations. Ideally the neural network algorithm in PCI would have an adaptive learning rate that would automatically adjust according to the success of network convergence, similar to neural network software utilized by other researchers (Paola, 1994).

- Imagery dates which correspond with the main growing season (July and August) for the dominant vegetation cover types provide superior classification results. In addition, the July and August image dates also have higher solar angles which minimize the effect of shadow. Shadow is strongly influential in this study area due to the nature of the topography in the region.
- The geometric and atmospheric corrections do not distort the inherent spectral properties of the image to the extent that they result in an inferior classification. On the contrary, the classification results obtained from the corrected image data are preferable to those obtained from the uncorrected image data, having achieved a higher level of overall correspondence. In addition, the use of image data corrected to a projection which is common for all of the data sources used in the study, is necessary to facilitate the incorporation of ancillary and training data.
- The neural network provides a superior classification result when compared to the maximum likelihood classifier. The MLC appears to be much more strongly influenced by the presence of shadow in the image.

5.6 Level 2: Land Cover Type (Treed and Non-Treed)

Twelve classification trials were processed for level two as outlined in Appendix D. There is a dichotomy in the correspondence matrix similar to that found at level one of the BCLCCS. There were only 39 test points available for the vegetated non-treed

(VN) class, compared to 719 test points for the vegetated treed (VT) class, thereby negating a statistically valid comparison at this level of the BCLCCS. The measures of correspondence and the results of the comparison between the classification trials are nonetheless reported here because they reinforce the conclusions made at level one. A summary of the descriptive statistics for P_c is included in Table 5.14. Values for P_c ranged from 39.06% to 95.17% with a standard deviation of 18.79%. The mean value for P_c was 73.29%.

Table 5.14 Level 2 Descriptive Statistics for All Trials.

DESCRIPTIVE STATISTICS			
Mean	73.292583	Range	56.103
Standard Error	5.4251533	Minimum	39.067
Median	76.99	Maximum	95.17
Mode	95.17	Sum	879.511
Standard Deviation	18.793282	Count	12
Sample Variance	353.18745	Confidence Level (95.0%)	11.940688

Table 5.15 Level 2 Descriptive Statistics, Producers' and Consumers' Correspondence Measures for All Trials.

Descriptive Statistics	Producer (VT)	Producer (VN)	Consumer (VT)	Consumer (VN)
Mean	96.889	4.750	75.410	31.267
Standard Error	0.551	0.884	6.044	6.984
Median	96.065	5.756	79.355	30.000
Mode	100.000	0.000	100.000	0.000
Standard Deviation	1.910	3.063	20.937	24.194
Sample Variance	3.648	9.379	438.349	585.343
Range	5.100	8.547	62.500	70.000
Minimum	94.900	0.000	37.500	0.000
Maximum	100.000	8.547	100.000	70.000
Sum	1162.663	56.999	904.925	375.200
Count	12.000	12.000	12.000	12.000
Confidence Level (95.0%)	1.214	1.946	13.303	15.372

A summary of the descriptive statistics for the producers' and consumers' measures of correspondence for both VT and VN classes is presented in Table 5.15. An

examination of these descriptive statistic indicates that there is a large discrepancy in the measured correspondence between the two classes which is not surprising given the disparity in the distribution of the test data. The P_P for the vegetated class is 96.89% compared to 4.73% for the vegetated non-treed class. The mean P_c value for the latter is 31.27% and 75.41% for vegetated treed.

The first pairwise comparison (comparison A) was made between the original neural network as performed on the corrected imagery (trial 1) and the same neural network processed over the uncorrected imagery (trial 6). In comparisons B and C, trial 1 is compared to the maximum likelihood classifications of trial 5 and trial 12. The results of these comparisons are included in Table 5.16.

Table 5.16 Comparative Statistics: Corrected vs. Uncorrected and ANN vs. MLC.

	Comparison A		***Comparison B***		***Comparison C***	
Structure	Corrected	Uncorrected	Neural Network	MLC	Neural Network	MLC
Trial	1	6	1	5	1	12
overall	0.7733	0.7697	0.7733	0.3906	0.7733	0.4196
comparative z statistic P_C	**0.2242**		**25.8490**		**23.8831**	
kappa	0.03791	0.03408	0.03791	0.01117	0.03791	0.01640
individual z statistic K	0.61123	0.54283	0.61123	0.14741	0.61123	0.21874
comparative z statistic K	**0.0434**		**0.2730**		**0.2211**	
modified kappa	0.5466	0.5394	0.5466	-0.2187	0.5466	-0.1608

The differences between the classifications produced from the corrected and uncorrected images in comparison A are not significant at the 95% confidence level, for overall accuracy or for kappa and neither of the error matrices from trial 1 and trial 6 are individually significant. The relatively high P_C for both of these trials is a contrast to the kappa values for these trials. Visually, it is very difficult to discern any difference

between trials 1 and 6 and both have delineated much more VN than exists in the training data, as reflected in the kappa and P_P values for this class. The trial 1 classification result is more heterogeneous in appearance and has a clearer discrimination of linear features. The important result is that the atmospheric and geometric corrections have made no significant difference in the classification results of these two trials.

Although the level of correspondence for trials 1 and 6 is high, both can be seen to have more non-treed areas delimited than are indicated by the original training data. This may however not be indicative of misclassification, since an examination of the image data suggests that a number of these non-treed areas identified in trials 1 and 6 appear as recent harvesting cutblocks on the image. This would seem to indicate that the image and thereby the classification, may be reflecting changes to the landscape which had not yet been recorded in the forest inventory data used for this project.

Comparisons B and C both had significant differences in P_c at the 95% confidence level; however, differences in kappa for these trials are not significant. The overall levels of correspondence for the maximum likelihood classifications (trials 5 and 12) are very low in comparison to the neural network. The negative modified kappa values for these trials may be explained as follows: any classification which has less than the probability of any class (in this case, 50%) will be negative (refer to section 5.3.3). Visual inspection of trial 5 and 12 indicate that the MLC was *significantly* influenced by the existence of shadows in the image. Areas which have been designated as non-treed in trials 5 and 12 are in reality, areas of the TM image dominated by deep shadow. Trial 12 was processed with agricultural and urban development masked out in an effort to improve the MLC results, however differences between trial 5 and trial 12 are very minimal. The results of

comparisons B and C indicate that the neural network is more successful at discriminating treed and non-treed areas than the MLC. This confirms the findings from level one of the BCLCCS.

Further comparisons were made between the original neural network configuration and two additional configurations which were developed in an effort to improve on the classification results achieved in trial 1. The first of these (trial 8) involved the use of the heuristic developed by Kanellopoulos and Wilkinson (1997), with a 6-24-6-2 network architecture and which was used in level one of the BCLCCS. The second trial (trial 10) used a 6-6-2 network architecture with the six original Thematic Mapper bands (trial 10). This configuration was selected on the basis of its successful implementation at BCLCCS level one trials. These comparisons are summarized in Table 5.17.

Table 5.17 Comparative Statistics: Different ANN Architecture.

	Comparison D		***Comparison E***	
Type	Corrected	Revision	Neural Network	Revision
Trial	1	8	1	10
overall	0.7733	0.9517	0.7733	0.7958
comparative z statistic P_C	**-12.280**		**-1.551**	
kappa	0.03791	0.0000	0.03791	0.06421
individual z statistic K	0.61123	0.0000	0.61123	1.04984
comparative z statistic K	**0.3740**		**-0.3019**	
modified kappa	0.5466	-0.9035	0.5466	0.5916

The difference between trials 1 and 10 (comparison D) was not statistically significant at the 95% confidence level, however the difference between trials 1 and trial 8 (comparison E) was significant. Visual inspection of trial 8 indicates that its high level

of overall correspondence is achieved at the expense of very low correspondence for the non-treed class. A similar situation exists here as that found at level one of the BCLCCS, whereby the dominance of one class in the test sample can significantly misconstrue the results of the correspondence assessment. The trials in comparison E appear to be very similar upon visual inspection. A final observation at level two concerns the extremely poor performance of trials 7, 8 and 9, which seem to exemplify the phenomena of the neural network stalling at a local minimum. Increasing the number of training iterations has no impact on the overall level of accuracy.

5.6.1 Summary

The results of the level 2 classification trials may be summarized as follows:

- Mean P_c = 73.29% and standard deviation = 18.79%
- None of the trials had significant individual z-statistics.
- There was no statistically significant difference between results derived from the corrected versus the uncorrected image.
- The MLC performance was very poor relative to the neural network and image masks did not improve the classification results significantly.
- The 6-6-2 network structure with the original TM bands (trial 10) gave the best result.
- The dichotomous error matrix dominated by the VT class and the low number of reference points for the VN class compromise the validity of the results at this level.

5.7 Level 3: Landscape Position (Wetland, Upland and Alpine)

Two distinct approaches have been taken to the classification of wetland, upland and alpine areas for non-vegetated land (NL) and vegetated treed (VT) area. Initially, classifications based purely on spectral properties were generated, similar to the neural network and MLC trials of levels one and two. Secondly, the flexibility of the neural network was explored with the inclusion of slope, aspect and elevation information along with the spectral information as inputs to the neural network.

5.7.1 Vegetated Treed

A total of twenty-one trials were processed for level 3 VT as outlined in Appendix D. There are two potential classes for consideration at this level: vegetated treed wetland and vegetated treed upland. The vegetated treed wetland class, has a small areal extent, resulting in an insufficient reference sample size to assess the correspondence of this class. A summary of the descriptive statistics for these trials is included in Table 5.18. Values of overall correspondence ranged from 34.87% to 99.47% with a mean value of 78.08%.

Table 5.18 Level 3VT Descriptive Statistics for All Trials.

DESCRIPTIVE STATISTICS			
Mean	78.085	Range	64.591
Standard Error	4.102	Minimum	34.875
Median	80.070	Maximum	99.466
Mode	99.466	Sum	1639.794
Standard Deviation	18.798	Count	21.000
Sample Variance	353.361	Confidence Level (95.0%)	8.557

The summary statistics for P_P and P_U for the upland class are included in Table 5.19. Only the values for the upland class are included. The mean value of P_P was 99.48% and the standard deviation was 0.52%, compared to a mean of 78.36% for P_U and a standard deviation of 19.00%.

Table 5.19 Level 3VT Descriptive Statistics, Producers' and Consumers' Correspondence Measures for All Trials.

Descriptive Statistics	Producer (VTU)	Consumer (VTU)
Mean	99.480	78.363
Standard Error	0.034	4.146
Median	99.460	80.300
Mode	99.460	100.000
Standard Deviation	0.157	19.001
Sample Variance	0.025	361.054
Range	0.520	65.300
Minimum	99.250	34.700
Maximum	99.770	100.000
Sum	2089.083	1645.628
Count	21.000	21.000
Confidence Level (95.0%)	0.071	8.649

Similar to comparisons made at the previous two levels of the BCLCCS hierarchy, comparisons were made both between the uncorrected and corrected image data and the ANN and the MLC and these results are summarized in Table 5.20. The stronger level of correspondence achieved by trial 1 (conducted on the corrected image) in comparison A is significant at the 95% confidence level and a visual comparison between trials 1 and 6 indicates that the classification completed on the uncorrected image has identified more wetland areas, thereby contributing to its overall lower level of P_C. The P_P is very similar for trial 1 (99.4%) and trial 6 (99.3%), however the P_U values for these two trials are significantly different, trial 1 having a P_U of 83.9% and trial 6 having a P_U of 77.9%.

Table 5.20 Comparative Statistics: Corrected vs. Uncorrected and ANN vs. MLC.

	Comparison A		*Comparison B*		*Comparison C*	
Structure	Corrected	Uncorrected	ANN	MLC	ANN	MLC
Trial	1	6	1	5	1	21
overall	0.8345	0.7752	0.8345	0.3487	0.8345	0.3576
comparative z statistic P_C	**4.7841**		**36.9158**		**36.2334**	
kappa	-0.01044	-0.01064	-0.01044	0.00022	-0.01044	0.00037
individual z statistic K	-0.12921	-0.12871	-0.12921	0.00266	-0.12921	0.00266
comparative z statistic K	**1.5178**		**1.4125**		**0.0928**	
modified kappa	0.6690	0.5504	0.6690	-0.3025	0.6690	-0.2847

In comparison B, the extremely poor degree of correspondence achieved by the MLC in trials 5 and 21 is confirmed by the visual appearance of the classification. Trial 21 was generated with additional areas of agricultural and urban development masked out (see Section 4.7), however there was no significant improvement in classification accuracy. In these classification trials, there is obvious confusion with the deep shadows in the image and the MLC has discriminated shadow and non-shadow into two separate classes. The multi-modal nature of the wetland class contributed to the confusion between the two classes and the poor discrimination of the wetland class, particularly by the MLC which is incapable of handling non-normally distributed image data. For the ANN, this is less problematic. The negative kappa values also result from this imbalance in the error matrix and will therefore not be considered appropriate measures of correspondence for level 3.

Additional comparisons were made between the original ANN and alternate ANN architectures and these results are summarized in Table 5.21. Trials 9 and 12 in

comparisons D and E involved the use of a 6-24-6-2 network structure and trial 9 used the selected set of input bands while trial 12 included slope, aspect and elevation amongst its inputs. Both trials 9 and 12 have high levels of correspondence as a result of the limited wetland area delineated in their classifications and their ability to discern the upland class. These high overall accuracies mask the reality that wetland classes are not being accurately located relative to the training data, or in the case of trial 12, are not being delineated at all.

Table 5.21 Comparative Statistics: ANN Configurations.

	Comparison D		**Comparison E**	
Trial	1	9	1	12
Overall	0.8345	0.9893	0.8345	0.9946
comparative z statistic P_C	**-12.0321**		**-12.4492**	
kappa	-0.010440	-0.005370	-0.010440	0.0
individual z statistic K	-0.129210	-2.502080	-0.129210	0.0
comparative z statistic K	**2.1059**		**1.4111**	
modified kappa	0.6690	0.9786	0.6690	0.9893

Additional alternate ANNs were considered. Trials 14 and 16 in comparisons F and G are developed from Kanellopoulos and Wilkinson (1994) heuristic (see Section 4.8.1.1) and have an architecture where the number of hidden units are identical to the number of input units. Both trial 14 and trial 16 are significantly different from the results achieved by the original network architecture at the 95% confidence level. Trial 14 involved the use of the six original Thematic Mapper spectral bands while trial 16 also included slope, aspect and elevation information.

Table 5.22 Comparative Statistics: ANN Configurations.

	Comparison F		*Comparison G*	
Trial	1	14	1	16
overall	0.8345	0.6263	0.8345	0.8737
comparative z statistic P_C	**15.799659**		**-3.036082**	
kappa	-0.01044	-0.00110	-0.01044	-0.01033
individual z statistic	-0.12921	-0.01324	-0.12921	-0.13069
comparative z statistic K	**1.4299**		**1.5497**	
modified kappa	0.6690	0.2527	0.6690	0.7473

Trial 14 was one of the few neural network classifications which exhibited confusion resulting from the presence of shadow and light caused by the varying topography in the image. Trial 16 benefits from the addition of slope, aspect and elevation, however the presence of shadow have confused the results of these trials as well. What is perhaps most interesting about the comparison between these two trials, given that their structure is identical except for the addition of the ancillary data, is the significant differences in the classification results.

Additional trials which used the elevation (trial 18) and slope (trial 20) information separately in combination with the original TM bands were also attempted with poor results, and neither achieved the level of correspondence obtained by trial 1. The differences in overall classification accuracy for both comparisons H and I were significant at the 95% confidence level. The results of these classifications indicate that they are discriminating between topographic classes of lowland and upland as opposed to wetland and upland based on the premise that most of the training data for the wetland and upland classes falls within a certain elevation or slope range. These results would suggest that the use of elevation information alone may be preferable to the use of slope

information alone or the use of slope, aspect and elevation combined, if more training data were available for the wetland class.

Table 5.23 Comparative Statistics: Alternate Neural Network Configurations.

	Comparison H		*Comparison I*	
Trial	1	18	1	20
overall	0.8345	0.7724	0.8345	0.8007
comparative z statistic P_C	**4.7216**		**2.5632**	
kappa	-0.01044	0.02008	-0.01044	0.02434
individual z statistic	-0.12921	0.24636	-0.12921	0.29936
comparative z statistic K	**1.2614**		**1.2259**	
modified kappa	0.6690	0.5445	0.6690	0.6014

5.7.2 Non-Vegetated

There are three potential classes for consideration: non-vegetated land wetland (NVLW), non-vegetated land upland (NVLU) and non-vegetated land alpine (NVLA). Once again, there were only sufficient reference samples available to assess the correspondence of the non-vegetated upland class and as the error matrices are dominated by this class, the measures of overall correspondence must be viewed critically. A total of twenty-one trials were processed at this level and these trials are summarized in Appendix D. A summary of the descriptive statistics for these trials is included in Table 5.24. The mean value P_c was 47.88% with standard deviation of 21.63%. Overall, the results were poorer for level 3 NV in comparison to level 3 VT.

Table 5.24 Level 3NV Descriptive Statistics for All Trials.

DESCRIPTIVE STATISTICS			
Mean	47.883	Range	66.668
Standard Error	4.656	Minimum	22.220
Median	44.444	Maximum	88.888
Mode	50.000	Sum	1005.552
Standard Deviation	21.334	Count	21.000
Sample Variance	455.151	Confidence Level (95.0%)	9.711

The summary statistics for P_P and P_U for the upland class are included in Table 5.25. Confusion with the wetland class is demonstrated by the difference between the mean values for P_P at 87.74% and P_U at 44.59%.

Table 5.25 Level 3 Descriptive Statistics, Producers' and Consumers' Correspondence Measures for All Trials.

Descriptive Statistics	Producer (NVLU)	Consumer (NVLU)
Mean	87.744	44.872
Standard Error	4.633	5.613
Median	91.300	41.177
Mode	100.000	23.500
Standard Deviation	21.231	25.720
Sample Variance	450.741	661.519
Kurtosis	16.253	-0.674
Skewness	-3.829	0.275
Range	100.000	94.100
Minimum	0.000	0.000
Maximum	100.000	94.100
Sum	1842.615	942.309
Count	21.000	21.000
Confidence Level (95.0%)	9.664	11.708

Comparisons were made as for the previous levels of the BCLCCS considered thus far. In comparison A, there is a significant difference in the measures of overall correspondence between trial 1 (corrected) and trial 6 (uncorrected) at the 95% confidence level, with trial 1 having a higher correspondence for the upland class – contributing to its higher level of overall correspondence. Visual inspection would indicate that trial 6 is more successful at identifying alpine areas than the corrected image,

suggesting that the atmospheric correction may have removed some of the contrast for pixels in this class, for example in areas of snow and ice. Neither of these matrices are significant individually. The low level of overall correspondence achieved by the MLC in trials 5 and 21 is verified by visual inspection and appears to be the result of confusion between the wetland and upland classes. Ice and snow clearly make the alpine region distinct and spectrally discernable for all of the classifiers. However, the spectral extent of the alpine is less than the spatial extent of the training data, which is based more on an elevational classification. These trials indicate that the ANN was able to accommodate the spectral complexity of the alpine class and also to discriminate more effectively between wetland and upland areas.

Table 5.26 Comparative Statistics: Corrected vs. Uncorrected and ANN vs. MLC.

	Comparison A		***Comparison B***		***Comparison C***	
Structure	Corrected	Uncorrected	ANN	MLC	ANN	MLC
Trial	1	6	1	5	1	21
Overall	0.500000	0.444400	0.500000	0.22200	0.50000	0.27778
comparative z statistic P_C	**3.1573**		**15.855431**		**12.6597**	
Kappa	0.060760	0.090910	0.060760	0.011760	0.060760	0.016807
Individual z statistic	0.467390	0.280130	0.467390	0.028790	0.467390	0.041850
comparative z statistic K	**-0.0862**		**0.1143**		**-0.1041**	
modified kappa	0.2500	0.1667	0.2500	-0.1667	0.2500	-0.0830

Trials 9 and 12 involved the use of the 6-24-6-2 structure. Trial 9 used the selected set of input bands while trial 12 included the addition of slope, aspect and elevation. The differences in overall classification accuracy in both comparisons D and E are significant at the 95% confidence level. Visually, both trial 9 and particularly trial 12, appear to have very poor correspondence with the training data. Trial 9 has more

confusion between all of the classes and no alpine data has been discriminated in trial 12. The four layer ANN is clearly inappropriate for this classification.

Table 5.27 Comparative Statistics: Alternate Neural Network Configurations.

	Comparison D		*Comparison E*	
Trial	1	9	1	12
overall	0.500000	0.5556	0.500000	0.4444
comparative z statistic P_C	**-3.157362**		**3.157362**	
kappa	0.060760	-0.082710	0.060760	-0.052630
individual z statistic	0.467390	-0.238420	0.467390	-0.118260
comparative z statistic K	**0.3873**		**0.2446**	
modified kappa	0.2500	0.3333	0.2500	0.1667

Trials 14 and 16 are developed from experiments with network architecture at level one and from Kanellopoulos and Wilkinson (1997) heuristic (see Section 4.8.1.1). These trials have an architecture where the number of hidden units are identical to the number of input units. Trial 14 involves the use of the six original Thematic Mapper Bands while trial 16 also included slope, aspect and elevation information. Visual inspections indicate that both trials are very poor.

Table 5.28. Comparative Statistics: Alternate Neural Network Configurations.

	Comparison F		*Comparison G*	
Trial	1	14	1	16
overall	0.500000	0.3888	0.500000	0.3333
comparative z statistic P_C	**6.314161**		**9.482353**	
Kappa	0.060760	-0.081970	0.060760	-0.053660
Individual z statistic	0.467390	-0.222330	0.467390	-0.117050
comparative z statistic K	**0.3651**		**0.2401**	
modified kappa	0.2500	0.0833	0.2500	0.0000

Further additional trials which used the slope (trial 17) and elevation (trial 20) bands separately in combination with the original Thematic Mapper bands, as for level 3 VT, were also attempted with consistently fair results. The differences in P_C for both comparisons H and I were significant at the 95% confidence level. Trial 20 which included only the elevation information, is superior in appearance to trial 17 and indeed superior to most of the other classification trials that were performed at this level. There is however some confusion remaining between the wetland and the upland class and between the upland and the alpine class.

Table 5.29 Comparative Statistics: Alternate Neural Network Configurations.

	Comparison H		***Comparison I***	
Trial	1	17	1	20
overall	0.500000	0.72222	0.500000	0.77770
comparative z statistic P_C	**-12.657980**		**-15.853700**	
Kappa	0.060760	-0.071430	0.060760	0.142860
individual z statistic	0.467390	-0.253400	0.467390	0.586400
comparative z statistic K	**0.4258**		**-0.2973**	
modified kappa	0.2500	0.5833	0.2500	0.6667

5.7.3 Summary

The results of the level 3 classification trials may be summarized as follows:

- Mean P_c VT= 78.08%, NV= 47.88%;
- Trial 9 was the only trial with a significant individual z-statistic.
- Classification results for the corrected image had a higher P_c for both VT and NV.
- The MLC produced poor results for both VT and NV and masking did not improve these results significantly.

- The trials which used the Kanellopoulos and Wilkinson (1997) heuristic and the Paola (1994) heuristic for both VT and NV with the original bands, the selected bands or with the addition of the elevation, slope and aspect information gave poor results.
- The trials which used the Paola (1994) heuristic with the single addition of elevation provided the best results. In the case of VT, these results may have been improved if more training data had been available for the wetland class. In the case of NV, the trial developed with the Kanellopoulos and Wilkinson (1997) heuristic provided the superior result.

5.8 Level 4: Vegetation Type (Coniferous, Broadleaf and Mixed)

Vegetation type is considered in three broad classes: coniferous, broadleaf and mixed. Membership in each of these classes is a product of the proportion of species it contains. The neural networks and MLC were expected to perform better at level four because the level four classes are arguably more spectral in their nature than the landscape position classes of level three. In addition, the separation of coniferous and broadleaf classes has been well documented in the literature.

5.8.1 Vegetated Treed Upland

Twelve classification trials were processed for level 4 VTU, as presented in Appendix D. There were only sufficient reference samples to adequately test the correspondence of the coniferous (VTUCO) and mixed (VTUM) classes. A summary of the descriptive statistics for P_c are included in Table 5.30. The mean value for P_c was 41.65% with a standard deviation of 18.97% .

Table 5.30 Level 4 VTU Descriptive Statistics for All Trials.

DESCRIPTIVE STATISTICS			
Mean	41.654	Range	47.619
Standard Error	5.476	Minimum	24.725
Median	34.798	Maximum	72.344
Mode	72.344	Sum	499.846
Standard Deviation	18.969	Count	12
Sample Variance	359.858	Confidence Level (95.0%)	12.053

A summary of the descriptive statistics for P_P and P_U for VTUCO and VTUM are included in Table 5.31. The mean values of P_P for VTUCO were 18.37% and 73.77% for VTUM. On the contrary, the mean values for P_U were 36.06% for VTUCO and 43.98% for VTUM.

Table 5.31 Level 3NV Descriptive Statistics, Producers' and Consumers' Correspondence Measures for All Trials.

Descriptive Statistics	Producer (VTUCO)	Producer (VTUM)	Consumer (VTUCO)	Consumer (VTUM)
Mean	18.374	73.768	36.057	43.988
Standard Error	3.257	1.131	7.078	9.982
Median	24.845	72.345	42.41	29.78
Mode	0	72.34	0	100
Standard Deviation	11.284	3.917	24.519	34.578
Sample Variance	127.334	15.342	601.189	1195.647
Range	27.2	13.36	61.3	86.84
Minimum	0	67.89	0	13.16
Maximum	27.2	81.25	61.3	100
Sum	220.487	885.217	432.68	527.857
Count	12	12	12	12
Confidence Level (95.0%)	7.169	2.489	15.578	21.969

As for previous levels of the BCLCCS, comparisons were made between corrected, uncorrected and MLC results. The poor performance of all of the trials generated at level four was surprising. Poor overall correspondence is undoubtedly linked to the difficulties in trying to separate the mixed and coniferous classes in which membership for the BCLCCS is established by arbitrary percentage thresholds. The

differences in comparison A between the corrected and uncorrected imagery were not significant at the 95% confidence level, however a comparison of the producer's and user's measures for the two trials indicates that there are some fundamental differences between the two classifications. For example, the level of the user's correspondence for the coniferous class is 54.5% for the trial 1 and only 27.3% for trial 6. This may be explained by the smaller amount of mixed class found in trial 1 (thereby facilitating greater correspondence for the coniferous class by default).

Table 5.32 Comparative Statistics: Corrected vs. Uncorrected and ANN vs. MLC.

	Comparison A		***Comparison B***		***Comparison C***	
Structure	Corrected	Uncorrected	Neural Network	MLC	Neural Network	MLC
trial	1	6	1	5	1	12
overall	0.3113	0.2989	0.3114	0.3699	0.3114	0.3717
comparative z statistic P_C	**0.7651**		**-3.5927**		**-3.7048**	
kappa	0.00671	-0.01864	0.00671	-0.01356	0.00671	-0.00990
individual z statistic	0.12906	-0.3929	0.12906	-0.22362	0.12906	-0.16307
comparative z statistic K	**0.3601**		**-0.2078**		**-0.2538**	
modified kappa	-0.0330	-0.0517	-0.0330	-0.4698	-0.0330	0.0577

The differences between both of the MLC trials and trial 1 are significant at the 95% confidence level. Upon visual inspection the maximum likelihood classification is fundamentally different from the two neural network classifications. Interestingly, it has placed mixed stands on the opposite aspect to where the neural networks have placed mixed stands. Shadow and light seem to have the strongest impact on trials 5 and 12. All of these three trials have more broadleaf than is indicated by the training data, and visual inspection indicates that all correspond poorly to the training data.

Conceptually, it is possible to rationalize why these image based classifications do not correspond well with the training data. A photointerpreter examines discrete polygons and defines their attribute composition for the type and percent composition of each species. The classification algorithms considered in these trials are operating pixel by pixel and attempting to distinguish essentially between two pure classes (coniferous and broadleaf) and then between a mixture of these two classes. The vegetation inventory defines these pure stands as being greater than seventy-five percent covered by either coniferous or broadleaf or else they are mixed. It is fairly apparent that very few classification techniques will be able to interpret that percentage threshold very effectively.

Trial 9 used a 6-24-6-2 network architecture and although this trial appears to have a high level of correspondence with the training data (based on its P_C), visual inspection indicates that the neural network used in this trial has essentially classified the entire area as mixed, thereby contributing to the misleadingly high level of overall correspondence. Trial 11 involved the use of a 6-6-2 architecture with the six TM bands and it has a poor level of correspondence with the training data.

Table 5.33 Comparative Statistics: Alternate Neural Network Configurations.

	Comparison C		*Comparison D*	
Trial	1	9	1	11
overall	0.3114	0.7234	0.3115	0.2582
comparative z statistic P_C	**-25.322214**		**3.268032**	
kappa	0.00671	0.0	0.00671	0.00811
individual z statistic	0.12906	0.0	0.12906	0.15388
comparative z statistic K	**0.0682**		**-0.0188**	
modified kappa	-0.0330	0.5852	-0.0330	-0.1126

5.8.2 Summary

The results of the level 4 classification trials may be summarized as follows:

- Mean P_c was 38.72% with a standard deviation of 22.17%.
- None of the classification trials had significant individual z-statistics.
- The differences between the corrected and uncorrected images were not significant.
- The MLC performed significantly better in comparison to the original neural network.
- The trials using the alternate 6-24-6-2 and 6-6-2 structure produced very poor results.

5.9 Level 5: Density Classes (Dense, Open and Sparse)

The final level of the BCLCCS considered in this study is that of density classes. Percentage thresholds for crown closure are used by photointerpreters to assess which density classes are appropriate for each polygon. There are three possible density classes: dense, open and sparse. There were sufficient training data to assess the correspondence of all three classes for the vegetated treed upland coniferous (VTUCO) areas and the vegetated treed upland mixed (VTUM) areas.

5.9.1 Vegetated Treed Upland Coniferous (VTUCO)

Twelve classification trials were processed for level 5 VTUCO and they are presented in Appendix D. There were sufficient test points to adequately test the dense and open density classes at this level but not the sparse class. A summary of the descriptive statistics for P_c are included in Table 5.34. The mean P_c was 30.22% with a range of 6% to 54% and a standard deviation of 16.27%.

Table 5.34 Level 5 VTUCO Descriptive Statistics for All Trials.

DESCRIPTIVE STATISTICS			
Mean	30.221	Range	48.000
Standard Error	4.698	Minimum	6.000
Median	34.000	Maximum	54.000
Mode	34.000	Sum	362.648
Standard Deviation	16.274	Count	12
Sample Variance	264.842	Confidence Level (95.0%)	10.340

A summary of the descriptive statistics for P_U and P_P for the dense (VTUCODE), open (VTUCOOP) and sparse (VTUCOSP) are included in Table 5.35. The mean P_P for the dense class was only 36.17% compared to 51.24% for the open class. Conversely, the mean P_U was 48.68% for the dense class compared to only 18.45% for the open class.

Table 5.35 Level 5 VTUCO Descriptive Statistics, Producers' and Consumers' Correspondence Measures for All Trials.

Descriptive Statistics	Producer (VTUCODE)	Producer (VTUCOOP)	Producer (VTUCOSP)	Consumer (VTUCODE)	Consumer (VTUCOOP)	Consumer (VTUCOSP)
Mean	36.176	51.240	10.663	48.680	18.466	36.748
Standard Error	7.865	11.424	4.532	12.596	9.877	11.773
Median	40.295	52.655	5.940	63.155	3.571	33.315
Mode	0.000	100.000	0.000	0.000	0.000	0.000
Standard Deviation	27.246	39.575	15.699	43.635	34.216	40.783
Sample Variance	742.363	1566.144	246.468	1904.023	1170.745	1663.219
Range	100.000	100.000	50.000	94.736	92.850	100.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	100.000	100.000	50.000	94.736	92.850	100.000
Sum	434.108	614.880	127.960	584.155	221.592	440.980
Count	12.000	12.000	12.000	12.000	12.000	12.000
Confidence Level (95.0%)	17.312	25.144	9.975	27.724	21.740	25.912

Comparisons were made between classifications performed on the corrected and uncorrected imagery (comparison A) and between the ANN and the MLC classifications (comparisons B and C). The differences in overall classification accuracy found in all three of these comparisons were significant at the 95% confidence level. A visual comparison indicates that trial 6 has a closer correspondence with the training data and a

higher P_C., although it does have a lower kappa value. The classification results of trial 1 are more heterogeneous in appearance than trial 6 and although both of the MLC trials have a higher P_C than the ANN, their classifications appear inferior in overall appearance than either of the neural networks (trials 1 and 6), relative to the inventory training data. The predominance of the dense category in this MLC classification guarantees a high level of correspondence for the dense category and therefore a high level of correspondence overall.

Table 5.36 Comparative Statistics: Corrected vs. Uncorrected and ANN vs. MLC.

	Comparison A		***Comparison B***		***Comparison C***	
Structure	Corrected	Uncorrected	Neural Network	MLC	Neural Network	MLC
trial	1	6	1	5	1	12
overall	0.3400	0.3958	0.3400	0.4400	0.3400	0.3932
comparative z statistic P_C	**-3.3451**		**-5.9885**		**-3.594760**	
kappa	0.085370	0.046580	0.085370	0.069770	0.085370	0.056010
individual z statistic	0.592570	0.255470	0.592570	0.318040	0.592570	0.249860
comparative z statistic K	**0.1669**		**0.0594**		**0.1102**	
modified kappa	0.0100	0.0937	0.0100	0.3100	0.0100	0.1000

Additional comparisons were made using a 6-24-9-3 network structure (trial 9) and a 6-6-3 structure with the raw TM bands (trial 11). The poor level of correspondence for trial 9 resulted from the majority of the image area being classified as sparse. Trial 11, which was developed with heuristics from Kanellopoulos and Wilkinson (1997) and the experimental trials generated at level one of the BCLCCS, has produced a higher level of correspondence than trial 1 due to the predominance of the dense class as well, however visually, the classification is inferior.

Table 5.37 Comparative Statistics: Alternate Neural Network Configurations.

	Comparison C		*Comparison D*	
Trial	1	9	1	11
overall	0.3400	0.0706	0.3400	0.4200
comparative z statistic P_C	**16.3421**		**-7.1851**	
kappa	0.085370	0.00	0.085370	0.121740
Individual Z statistic	0.592570	0.00	0.592570	0.729180
comparative z statistic K	**0.4659**		**-0.1650**	
modified kappa	0.0100	-0.3940	0.0100	0.1300

5.9.2 Vegetated Treed Upland Mixed (VTUM)

Twelve classification trials were processed for level 5 VTUM as outlined in Appendix D. There were sufficient test points available at this level to adequately test the correspondence of the dense, open and sparse classes. A summary of the descriptive statistics for P_c are included in Table 5.38. The mean P_c was 36.41%, which was higher than for VTUCO and the standard deviation was 16.95%. The values for overall correspondence ranged from a low of 16.95% to a high of 77.5%.

Table 5.38 Level 5 VTUCO Descriptive Statistics for All Trials.

DESCRIPTIVE STATISTICS			
Mean	36.414	Range	60.569
Standard Error	4.705	Minimum	16.949
Median	38.136	Maximum	77.518
Mode	16.949	Sum	436.970
Standard Deviation	16.299	Count	12.000
Sample Variance	265.669	Confidence Level (95.0%)	10.356

A summary of the descriptive statistics for P_P and P_U are included in Table 3.9. The mean values for P_P for the dense and open classes were 37.53% and 39.77%

respectively, while the sparse class was a low 16.68%. Conversely, the mean P_U for the sparse class was 31.36% and 28.04% for the dense and open classes respectfully.

Table 5.39 Level 5 VTUM Descriptive Statistics, Producers' and Consumers' Correspondence Measures for All Trials.

Descriptive Statistics	Producer (VTUMDE)	Producer (VTUMOP)	Producer (VTUMSP)	Consumer (VTUMDE)	Consumer (VTUMOP)	Consumer (VTUMSP)
Mean	37.528	39.771	16.680	31.364	28.043	49.000
Standard Error	8.093	8.304	2.666	8.063	8.447	10.155
Median	42.830	47.220	17.665	39.015	25.405	39.000
Mode	0.000	0.000	16.940	0.000	0.000	100.000
Standard Deviation	28.033	28.764	9.235	27.930	29.262	35.177
Sample Variance	785.866	827.388	85.290	780.086	856.276	1237.455
Range	100.000	99.300	29.300	75.600	90.160	100.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	100.000	99.300	29.300	75.600	90.160	100.000
Sum	450.340	477.250	200.160	376.370	336.520	588.000
Count	12.000	12.000	12.000	12.000	12.000	12.000
Confidence Level (95.0%)	17.812	18.276	5.868	17.746	18.592	22.351

Comparisons between trials were made as for the previous BCLCCS levels considered thus far. For comparison A, there was a significant difference between the corrected and uncorrected image classifications, with the latter having a higher overall level of correspondence, although it does have a lower kappa value. In comparison B, the difference between the neural network and MLC trial is significant at the 95% confidence level, with the MLC having a larger P_c and a much lower khat and also a lower modified kappa statistic. Trial 5 is predominantly covered by the dense class and obviously has been unable to discriminate between open and dense classes. The MLC appears poorer visually than either of the neural network classifications. Trial 12 was generated with agricultural and urban development masked out, however this produced an inferior result to trial 5 and is not significantly different from trial 1 at the 95% confidence level.

Table 5.40 Comparative Statistics: Corrected vs. Uncorrected and ANN vs. MLC.

	Comparison A		***Comparison B***		***Comparison C***	
Structure	Corrected	Uncorrected	Neural Network	MLC	Neural Network	MLC
Trial	1	6	1	5	1	12
overall	0.4203	0.7751	0.4203	0.4842	0.4203	0.3932
comparative z statistic P_C	**-20.5118**		**-3.6747**		**1.5583**	
kappa	0.09075	0.06601	0.09075	-0.01416	0.09075	0.03977
individual z statistic	1.85220	1.24796	1.85220	-0.11901	1.85220	0.62910
comparative z statistic K	**0.3431**		**0.8153**		**0.6374**	
modified kappa	0.1305	0.6628	0.1305	0.1203	0.1305	0.0898

A further comparison was made between a 6-24-9-3 network developed from the Kanellopoulos and Wilkinson (1997) heuristic (trial 9) and the original neural network. As was the case for VTUCO, this trial is significantly inferior to the original neural network architecture, having identified the entire area as sparse. Comparison D examined a 6-6-3 network and it was found that this difference was not significant at the 95% confidence level.

Table 5.41 Comparative Statistics: Alternate Neural Network Configurations.

	Comparison C		*Comparison D*	
Trial	1	9	1	11
overall	0.4203	0.1694	0.4203	0.3966
comparative z statistic P_C	**14.5304**		**1.3631**	
kappa	0.090750	0.0	0.090750	-0.001070
individual z statistic	1.852200	0.0	1.852200	-0.010980
comparative z statistic K	**0.8874**		**0.8416**	
modified kappa	0.1305	-0.2458	0.1305	0.0054

5.9.3 Summary

The results of the level 5 classification trials may be summarized as follows:

- Mean P_c VTUCO = 30.22%, VTUM = 36.41%;
- Trial 1 (the original ANN configuration) was the only trial which had a significant individual z-statistic.
- Classification results for the uncorrected image had a higher value for P_c for both VTUCO and VTUM, however both had lower kappa values and a poorer overall visual appearance.
- The MLC performed better in comparison to the original neural network architecture for both VTUCO and VTUM.
- The 6-24-9-3 structure produces a very poor result for both VTUCO and VTUM.
- The 6-6-3 structure has a lower P_c than the original neural network structure but the difference is not significant at the 95% confidence level.

5.10 Summary of Results

While it is difficult to generalize the range of results achieved by the classification trials produced for this project, a number of summary statements may be made:

- The architecture of the neural network is extremely important to the success of classification trials. Certain specific parameters are crucial:

    1. Three layer networks are more suitable for these simple classification schemes;

    2. Band enhancements are not necessarily more suitable than the basic TM bands;

    3. Hidden layers should be the same size or similar in size to the input layer (Kannellopoulos and Wilkinson, 1997);

    4. Conservative learning and momentum rates are preferable, as are moderate numbers of training samples. Other researchers have demonstrated that neural networks perform well with limited training data (something that the MLC does not do).

    5. The addition of ancillary data is simple and beneficial when using an ANN, where warranted by the nature of the classes. Ancillary data however may also reduce classification correspondence by introducing confusion and noise. Ancillary data should be used with caution.

- The MLC performs poorly compared to the results of the neural network. The MLC is strongly influenced by topography and the presence of shadows. The mutually exclusive training property of the neural network allows the ANN to produce superior classification results.
- The difference between classifications performed on the corrected imagery versus those performed on the uncorrected imagery were rarely significant. This is important, because the geometric correction is useful for ensuring that all of the project data is georeferenced to the same projection. This helps to facilitate the

process of selecting training data and referencing other ancillary data. The process of atmospheric correction and conversion of DNs to reflectance serves to normalize the response of various cover types regardless of image date or atmospheric phenomena.

- The ability of the neural network to adequately discriminate the BCLCCS classes varied with the level under consideration. Certainly this study indicates that there is a strong potential for the neural network method to be successful, but a more rigorous assessment of all of the considered classes would be required before the utility of the neural network could be reasonably assessed. The information content of the Landsat Thematic Mapper relevant to the BCLCCS classes considered, as evaluated by the classification results, would suggest that the original hypothesis put forward in Chapter 1 has been conditionally validated. Other methods *may* have been more successful at achieving consistent results (e.g. the use of phenological information from multi-date image sequences), although given the topography, it may have been difficult to consistently achieve acceptable results at level four, regardless of the classification method used.
- It must be reiterated that the results of this study provide an indication of the degree to which the classifications correspond to the generalized forest inventory attributes. It provides no indication of how well the classifications correspond to the reality on the ground. This is important because at some of the levels of the BCLCCS it was found that the classification was actually detecting areas in the imagery with a particular attribute that had not yet been incorporated into the forest inventory (i.e. recent harvesting). In addition, while it may appear that the classification corresponds poorly to the homogenous classes of the generalized forest inventory, in reality, the

vegetation cover is much more heterogeneous than the forest inventory would suggest and the spectral information in the Landsat TM image reflects that. This further reinforces the requirement for an independent ground truth data set to enable a rigorous assessment of the correspondence of these classes.

CHAPTER SIX

CONCLUSION

Remote sensing image classification, particularly as it pertains to vegetation and land cover, is an extremely broad subject area. There are an endless number of classification tools, methods and strategies which have been applied, somewhat haphazardly, to an equally endless number of geographic locations, as the literature review in Chapter 2 has demonstrated. Research projects are often conceived as "proof of concept" studies, much like the research conducted for this thesis. The conclusions of these studies often suggest numerous possibilities for further research, however follow-up to these studies is uncommon and systematic research in image classification is rarely conducted.

Neural networks have become an extremely popular subject of remote sensing research in this decade. Several synoptic papers in this field have recently been published along with heuristics for successful neural network implementation (Atkinson and Tatnall, 1997; Kanellopoulos and Wilkinson, 1997; Paola and Schowengerdt, 1997). Neural networks are being implemented in commercial image processing packages and improvements to the algorithms such as adaptive learning and momentum rates are reducing some of the trial and error associated with the development of appropriate neural network architectures for specific applications. Artificial neural networks have proven to be powerful classification techniques, largely as a result of their inherent flexibility. The disadvantages associated with neural networks, as outlined in Chapter 3, will undoubtedly encumber their use beyond a research context.

Table 5.42 and Figure 5.2 summarizes the overall measure of correspondence for the classification trials at each of the five levels. The overall mean values include both the ANN and MLC trials and then the results for each method are reported separately.

Table 6.1 A Summary of Classification Results.

	ANN		MLC	
	MEAN	STD. DEV.	MEAN	STD. DEV.
Level 1	79.85	12.03	81.78	0.01
Level 2	72.71	7.82	40.51	2.05
Level 3 VT	79.62	16.59	35.32	0.63
Level 3 NV	50.29	20.99	25	2.78
Level 4	42.57	20.84	37.09	0.13
Level 5 VTUCO	26.86	15.42	43	1.41
Level 5 VTUM	35.63	17.9	40.33	1.44

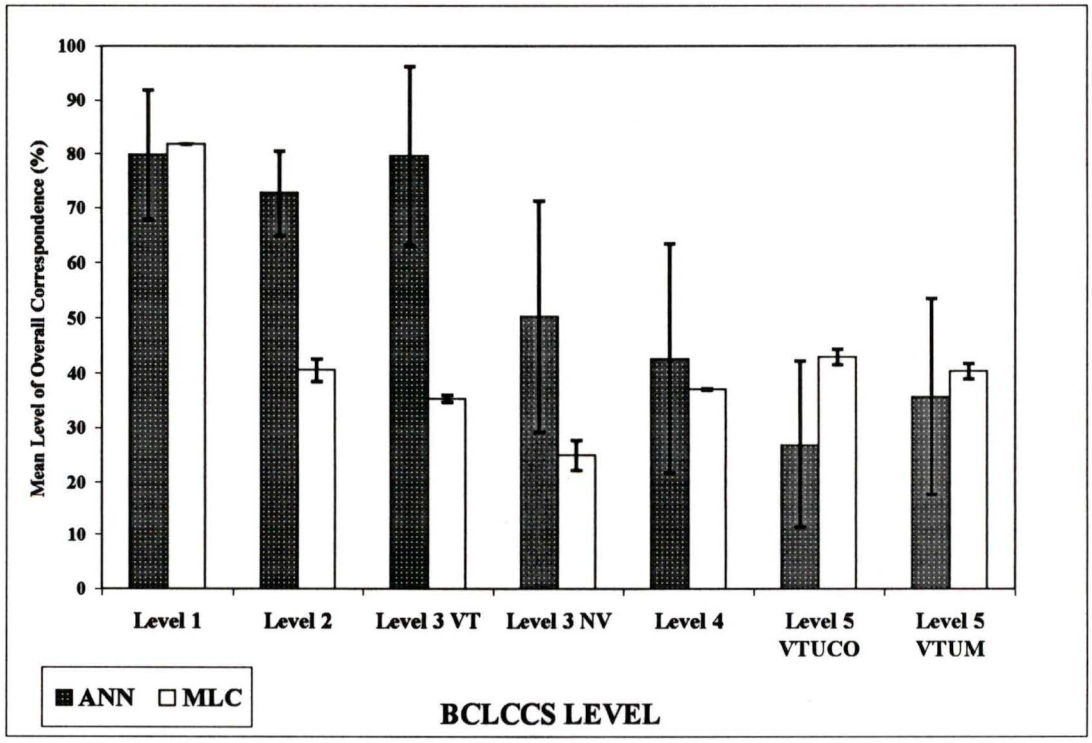


Figure 6.1 Mean Values for Overall Measures of Correspondence.

The first objective of this thesis was to assess the information content of the Landsat Thematic Mapper imagery relevant to the BCLCCS. The results of the classification trials would suggest that this information is limited, with consistent results achievable only to level 3 of the BCLCCS hierarchy (for vegetated treed polygons). This conclusion is not suprising and supports the hypothesis put forward in Chapter 1. This conclusion is not definitive however because the data used in the correspondence assessment is not an appropriate sample on which to unequivocally base such a conclusion. A study site with a more equal distribution of wetland, upland and alpine areas for level 3 and more broadleaf, coniferous and mixed areas for level 4 would be a more appropriate setting to further test this hypothesis. Given past research with Landsat TM however, it is not entirely unexpected that distinguishing mixed and coniferous stands was difficult, and as was suggested in Chapter 5, a greater use of the phenological information in multi-date imagery may be beneficial in a site with more equal distributions of the various classes. Only a small portion of the BCLCCS was considered in this study. Other vegetation types, such as shrubs, would probably be even more difficult, if not impossible, to characterize with Landsat TM data, particularly if they are concealed beneath the forest canopy. The practical limitations of Landsat TM for the vegetation inventory are therefore acknowledged. As was suggested in Chapter 1, TM data is a useful synoptic tool and its most pragmatic application will be in the context of strategic level planning and the study of landscape-level processes.

The second objective of this thesis was to compare the performance of an artificial neural network against that of a maximum likelihood classifier. It was hypothesized in Chapter 1 that given the complexity of the vegetation in the study area,

coupled with the varying topography that the ANN would be able to outperform the MLC in all of the classification trials. Figure 5.2 indicates that this is not the case at level five where the MLC performed better than the ANN in the discrimination of density classes. At levels 1, 2 and 3VT, the ANN produced superior results to the MLC. The most interesting difference between the two classification methods was that the neural network classifications were not influenced by the presence of shadow in the image. Although a qualitative assessment is not valid, the ANN classifications were often visually more logical than the MLC results. Due to the skewed distribution of the test data, the overall measures of correspondence are often distorted by the dominance of a particularly large class which is correctly identified, and hence the importance of using multiple measures of correspondence to provide a more complete assessment of the classification results.

A secondary objective of the thesis was to compare the heuristics developed by Paola (1994) and by Kanellopoulos and Wilkinson (1997). It was demonstrated in the classification trials performed in this research that the architecture of the neural network is extremely important in determining the classification results. The opacity of neural networks however, makes it extremely difficult to determine the reasons for the relative successes or failures of the classification trials. The heuristic developed by Kannellopoulos and Wilkinson (1997) regarding hidden layer size was more effective for the neural networks used in this study area.

Some consideration must be given to the correlation between the informational attributes from the BCLCCS and the spectral information in the TM image. If the informational classes fail to be strongly correlated to the spectral information in the image, no classification method, regardless of its sophistication, will produce meaningful

results. (Duggin and Robinove, 1990). If such a correlation does not exist a different source of image data must be used or conversely, the informational classes may be restructured to correspond to the spectral information in the image.

Overall, the conclusions of this research are not unique. New classification techniques are often over-sold to the remote sensing community and this is not the first study to suggest that artificial neural networks, as a tool for remote sensing image classification, should be regarded with some degree of skepticism (Skidmore et al., 1997). The use of artificial neural networks as classification tools is most appropriate where the assumptions associated with conventional methods are seriously violated, such as when the distribution of the image data is not normal or the number of training samples is severely limited relative to the number of image bands to be used for the classification. In these situations, the neural network will almost assuredly produce superior results and the time required to configure and train the network will be well afforded by the analyst. As was previously stated, on-going research in the development of artificial neural networks in other disciplines will undoubtedly influence the continued use of these techniques for remote sensing image classification.

LITERATURE CITED

Adams, J.B., Sabol, D.E., Kapos, V., Filho, R.A., Roberts, D.A., Smith, M.O., and Gillespie, A.R. 1995. Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon. *Remote Sensing of Environment*, 52:137-154.

Ahern, F.J., and Archibald, P.D. 1986. Thematic Mapper information about Canadian forests: Early results across the country. In D. Thompson and R. Brown (Eds.), *Tenth Canadian Symposium on Remote Sensing.* 1: 683-691.

Alonso, F.G., and Soria, S.L. 1991. Using contextual information to improve land use classification of satellite images in central Spain. *International Journal of Remote Sensing*, 12(11):2227-2235.

Apan, A.A. 1997. Land cover mapping for tropical forest rehabilitation planning using remotely-sensed data. *International Journal of Remote Sensing*, 18(5):1029-1049.

Arai, K. 1997. Multi-temporal texture analysis in TM classification. *Canadian Journal of Remote Sensing*, 17(3):263-270.

Argialas, D.P. and C.A. Harlow, 1990. Computational image processing models: an overview of perspectives. *Photogrammetric Engineering and Remote Sensing*, 56(6):871-886.

Aronoff, S. 1982. Classification accuracy: A user approach. *Photogrammetric Engineering and Remote Sensing*, 48(8):1299-1307.

Atkinson, P.M., and Tatnall, A.R.L. 1997. Neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4):699-709.

Avery, T.E. and Berlin, G.L. 1992. *Fundamentals of Remote Sensing and Airphoto Interpretation, 5th Edition.* Macmillan Publishing. New York, NY.

Baker, W.L., Honaker, J.J., and Weisberg, P.J. 1995. Using aerial photography and GIS to map the forest-tundra ecotone in Rocky Mountain National Park, Colorado, for global change research. *Photogrammetric Engineering and Remote Sensing*, 61(3):313-320.

Bancroft, B. 1995. *The B.C. Land Cover Classification: Background and Definitions.* Resources Inventory Committee, Victoria, B.C.

Baraldi, A. and Parmiggiani, F. 1995. A Neural Network for Unsupervised Categorization of Multivalued Input Patterns: An Application to Satellite Image Clustering. *IEEE Transactions of Geosciences and Remote Sensing.* 33(2):305-316

Barker, J. 1983. Relative radiometric calibration of TM reflective bands, *Landsat 4 Science Characterization Early Results*, NASA Conference Publication 2355, 3: 1-29.

Bauer, M.E., Burk, T.E., Ek, A.R., Coppin, P.R., Lime, S.D., Walsh, T.A., Walters, D.K., Befort, W., and Heinzen, D.F. 1994. Satellite inventory of Minnesota forest resources. *Photogrammetric Engineering and Remote Sensing*, 60(3):287-298.

Beaubien, J., 1994. Landsat TM satellite images of forests: from enhancement to classification. *Canadian Journal of Remote Sensing*, 20(1): 17-26.

Befort, W. 1988. Controlled-scale aerial sampling photography: Development and implications for multi-resource inventory. *Journal of Forestry*, 86(11):21-28.

Belward, A.S., Taylor, J.C., Stuttard, M.J., Bignal, E., Mathews, J., and Curtis, D. 1990. An unsupervised approach to the classification of semi-natural vegetation from Lansat Thematic Mapper data: A pilot study on Islay. *International Journal of Remote Sensing*, 11(3):429-445.

Benedetti, R., Rossini, P., and Taddei, R. 1994. Vegetation classification in the Middle Mediterranean area by satellite data. *International Journal of Remote Sensing*, 15(3):583-596.

Benediktsson, J.A., Ersoy, O.K., Swain, P.H., and Lafayette, W. 1991. A consensual neural network. *International Geoscience and Remote Sensing Symposium*, 4:2219-2222.

Benediktsson, J.A., and Sveinsson, J.R. 1997. Feature extraction for multisource data classification with artifical neural networks. *International Journal of Remote Sensing*, 18(2):727-740.

Benediktsson, J.A., Swain, P.H., and Ersoy, O.K. 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):540-552.

Benediktsson, J.A., Swain, P.H., and Ersoy, O.K. 1993. Conjugate-gradient neural networks in classification of multisource and very-high-dimensional remote sensing data. *International Journal of Remote Sensing*, 14(15):2883-2903.

Benson, A.S. and DeGloria, S.D. 1985. Interpretation of Landsat 4 Thematic Mapper and Multispectral Scanner Data for Forest Surveys. *Photogrammetric Engineering and Remote Sensing,* 51(9): 1281-1289.

Bezdek, J.C., Ehrlick, R., and Full, W. 1984. FCM: the fuzzy c-means clustering algorithm. *Computers and Geosciences,* 10: 191-203.

Bhattacharya, U. and Parui, S.K. 1997. An improved backpropagation neural network for detection of road-like features in satellite imagery. *International Journal of Remote Sensing*, 18(16): 3379-3394.

Binaghi, E., Madella, P., Grazia Montesano, M., and Rampini, A. 1997. Fuzzy contextual classification of multisource remot sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 35(2):326-339.

Bischof, H., Schneider, W., and Pinz, A.J. 1992. Multispectral classification of Landsat images using neural networks. *IEEE Transactions on Geoscience and Remote Sensing,* 30(3):482-490.

Bishop, Y.M.M., Feinberg, S.E., Holland, P.W. 1975. *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, MA, 557 p.

Blamire, P.A. 1996. The influence of relative sample size in training artificial neural networks. *International Journal of Remote Sensing*, 17(1):223-230.

Blonda, P., Bennardo, A., la Forgia, V., and Satalino, G. 1995. Modular neural system, based on a fuzzy clustering network, for classification. *International Geoscience and Remote Sensing Symposium*, 1:449-451.

Blonda, P., la Forgia, V., Pasquariello, G., and Satalino, G. 1994. Multispectral classification by a modular neural network architecture. *International Geoscience and Remote Sensing Symposium*, 4:1873-1876.

Blonda, P., Pasquariello, G., Losito, S., Mori, A., Posa, F., and Ragno, D. 1991. An experiment for the interpretation of multitemporal remotely sensed images based on a fuzzy logic approach. *International Journal of Remote Sensing*, 12(3):463-476.

Bolstad, P.V., and Lillesand, T.M. 1992. Improved classification of forest vegetation in Northern Wisconsin through a rule-based combination of soils, terrain, and Landsat Thematic Mapper Data. *Forest Science*, 38(1):5-20.

Bones, J.T, 1993. Report for the Consultancy on Forest Definitions and Classifications to be Employed for Global Forest Resources Assessment, 2000. p. 175-194, in A.

Nyyssonen (ed.) *Proceeding of FAO/ECE Meeting of Experts on Global Forest Resources Assessment in coorperation with UNEP and with the support of FINNIDA*, The Finish Forest Research Institute, Research Papers 469, Helsinki,Finland.

Brennan, R.L. and Prediger, D.J. 1981. Coefficient kappa: some uses, misuses and alternatives. *Education Psychological Measure*. 41: 687-699.

Brondizio, E., Moran, E., Mausel, P.W., and Wu, Y. 1996. Land cover in the amazon estuary: linkin of the Thematic Mapper with botanical and historical data. *Photogrammetric Engineering and Remote Sensing*, 62(8):921-929.

Bruzzone, L., Conese, C., Maselli, F., and Roli, F. 1997. Multisource classification of complex rural areas by statistical and neural-network approaches. *Photogrammetric Engineering and Remote Sensing*, 63(5):523-533.

Bruzzone, L., Roli, F., and Serpico, S.B. 1995. An experimental comparison of neural networks for the classification of multi-sensor remote sensing images. *International Geoscience and Remote Sensing Symposium*, 1:452-454.

Butera, M.K., 1986. A correlation and regression analysis of percent canopy closure versus Thematic Mapper spectral response for selected forest sites in San Juan National Forest, Colorado. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(1):122-129.

Campbell, J.B. 1987. *Introduction to Remote Sensing*, Guilford, New York

Canters, F. 1997. Evaluating the uncertainty of area estimates drived from fuzzy land-cover classification. *Photogrammetric Engineering and Remote Sensing*, 63(4):403-414.

Carbone, G.J., Narumalani, S., and King, M. 1996. Application of remote sensing and GIS technologies with physiological crop models. *Photogrammetric Engineering and Remote Sensing*, 62(2):171-179.

Card, D. 1982. Using known map category margina frequencies to improve estimates of thematic map accuracy. *Photogrammetric Engineering and Remote Sensing*, 48(3):431-439.

Carpenter, G.A., Gjaja, M.N., Gopal, S., and Woodcock, C.E. 1997. ART neural networks for remote sensing: vegetation classification from Landsat TM and terrain data. *IEEE Transactions on Geoscience and Remote Sensing*, 35(2):308-325.

Caudill, M. 1988. Neural networks primer: Part III. *AI Expert*, 53-59.

Chae, H.S., Kim, S.J. and Ryu, J.A. 1997. A classification of multitemporal Landsat TM data using principal component analysis and an artificial neural network. *Proceedings, 17th International Geoscience and Remote Sensing Symposium,* Singapore.

Chavez, P. 1992. Comparison of spatial variability in visible and near-infrared spectral images. *Photogrammetric Engineering and Remote Sensing*, 58(7):957-964.

Chavez, P. 1997. Image based atmospheric corrections - revisited and improved. *Photogrammetric Engineering and Remote Sensing*, 62(9):1025-1036.

Chavez, P., and Kwarteng, A. 1989. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogrammetric Engineering and Remote Sensing*, 55(3):339-348.

Chavez, P., and MacKinnon, D.J. 1994. Automatic detection of vegetation changes in the southwestern United States using remotely sensed images. *Photogrammetric Engineering and Remote Sensing*, 60(5):571-583.

Chen, J.M., and Guilbeault, M.A. 1996. Evaluation of vegetation indicies for retreiving biophysical parameters of boreal conifer forests. *26th International Symposium on Remote Sensing of Environment*, 1:168-171.

Chen, K.S., Kao, W.L., and Tzeng, Y.C. 1995. Retrieval of surface parameters using dynamic learning neural network. *International Journal of Remote Sensing*, 16(5):801-809.

Cheng, P. & Toutin, Th. 1995. High Accuracy Data Fusion of Satellite and Airphoto Images, *1995 ACSM/ASPRS Annual Convention & Exposition Technical Papers*, Charlotte, North Carolina, Feb 27 - Mar 2, 1995.

Cialella, A.T., Dubayah, R., Lawrence, W., and Levine, E. 1997. Predicting soil drainage class using remotely sensed and digital elevation data. *Photogrammetric Engineering and Remote Sensing*, 63(2):171-178.

Civco, D. 1993. Artificial neural networks for land-cover classification and mapping. *International Journal of Geographical Information Systems*, 7(2):173-186.

Civco, D. and Wang, Y. 1994. Classification of multispectral, multitemporal, multisource spatial data using neural networks in *Proceedings of 1994 ASPRS/ACSM Convention,* Reno, Nevada.

Civco, D.L. 1989. Topographic normalization of Landsat Thematic Mapper digital imagery. *Photogrammetric Engineering and Remote Sensing,* 55: 1303-1309.

Clark, C., and Canas, A. 1995. Spectral identification by artifical neural network and genetic algorithm. *International Journal of Remote Sensing*, 16(12):2255-2276.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 20(1): 37-46.

Cohen, W.B. and Spies, T.A. 1992. Estimating attributes of Douglas fir/Western hemlock forest stands from satellite imagery. *Remote Sensing of Environment,* 4(1):1-18.

Cohen, W.B. Fiorella, M., Gray, J., Helmer, E. and Anderson, K. 1998. An efficient and accurate method for mapping forest clearcuts in the Pacific Northwest using Landsat imagery. *Photogrammetric Engineering and Remote Sensing.* 64(4):293-300.

Colby, J.D. 1991. Topographic normalization in rugged terrain. *Photogrammetric Engineering and Remote Sensing*, 57(5):531-537.

Collins, J.B., and Woodcock, C.E. 1994. Change detection using the Gramm-Schmidt Transformation applied to mapping forest mortatlity. *Remote Sensing of Environment*, 50:267-279.

Collins, J.B., and Woodcock, C.E. 1997. An assessment of several linear change detection techniques for mapping forest mortality using multtemporal Landsat TM data. *Remote Sensing of Environment*, 57:161-166.

Colwell, J.E. 1974. Vegetation canopy reflectance, *Remote Sensing Environment*, 3 (175).

Colwell, R.N. 1983. *Manual of Remote Sensing.* Falls Church: American Society of Photogrammetry.

Conese, C., Gilabert, M.A., Maselli, F., and Bottai, L. 1993. Topographic normalization of TM scenes through the use of an atmospheric correction method and digital terrain models. *Photogrammetric Engineering and Remote Sensing*, 59(12):1745-1753.

Conese, C., and Maselli, F. 1993. Selection of optimum bands from TM scenes through mutual information analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 48(3):2-11.

Congalton, R.G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37: 35-46.

Congalton, R.G., and Biging, G.S. 1992. A pilot study evaluating ground reference data collection efforts for use in forest inventory. *Photogrammetric Engineering and Remote Sensing*, 58(12):1669-1671.

Congalton, R.G., and Green, K. 1993. A practical look at the sources of confusion in error matrix generation. *Photogrammetric Engineering and Remote Sensing*, 59(5):641-644.

Congalton, R.G., Green, K., and Teply, J. 1993. Mapping old growth forests on national forest and park lands in the Pacific Northwest from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 59(4):529-535.

Congalton, R.G., and Mead, R.A. 1986. A review of three discrete multivariate analysis techniques used in assessing the accuracy of remotely sensed data from error matrices. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(1):169-174.

Congalton, R.G., Oderwald, R.G. and Mead, R.A. (1983) Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering and Remote Sensing*, 49:1671-1678.

Cortijo, F.J., and Perez De La Blanca, N. 1997. A comparative study of some non-parametric spectral classifiers. Applications to problems with high-overlapping training sets. *International Journal of Remote Sensing*, 18(6):1259-1275.

Côté, S. and Tatnall, A.R.L., 1995. A neural network-based method for tracking feaatures from satellite sensor images. *International Journal of Remote Sensing*, 16: 3695-3701.

Crist, E., and Cicone, R.C. 1984. Application of the tasseled cap concept to simulated Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing*, 50(3):343-352.

Curran, P.J. 1980. Multispectral remote sensing of vegetation amount. *Progress in Physical Geography*, 4: 315-341.

Curran, P.J. 1985. *Principles of Remote Sensing*. Longman Scientific and Technical, New York, NY.

Curtis, P. 1980. Remote sensing systems for inventory crops and vegetation. *Progress in Physical Geography*. 4: 315-341.

Dawson, M.S., Amar, F., Manry, M.T., Rawat, V., and Fung, A.K. 1994. Classification of remote sensing data using fast learning neural networks and topology selection algorithms. *International Geoscience and Remote Sensing Symposium*, 3:1410-1412.

Decatur, S.E. 1989a. Application of neural networks to terrain classification. *Proceedings of IJCNN'89*, 1:283-288.

Decatur, S.E. 1989b. Application of Neural Networks to Terrain Classification, M.S. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Massachusetts.

Deppe, F. 1998. Forest area estimation using sample surveys and Landsat MSS and TM data. *Photogrammetric Engineering and Remote Sensing*, 64(4):285-292.

Dobbertin, M., and Biging, G.S. 1997. A simulation study of the effect of scene autocorrelation, training sample size and sampling method on classification accuracy. *Canadian Journal of Remote Sensing*, 22(4):360-367.

Dreyer, P. 1993. Classification of land cover using optimized neural nets on SPOT data. *Photogrammetric Engineering and Remote Sensing*, 59(5):617-621.

Duggin, M.J., and Robinove, C.J. 1990. Assumptions implicit in remote sensing data acquisition and analysis. *International Journal of Remote Sensing*, 11(10):1009-1094.

Duguay, C.R., and Leverington, D.W. 1996. Classification of alpine tundra soil-plant communities using an artificial neural network. *26th International Symposium on Remote Sensing of Environment*, 1:75-78.

Duguay, C.R., and Peddle, D.R. 1996. Comparison of evidential resoning and neural network approaches in a multi-source classification of alpine tundra vegetation. *Canadian Journal of Remote Sensing*, 22(4):433-440.

Dutra, L.V., and Mascarenchas, N.D.A (1984) Some experiments with spatial feature extraction methods in multi-spectral classification. *International Journal of Remote Sensing*, 5(2):303-313.

Earth Observation Satellite Company. 1994. Landsat system status report--September 1994: Lanham, Md., Earth Observation Satellite Company, p. 1-11.

Earth Observation Satellite Company. 1994. Radiometric Correction. *EOSAT Notes*, 9(2).

Ediriwickrema, J. and Khorram, S. 1997. Hierarchical maximum likelihood classification for improved accuracies. *IEEE Transactions on Geosciences and Remote Sensing*, 35(4): 810-816.

Ekstrand, S. 1996. Landsat TM based forest damage assessment: Correction for topographic effects. *Photogrammetric Engineering and Remote Sensing*, 62(2):151-161.

Ersoy, O.K and Hong, D. 1990. Parallel, self-organizing, hierarchical neural networks. *IEEE Transactions of Neural Networks*, 1:167-178.

Excell, P.S. and Studholme, C. 1990. A neural network for spatial feature recognition. *Proceedings of 16th Annual Conference of the Remote Sensing Society*, 19-21 September, 1990, Swansea, UK. pp. 224-231.

Fang, W.C. 1992. A VLSI neural processor for image data compression using self-organization networks. *IEEE Transactions on Neural Networks*, 3(3):506-518.

Fausett, L. 1994. *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications.* Prentice-Hall Inc., Englewood Cliffs, NJ, 461 p.

Fierens, F., Kanellopoulos, I., Wilkinson, G.G., and Megier, J. 1994. Comparison and visualization of feature space behaviour of statistical and neural classifiers of satellite image. *International Geoscience and Remote Sensing Symposium*, 4:1880-1882.

Fiorella, M., and Ripple, W. 1993a. Determining successional stage of temperate coniferous forests with Landsat satellite data. *Photogrammetric Engineering and Remote Sensing*, 59(2):239-246.

Fiorella, M., and Ripple, W. 1993b. Analysis of conifer forest regeneration using Landsat Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing*, 59(9):1383-1388.

Fischer, M.M. and Gopal, S. 1993. Neurocomputing – a new paradigm for geographic information processing. *Environment and Planning*, 23:757-760.

Fitzgerald, R.W., Lees, B.G. 1994. Assessing the classification accuracy of multi-source remote sensing data. *Remote Sensing of Environment*, 47:362-368

Fleiss, J.L. 1981. *Statistical Methods for Rates and Proportions* (New York, NY: John Wiley & Sons, Inc.)

Fleiss, J.L., Cohen, J., Everitt, B.S. 1969. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 72(5):323-327.

Fleming, M. and Hoffer, R. 1977. *Computer-Aided Analysis Techniques for an Operational System to Map Forest Land Utilizing Landsat MSS Data,* LARS/Purdue University, West Lafayette, Indiana, LARS Technical Report 112277, 236 pp.

Fogelman Soulie, F. 1991. Neural network architectures and algorithms: a perspective. *Proceedings of the 1991 International Conference on Artificial Neural Networks (ICANN-91)* 19-24 June, 1991; Espoo, Finland.

Foody, G.M. 1992a. Fuzzy sets approach to the representation vegetation continua from remotely sensed data: An example from lowland heath. *Photogrammetric Engineering and Remote Sensing*, 58(2):221-225.

Foody, G.M. 1992b. On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, 58(10):1459-1460.

Foody, G.M. 1995a. Using prior knowledge in artificial neural network classification with a minimal training set. *International Journal of Remote Sensing*, 16(2):301-312.

Foody, G.M. 1995b. Land cover classification using an artificial neural network with ancillary information. *International Journal of Geographical Information Systems,* 9: 527-542.

Foody, G.M. 1996. Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, 17(2):1317-1340.

Foody, G.M., and Arora, M.K. 1997. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *International Journal of Remote Sensing*, 18(4):799-810.

Foody, G.M., Campbell, N.A., Trodd, N.M. and Wood, T.F. 1992. Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. *Photogrammetric Engineering and Remote Sensing*, 58(9):1335-1341.

Foody, G.M., McCulloch, M.B., and Yates, W.B. 1995a. The effect of training set size and composition on artificial neural network classification. *International Journal of Remote Sensing*, 16(9):1707-1723.

Foody, G.M., McCulloch, M.B., and Yates, W.B. 1995b. Classification of remotely sensed data by an artificial neural network: issues related to training data characteristics. *Photogrammetric Engineering and Remote Sensing*, 61:391-401.

Foody, G.M., and Trodd, N.M. 1996. Representation of ecological trends in remotely sensed data: relating the probability of class membership to canopy composition and a vegetation ordination. *Geocarto International*, 11(1):3-11.

Franklin, J., Logan, T.L., Woodcock, C.E., and Strahler, A.H. 1986. Coniferous forest classification and inventory using Landsat and digital terrain data. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(1):139-149.

Franklin, S.E. 1994. Discrimination of subalpine forest species and canopy density using digital CASI, SPOT PLA, and Landsat TM data. *Photogrammetric Engineering and Remote Sensing*, 60(10):1233-1241.

Franklin, S.E., Connery, D.R., and Williams, J.A. 1994. Classification of alpine vegetation using Landsat Thematic Mapper Spot HRV and DEM data. *Canadian Journal of Remote Sensing*, 20(1):49-56.

Franklin, S.E., and Luther, J.E. 1997. Satellite remote sensing of balsam fir forest stucture, growth and cumulative defoliation. *Canadian Journal of Remote Sensing*, 21(4):400-411.

Franklin, S.E., and Peddle, D.R. 1990. Classification of SPOT HRV imagery and texture features. *International Journal of Remote Sensing*, 11(3):551-556.

Friedl, M.A., and Brodley, C.E. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61:399-409.

Fukushima, K. 1975. Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, 20:121-136.

Fung, T. 1990. An assessment of TM imagery for land cover change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):681-684.

Fung, T. and LeDrew, E. 1988. The determination of optimal threshold levels for change detections using various accuracy indices. *Photogrammetric Engineering and Remote Sensing*, 56:597-603.

Gastellu-Etchegorry, J.P. 1990. An assessment of SPOT XS and Landsat MSS data for digital classification of near-urban land cover. *International Journal of Remote Sensing*, 11(2):225-235.

Gates, D.M. 1970. Physical and physiological properties of plants in *Remote Sensing With Special Reference to Agriculture and Forestry*. pp. 224-252. National Research Council, Agriculture Board, Washington: National Academy of Sciences.

Gates, D.M., Keegan, H.J., Schletere, J.C. and Weidner, V.R. 1965. Spectral properties of plants. *Applied Optics*, 4(11).

Gath, I., and Geeva, A.B. 1989. Unsupervised optimal fuzzy clustering. *Pattern Analysis and Machine Intelligence*, 11(7):773-781.

Gausman, H.W., Escobar, D.E., Everitt, J.H., Richardson, A.J. and Rodruguez, R.R. 1978. Distinguishing succulent plants from crop and woody plants. *Photogrammetric Engineering and Remote Sensing*, 44:487-491.

Gaydos,L., and Newland, W. 1978. Inventory of Land Use and Land Cover of the Puget Sound Region Using Landsat Digital Data. *General Research*, USGS, 6(6):807-814.

Geographic Data B.C. 1996. *Gridded DEM Specification, Release 1.1*. Ministry of Environment, Lands and Parks, Province of British Columbia. Victoria, B.C.

Gilbert, D. 1998. The Challenge and the Reality. *Presentation to the International Forum: Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, February 10-12, 1998, Pacific Forestry Centre, Victoria, B.C.

Gong, P., and Howarth, P.J. 1992. Frequency based contextual classification and grey-level vector reduction for land use identification. *Photogrammetric Engineering and Remote Sensing*, 58(4):423-437.

Gong, P., Pu, R., and Chen, J. 1996. Mapping ecological land systems and classification uncertainties from digital elevation and forest-cover data using neural networks. *Photogrammetric Engineering and Remote Sensing*, 62(11):1249-1260.

Gopal, S., and Woodcock, C.E. 1994. Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing*, 60(2):181-188.

Gopal, S., and Woodcock, C.E. 1996. Remote sensing of forest change using artifical neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 34(2):398-404.

Green, E.J., Strawderman, W.E., and Airola, T.M. 1993. Assessing classification probabilities for thematic maps. *Photogrammetric Engineering and Remote Sensing*, 59(5):635-639.

Green, R.N. and Klinka, K. 1994. *A Field Guide to Site Identiciation and Interpretation for the Vancouver Forest Region*. Research Branch, Ministry of Forests, Victoria, British Columbia.

Grignetti, A., Salvatori, R., Casacchia, R., and Manes, F. 1997. Mediterranean vegetation analysis by multi-temporal satellite sensor data. *International Journal of Remote Sensing*, 18(6):1307-1318.

Groom, G.B., Fuller, R.M., and Jones, A.R. 1996. Contextual correction: techniques for improving land cover mapping from remotely sensed images. *International Journal of Remote Sensing*, 17(1):69-89.

Grossberg, S. 1988. Competitive learning: from interactive activation to adaptive resonance. *Cognitive Science*, 11:23-63.

Grossberg, S., Mingolla, E and Willamson, J.R. 1995. Synthetic aperture radar processing by a mutilple scale neural system for boundary and surface representation. *Neural Networks*, 8:1005-1028.

Guo, L.J., and Haigh, J.D. 1994. A three-dimensional feature space iterative clustering method for multi-spectral image classification. *International Journal of Remote Sensing*, 15(3):633-644.

Gurney, C.M., and Townshend, J.R.G. 1983. The use of contextual information in the classification of remotely sensed data . *Photogrammetric Engineering and Remote Sensing*, 49(1):55-64.

Hadipriono, F.C., Lyon, J.G., and Li, T. 1991. Expert opinion in satellite data interpretation. *Photogrammetric Engineering and Remote Sensing*, 57(1):75-78.

Hagner, O. and Rigina, H. 1998. Detection of forest decline in Monchegorsk area. *Remote sensing of Environment*, 63(1):11-23.

Hall, R.J., Crown, P.H., Titus, S.J., and Volney, W.J.A. 1997. Evaluation of LANDSAT Thematic Mapper Data for mapping top kill caused by jack Pine budworm defoliation. *Canadian Journal of Remote Sensing*, 21(4):388-399.

Hansen, M., Dubayah, R., and DeFries, R. 1996. Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, 17(5):1075-1081.

Haralick, R.M. 1983. *Pattern Recognition Of Remotely Sensed Data.* Germany: Springer-Verlag.

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

Hassoun, M.H. 1997. *Fundamentals of Artificial Neural Networks.* Massachusetts Institute of Technology.

Hay, A.M. 1979. Sampling designs to test land use map accuracy. *Photogrammetric Engineering and Remote Sensing*, 45(4):529-533.

Heermann, P.D. and Khazenie, N. 1990. Application of neural networks for classification of multi-source multi-spectral remote sensing data. In *Proceedings, 10th Annual International Geoscience and Remote Sensing Symposium,* College Park, Md. 1:1277-1280.

Heermann, P.D., and Khazenie, N. 1992. Classification of multispectral remote sensing data using a back-propagation neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 30(1):81-88.

Hegyi, F., and Quenet, R.V. 1982. Updating the forest inventory data base in British Columbia. In C.J. Johansen and J.L. Sanders (Eds.), *Remote Sensing for Resource Management.* pp. 512-518). Ankeny: Soil Conservation Society of America.

Hepner, G. and Ritter, N. 1989. *Application of an Atificial Neural Netowork to Land Cover Classification of Thematic Mapper Imagery,* JPL Internal Technical Report.

Hepner, G.F., Logan, T., Ritter, N., and Bryant, N. 1990. Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification. *Photogrammetric Engineering and Remote Sensing*, 56(4):469-473.

Hill, J., and Sturm, B. 1991. Radiometric correction of multitemporal Thematic Mapper data for use in agricultural land-cover classification and vegetation monitoring. *International Journal of Remote Sensing*, 12(7):1471-1491.

Hoffer, R.M. 1978. Biological and physical considerations in applying computer-aided analysis techniques to remote sensor data *Remote Sensing: The Quantitative Approach* ed. P.H. Swain and S.M. Davis, McGraw-Hill, Inc., New York, chapter 5.

Hoffer, R.M. 1979. Computer-Aided Analysis of Remote Sensing Data – Magic, Mystery or Myth? *Proceedings, Remote Sensing for Natural Resources,* University of Idaho, Moscow, pp. 156-179.

Holben, B.N. and C.O. Justice. 1980. The topographic effect on spectral response from nadir-pointing sensors. *Photogrammetric Engineering and Remote Sensing,* 46(9):1191-1200.

Hojsgaard, S., Caccetta, P., and Kiiveri, H. 1997. Pixel allocation using remotely-sensed data and ground data. *International Journal of Remote Sensing*, 18(2):417-433.

Hopfield, J.J. 1982. Neural networks and physical systems with emergent collective computational properties. *Proceedings of the National Academy of Sciences in the USA,* 79:2554-2588.

Hord, R.M. 1982. *Digital Image Processing of Remotely Sensed Data* (New York, NY: Academic Press).

Hord, R.M., and Brooner, W. 1976. Land use map accuracy criteria. *Photogrammetric Engineering and Remote Sensing*, 42(5):671-677.

Horler, D.N.H. and Ahern, F.J. 1986. Forestry information content of thematic mapper data. *International Journal of Remote Sensing*, 7:405-428.

Hudson, W.D., and Ramm, C.W. 1987. Correct formulation of the Kappa Coefficient of Agreement. *Photogrammetric Engineering and Remote Sensing*, 53(4):421-422.

Hutchinson, C.F. 1982. Techniques for combining Landsat and ancillary data for digital classification improvement. *Photogrammetric Engineering and Remote Sensing,* 48(1):123-130.

Hyppanen, H. 1997. Spatial autocorrelation and optimal spatial resolution of optical remote sensing data in boreal forest environment. *International Journal of Remote Sensing*, 17(17):3441-3452.

Iisaka, J., Amano, E., and Sakuri-Amano, T. 1996. Automated forest clearcut detection from TM data. *26th International Symposium on Remote Sensing of Environment,* 1:67-70.

Isaacs, R.G., and Vogelmann, A.M. 1988. Multispectral sensor data simulation modelling based on the multiple scattering LOWTRAN code. *Remote Sensing of Environment*, 26:75-99.

Ito, Y., and Omatu, S. 1997. Category classification method using a self-organizing neural network. *International Journal of Remote Sensing*, 18(4):829-845.

Janssen, L.L.F., Jarrsmar, M.N., van der Linden, E.T.M. 1990. Integrating topographic data with remote sensing for land-cover classification. *Photogrammetric Engineering and Remote Sensing*, 56(11):1503-1506.

Janssen, L.L.F. and van der Wel, F.J.M. 1994. Accuracy assessment of satellite derived land-cover data: a review. *Photogrammetric Engineering and Remote Sensing*, 60(4): 419-426.

Jasinski, M.F. 1997. Estimation of subpixel vegetation density of natural regions using satellite multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 34(3):804-813.

Jensen, J.R. 1979. Spectral and textural features to classifiy elusive land cover at the urban fringe. *The Professional Geographer*, 39:400-409.

Jensen, J.R. 1983. Biophysical Remote Sensing. *Annals of the Association of American Geographers*, 73(1):111-132.

Jones, H.G. 1992. *Plants and Micro-climate: A quantitative approach to environmental plant physiology, 2nd Edition.* Cambridge University Press, New York, NY.

Joria, P.E., and Jorgenson, J.C. 1996. Comparison of three methods for mapping tundra with Landsat digital data. *Photogrammetric Engineering and Remote Sensing*, 62(2):163-169.

Justice, C.O., Wharton, S.W. and Holben, B.N. 1981. Application of digital terrain data to quantify and reduce the topographic effect on Landsat data. *International Journal of Remote Sensing*, 2(3):213-230.

Kanellopoulos, I., Varfis, A., Wilkinson, G.G., and Megier, J. 1991. Neural network classification of multi date satellite imagery. *International Geoscience and Remote Sensing Symposium*, 4:2215-2218.

Kanellopoulos, I., Varfis, A., Wilkinson, G.G., and Megier, J. 1992. Land cover discrimination in SPOT HRV imagery using an artificial neural network: A 20-class experiment. *International Journal of Remote Sensing*, 13(5):917-924.

Kanellopoulos, I., and Wilkinson, G.G. 1997. Strategies and best practice for neural network image classification. *International Journal of Remote Sensing*, 18(4):711-725.

Kartikeyan, B., Gopalakrishna, B., Kalubarme, M., and Majumder, K.L. 1994. Contextual techniques for classification of hgih and low resolution remote sensing data. *International Journal of Remote Sensing*, 15(5):1037-1051.

Kattenborn, G. 1997. Atmospheric correction of Landsat TM data over mountainous terrain. *International Archives of Remote Sensing*, 29(7):891-896.

Kauth, R.J. and Thomas, G.S. 1976. The Tasslled Cap – a graphic description of the spectral temporal development of agricultural crops as seen by Landsat. *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data.* Purdue University, West Lafayette, Indiana, pp. 4B41-4B51.

Kenk, E., Sondheim, M., and Yee, B. 1988. Methods for improving accuracy of thematic mapper ground cover classifications. *Canadian Journal of Remote Sensing*, 14(1):17-31.

Kent, M. and Coker, P. 1992. *Vegetation Description and Analysis.* Bellhaven Pres, London, U.K.

Key, J.R., Maslanik, J.A., and Barry, R.G. 1989. Cloud classification from satellite data using a fuzzy sets algorithm: A polar example. *International Journal of Remote Sensing*, 10(12):1823-1842.

Khorram, S., Brockhaus, J.A., Bruck, R.I. and Campbell, M.V. 1990. Modelling and multitemporal evaluation of forest decline with Landsat TM digital data, *IEEE Transactions on Geoscience and Remote Sensing,* 28(4):746-748.

Kiang, R.K. 1992. Classification of remotes sensed data using OCR-inspiered neural network techniques, *Proceedings, 12th Annual Interatinal Gescience and Remote Sensing Symposium,* College Park, Md. 2:1277-1280.

Kim.K., Yang, Y., Lee, J., Choi, K., and Kim, T. 1995. Classification of multispectral image using neural network. *International Geoscience and Remote Sensing Symposium*, 1:446-448.

Kimes, D.S., Holben, B.N., Nickeson, J.E., and McKee, W.A. 1996. Extracting forest age in a Pacific Northwest forest from Thematic Mapper and topographic data. *Remote Sensing of Environment*, 56:133-140.

Knipling, E.B. 1970. Physical and physiological basis for the reflectance of visible and near-infrared radiation of vegetation. *Remote Sensing of Environment,* 1:155-159.

Kohohen, T. 1982. Self organized formation of topologically correct feature maps. *Biological Cybernetics,* 43:59-69.

Kriegler, F. J., Malila, W. A., Nalepka, R. F., and Richardson, W. 1969. Preprocessing transformations and their effects on multispectral recognition, in *Proceedings of the Sixth*
International Symposium on Remote Sensing of Environment, University of Michigan, Ann Arbor, MI, pp.97-131.

Kuchler, A.W. and Zonneveld, I.S. (editors). 1988. *Vegetation Mapping (Handbook Vegetation Science, 10).* Dordrecht: Kluwer Academic Publishers.

Kushwaha, S.P.S., Kuntz, S., and Oesten, G. 1994. Applications of image texture in forest classification. *International Journal of Remote Sensing*, 15(11):2273-2284.

Langford, M., and Bell, W. 1997. Land cover mapping in a tropical hillsides environment: a case study in the Cauca region of Colombia. *International Journal of Remote Sensing*, 18(6):1289-1306.

Lark, R.M. 1995. Components of accuracy of maps with special refernce to disciminant analsyis on remote sensor data. *International Journal of Remote Sensing*, 16(8):1461-1480.

Lau, C. 1992. *Neural Networks: Theoretical Foundations and Analysis* (New York, NY: IEEE Press).

Lauver, C.L., and Whistler, J.L. 1993. A Hierarchical Classification of Landsat TM Imagery to Identify Natural Grassland Areas and Rare Species Habitit. *Photogrammetric Engineering and Remote Sensing*, 59(5):627-634.

Leckie, D., Gillis, M., Gougeon, F., Lodin, M., Wakelin, J. and Yuan, X. 1998. Computer Assisted Photo Interpretation Aids to Forest Inventory Mapping: Some Approaches. *Presentation to the International Forum: Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, February 10-12, 1998, Pacific Forestry Centre, Victoria, B.C.

Leiss, I.A., Sandmeier, S., Itten, K., and Kellenberger, T.W. 1995. Improving Land Use Classification in Rugged Terrain Using Radiometric Corrections and a Possibility Based Classification Approach. *International Geoscience and Remote Sensing Symposium*, 3:1924-1926.

Leprieur, C.E., Durand, J.M., and Peyron, J.L. 1988. Influence of topography on forest reflectance using Landsat Thematic Mapper and digital terrain data. *Photogrammetric Engineering and Remote Sensing*, 54(4):491-496.

Leverington, D.W., and Duguay, C.R. 1996. Evaluation of neural network performance in land cover classification in *Proceedings, 26th International Symposium on Remote Sensing of Environment*, 1:83-86.

Lillesand, T.M. and Kiefer, R.W. 1987. *Remote Sensing and Image Interpretation, 2nd Edition*. John Wiley and Sons, New York, NY.

Lins, H. 1976. Land-Use Mapping from Skylab S-190B Photography
The accuracy of a land-use map produced from a Skylab S-190B color photograph was determined by comparing it with a field-checked land-use map made from high-altitude aircraft photography. *Photogrammetric Engineering and Remote Sensing*, 42(3):301-307.

Lippmann, R.P. 1987. An introduction to computing with neural nets. *IEEE ASSP Magazine*, 2:4-22.

Liu, Z.K. and Wilkinson, G.G. 1992. *A neural network approach to geometrical rectification of remotely sensed imagery*. Institute for Remote Sensing Applications, JRC, Technical Note I.92.118.

Liu, Z.K., and Xiao, J.Y. 1991. Classification of remotely-sensed image data using artificial neural networks. *International Journal of Remote Sensing*, 12(11):2433-2438.

Loung, G. and Tan, Z. 1992. Stereo matching using artificial neural networks. International Archives of Photogrammetry and Remote Sensing, 29(B3):417-421.

McClelland, G.E., DeWitt, R.N., Hemmer, T.H., Matheson, L.H. and Moe, G.H. 1989. Multispectral image processing with a three-layer backpropagation network. *Proceedings of IJCNN '89*, 1:151-153. (New York; IEEE Press).

McCulloch, W. and Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 7:115-133.

Ma, Z. and Redmond, R.L. 1995. Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, 61:435-439.

Manikopoulos, C.N. 1992. Neural network approach to DPCM system design for image coding. *IEE Proceedings-I: Communications, Speech and Vision*, 139(5):501-507.

Markham, B.L., and Barker, J.L. 1985. Spectral characterization of the LANDSAT Thematic Mapper sensors. *International Journal of Remote Sensing*, 6(5):697-716.

Markham, B.L., and Barker, J.L. 1986. Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures. *EOSAT Landsat Data User Notes*, :3-9.

Maselli, F., Rodolfi, A., and Conese, C. 1996. Fuzzy classification of spatially degraded Thematic Mapper data for estimation of sub-pixel components. *International Journal of Remote Sensing*, 17(3):537-551.

Matejek, S. and J.M.M. Dubois, 1988. Determining the age of cutting areas by Landsat 5 TM. *Photo Interpretation*, 27:15-22.

Mathieu-Marni, S., Leymarie, P., and Berthod, M. 1995. Removing ambiguities in a mutispectral image classification. *International Geoscience and Remote Sensing Symposium*, 3:1918-1920.

Maslanik, J., Key, J. and Schweiger, A. 1990. Neural network identificaiton of sea ice seasons in passive microwave data. *Proceedings of IGARSS '90*. 2:1281-1284. (New York: IEEE Press).

Masters, T. 1994. *Signal and Image Processing with Neural Networks* (New York, NY: John Wiley & Sons, Inc.)

Mausel, P.W., Kramber, W.J., and Lee, J.K. 1990. Optimum band selection for supervised classification of multispectral data. *Photogrammetric Engineering and Remote Sensing*, 56(1):55-60.

Mayer, K. and Fox, L. 1981. Identification of conifer species groupings from Landsat digital classifications. *Photogrammetric Engineering and Remote Sensing*, 48(11): 1607-1614.

Mayer, H.A., Schwaiger, R., and Huber, R. 1996. Evolving topologies of artificial neural networks adapted to image processing tasks. *26th International Symposium on Remote Sensing of Environment*, 1:71-74.

Medler, M.J., and Yool, S.R. 1997. Improving Thematic Mapper based classificatio of wildfire induced vegetation mortality. *Geocarto International*, 12(1):49-58.

Miller, D.M., Kaminsky, E.J., and Rana, S. 1995. Neural network classification of remote sensor data. *Computers and Geosciences*, 21(3):377-386.

Minnaert, M. 1941. The reciprocity principle in lunar photometry. *Astrophysical Journal*, 93:403-10.

Moody, A., Gopal, S., and Strahler, A.H. 1996. Artifical neural network response to mixed pixels in coarse-resolution satellite data. *Remote Sensing of Environment*, 58:329-343.

Moody, A., Gopal, S., Strahler, A.H., Borak, J., and Fisher, P. 1994. A combination of temporal thresholding and neural network methods for classifying multiscale remotely sensed image data. *International Geoscience and Remote Sensing Symposium*, 5:1877-1879.

Moore, M.A., and Bauer, M.E. 1990. Classification of forest vegetation in North-Central Minnesota using Landsat Multispectral Scanner and Thematic Mapper data. *Forest Science*, 36(2):330-342.

Moran, M.S., Jackson, R.D., Clarke, T.R., Qi, J., Cabot, F., Thome, K.J., and Markham, B.L. 1995. Reflectance factor retrieval from Landsat TM and SPOT HRV data for bright and dark targets. *Remote Sensing of Environment*, 52:218-230.

Mouchot, M.C., Solaiman, B., and Parra, A. 1994. Use of fuzzy logic for unsupervised classification of remotely sensed images. *International Geoscience and Remote Sensing Symposium*, 2:1163-1165.

Mozer, M.C. and Smolensky, P. 1989. Using relevance to reduce network size automatically. *Connection Science*, 1(1): 3.

Mulder, N.J., Twente, U., and Spreeuwers, L. 1991. Neural networks applied to the classification of remotely sensed data. *International Geoscience and Remote Sensing Symposium*, 4:2211-2213.

Murai, H., and Omatu, S. 1997. Remote sensing image analysis using a neural network and knowledge-based processing. *International Journal of Remote Sensing*, 18(4):811-828.

Murtha, P.A. 1978. Remote sensing and vegetation damage: a theory for detection and assessment. *Photogrammetric Engineering and Remote Sensing*, 44:1147-1158.

Narendra, P.M., and Fukunaga, K. 1977. A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers*, 26(9):917-922.

Nelson, R.F., R.S. Latty and G. Mott, 1984. Classifying northern forests with Thematic Mapper Simulator data. *Photogrammetric Engineering and Remote Sensing,* 50:607-617.

Niemann, K.O. 1989. Consideration of Geographical Information Systems as tools for classification of remotely sensed data and spatial modelling of landforms. . *Proceedings of IGARSS'89*. 1:51-55. (New York: IEEE Press).

Niemann, K.O. 1991. Landscape drainage modelling to enhance Landsat classification accuracies. *Geocarto International*, 1:13-30.

Niemann, K.O. 1993. Automated forest cover mapping using Thematic Mapper images and ancillary data. *Applied Geography,* 13:86-95.

Norwine, J. and Greegor, D.H. 1983. Vegetation classification based on Advanced High Resolution Radiometer (AVHRR) satellite imagery. *Remote Sensing of Environement,* 13:69-87.

Pack, K.N., Song, Y.S., Chae, H.S., Kim, K.E. 1997. Study on the characteristics of the supervised classification of remotely sensed data using artificial neural networks. *Proceedings, 17th Annual International Geoscience and Remote Sensing Symposium*, Singapore.

Pal, S.K. 1994. Fuzzy sets in image processing and recognition. In R.J. Marks (Ed.), *Fuzzy Logic Technology and its Applications*. New York: IEEE, pp. 33-40.

Palubinskas, G., Lucas, R.M., and Curran, P.J. 1995. An evaluation of fuzzy and texture-based classification approaches for mapping regenerating tropical forest classes from Landsat-TM data. *International Journal of Remote Sensing*, 16(4):747-755.

Pao, Y.H. 1989. *Adaptive Pattern Recognition and Neural Networks.* Addison-Wesley Publishing Company, Inc.

Paola, J.D. 1994. Neural Network Classification of Multispectral Imagery. University of Arizona. Unpublished Master's Thesis.

Paola, J.D., and Schowengerdt, R.A. 1994. Comparisons of neural networks to standard techniques for image classification and correlation. *International Geoscience and Remote Sensing Symposium*, 3:1404-1406.

Paola, J.D., and Schowengerdt, R.A. 1995. Searching for patterns in remote sensing image databases using neural networks. *International Geoscience and Remote Sensing Symposium*, 1:443-445.

Paola, J.D., and Schowengerdt, R.A. 1995. A review and analysis of backpropagation neural netowrks for classification of remotely-sensed multi-spectral imagery. *International Journal of Remote Sensing*, 16(16):3033-3058.

Paola, J.D., and Schowengerdt, R.A. 1997. The effect of neural-network structure on a multispectral land-use/land-cover classification. *Photogrammetric Engineering and Remote Sensing*, 63(5):535-544.

Peddle, D.R. 1993. An empirical comparison of evidential reasoning, linear discriminant analysis and maximum likelihood algorithms for alpine land cover classification. *Canadian Journal of Remote Sensing,* 61(4):31-44.

Peddle, D.R., Foody, G.M., Zhang, A., Franklin, S.E., and Ledrew, E.F. 1994. Multi-source image classification II: An empirical comparison of evidential reasoning and neural network approaches. *Canadian Journal of Remote Sensing*, 20(4):396-407.

Peddle, D.R., and Franklin, S.E. 1991. Image texture processing and data integration for surface pattern discrimination. *Photogrammetric Engineering and Remote Sensing*, 57(4):413-420.

Peterson, D.L., Westman, W.E., Stephenson, N.J., Ambrosia, V.G., Brass, J.A., and Spanner, M.A. 1986. Analysis of forest structure using Thematic Mapper simulator data. *IEEE Transactions on Geoscience and Remote Sensing*, 24(1):113-120.

Piekutowski, T. 1996. Personal communication.

Pierce, L.E., Sarabandi, K., and Ulaby, F.T. 1994. Application of an artificial neural network in canopy scattering inversion. *International Journal of Remote Sensing*, 15(16):3263-3270.

Pilon, P., and Wiart, R.J. 1990. Operational forest inventory applications using Landsat TM Data: The British Columbia experience. *Geocarto International*, 1:25-31.

Plummer, S.E., 1988. Exploring the relationships between leaf nitrogen content, biomass and near-infrared/red reflectance ratio. *International Journal of Remote Sensing*, 9(1): 177-183.

Pons, X., and Sole-Sugranes, L. 1994. A simple radiometric correction model to improve automatic mapping of vegetation from multispectral satellite data. *Remote Sensing of Environment*, 48:191-204.

Price, J.C. 1987. Calibration of satellite radiometers and the comparison of vegetation indices, *Remote Sensing of Environment*, 21:15-27.

Resources Inventory Branch. 1996. *The Preparation and Creation of FRGIS Data Files*. Ministry of Forests, Province of British Columbia. Victoria, B.C.

Resources Inventory Committee. 1996a. *Photo Interpretation Procedures Phase I Vegetation Resources Inventory, Version 2.0*. Province of British Columbia. Victoria, B.C.

Resources Inventory Committee. 1996b. *Corporate Land Use Classification System for British Columbia: Justification and Specification.* Province of British Columbia. Victoria, B.C.

Resources Inventory Committee. 1997. *Vegetation Resources Inventory, B.C. Land Cover Classification Scheme*. Province of British Columbia. Victoria, B.C.

Rhyherd, S., and Woodcock, C.E. 1996. Combining spectral and texture data in the segmentation of remotely sensed images. *Photogrammetric Engineering and Remote Sensing*, 62(2):181-194.

Richards, J.B. 1996. Classifier performance and map accuracy. *Remote Sensing of Environment*, 57:161-166.

Ritter, N., and Hepner, G.F. 1990. Application of an artificial neural network to land-cover classification of thematic mapper imagery. *Computers and Geosciences*, 16(6):873-880.

Rosenblatt, F. *Principles of Neurodynamics.* Spartan Books, 1962.

Rosenfield, G.H. 1982. Sample design for estimatig change in land use and land cover. *Photogrammetric Engineering and Remote Sensing*, 48(5):793-801.

Rosenfield, G.H. 1986. Analysis of thematic map classification error matricies. *Photogrammetric Engineering and Remote Sensing*, 52(2):681-686.

Rosenfield, G.H. and Fitzpatrick-Lins, K. 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetry Engineering and Remote Sensing*, 52(2):223-227.

Rosenfield, G.H., Fitzpatrick-Lins, K., and Ling, H.S. 1982. Sampling for thematic map accuracy testing. *Photogrammetric Engineering and Remote Sensing*, 48(1):131-137.

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. 1973. Monitoring vegetation systems in the great plains with ERTS, *Third ERTS Symposium*, NASA SP-351, 1:309-317.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. 1986. Learning internal representations by error propagation. In D.E. Rumelhart and J.L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations.* Cambridge, Massachusetts: MIT Press, pp. 318-362).

Ryan, J.W., Sementilli, P.J., Yuen, P., and Hunt, B.R. 1991. Extraction of shoreline features by neural nets and image processing. *Photogrammetric Engineering and Remote Sensing*, 57(7):947-955.

Sader, S.A. 1995. Spatial characteristics of forest clearing and vegetation regrowth as detected by Landsat Thematic Mapper imagery. *Photogrammetric Engineering and Remote Sensing*, 61(9):1145-1151.

Salkeld, T. 1998. The British Columbia Vegetation Resources Inventory. *Presentation to the International Forum: Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, February 10-12, 1998, Pacific Forestry Centre, Victoria, B.C.

Salu, Y., and Tilton, J. 1996. Classification of multispectral image data by the binary diamond neural network and be nonparametric pixel-by-pixel methods. *IEEE Transactions on Geoscience and Remote Sensing*, 31(3):606-617.

San Miguel-Ayanz, J., and Biging, G.S. 1997. Comparison of single-stage and lulti-stage classification approaches for covr type mapping with TM and SPOT data. *Remote Sensing of Environment*, 59:92-104.

Sant'Anna, S., Yanasse, C., Filho, P., Kuplich, T., Dutra, L., Frery, A., and Santos, P. 1995. Secondary forest age mapping in Amazonia using multi-temporal Landsat TM imagery. *International Geoscience and Remote Sensing Symposium*, 1:323-325.

Saxena, K.G., Tiwari, A.K., Porwal, M.C., and Menon, A.R.R. 1992. Vegetation maps, mapping needs and scope of digital processing of Landsat Thematic Mapper data in tropical region of southwest India. *International Journal of Remote Sensing*, 13(11):2017-2037.

Schalkoff, R. 1992. *Pattern Recognition: Statistical, Structural and Neural Approaches* (New York: Wiley).

Scheffner, E.J. 1996. The Landsat Program and Landsat Science. *Presentation to the Landsat Science Working Group,* Greenbelt, Maryland.

Schriever, J.R., and Congalton, R.G. 1995. Evaluating seasonal variability as an aid to cover-type mapping from Landsat Thematic Mapper data in the northeast. *Photogrammetric Engineering and Remote Sensing*, 61(3):321-327.

Schowengerdt, R.A. 1983. *Techniques for Image Processing and Classification in Remote Sensing* (New York: Academic Press).

Seftor, J.L., and Larch, D. 1995. The use of the genetic algorithm to optimize rule-based classifiers for land cover categorization. *Canadian Journal of Remote Sensing*, 21(4):412-420.

Sergi, R., Solaiman, B., Mouchot, M.C., Pasquariello, G., and Posa, F. 1995. Landsat TM image classification using principal components analysis and neural networks. *International Geoscience and Remote Sensing Symposium*, 3:1927-1929.

Serpico, S.B., and Roli, F. 1995. Classification of multisensor remote sensing images by structured neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 33(3):562-578.

Shen, S.S., G.D. Badwhar and J.G. Carnes, 1985. Separability of Boreal Forest Species in the Lake Jennette Area, Minnesota. *Photogrammetric Engineering and Remote Sensing,* 51(11): 1775-1783.

Sinclair, T.R., Hoffer, R.M. and Schrieber, M.M. 1971. Reflectance and internal structure of leaves from several crops during a growing season. *Agronomy Journal,* 63: 864-868.

Skidmore, A.K. 1989. A supervised nonparametric classifier to improve forest cover mapping accuracy. *Proceedings of the Fourth Australasian Remote Sensing Conference,* (D. Bruce, ed.), Adelaide, Australia. Scantec: Adelaide, Australia., 2:573-583.

Skidmore, A.K., and Turner, B.J. 1988. Forest mapping accuracies are improved using a supervised nonparametric classifier with SPOT data. *Photogrammetric Engineering and Remote Sensing,* 54(10):1415-1421.

Skidmore, A.K., Turner, B.J., Brinkhof, W., and Knowles, E. 1997. Performance of a neural network: mapping forests using GIS and remotely sensed data. *Photogrammetric Engineering and Remote Sensing,* 63(5):501-514.

Slater, P.N. 1987. Reflectance and radiance based methods for the in-flight absolute calibration of multispectral sensors. *Remote Sensing of Environment,* 22:11-37.

Smith, J.A., Lin, T.L., and Ranson, K.J. 1980. The Lambertian assumption and Landsat data, *Photogrammetric Engineering and Remote Sensing,* 46(9):1183-1189.

Smith, J.M., 1988. Landsat TM study of afforestation in northern Scotland and its impact on breeding bird population. *1988 International Geoscience and Remote Sensing Symposium*. The Institute of Electrical and Electronics Engineers Inc., pp. 1369-1370.

Snedecor, G.W. and Cochran, W.G. 1980. *Statistical Methods*, 7th Edition, Iowa State University Press, Ames, IA.

Solaiman, B., and Mouchot, M.C. 1994. A comprative study of conventional and neural network classification of multispectral data. *International Geoscience and Remote Sensing Symposium*, 3:1413-1415.

Spanner, M.A., C.A. Hlavka and L.L. Pierce, 1989. Analysis of forest disturbance using TM and AVHRR data. *1989 International Geoscience and Remote Sensing Symposium,* The Institute of Electrical and Electronics Engineers Inc. pp. 1387-1390.

Spanner, M.A., L.L. Pierce, D.L Peterson and S.W. Running, 1990. Remote sensing of temperate coniferous forest leaf area index. The influence of canopy closure, understory vegetation, and background reflectance. *International Journal of Remote Sensing,* 11(1): 95-111.

Stehman, S.V. 1997. Use of auxiliary data to improve the precision of extimators of thematic map accuracy. *Remote Sensing of Environment*, 58:169-176.

Stenback, J.M. and Congalton, R.G. 1990. Using Thematic Mapper imagery to examine forest understory, 56: 1285-1290.

Story, M. and Congalton, R.G. 1986. Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing,* 52: 397-399.

Sui, D.Z. 1994. Recent applications of neural networks for spatial data handling. *Canadian Journal of Remote Sensing*, 20(4):368-380.

Suits, G.H. 1972. The calculation of the directional relfectance of a vegetative canopy, *Remote Sensing of Environment*, 2 (117).

Swain, P.H. and Davis, S.M.(eds.) 1978. *Remote Sensing: The Quantitative Approach*, McGraw-Hill, Inc., New York, 396 p.

Swain, P.H., Vardeman, S.B., and Tilton, J.C. 1981. Contextual classification of multispectral image data. *Pattern Recognition*, 13(6):429-441.

Tanre, D., Deroo, C., Duhaut, P., Herman, M., Morcrette, J.J., Perbos, J. and Deschamps, P.Y. 1986. Simulation of the satellite signal in the solar spectrum. *Remote Sensing of Environment*, 22:11-37.

Tanre, D., Deroo, C., Duhaut, P., Herman, M., Morcrette, J.J., Perbos, J., and Deschamps, P.Y. 1990. Description of computer code to simulate the satellite singal in the solar spectrum: The 5S code. *International Journal of Remote Sensing*, 11(4):659-668.

Teillet, P.M., Guindon ,B. and Goodenough, D.G. 1982. On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, 8(2):84-106.

Teillet, P.M. 1996. Personal communication.

Teillet, P.M., and Fedosejevs, G. 1995. On the dark target approach to atmospheric correction of remotely sensed data. *Canadian Journal of Remote Sensing*, 21(4):374-387.

Teillet, P.M., Slater, P.N., Ding, Y., Santer, R.P., Jackson, R.D. and Moran, M.S. 1990. Three methods for the absolute calibration of the NOAA AVHRR sensors in-flight. *Remote Sensing of Environment*, 31:105-120.

Thomas, I.L., Benning, V.M. and Ching, N.P. 1987. *Classification of Remotely Sensed Images* (Bristol, England: IOP Publishing Limited).

Thome, K.J., Biggar, S.F., Gellman, D.I. and Slater, P.N. 1994. Absolute Radiometric Calibration of Landsat-5 Thematic Mapper and the Proposed Calibration of the Advanced Spaceborne Thermal Emission and Reflection Radiometer, *Proceedings of IGARSS-94, Pasadena, California*, 4:2295-2297.

Thomson, A.G. and C. Jones, 1990. Effects of topography on radiance from upland vegetation in North Wales, *International Journal of Remote Sensing*, 11(5); 829-840.

Toussaint, G.T. 1977. The use of context in pattern recognition. *Pattern Recognition*, 10:189-204.

Treitz, P.M., Howarth, P.J., Shepherd, P., and Miller, D.M. 1996. Integrating remote sensing data and terrain variables to discriminate forest ecosystems in Northwestern Ontario. *26th International Symposium on Remote Sensing of Environment*, :515-518.

Tucker, C. 1977a. Resolution of grass canopy biomass classes. *Photogrammetric Engineering and Remote Sensing*, 43:1059-1067.

Tucker, C. 1977b. Spectral estimation of grass canopy variables. *Remote Sensing of Environment*, 6:11-26.

Tucker, C. 1978. A comparison of satellite sensor bands for vegetation monitoring. *Photogrammetric Engineering and Remote Sensing*, 44:1369-1380.

Tucker, C. and Garrett, M. 1977. Leaf optical system modeled as a stochastic process. *Applied Optics*, 16:635-642.

Tzeng, Y.C., Chen, K.S., Kao, W.L., and Fung, A.K. 1994. A dynamic learning neural network for remote sensing applications. *IEEE Transactions on Geoscience and Remote Sensing*, 32(5):1096-1102.

Vogelmann, J.E., and Rock, B.N. 1988. Assessing forest damage in high-elevation coniferous forests in Vermont and New Hampshire using Thematic Mapper data. *Remote Sensing of Environment*, 24:227-246.

Wang, T. 1990. Improving remote sensing image analysis through fuzzy information representation. *Photogrammetric Engineering and Remote Sensing*, 9(8):1163-1169.

Wang, Y., and Civco, D.L. 1994. Evidential reasoning based classification of multi-source spatial data for improved land cover mapping. *Canadian Journal of Remote Sensing*, 20(4):381-395.

Wang, Y., and Civco, D.L. 1996. Land cover classification from multisource spatial data by three artificial neural network paradigms. *26th International Symposium on Remote Sensing of Environment* 1:79-82.

Wharton, S.W. 1982. A contextual classification method for recognizing land use patterns in high resolution remotely sensed data. *Pattern Recognition,* 15:317-324.

White, J.D., Kroh, G.C., and Pinder, J.E. 1995. Forest mapping at Lassen Volanic National Park, California, using Landsat TM data and geographical information system. *Photogrammetric Engineering and Remote Sensing*, 61(3):299-305.

Widrow, B. and Hoff, M.E. 1960. Adaptive switching circuits. *IRE WESON Convention Record,* 4:96-104.

Wilkinson, G.G., Fierens, F., and Kanellopoulos, I. 1995. Integration of neural and statistical approaches in spatial data classification. *Geographical Systems*, 2:1-20.

Wilkinson, G.G., Folving, S., Kanellopoulos, I., McCormick, M., Fullerton, K., and Megier, J. 1995. Forest mapping from multi-source satellite data using neural network classifiers - an experiment in Portugal. *Remote Sensing Reviews*, 12:83-106.

Wilkinson, G.G., Kanellopoulos, I., Kontoes, C., and Megier, J. 1992. A comparison of neural network and expert system methods for analysis of remotely sensed imagery. *International Geoscience and Remote Sensing Symposium*, 1:62-64.

Wilkinson, G.G., and Megier, J. 1990. Evidential reasoning in a pixel classification hierarchy - a potential method for integrating image classifiers and expert system rules based on geographic context. *International Journal of Remote Sensing*, 11(10):1963-1968.

Williams, D.L., and Nelson, R.F. 1986. Use of remotely sensed data for assessing forest stand conditions in the eastern United States. *IEEE Transactions on Geoscience and Remote Sensing*, 24(1):130-138.

Williams, D.J., Royer, A., O'Neill, N.T., Achal, S., and Weale, G. 1996. Reflectance extraction from CASI spectra using radiative transfer simulations and a rooftop irradiance collector. *Canadian Journal of Remote Sensing*, 22(4):251-261.

Wooley, J.R. 1971. Reflectances and transmittance of light. *Plant Physiology*, 47:656-662.

Wrigley, R.C., Spanner, M.A., Slye, R.E., Pueschel, R.F., and Aggarwal, H.R. 1992. Atmospheric correction of remotely sensed image data by a simplified model. *Journal of Geophysical Research*, 97(17):18797-18814.

Yool, S.R. 1998. Land cover classification in rugged areas using simulated moderate resolution remote sensor data and an artificial neural network. *International Journal of Remote Sensing*, 19(1):85-96.

Yoshida, T., and Omatu, S. 1994. Neural network approach to land cover mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 32(5):1103-1109.

Yoshikawa, M., Shindo, H., Nishii, R., and Tanaka, S. 1995. A fully automated design of binary decision tree for land cover classification. *International Geoscience and Remote Sensing Symposium*, 3:1921-1923.

Zhang, L., and Hoshi, T. 1994. A fuzzy neural network model (FNN Model) for classification using Landsat TM image data. *International Geoscience and Remote Sensing Symposium*, 3:1416-1417.

Zhou, J., and Civco, D.L. 1996. Using genetic learning neural networks for spatial decision making in GIS. *Photogrammetric Engineering and Remote Sensing*, 62(11):1287-1295.

Zhuang, X., Engel, B.A., Lozana-Garcia, D.F., Fernandez, R.N., and Johannsen, C.J. 1994. Optimization of training data required for neuro-classification. *International Journal of Remote Sensing*, 15(16):3271-3277.

APPENDIX A

SUMMARY OF STATISTICS FOR OVERALL PROJECT IMAGERY

IMAGE DATE: August 29, 1993
IMAGE LOCATION: 46-026
IMAGE SIZE: 3500P x 2944L
PROCESSING: Precision Georeferenced

Band Number:	1	2	3	4	5	6	7
μm:	0.45-0.52	0.52-0.60	0.63-0.69	0.76-0.90	1.55-1.75	2.08-2.35	10.4-12.5

UNIVARIATE STATISTICS

Mean	*60.3848*	*24.0493*	*23.6507*	*52.6414*	*43.3654*	*117.1970*	*16.2492*
Std. dev.	*34.8579*	*19.0900*	*25.8228*	*33.8462*	*36.4478*	*50.2561*	*19.2554*
Variance	*1215.070*	*364.430*	*666.820*	*1146.78*	*1328.442*	*2525.680*	*370.7700*
Minimum	*52*	*8*	*0*	*0*	*0*	*0*	*0*
Maximum	*80*	*36*	*32*	*86*	*64*	*160*	*32*
Median	*60*	*22*	*19*	*54*	*36*	*131*	*11*

VARIANCE-COVARIANCE MATRIX

1	*1215.07*						
2	*635.81*	*364.43*					
3	*830.24*	*484.53*	*666.82*				
4	*767.46*	*421.93*	*514.10*	*1146.78*			
5	*956.87*	*539.60*	*726.21*	*1029.21*	*1328.44*		
6	*649.39*	*212.10*	*129.16*	*958.06*	*736.78*	*2525.68*	
7	*537.74*	*309.52*	*432.84*	*426.50*	*658.40*	*187.34*	*370.77*

CORRELATION MATRIX

1	*1.00*						
2	*0.96*	*1.00*					
3	*0.92*	*0.98*	*1.00*				
4	*0.69*	*0.65*	*0.58*	*1.00*			
5	*0.75*	*0.78*	*0.77*	*0.83*	*1.00*		
6	*0.37*	*0.22*	*0.09*	*0.56*	*0.40*	*1.00*	
7	*0.80*	*0.84*	*0.87*	*0.65*	*0.94*	*0.19*	*1.00*

IMAGE DATE: September 30, 1993
IMAGE LOCATION: 46-026
IMAGE SIZE: 3500P x 2944L
PROCESSING: Systematic Georeferenced

Band Number:	1	2	3	4	5	6	7
μm:	0.45-0.52	0.52-0.60	0.63-0.69	0.76-0.90	1.55-1.75	2.08-2.35	10.4-12.5

UNIVARIATE STATISTICS

Mean	*50.2463*	*18.8351*	*18.5292*	*36.9645*	*33.4490*	*108.5566*	*12.6490*
Std. dev.	*26.9806*	*13.4977*	*18.7088*	*25.3507*	*31.1189*	*42.4507*	*15.9068*
Variance	*727.95*	*182.18*	*350.019*	*640.38*	*968.39*	*1802.06*	*253.03*
Minimum	*32*	*16*	*0*	*0*	*0*	*96*	*0*
Maximum	*64*	*32*	*32*	*72*	*64*	*160*	*32*
Median	*50*	*18*	*15*	*38*	*26*	*119*	*8*

VARIANCE-COVARIANCE MATRIX

1	*727.95*						
2	*351.45*	*182.19*					
3	*461.25*	*247.17*	*350.02*				
4	*442.98*	*232.51*	*297.84*	*640.38*			
5	*610.13*	*334.27*	*474.29*	*648.55*	*968.39*		
6	*624.67*	*247.45*	*247.53*	*682.81*	*691.19*	*1802.07*	
7	*332.35*	*184.17*	*268.34*	*270.85*	*468.98*	*245.97*	*253.03*

CORRELATION MATRIX

1	*1.00*						
2	*0.97*	*1.00*					
3	*0.91*	*0.98*	*1.00*				
4	*0.65*	*0.68*	*0.62*	*1.00*			
5	*0.72*	*0.79*	*0.81*	*0.82*	*1.00*		
6	*0.55*	*0.43*	*0.31*	*0.63*	*0.52*	*1.00*	
7	*0.77*	*0.86*	*0.93*	*0.67*	*0.95*	*0.36*	*1.00*

IMAGE DATE: July 31, 1994
IMAGE LOCATION: 46-026
IMAGE SIZE: 3500P x 2944L
PROCESSING: Systematic Georeferenced

Band Number:	1	2	3	4	5	6	7
µm:	0.45-0.52	0.52-0.60	0.63-0.69	0.76-0.90	1.55-1.75	2.08-2.35	10.4-12.5

UNIVARIATE STATISTICS

Mean	*64.1327*	*25.0967*	*23.3159*	*63.7027*	*48.9714*	*199.3381*	*19.0118*
Std. dev.	*31.4818*	*17.5230*	*22.2256*	*37.8225*	*34.1047*	*80.0014*	*16.9139*
Variance	*991.10*	*307.05*	*493.97*	*1430.54*	*1163.13*	*6400.22*	*286.08*
Minimum	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Maximum	*90*	*40*	*40*	*96*	*80*	*255*	*40*
Median	*66*	*24*	*20*	*67*	*45*	*232*	*14*

VARIANCE-COVARIANCE MATRIX

1	*991.10*						
2	*511.71*	*307.05*					
3	*616.39*	*379.04*	*493.98*				
4	*672.48*	*351.38*	*364.94*	*1430.54*			
5	*607.27*	*293.74*	*368.43*	*974.31*	*1163.13*		
6	*1606.38*	*601.47*	*548.15*	*2117.01*	*1848.24*	*6400.22*	
7	*317.44*	*161.18*	*221.59*	*336.27*	*533.94*	*723.49*	*286.08*

CORRELATION MATRIX

1	*1.00*						
2	*0.93*	*1.00*					
3	*0.88*	*0.97*	*1.00*				
4	*0.56*	*0.53*	*0.43*	*1.00*			
5	*0.57*	*0.49*	*0.49*	*0.76*	*1.00*		
6	*0.64*	*0.43*	*0.31*	*0.70*	*0.68*	*1.00*	
7	*0.60*	*0.54*	*0.59*	*0.53*	*0.93*	*0.53*	*1.00*

IMAGE DATE: October 3, 1994
IMAGE LOCATION: 46-026
IMAGE SIZE: 3500P x 2944L
PROCESSING: Systematic Georeferenced

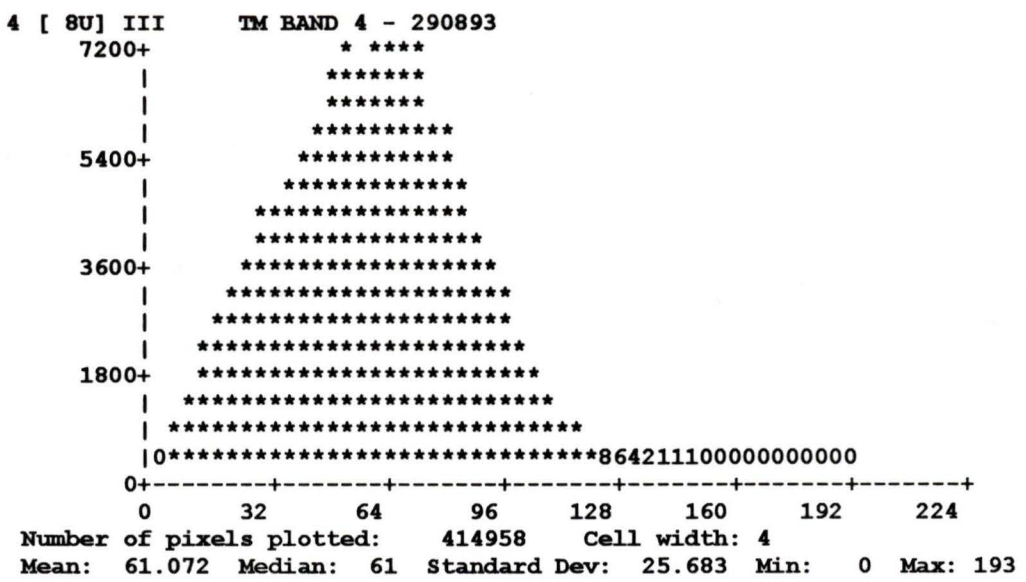
Band Number:	1	2	3	4	5	6	7
µm:	0.45-0.52	0.52-0.60	0.63-0.69	0.76-0.90	1.55-1.75	2.08-2.35	10.4-12.5
UNIVARIATE STATISTICS							
Mean	*52.1556*	*18.3554*	*16.954*	*35.6153*	*31.1679*	*96.8640*	*13.7532*
Std. dev.	*21.6268*	*9.2741*	*11.4048*	*24.6259*	*26.9904*	*88.6471*	*13.9110*
Variance	*467.72*	*86.01*	*130.07*	*606.43*	*728.48*	*7858.31*	*193.52*
Minimum	*45*	*0*	*0*	*0*	*0*	*0*	*0*
Maximum	*80*	*32*	*32*	*68*	*60*	*255*	*32*
Median	*54*	*18*	*14*	*37*	*25*	*83*	*9*
VARIANCE-COVARIANCE MATRIX							
1	*467.72*						
2	*192.12*	*86.01*					
3	*211.55*	*100.72*	*130.07*				
4	*312.28*	*148.71*	*161.55*	*606.44*			
5	*359.03*	*182.22*	*233.44*	*534.19*	*728.48*		
6	*988.86*	*431.23*	*437.53*	*1253.89*	*1367.70*	*7858.31*	
7	*188.53*	*96.71*	*131.07*	*215.59*	*354.96*	*592.21*	*193.51*
CORRELATION MATRIX							
1	*1.00*						
2	*0.96*	*1.00*					
3	*0.86*	*0.95*	*1.00*				
4	*0.59*	*0.65*	*0.58*	*1.00*			
5	*0.62*	*0.73*	*0.76*	*0.80*	*1.00*		
6	*0.52*	*0.52*	*0.43*	*0.57*	*0.57*	*1.00*	
7	*0.63*	*0.75*	*0.83*	*0.63*	*0.95*	*0.48*	*1.00*

APPENDIX B

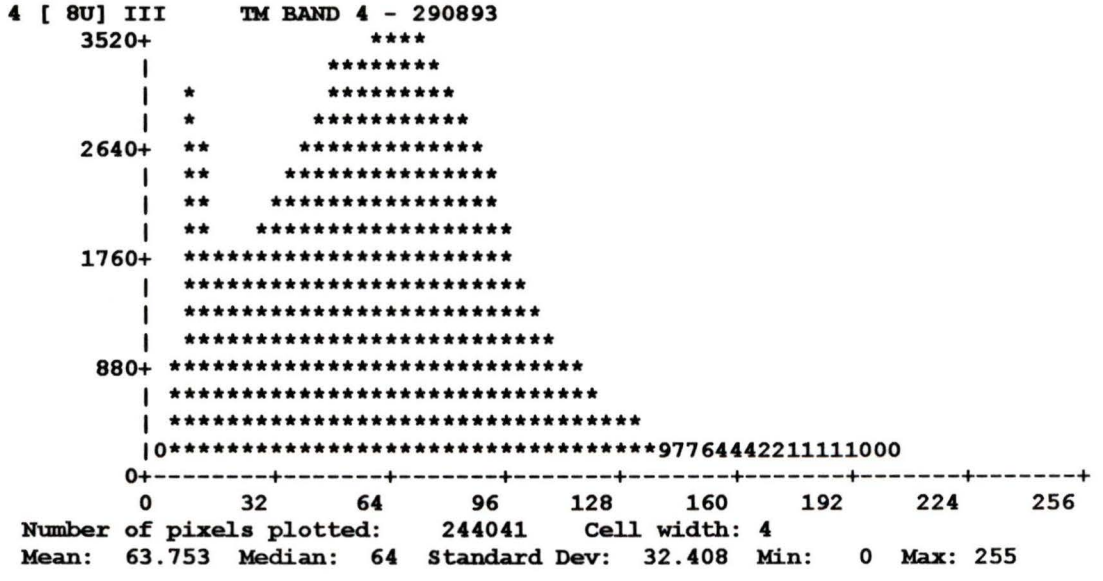
IMAGE HISTOGRAMS FOR TARGET LAND COVER CLASSES

(LANDSAT THEMATIC MAPPER BAND 4)

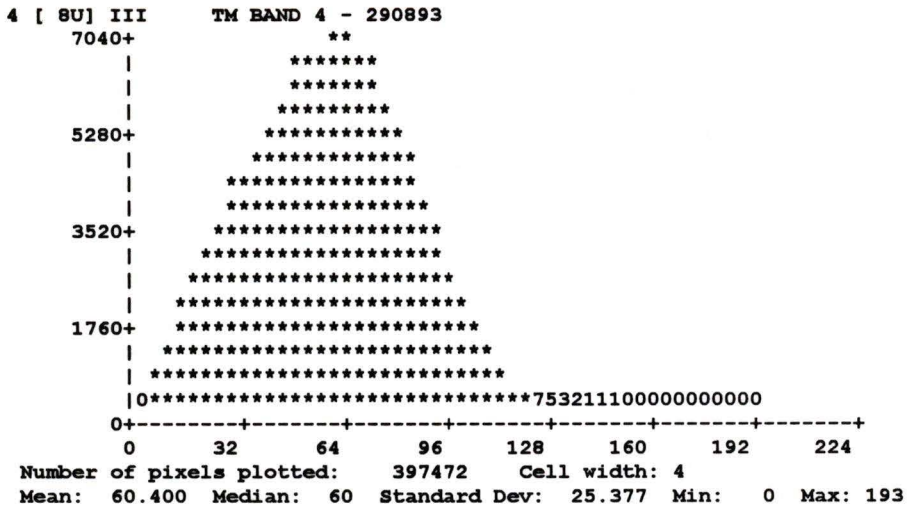
VEGETATED



NON-VEGETATED

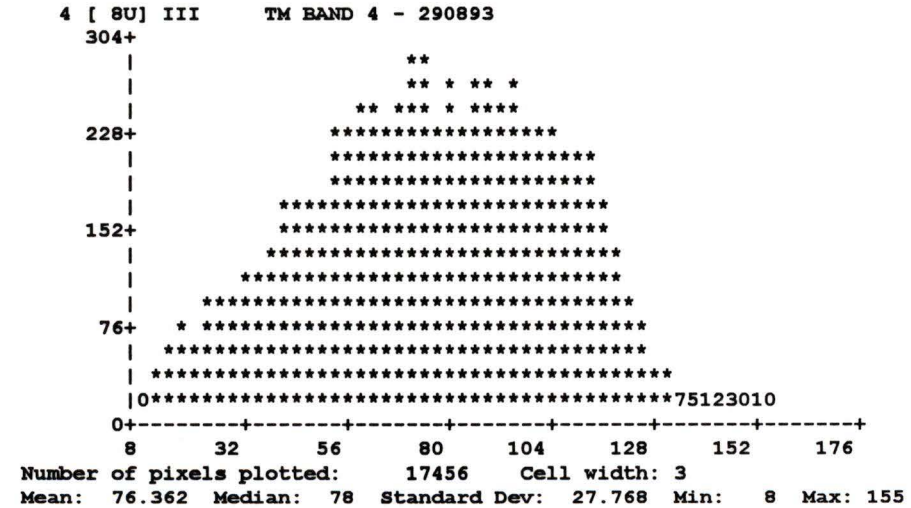


VEGETATED TREED



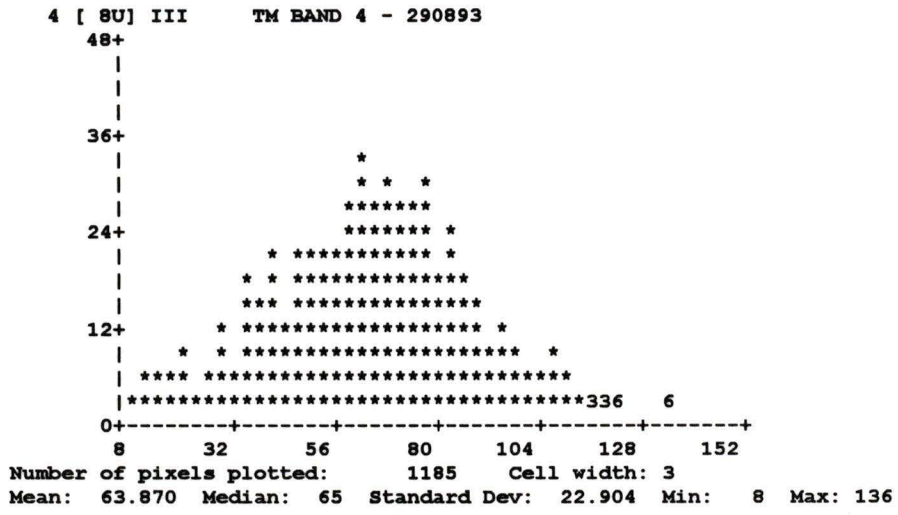
Number of pixels plotted: 397472 Cell width: 4
Mean: 60.400 Median: 60 Standard Dev: 25.377 Min: 0 Max: 193

VEGETATED NON-TREED

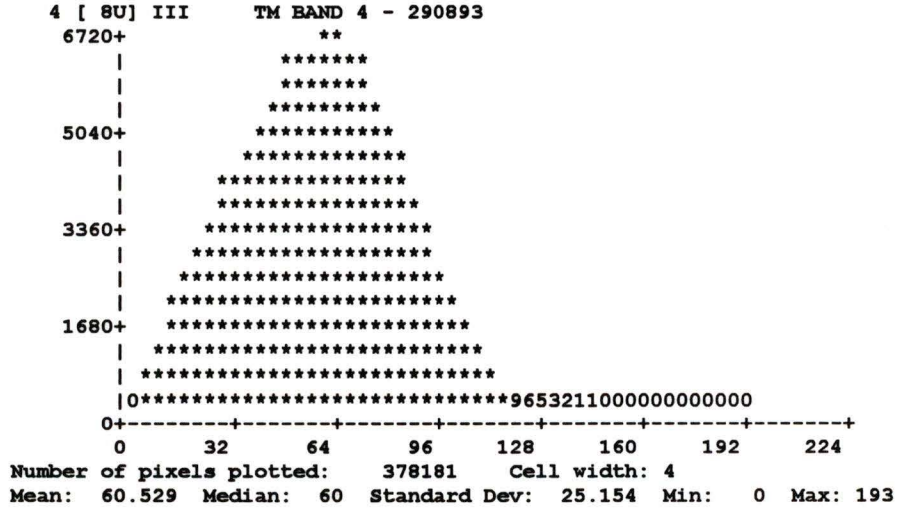


Number of pixels plotted: 17456 Cell width: 3
Mean: 76.362 Median: 78 Standard Dev: 27.768 Min: 8 Max: 155

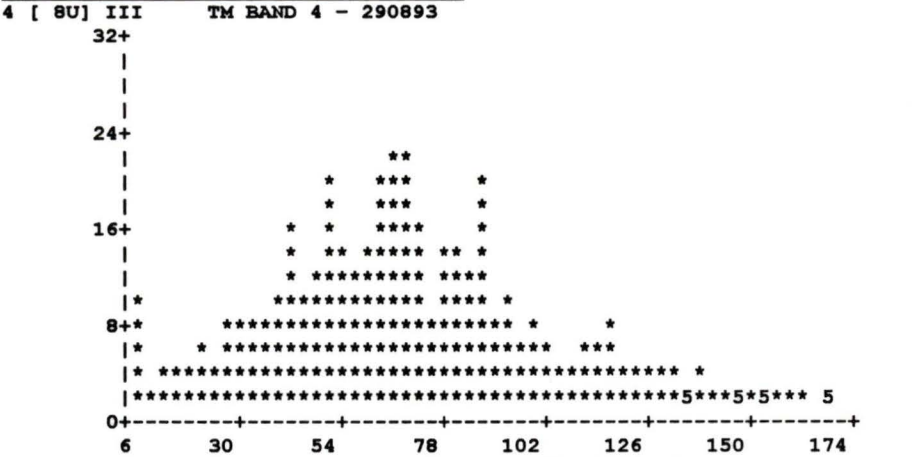
VEGETATED TREED WETLAND



VEGETATED TREED UPLAND

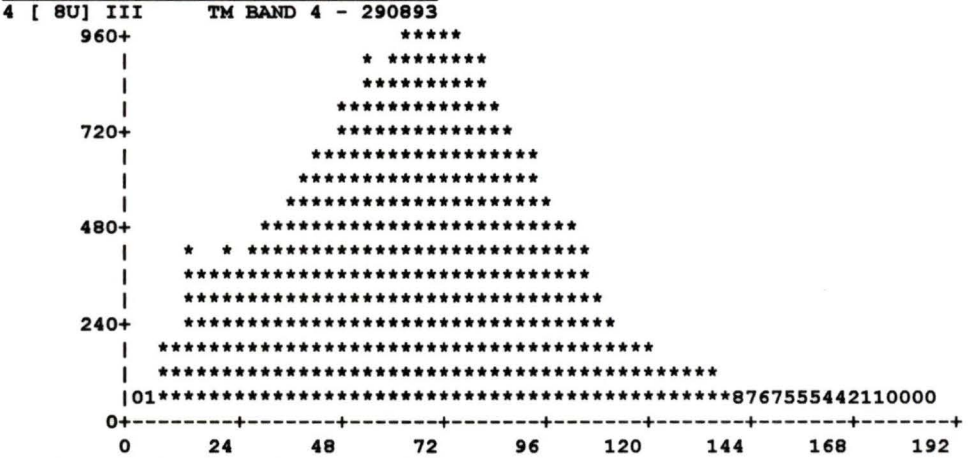


NON-VEGETATED LAND WETLAND



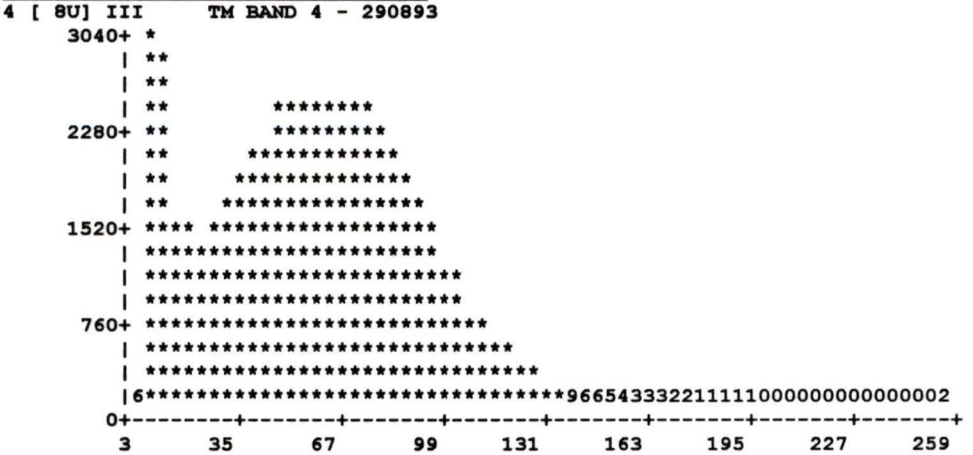
Number of pixels plotted: 853 Cell width: 3
Mean: 68.746 Median: 67 Standard Dev: 30.184 Min: 6 Max: 169

NON-VEGETATED LAND UPLAND



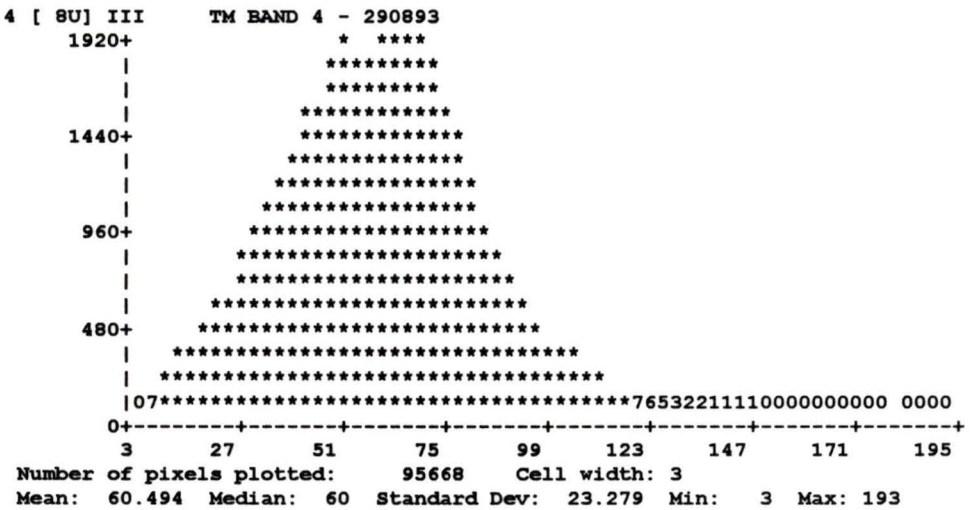
Number of pixels plotted: 59179 Cell width: 3
Mean: 68.047 Median: 68 Standard Dev: 29.769 Min: 0 Max: 188

NON-VEGETATED LAND ALPINE

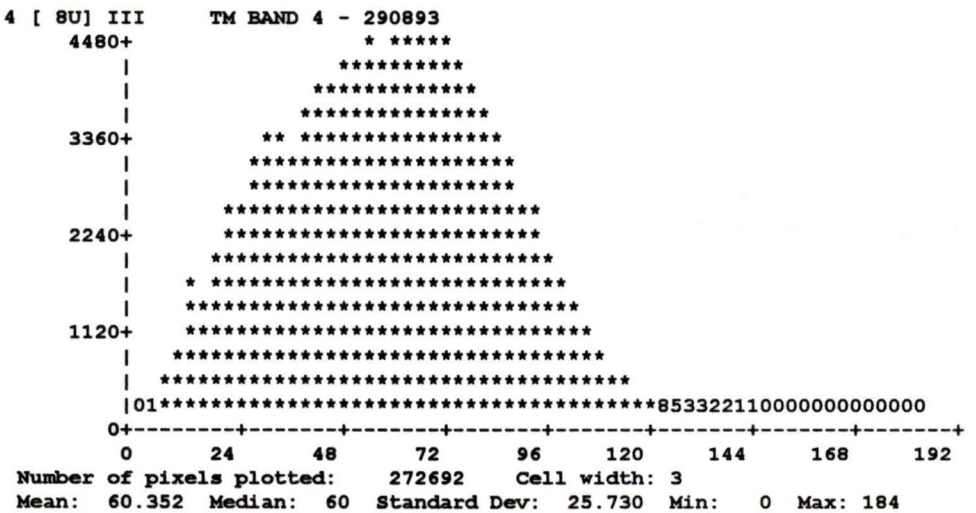


Number of pixels plotted: 183262 Cell width: 4
Mean: 62.375 Median: 61 Standard Dev: 33.128 Min: 3 Max: 255

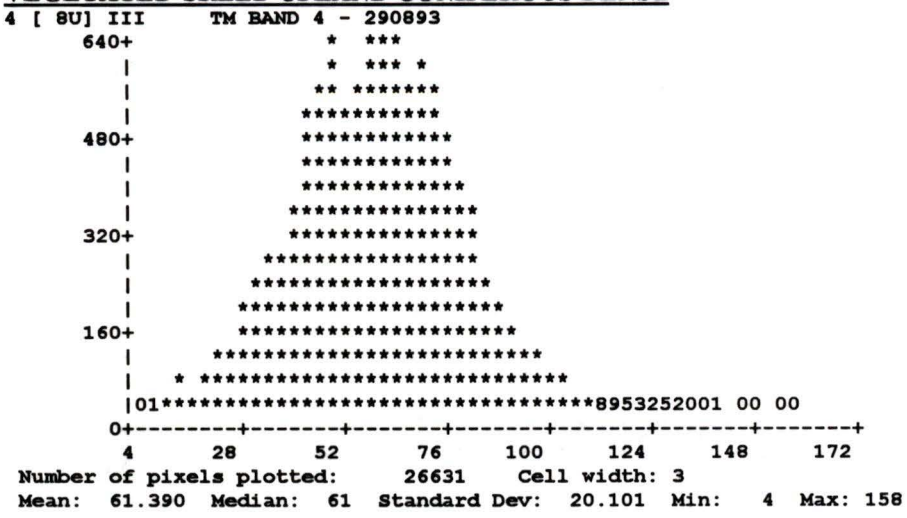
VEGETATED TREED UPLAND CONIFEROUS



VEGETATED TREED UPLAND MIXED

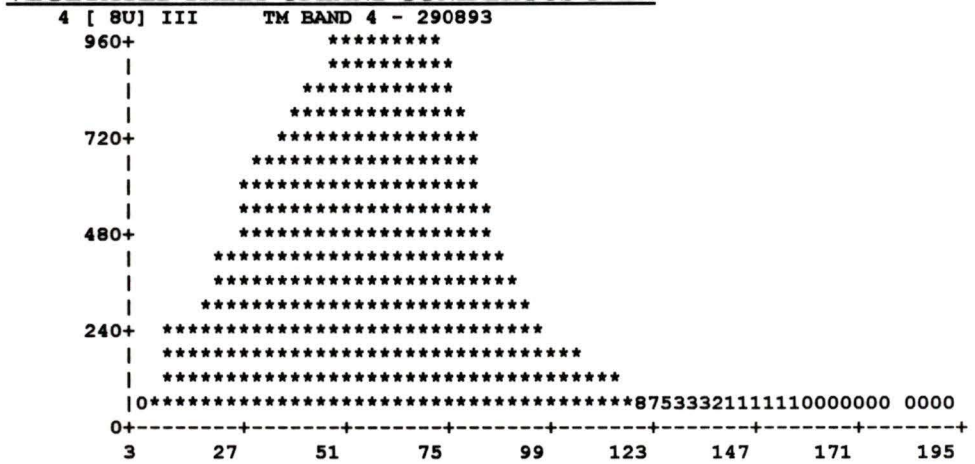


VEGETATED TREED UPLAND CONIFEROUS DENSE



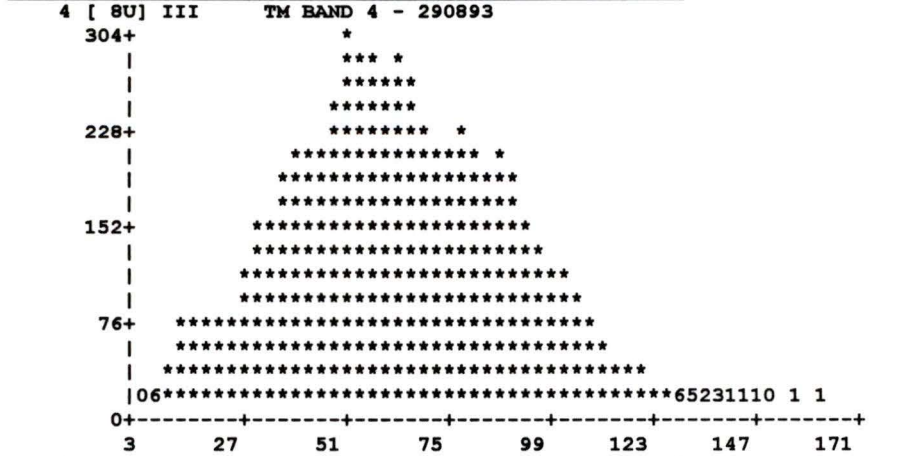
Number of pixels plotted: 26631 Cell width: 3
Mean: 61.390 Median: 61 Standard Dev: 20.101 Min: 4 Max: 158

VEGETATED TREED UPLAND CONIFEROUS OPEN



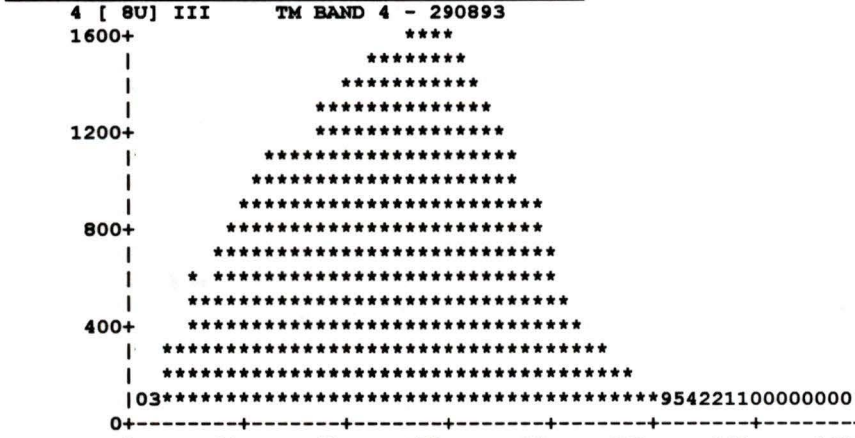
Number of pixels plotted: 54059 Cell width: 3
Mean: 58.865 Median: 59 Standard Dev: 23.850 Min: 3 Max: 193

VEGETATED TREED UPLAND CONIFEROUS SPARSE



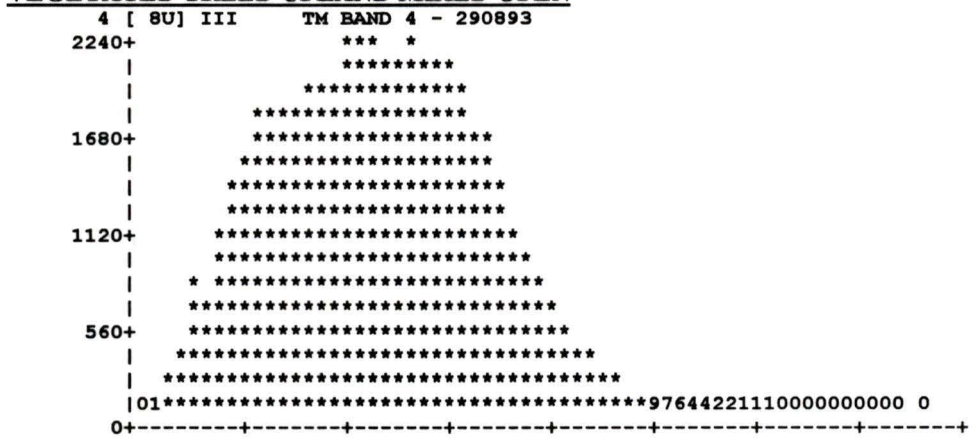
Number of pixels plotted: 14971 Cell width: 3
Mean: 64.777 Median: 64 Standard Dev: 25.701 Min: 3 Max: 164

VEGETATED TREED UPLAND MIXED DENSE



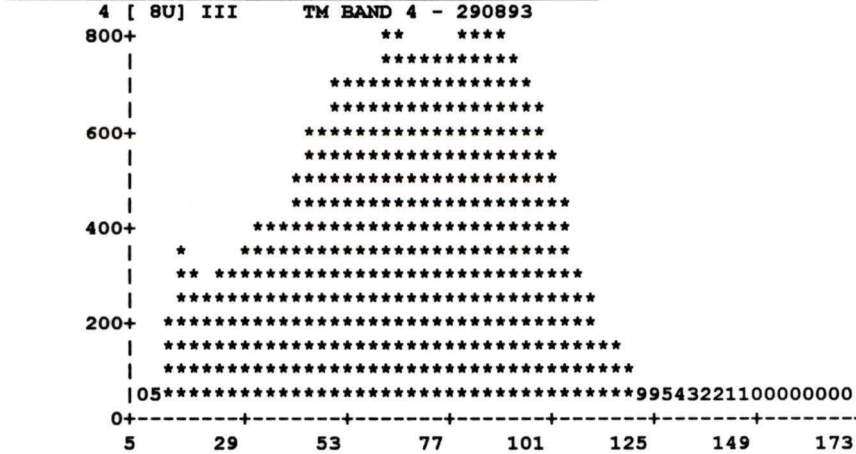
Number of pixels plotted: 92679 Cell width: 3
Mean: 60.310 Median: 61 Standard Dev: 25.370 Min: 0 Max: 165

VEGETATED TREED UPLAND MIXED OPEN



Number of pixels plotted: 130454 Cell width: 3
Mean: 57.344 Median: 57 Standard Dev: 25.133 Min: 0 Max: 183

VEGETATED TREED UPLAND MIXED SPARSE



Number of pixels plotted: 49503 Cell width: 3
Mean: 68.365 Median: 70 Standard Dev: 26.240 Min: 5 Max: 184

APPENDIX C

MEAN REFLECTANCE PLOTS AND STANDARD DEVIATION PLOTS FOR TARGET LAND COVER CLASSES

Vegetated

Image Date: August 29, 1993

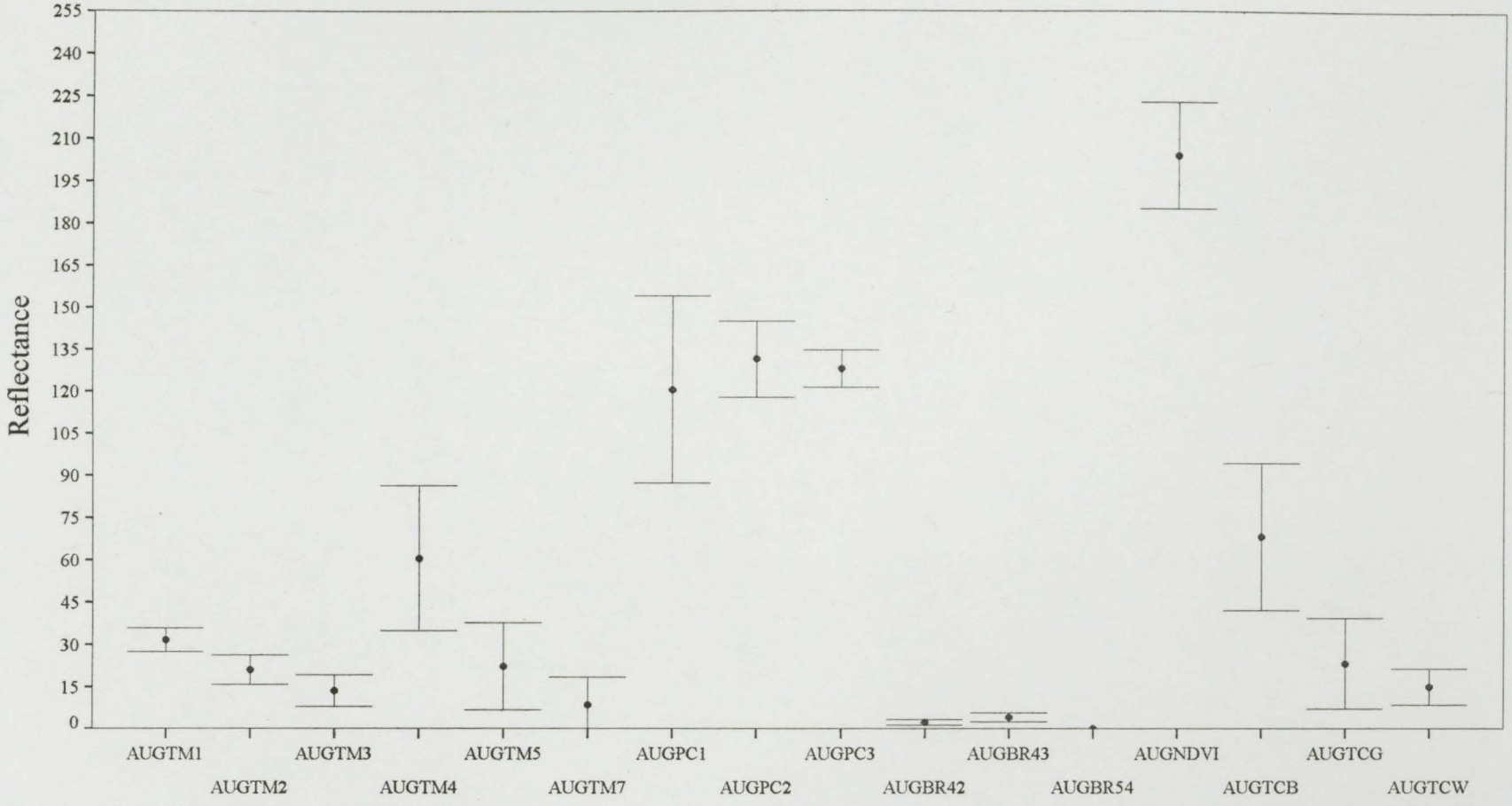


Image Channel

Non-Vegetated

Image Date: August 29, 1993

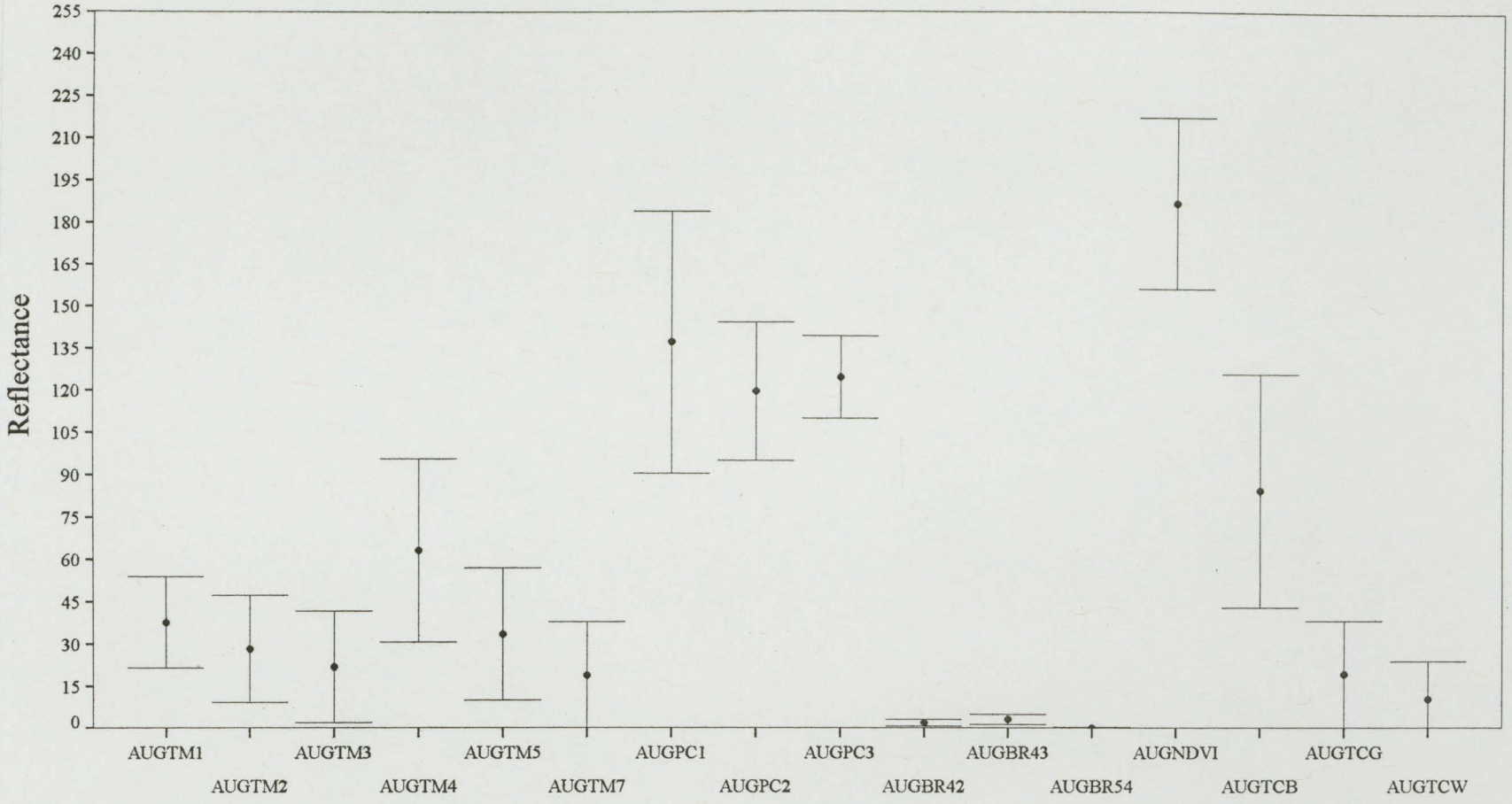


Image Channel

Vegetated Treed

Image Date: August 29, 1993

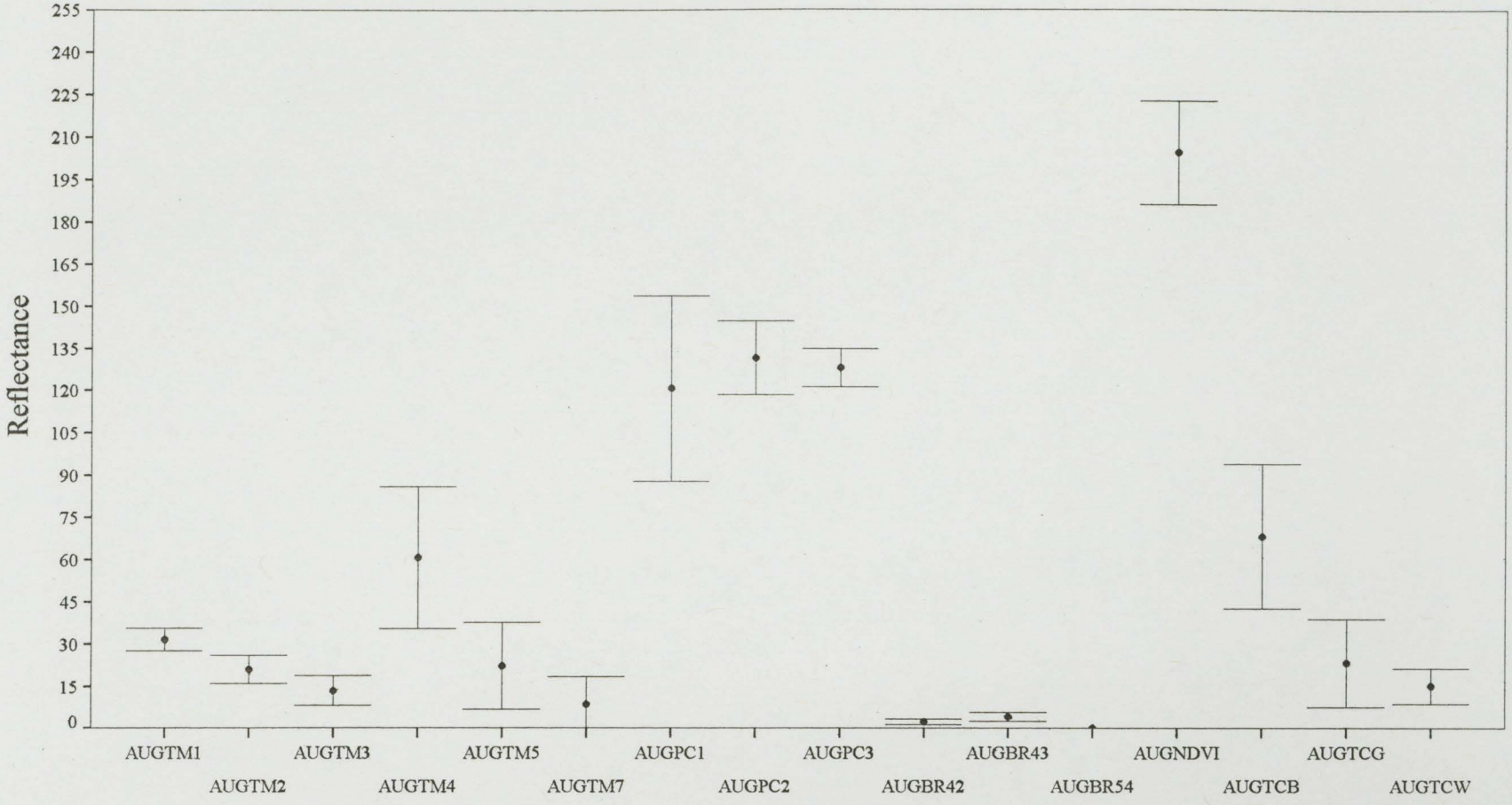


Image Channel

Vegetated Non-Treed

Image Date: August 29, 1993

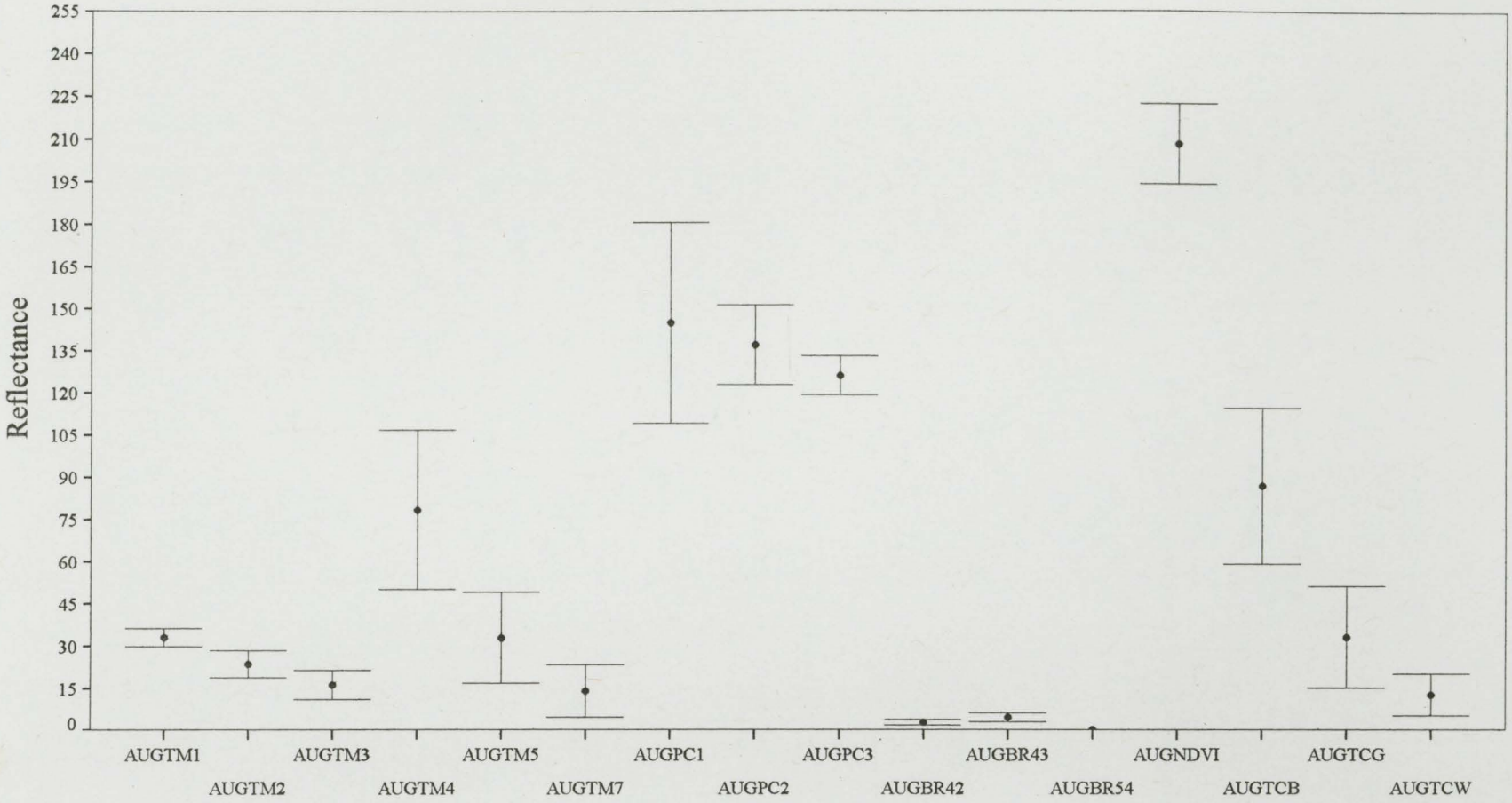
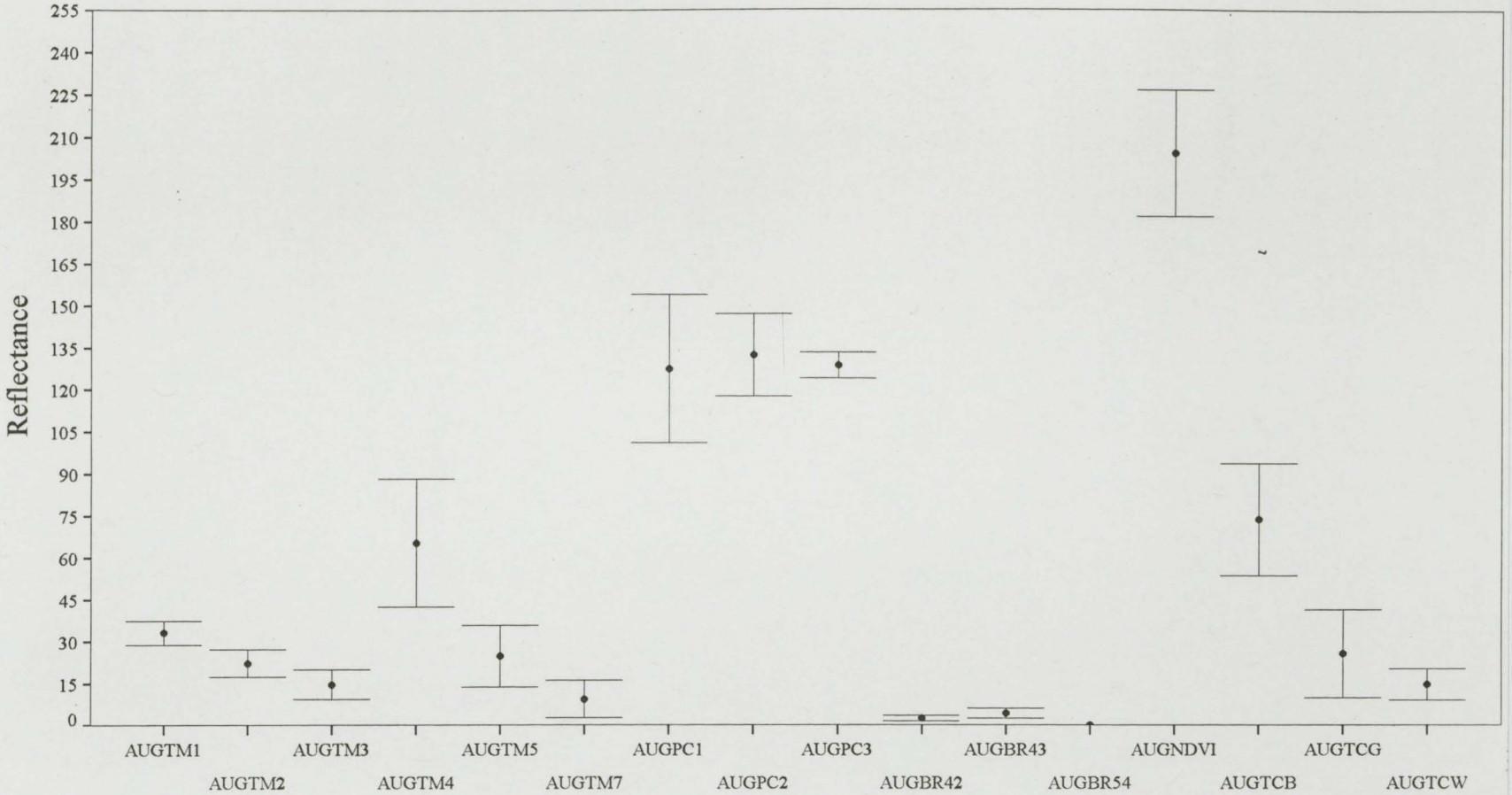


Image Channel

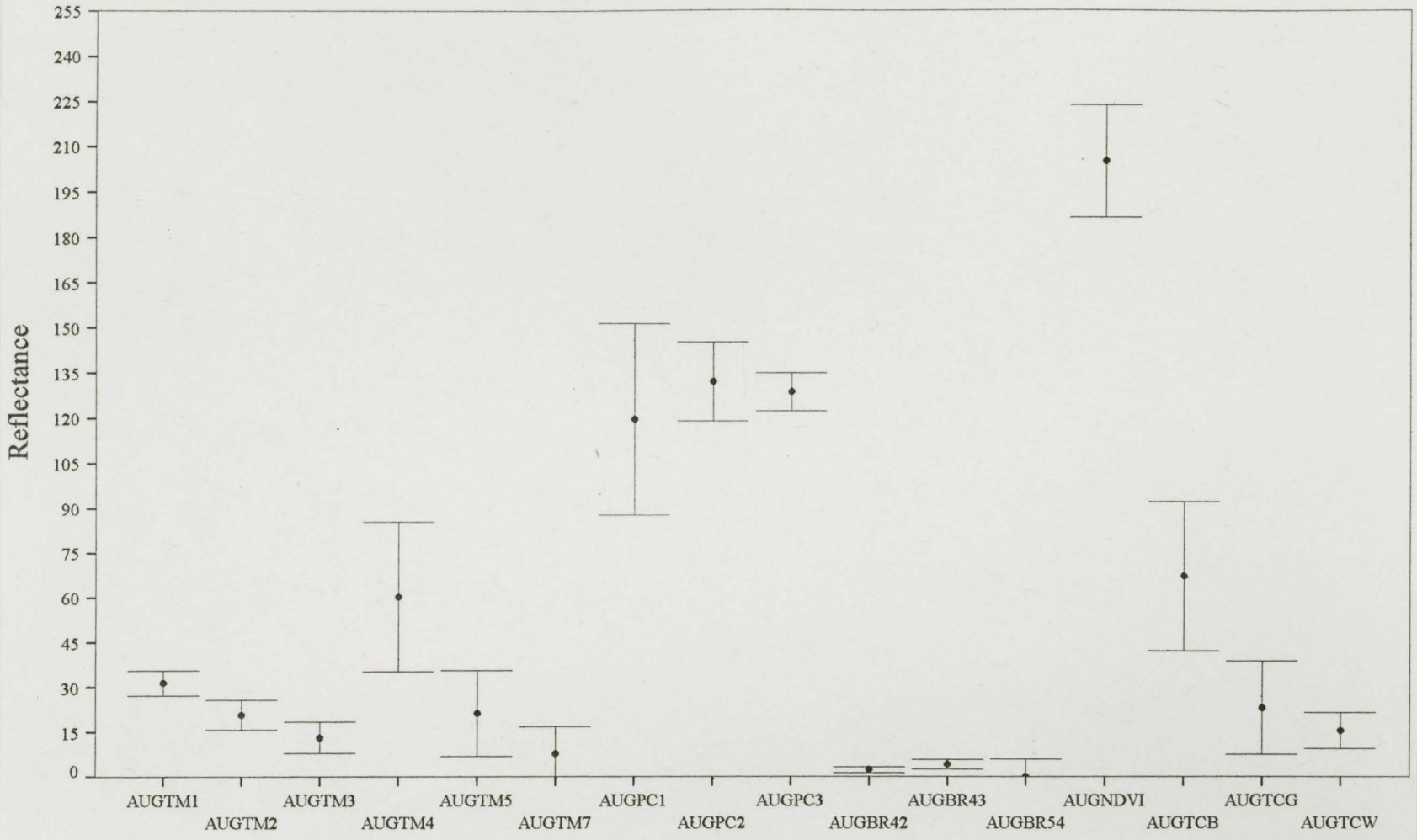
Vegetated Treed Wetland

Image Date: August 29, 1993



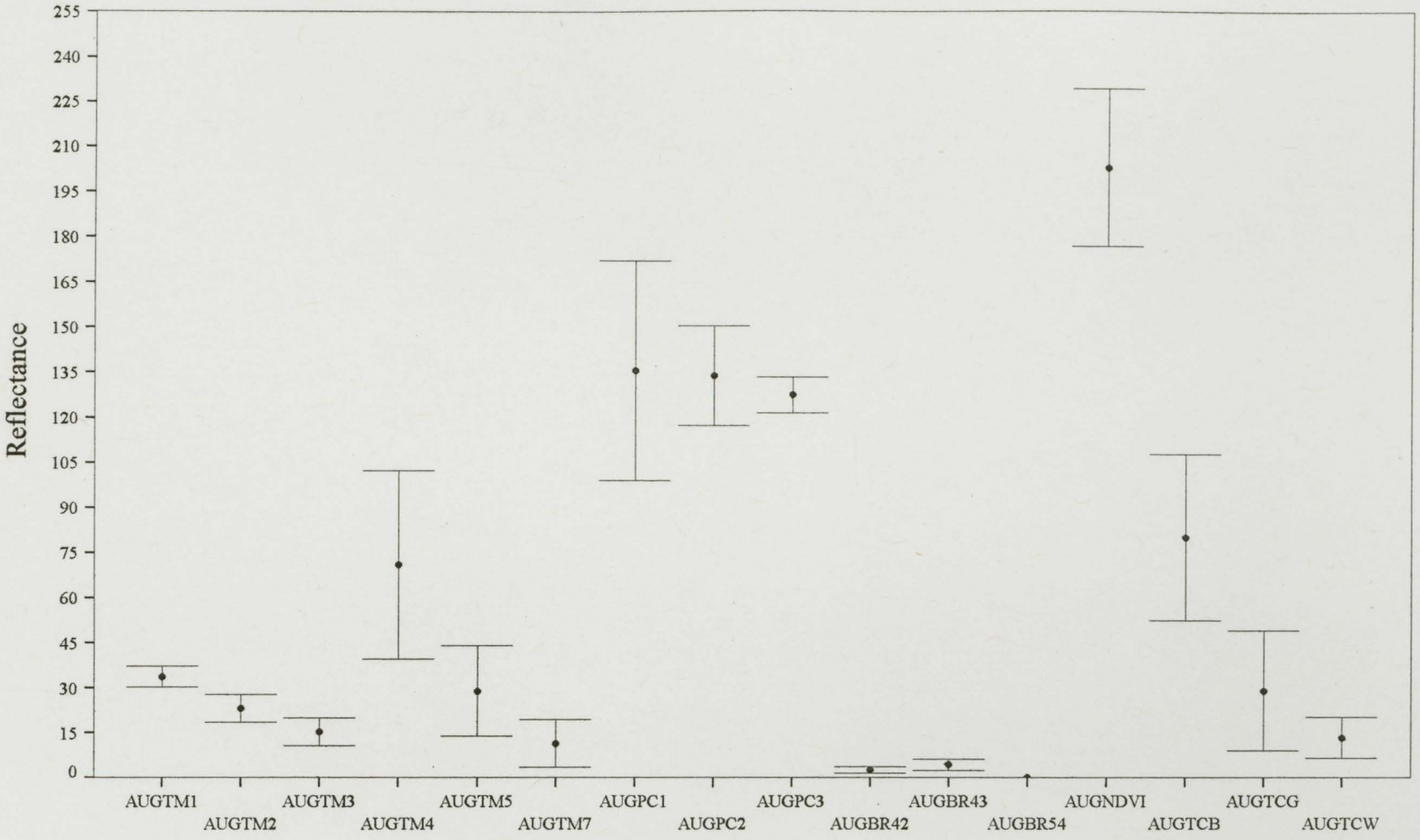
Vegetated Treed Upland

Image Date: August 29, 1993



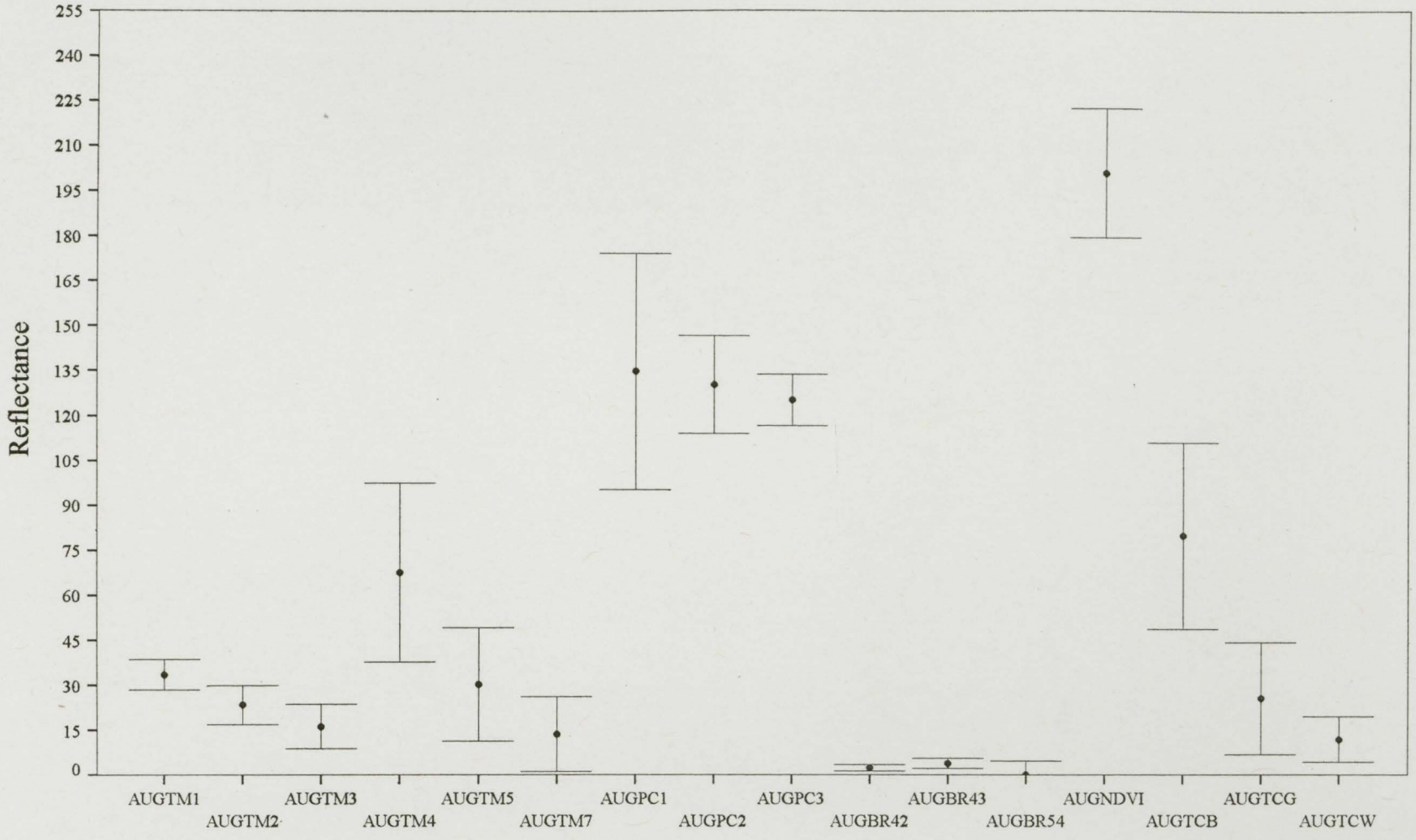
Non-Vegetated Land Wetland

Image Date: August 29, 1993



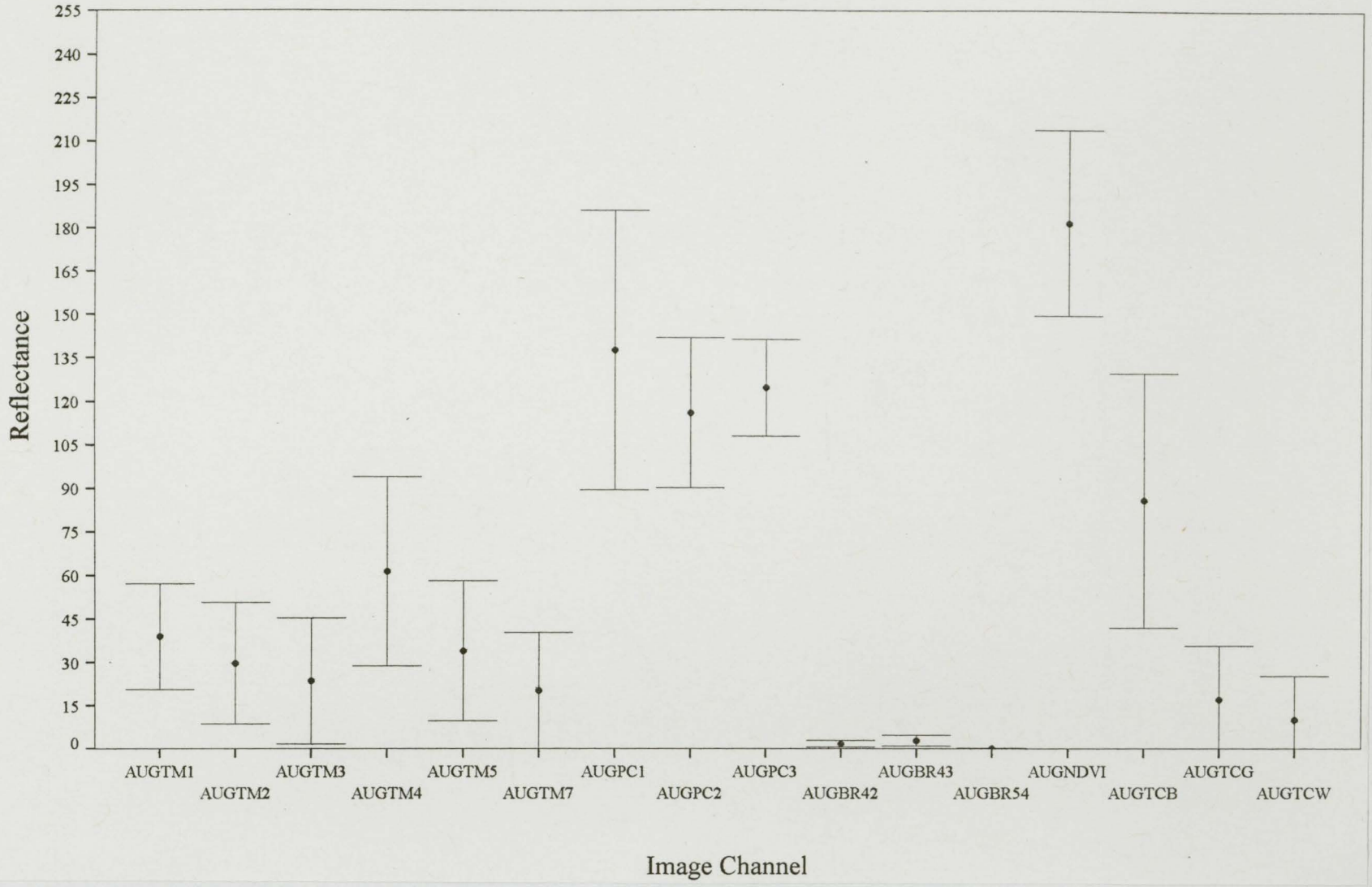
Non-Vegetated Land Upland

Image Date: August 29, 1993



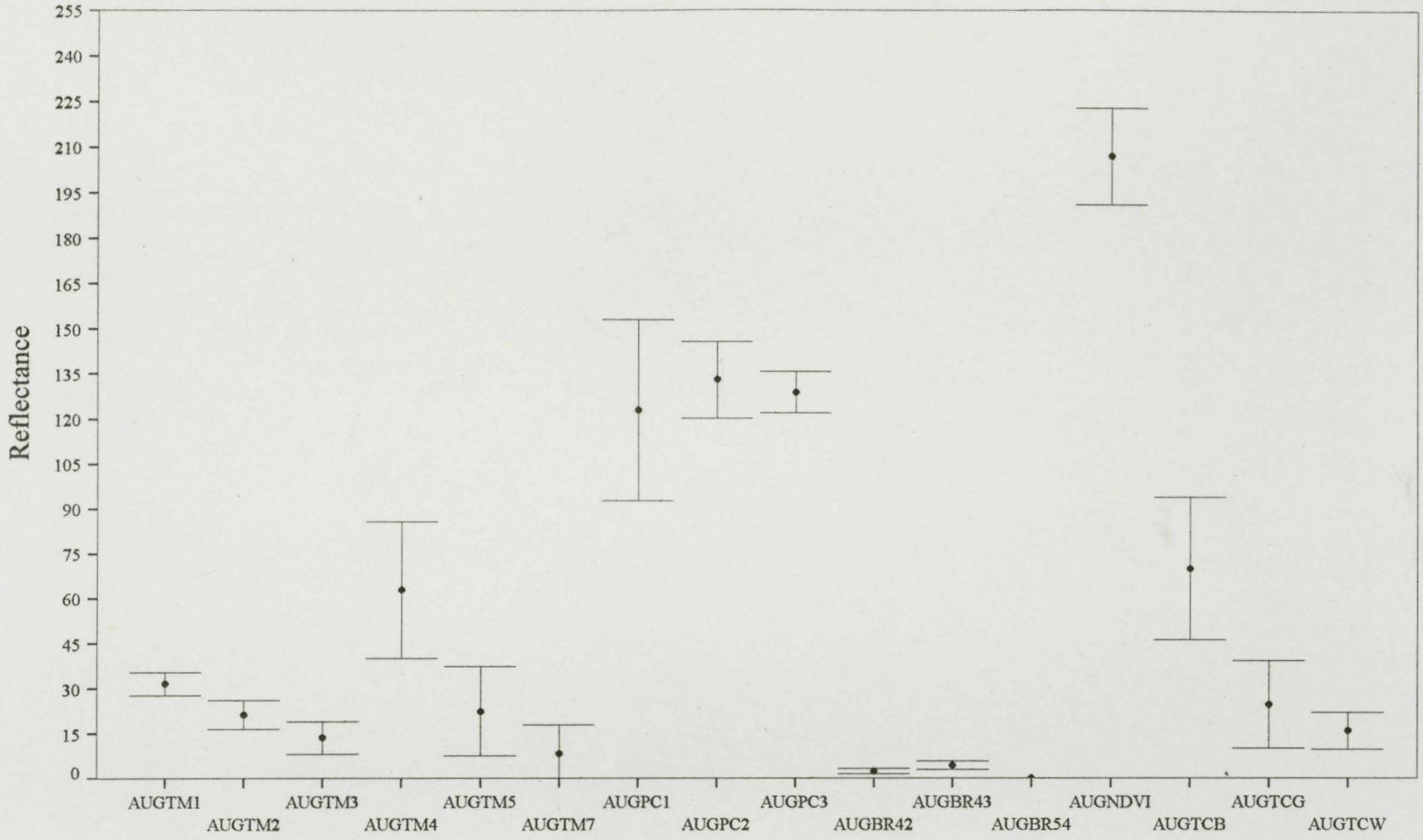
Non-Vegetated Land Alpine

Image Date: August 29, 1993



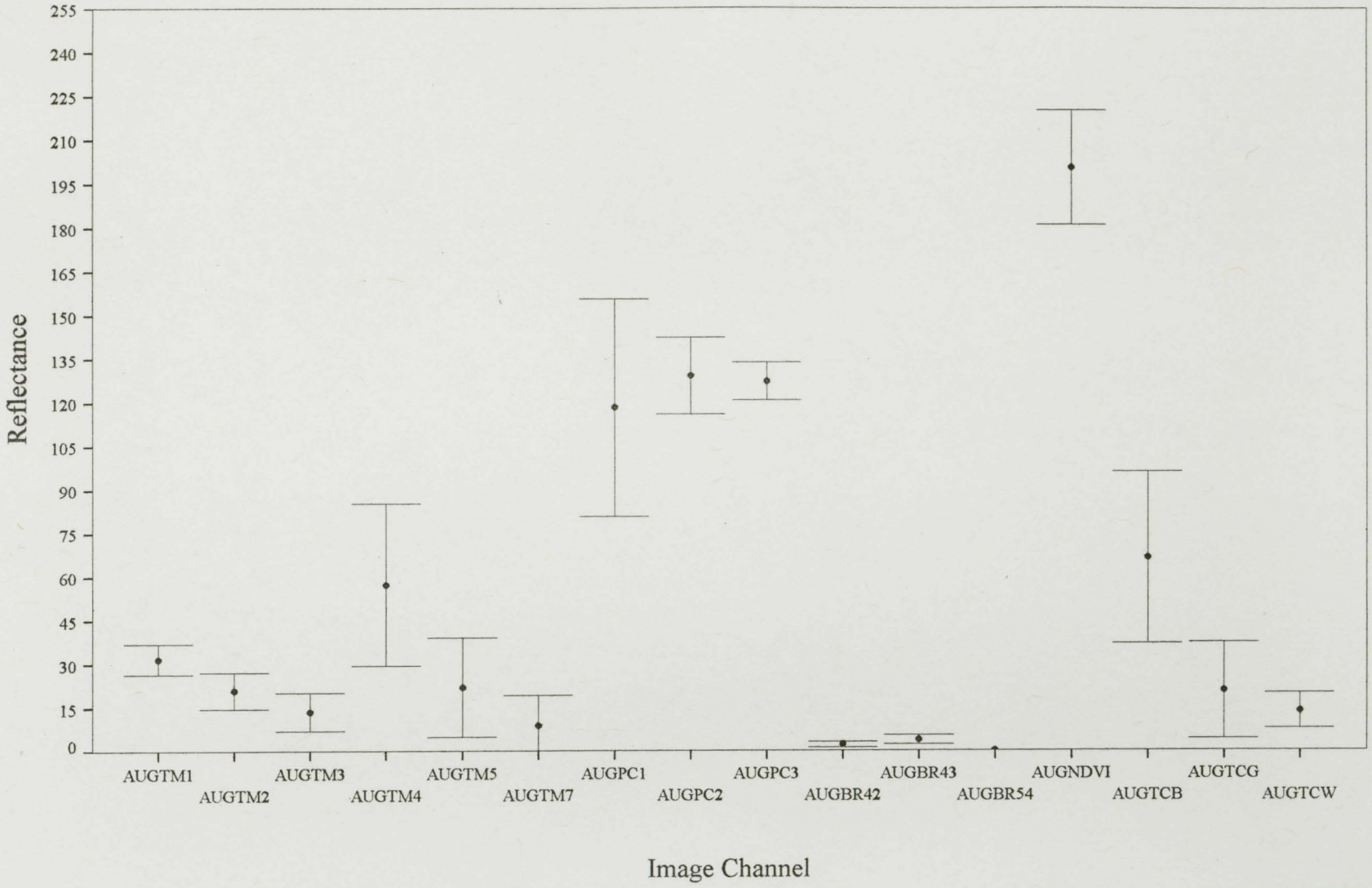
Vegetated Treed Upland Coniferous

Image Date: August 29, 1993



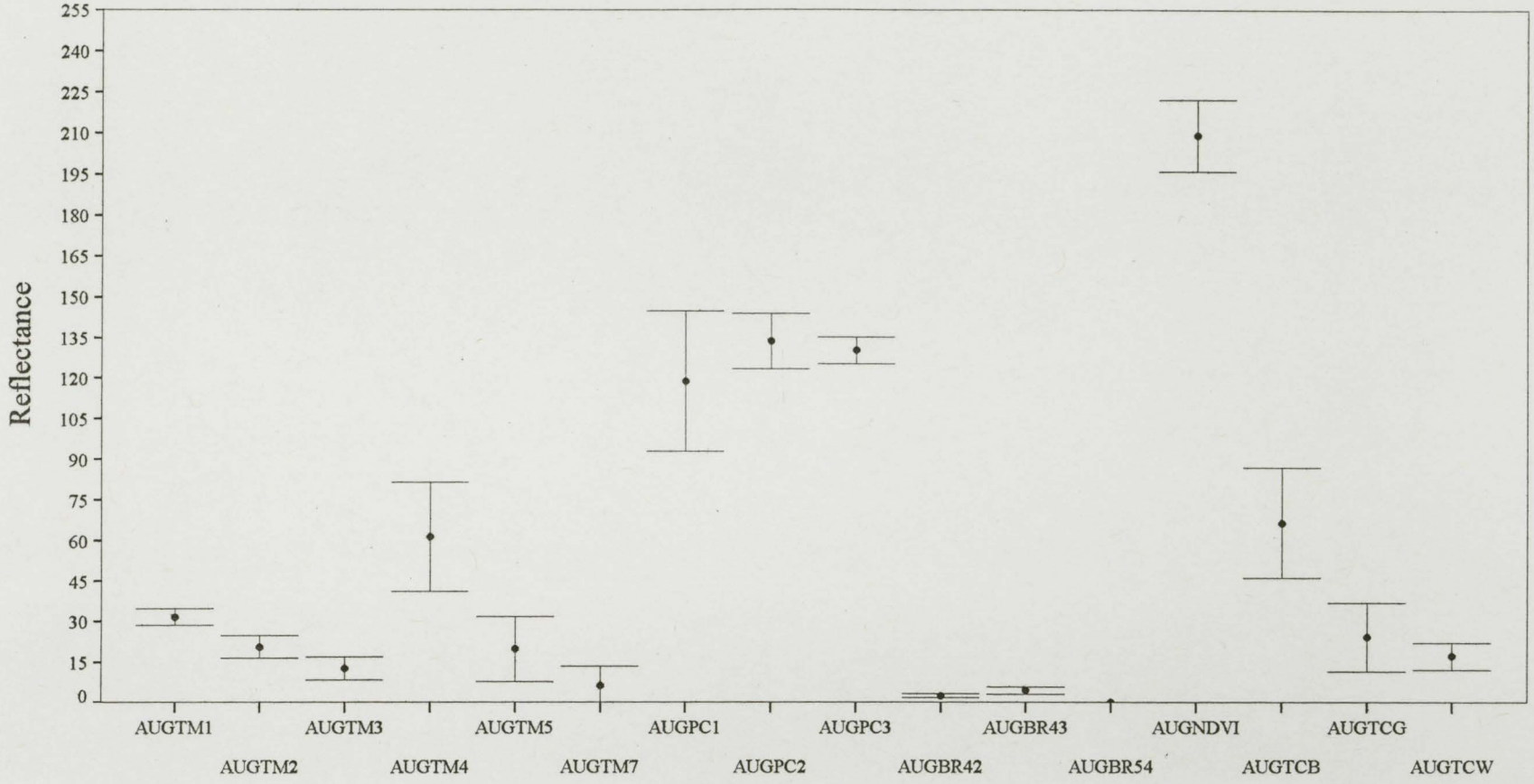
Vegetated Treed Upland Mixed

Image Date: August 29, 1993



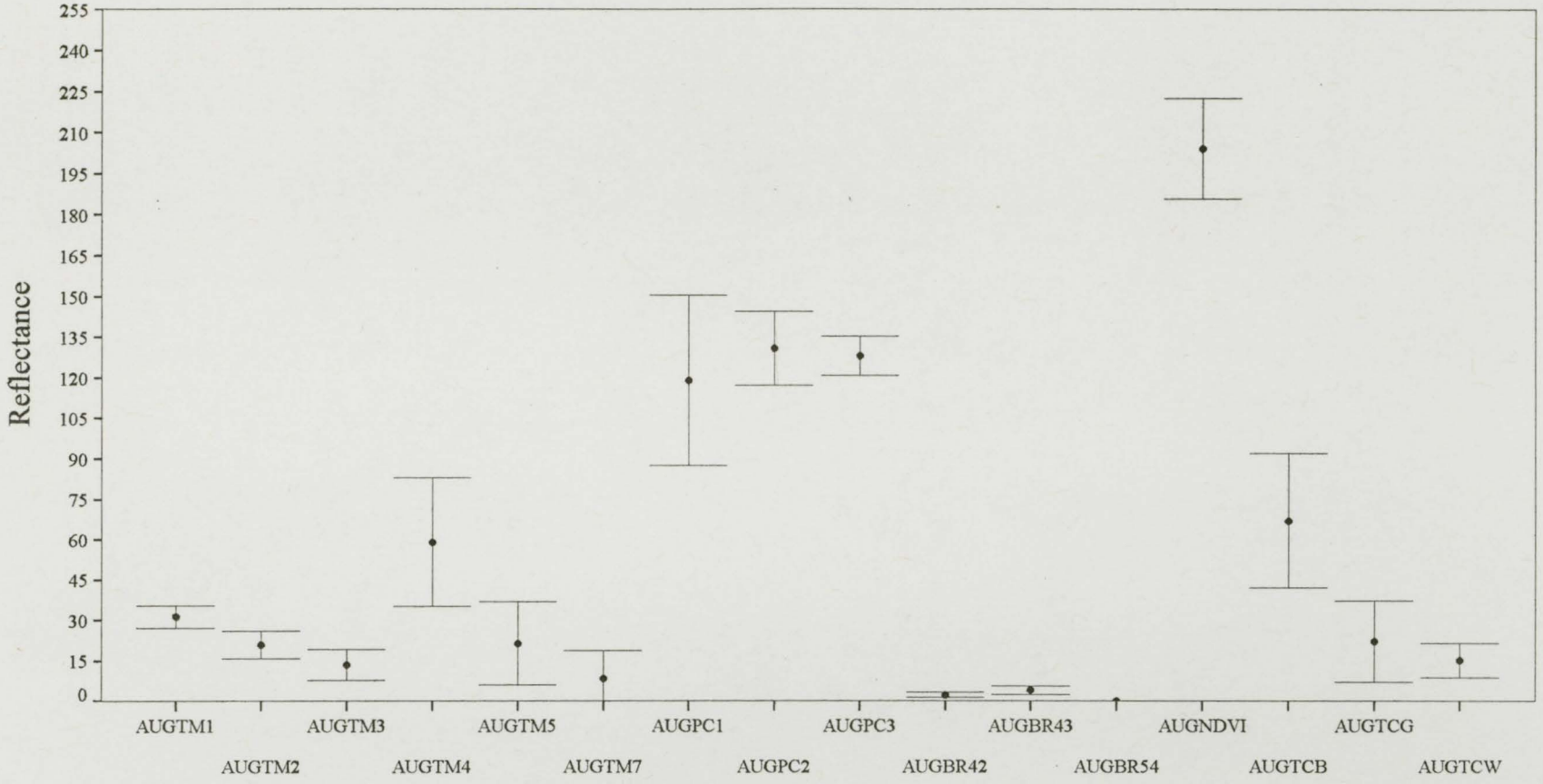
Vegetated Treed Upland Coniferous Dense

Image Date: August 29, 1993



Vegetated Treed Upland Coniferous Open

Image Date: August 29, 1993



Vegetated Treed Upland Coniferous Sparse

Image Date: August 29, 1993

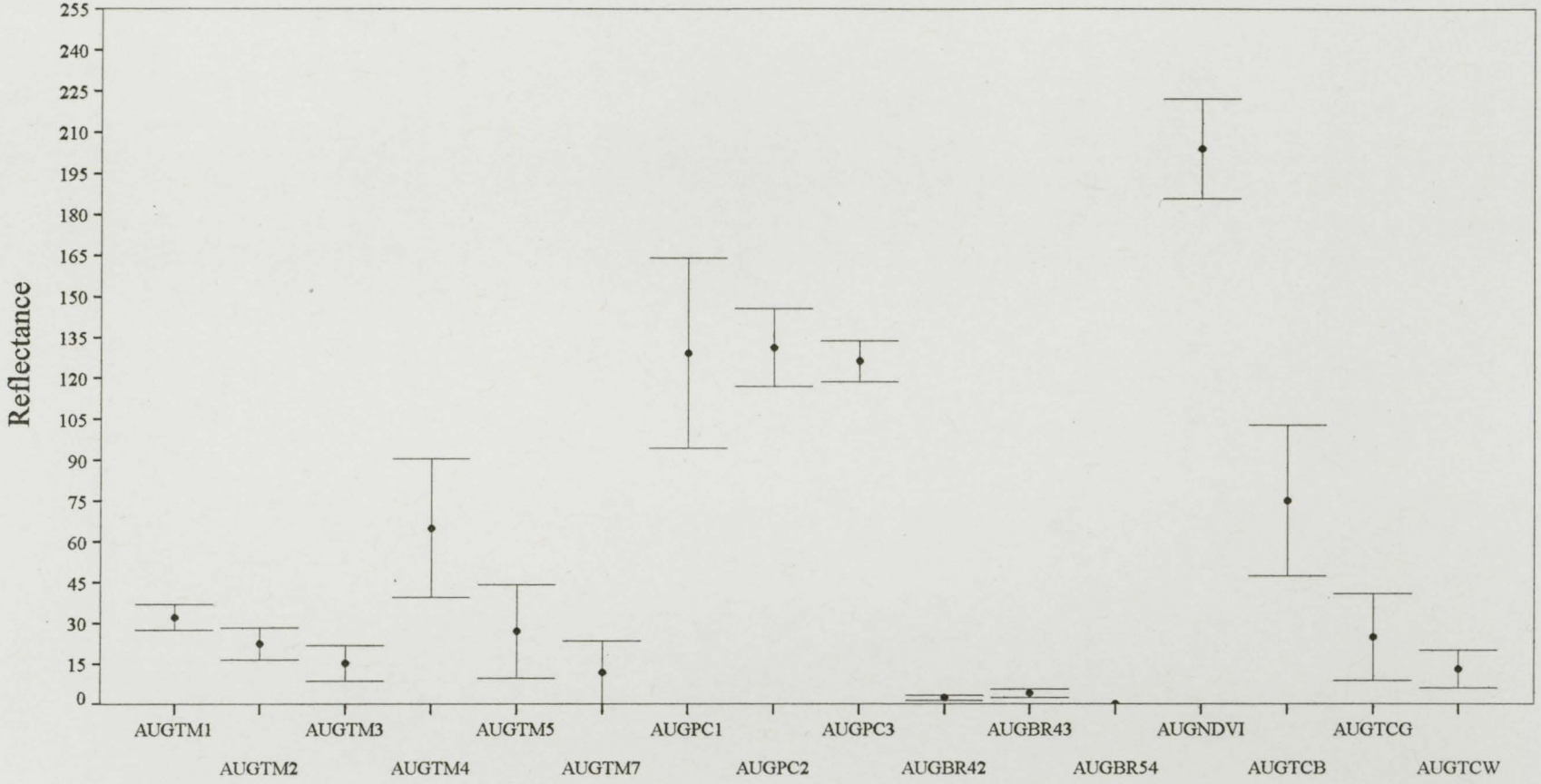
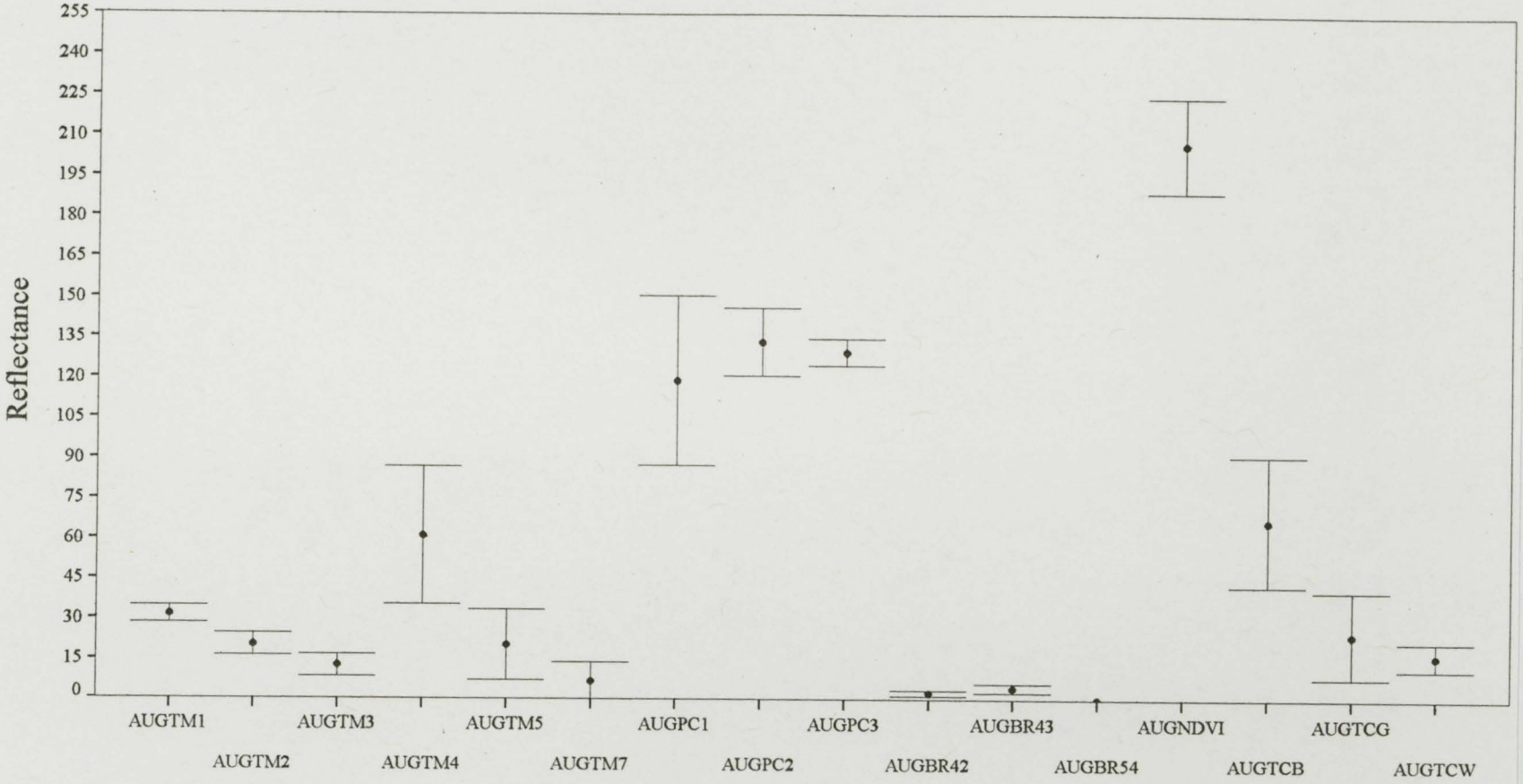


Image Channel

Vegetated Treed Upland Mixed Dense

Image Date: August 29, 1993



Vegetated Treed Upland Mixed Open

Image Date: August 29, 1993

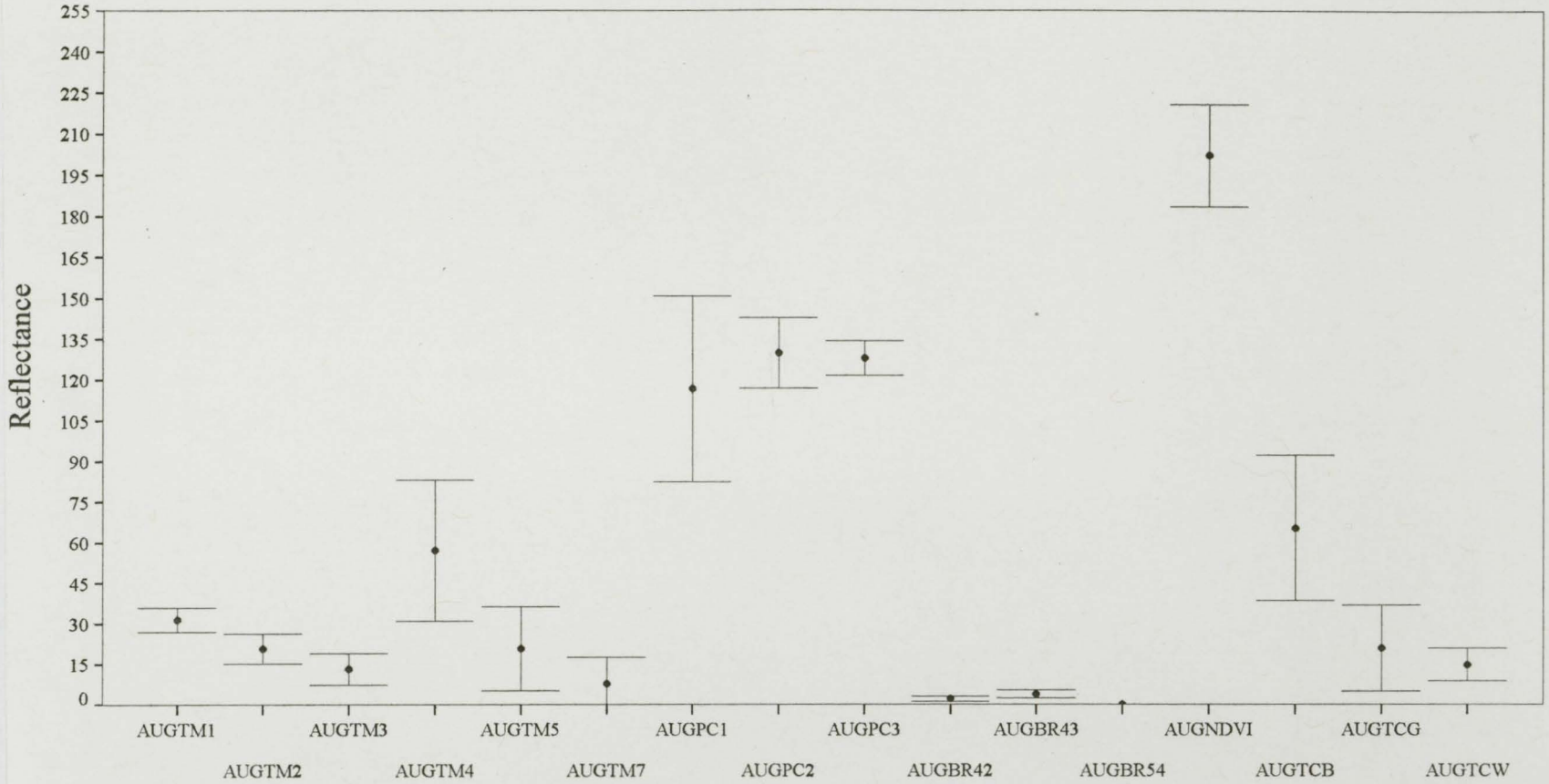
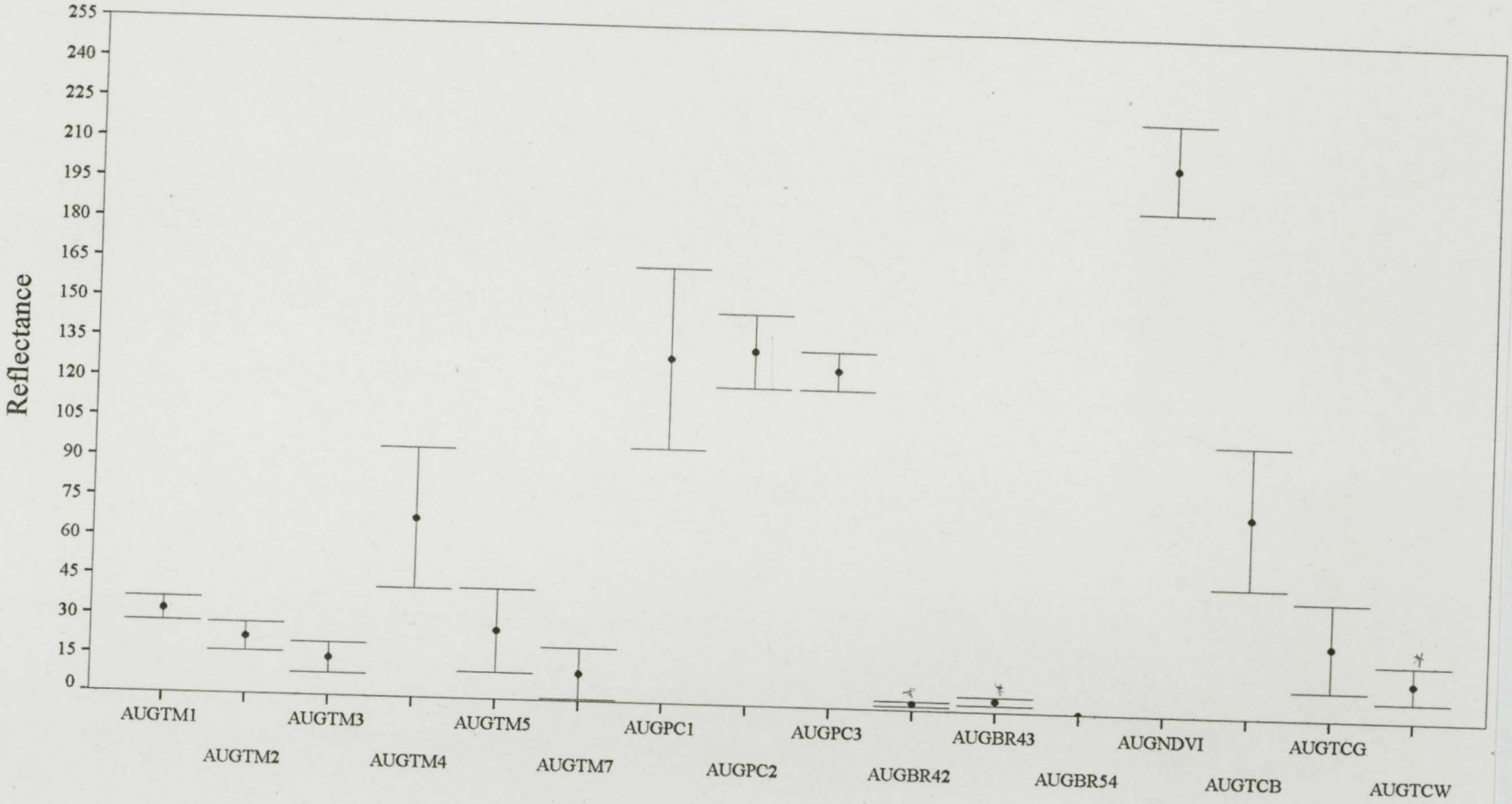


Image Channel

Vegetated Treed Upland Mixed Sparse

Image Date: August 29, 1993



APPENDIX D

SUMMARIES OF THE CONFIGURATION AND RESULTS OF THE CLASSIFICATION TRIALS

The "channels" listed in the tables included in Appendix D refer to the following "raw" and "synthetic" bands:

```
 1L[ 8U] III     TM BAND 1 - 290893
 2L[ 8U] III     TM BAND 2 - 290893
 3L[ 8U] III     TM BAND 3 - 290893
 4L[ 8U] III     TM BAND 4 - 290893
 5L[ 8U] III     TM BAND 5 - 290893
 6L[ 8U] III     TM BAND 7 - 290893
 7L[ 8U] III     Principle Component #1 Using TM Bands 1-7 - 290893
 8L[ 8U] III     Principle Component #2 Using TM Bands 1-7 - 290893
 9L[ 8U] III     Principle Component #3 Using TM Bands 1-7 - 290893
10L[ 8U] III     Band Ratio: TM4/TM3 - 290893
11L[ 8U] III     Band Ratio: TM4/TM2 - 290893
12L[ 8U] III     Band Ratio: TM5/TM4 - 290893
13L[ 8U] III     Normalized Difference Vegetation Index - 290893
14L[ 8U] III     Tasslled Cap Transformation: Brightness - 290893
15L[ 8U] III     Tasslled Cap Transformation: Greenness - 290893
16L[ 8U] III     Tasslled Cap Transformation: Wetness - 290893
```

LEVEL 1 LAND BASE
RESULTS OF CLASSIFICATION TRIALS

Trial	Channels	Architecture	Iterations	Overall	P_p(NV)	P_p(V)	P_u(NV)	P_u(V)	Khat	z-statistic	Foody Kappa
1	3,5,7,9,11,16	6-6-2	20000	0.73890	0.170	0.911	0.366	0.783	0.099840	2.244370	0.4778
2	3,5,7,9,11,16	6-6-2	20000	0.70540	0.145	0.905	0.355	0.747	0.063120	1.379790	0.4108
3	3,5,7,9,11,16	6-6-2	20000	0.74970	0.148	0.902	0.278	0.806	0.060960	1.433820	0.4994
4	3,5,7,9,11,16	6-6-2	20000	0.69940	0.136	0.902	0.333	0.744	0.046930	1.024470	0.3988
5	3,5,7,9,11,16	6-24-6-2	1000	0.75420	0.175	0.910	0.344	0.803	0.104020	2.373920	0.5084
6	3,5,7,9,11,16	6-24-6-2	1000	0.81530	0.153	0.897	0.156	0.895	0.050230	1.276950	0.6306
7	3,5,7,9,11,16	6-24-2	1000	0.74820	0.155	0.904	0.300	0.802	0.072630	1.684840	0.4964
8	3,5,7,9,11,16	6-24-2	1000	0.82613	0.156	0.907	0.169	0.899	0.065040	1.588350	0.6523
9	3,5,7,9,11,16	6-2-2	10000	0.76380	0.159	0.904	0.278	0.823	0.075620	1.788270	0.5276
10	3,5,7,9,11,16	6-2-2	20000	0.77098	0.161	0.904	0.267	0.832	0.076590	1.826020	0.5420
11	3,5,7,9,11,16	6-2-2	50000	0.67505	0.148	0.910	0.422	0.706	0.070430	1.471140	0.3501
12	3,5,7,9,11,16	6-6-2	10000	0.71822	0.166	0.913	0.400	0.757	0.096730	2.122120	0.4364
13	3,5,7,9,11,16	6-6-2	20000	0.71822	0.163	0.911	0.389	0.758	0.091250	2.006020	0.4364
14	3,5,7,9,11,16	6-6-2	50000	0.73621	0.165	0.909	0.356	0.782	0.091400	2.056730	0.4724
15	1,2,3,4,5,6	6-6-2	10000	0.77338	0.172	0.906	0.289	0.832	0.093130	2.188350	0.5468
16	3,5,7,9,11,16	6-6-2	1000	0.78657	0.172	0.904	0.256	0.851	0.087550	2.093300	0.5731
17	3,5,7,9,11,16	6-6-2	5000	0.85612	0.212	0.899	0.122	0.945	0.082410	1.759990	0.7122
18	3,5,7,9,11,16	6-6-2	10000	0.63309	0.142	0.912	0.478	0.652	0.063710	1.264710	0.2662
19	3,5,7,9,11,16	6-24-6-2	1000	0.50839	0.130	0.915	0.622	0.495	0.043760	0.771140	0.0168
20	3,5,7,9,11,16	6-24-6-2	5000	0.87170	0.207	0.896	0.067	0.969	0.050920	0.976350	0.7434
21	3,5,7,9,11,16	6-24-6-2	10000	0.63309	0.152	0.918	0.522	0.647	0.081340	1.619090	0.2662
22	1,2,3,4,5,6	6-24-6-2	1000	0.72662	0.158	0.908	0.356	0.772	0.082140	1.834050	0.4532
23	1,2,3,4,5,6	6-24-6-2	5000	0.67745	0.160	0.916	0.467	0.703	0.091940	1.922230	0.3549
24	1,2,3,4,5,6	6-24-6-2	10000	0.72420	0.153	0.907	0.344	0.770	0.074080	1.656180	0.4484
25	3,5,7,9,11,16	6-6-2	10000	0.67780	0.149	0.910	0.422	0.781	0.072510	1.519900	0.3556
26	3,5,7,9,11,16	n/a	n/a	0.81794	0.148	0.897	0.144	0.899	0.044210	1.134790	0.6359
27	3,5,7,9,11,16	6-24-2	10000	0.61631	0.136	0.909	0.477	0.633	0.052690	1.025150	0.2326
28	3,5,7,9,11,16	n/a	n/a	0.81774	0.147	0.897	0.144	0.899	0.044070	1.109942	0.6355

LEVEL 2 LAND COVER TYPE

RESULTS OF CLASSIFICATION TRIALS

Trial	Channels	Architecture	Iterations	Overall	P_p(VT)	P_p(VN)	P_u(VT)	P_u(VN)	Khat	z-statistic	Foody Kappa
1	1,3,5,7,8,15	6-6-2	20000	0.7733	0.9570	0.0697	0.7970	0.3000	0.037910	0.611230	0.5466
2	1,3,5,7,8,15	6-6-2	20000	0.7267	0.9610	0.0732	0.7430	0.4000	0.045910	0.707300	0.4534
3	1,3,5,7,8,15	6-6-2	20000	0.7701	0.9570	0.0687	0.7940	0.3000	0.036150	0.581000	0.5402
4	1,3,5,7,8,15	6-6-2	20000	0.7041	0.9490	0.0416	0.7280	0.2330	-0.012130	-0.182490	0.4082
5	1,3,5,7,8,15	n/a	n/a	0.3906	0.9610	0.0537	0.3750	0.7000	0.011170	0.147410	-0.2187
6	1,3,5,7,8,15	6-6-2	10000	0.7697	0.9583	0.0625	0.7931	0.2857	0.034080	0.542830	0.5394
7	1,3,5,7,8,15	6-24-6-2	1000	0.9517	0.9517	0.0000	1.0000	0.0000	0.000000	0.000000	-0.9035
8	1,3,5,7,8,15	6-24-6-2	5000	0.9517	0.9517	0.0000	1.0000	0.0000	0.000000	0.000000	-0.9035
9	1,3,5,7,8,15	6-24-6-2	10000	0.9517	0.9517	0.0000	1.0000	0.0000	0.000000	0.000000	-0.9035
10	1,2,3,4,5,6	6-6-2	5000	0.7958	0.9603	0.0855	0.8193	0.3333	0.064210	1.049840	0.5916
11	1,2,3,4,5,6	6-6-2	10000	0.5900	0.9591	0.0588	0.5945	0.5000	0.020740	0.292310	0.1801
12	1,3,5,7,8,15	n/a	n/a	0.4196	0.9639	0.0563	0.4054	0.7000	0.016400	0.218740	-0.1608

LEVEL 3 LANDSCAPE POSITION
(VEGETATED TREED)

RESULTS OF CLASSIFICATION TRIALS

Trial	Channels	Architecture	Iterations	Overall	$P_p(W)$	$P_p(U)$	$P_u(W)$	$P_u(U)$	Khat	z-statistic	Foody Kappa
1	3,5,6,9,10,16	6-6-2	20000	0.8345	0.000	0.994	0.000	0.839	-0.010440	-0.129210	0.6690
2	3,5,6,9,10,16	6-6-2	10000	0.8132	0.000	0.993	0.000	0.817	-0.010480	-0.128690	0.6264
3	3,5,6,9,10,16	6-6-2	10000	0.8132	0.010	0.995	0.333	0.816	0.008400	0.103230	0.6264
4	3,5,6,9,10,16	6-6-2	10000	0.7989	0.000	0.993	0.000	0.803	-0.010500	-0.128400	0.5978
5	3,5,6,9,10,16	n/a	n/a	0.3487	0.005	0.995	0.666	0.347	0.000220	0.002660	-0.3025
6	3,5,6,9,10,16	6-6-2	10000	0.7752	1.000	0.993	0.000	0.779	-0.010640	-0.128710	0.5504
7	3,5,6,9,10,16	6-24-6-2	1000	0.4644	0.006	0.996	0.667	0.463	0.002570	0.030760	-0.0712
8	3,5,6,9,10,16	6-24-6-2	5000	0.9893	0.000	0.995	0.000	0.995	-0.005370	-2.502080	0.9786
9	3,5,6,9,10,16	6-24-6-2	10000	0.9893	0.000	0.995	0.000	0.995	-0.005370	-2.502080	0.9786
10	3,5,6,9,10,16,167,168,169	9-36-6-2	1000	0.9946	0.000	0.995	0.000	1.000	0.000000	0.000000	0.9893
11	3,5,6,9,10,16,167,168,170	9-36-6-2	5000	0.9946	0.000	0.995	0.000	1.000	0.000000	0.000000	0.9893
12	3,5,6,9,10,16,167,168,171	9-36-6-2	10000	0.9946	0.000	0.995	0.000	1.000	0.000000	0.000000	0.9893
13	1,2,3,4,5,6	6-6-2	5000	0.5747	0.004	0.994	0.333	0.576	-0.002270	-0.027130	0.1495
14	1,2,3,4,5,6	6-6-2	10000	0.6263	0.005	0.994	0.333	0.628	-0.001100	-0.013240	0.2527
15	1,2,3,4,5,6,177,178,179	9-9-2	5000	0.4751	0.003	0.993	0.333	0.476	-0.003880	-0.046090	-0.0498
16	1,2,3,4,5,6,177,178,179	9-9-2	10000	0.8737	0.000	0.994	0.000	0.878	-0.010330	-0.130690	0.7473
17	1,2,3,4,5,6,elevation	7-7-2	5000	0.7437	0.014	0.998	0.667	0.744	0.016740	0.204810	0.4875
18	1,2,3,4,5,6,elevation	7-7-2	10000	0.7722	0.016	0.998	0.667	0.773	0.020080	0.246360	0.5445
19	1,2,3,4,5,6,slope	7-7-2	5000	0.7989	0.018	0.998	0.666	0.800	0.024040	0.295630	0.5979
20	1,2,3,4,5,6,slope	7-7-2	10000	0.8007	0.018	0.997	0.666	0.801	0.024340	0.299360	0.6014
21	3,5,6,9,10,16	n/a	n/a	0.3576	0.005	0.995	0.667	0.356	0.000374	0.004460	-0.2847

LEVEL 3 LANDSCAPE POSITION (NON-VEGETATED)

RESULTS OF CLASSIFICATION TRIALS

Trial	Channels	Architecture	Iterations	Overall	P_p (W)	P_p (U)	P_p (A)	P_u (W)	P_u (U)	P_u (A)	Khat	z-statistic	Foody Kappa
1	3,5,6,9,10,16	6-6-3	10000	0.50000	0.000	0.913	0.214	0.000	0.532	0.273	0.060760	0.467390	0.2500
2	3,5,6,9,10,16	6-6-3	20000	0.72220	0.000	1.000	0.200	0.000	0.705	1.000	0.217390	0.681950	0.5833
3	3,5,6,9,10,16	6-6-3	20000	0.55560	0.000	1.000	0.167	0.000	0.529	1.000	0.127270	0.401220	0.3334
4	3,5,6,9,10,16	6-6-3	20000	0.22230	0.000	0.800	0.000	0.000	0.235	0.000	-0.086210	-0.226820	-0.1665
5	3,5,6,9,10,16	n/a	n/a	0.22220	0.000	1.000	0.000	0.000	0.235	0.000	0.011760	0.028790	-0.1667
6	3,5,6,9,10,16	6-6-3	10000	0.44444	0.090	0.000	1.000	0.143	0.000	0.418	0.090910	0.280130	0.1667
7	3,5,6,9,10,16	6-24-9-3	1000	0.50000	0.000	0.900	0.000	0.000	0.529	0.000	-0.065790	-0.174560	0.2500
8	3,5,6,9,10,16	6-24-9-3	5000	0.27778	0.000	0.833	0.000	0.000	0.294	0.000	-0.063640	-0.153010	-0.0833
9	3,5,6,9,10,16	6-24-9-3	10000	0.55556	0.000	0.909	0.000	0.000	0.588	0.000	-0.082710	-0.238420	0.3333
10	3,5,6,9,10,16,167,168,170	9-36-9-3	1000	0.22222	0.000	1.000	0.000	0.000	0.235	0.000	0.015630	0.036300	-0.1667
11	3,5,6,9,10,16,167,168,171	9-36-9-3	5000	0.22222	0.000	0.800	0.000	0.000	0.235	0.000	-0.054390	-0.116200	-0.1667
12	3,5,6,9,10,16,167,168,172	9-36-9-3	10000	0.44444	0.000	0.889	0.000	0.000	0.471	0.000	-0.052630	-0.118260	0.1667
13	1,2,3,4,5,6	6-6-3	5000	0.33333	0.000	0.857	0.000	0.000	0.353	0.000	-0.085430	-0.227740	0.0000
14	1,2,3,4,5,6	6-6-3	10000	0.38888	0.000	0.875	0.000	0.000	0.412	0.000	-0.081970	-0.222330	0.0833
15	1,2,3,4,5,6,177,178,179	9-9-3	5000	0.88888	0.000	0.941	0.000	0.000	0.941	0.000	-0.028570	-0.115830	0.8333
16	1,2,3,4,5,6,177,178,179	9-9-3	10000	0.33330	0.000	0.857	0.000	0.000	0.353	0.000	-0.053660	-0.117050	0.0000
17	1,2,3,4,5,6,elevation	7-7-3	5000	0.72220	0.000	0.929	0.000	0.000	0.765	0.000	-0.071430	-0.253400	0.5833
18	1,2,3,4,5,6,elevation	7-7-3	10000	0.66666	0.000	0.923	0.000	0.000	0.071	0.000	-0.080000	-0.256430	0.5000
19	1,2,3,4,5,6,slope	7-7-3	5000	0.77777	0.000	1.000	0.000	0.000	0.824	0.000	0.132530	0.502210	0.6667
20	1,2,3,4,5,6,slope	7-7-3	10000	0.77777	0.000	1.000	0.000	0.000	0.824	0.000	0.142860	0.586400	0.6667
21	3,5,6,9,10,16	n/a	n/a	0.27778	0.000	1.000	0.000	0.000	0.294	0.000	0.016807	0.041850	-0.0830

LEVEL 4 VEGETATION TYPE
(VEGETATED TREED UPLAND)

RESULTS OF CLASSIFICATION TRIALS

Trial	Channels	Architecture	Iterations	Overall	P_p(C)	P_p(B)	P_p(M)	P_u(C)	P_u(B)	P_u(M)	Khat	z-statistic	Foody Kappa
1	2,4,5,8,13,15	6-6-3	10000	0.3113	0.272	0.015	0.724	0.545	0.333	0.225	0.006710	0.129060	-0.0330
2	2,4,5,8,13,15	6-6-3	20000	0.3260	0.189	0.019	0.679	0.324	0.333	0.327	-0.074810	-1.534290	-0.0110
3	2,4,5,8,13,15	6-6-3	20000	0.2747	0.255	0.014	0.767	0.566	0.333	0.167	0.002540	0.046440	-0.0879
4	2,4,5,8,13,15	6-6-3	20000	0.3699	0.246	0.007	0.736	0.331	0.167	0.387	-0.005360	-0.120470	0.0549
5	2,4,5,8,13,15	n/a	n/a	0.3699	0.259	0.000	0.711	0.614	0.000	0.286	-0.013560	-0.223620	-0.4698
6	2,4,5,8,13,15	6-6-3	10000	0.2989	0.222	0.010	0.729	0.273	0.400	0.307	-0.018640	-0.392900	-0.0517
7	2,4,5,8,13,15	6-24-9-3	1000	0.7234	0.000	0.000	0.723	0.000	0.000	1.000	0.000000	0.000000	0.5852
8	2,4,5,8,13,15	6-24-9-3	5000	0.7234	0.000	0.000	0.723	0.000	0.000	1.000	0.000000	0.000000	0.5852
9	2,4,5,8,13,15	6-24-9-3	10000	0.7234	0.000	0.000	0.723	0.000	0.000	1.000	0.000000	0.000000	0.5852
10	1,2,3,4,5,6	6-6-3	5000	0.2473	0.254	0.018	0.813	0.552	0.500	0.132	0.007750	0.143950	-0.1291
11	1,2,3,4,5,6	6-6-3	10000	0.2582	0.251	0.018	0.808	0.517	0.500	0.159	0.008110	0.153880	-0.1126
12	2,4,5,8,13,15	n/a	n/a	0.3717	0.257	0.022	0.717	0.607	0.167	0.289	-0.009900	-0.163070	0.0577

LEVEL 5 DENSITY CLASSES
(VEGETATED TREED UPLAND CONIFEROUS)

CLASSIFICATION TRIAL RESULTS

Trial	Channels	Architecture	Iterations	Overall	$P_p(D)$	$P_p(O)$	$P_p(S)$	$P_u(D)$	$P_u(O)$	$P_u(S)$	Khat	z-statistic	Foody Kappa
1	3,4,6,10,13,15	6-6-3	10000	0.3400	0.484	0.500	0.059	0.789	0.036	0.333	0.085370	0.592570	0.0100
2	3,4,6,10,13,15	6-6-3	20000	0.2400	0.313	0.182	0.000	0.526	0.071	0.000	-0.215610	-1.194480	-0.1400
3	3,4,6,10,13,15	6-6-3	20000	0.3000	0.424	1.000	0.000	0.737	0.036	0.000	0.026150	0.167260	-0.0500
4	3,4,6,10,13,15	6-6-3	20000	0.3400	0.415	0.000	0.000	0.895	0.000	0.000	-0.003650	-0.018380	0.0100
5	3,4,6,10,13,15	n/a	n/a	0.4400	0.400	0.800	0.000	0.947	0.143	0.000	0.069770	0.318040	0.3100
6	3,4,6,10,13,15	6-6-3	10000	0.3958	0.047	0.447	0.500	0.000	0.895	0.077	0.046580	0.255470	0.0937
7	3,4,6,10,13,15	6-24-9-3	1000	0.0600	0.000	0.000	0.060	0.000	0.000	1.000	0.000000	0.000000	-0.4100
8	3,4,6,10,13,15	6-24-9-3	5000	0.0600	0.000	0.000	0.060	0.000	0.000	1.000	0.000000	0.000000	-0.4100
9	3,4,6,10,13,15	6-24-9-3	10000	0.0706	1.000	1.000	0.000	0.000	0.000	1.000	0.000000	0.000000	-0.3940
10	1,2,3,4,5,6	6-6-3	5000	0.4600	0.474	0.800	0.143	0.947	0.143	0.333	0.165120	1.044090	0.1900
11	1,2,3,4,5,6	6-6-3	10000	0.4200	0.462	0.667	0.125	0.947	0.071	0.333	0.121740	0.729180	0.1300
12	3,4,6,10,13,15	n/a	n/a	0.3932	0.391	1.00	0.333	0.947	0.036	0.333	0.056010	0.249860	0.1000

LEVEL 5 DENSITY CLASSES
(VEGETATED TREED UPLAND MIXED)

CLASSIFICATION TRIAL RESULTS

Trial	Channels	Structure	Iterations	Overall	P_p(NVLW)	P_p(NVLU)	P_p(NVLA)	P_u(NVLW)	P_u(NVLU)	P_u(NVLA)	Khat	z-statistic	Foody Kappa
1	3,4,6,10,13,15	6-6-3	10000	0.4203	0.442	0.500	0.281	0.585	0.279	0.360	0.090750	1.852200	0.1305
2	3,4,6,10,13,15	6-6-3	20000	0.3694	0.432	0.507	0.192	0.439	0.295	0.380	0.054580	1.185480	0.0542
3	3,4,6,10,13,15	6-6-3	20000	0.3559	0.421	0.444	0.184	0.496	0.230	0.320	0.019320	0.411870	0.0339
4	3,4,6,10,13,15	6-6-3	20000	0.4000	0.460	0.523	0.227	0.610	0.189	0.400	0.087280	1.765780	0.1000
5	3,4,6,10,13,15	n/a	n/a	0.4842	0.550	0.400	0.000	0.846	0.065	0.000	-0.014160	-0.119010	0.1203
6	3,4,6,10,13,15	6-6-3	10000	0.7751	1.000	0.993	0.000	0.000	0.779	0.000	0.066010	1.247960	0.6628
7	3,4,6,10,13,15	6-24-9-3	1000	0.1694	0.000	0.000	0.169	0.000	0.000	1.000	0.000000	0.000000	-0.2458
8	3,4,6,10,13,15	6-24-9-3	5000	0.1694	0.000	0.000	0.169	0.000	0.000	1.000	0.000000	0.000000	-0.2458
9	3,4,6,10,13,15	6-24-9-3	10000	0.1694	0.000	0.000	0.169	0.000	0.000	1.000	0.000000	0.000000	-0.2458
10	1,2,3,4,5,6	6-6-3	5000	0.3966	0.538	0.506	0.205	0.398	0.361	0.480	0.115520	2.499450	0.0949
11	1,2,3,4,5,6	6-6-3	10000	0.3966	0.538	0.506	0.205	0.398	0.361	0.480	-0.001070	-0.010980	0.0054
12	3,4,6,10,13,15	n/a	n/a	0.3932	0.425	0.333	0.293	0.756	0.049	0.340	0.039770	0.629100	0.0898

APPENDIX E

PLOTS OF THE CLASSIFICATION OUTPUT

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRAINING DATA

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 1
(August, 1993; 6-6-2 architecture ; 20000 iterations; P_C = 73.89%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 2
(September, 1993; 6-6-2 architecture ; 20000 iterations; P_C = 70.54%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 3
(July, 1994; 6-6-2 architecture ; 20000 iterations; P_C = 74.97%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 4
(October, 1994; 6-6-2 architecture ; 20000 iterations; P_C = 69.94%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 5
(August, 1993; 6-24-6-2 architecture ; 1000 iterations; P_C = 75.42%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 6
(August, 1993; 6-24-6-2 architecture ; 1000 iterations; P_C = 81.53%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 7
(August, 1993; 6-24-2 architecture ; 1000 iterations; P_C = 74.82%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 8
(August, 1993; 6-24-2 architecture ; 1000 iterations; P_C = 82.61%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 9
(August, 1993; 6-6-2 architecture ; 10000 iterations; P_C = 76.38%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 10
(August, 1993; 6-6-2 architecture ; 20000 iterations; P_C = 77.09%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 11

(August, 1993; 6-6-2 architecture ; 50000 iterations; P_C = 67.50%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 12
(August, 1993; 6-2-2 architecture ; 10000 iterations; P_C = 71.82%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 13
(August, 1993; 6-2-2 architecture ; 20000 iterations; P_C = 71.82%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 14
(August, 1993; 6-2-2 architecture ; 50000 iterations; P_C = 73.62%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 15
(August, 1993; 6-6-2 architecture ; 10000 iterations; P_C = 77.33%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 16
(August, 1993; 6-6-2 architecture ; 1000 iterations; P_C = 78.65%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 17
(August, 1993; 6-6-2 architecture ; 5000 iterations; P_C = 85.61%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 18
(August, 1993; 6-6-2 architecture ; 10000 iterations; P_C = 63.31%)

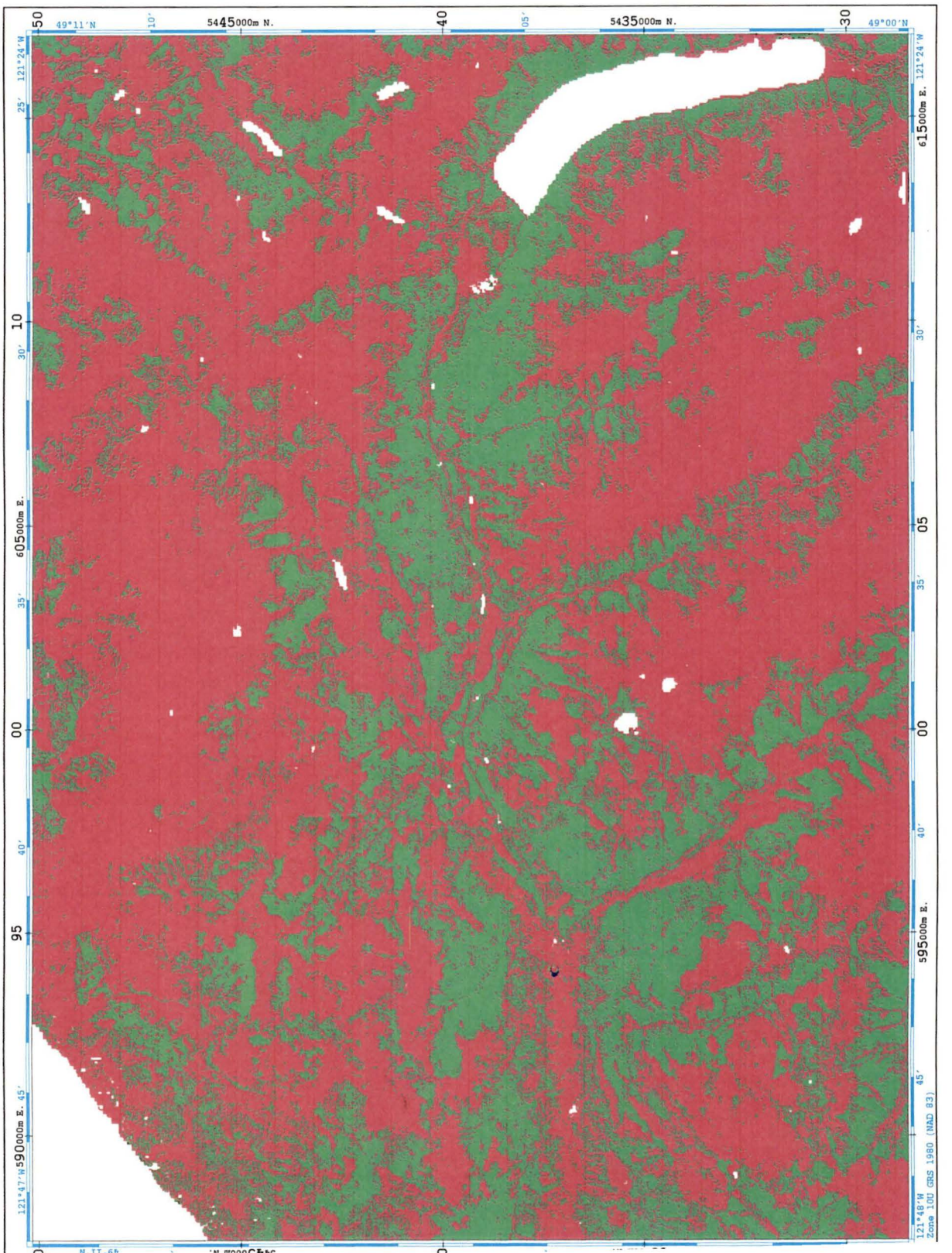
(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 19
(August, 1993; 6-24-6-2 architecture ; 1000 iterations; P_C = 50.83%)

(RED = NON-VEGETATED; GREEN = VEGETATED)



LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 20
(August, 1993; 6-24-6-2 architecture ; 5000 iterations; P_C = 87.17%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 21
(August, 1993; 6-24-6-2 architecture ; 10000 iterations; P_C = 63.31%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 22
(August, 1993; 6-24-6-2 architecture ; 1000 iterations; P_C = 72.66%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 23

(August, 1993; 6-24-6-2 architecture ; 5000 iterations; P_C = 67.74%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 24
(August, 1993; 6-24-6-2 architecture ; 10000 iterations; P_C = 72.42%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 25
(August, 1993; 6-6-2 architecture ; 10000 iterations; P_C = 67.78%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 26
(August, 1993; maximum likelihood classification; P_C = 81.79%)

(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL ONE: LAND BASE

NON-VEGETATED,VEGETATED

TRIAL 28
(August, 1993; maximum likelihood classification; P_C = 81.77%)

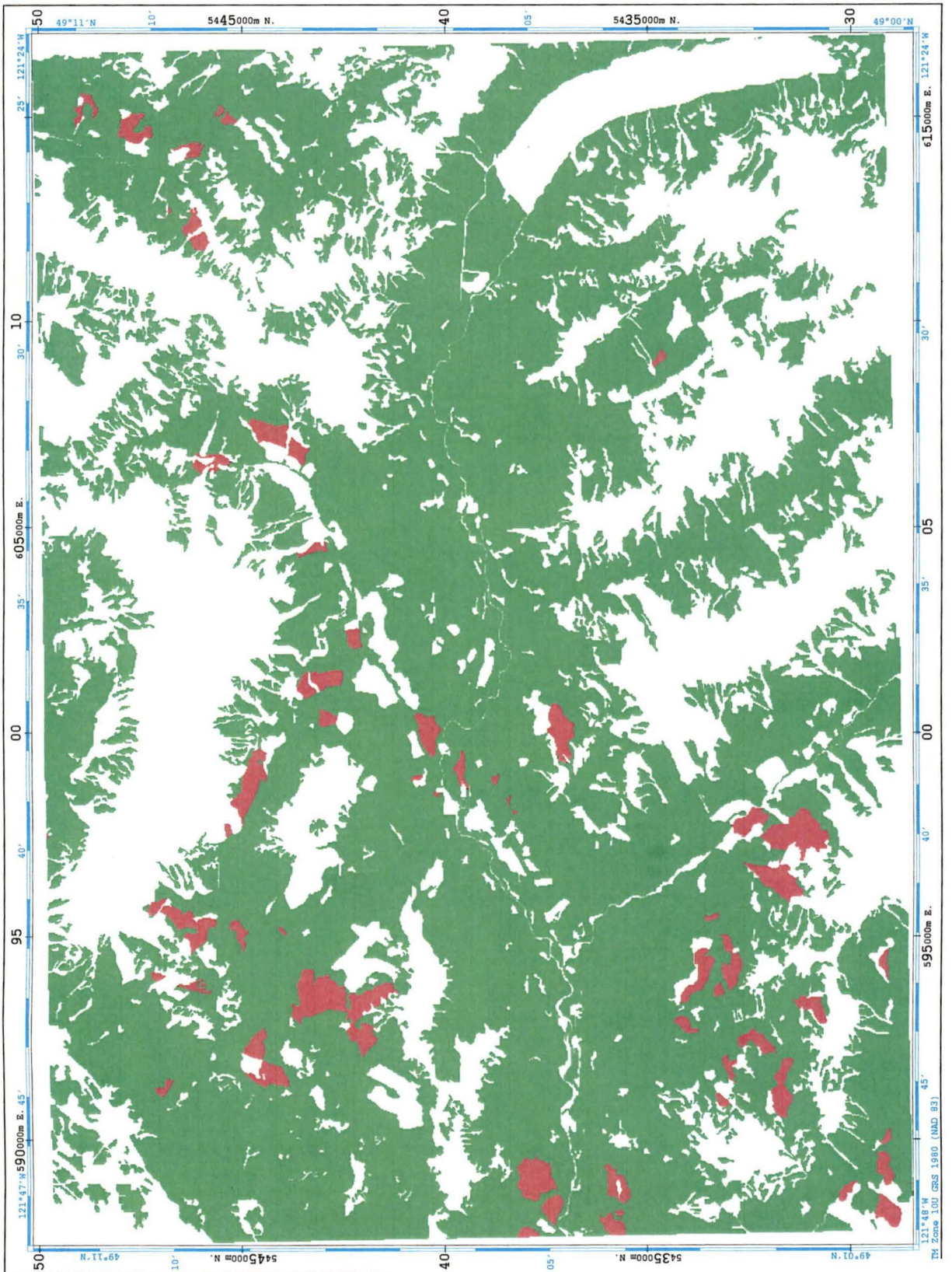
(RED = NON-VEGETATED; GREEN = VEGETATED)

LEVEL TWO: LAND COVER TYPE

NON-TREED, TREED

TRAINING DATA

(RED = NON-TREED; GREEN = TREED)

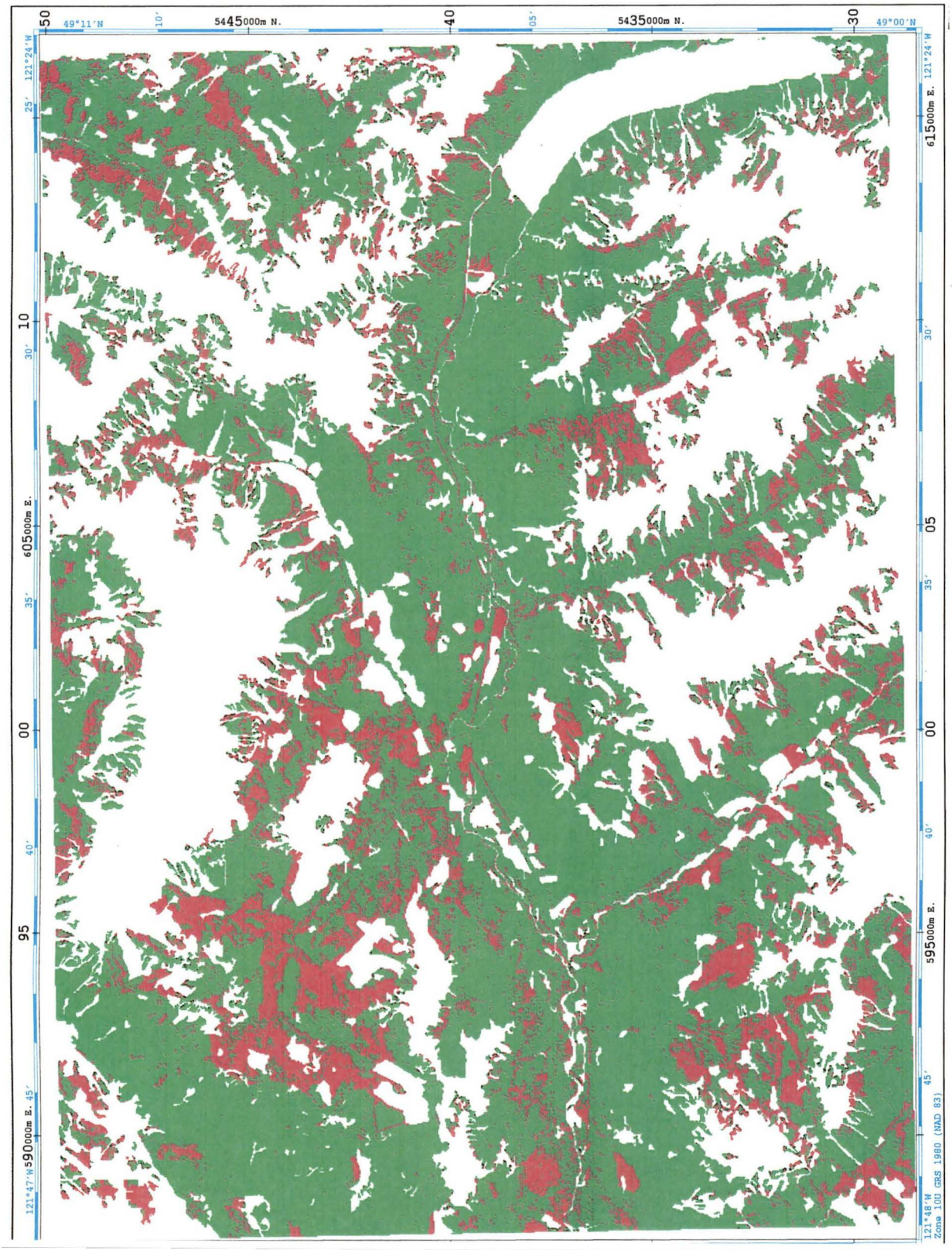


LEVEL TWO: LAND COVER TYPE

NON-TREED, TREED

TRIAL 1
(August, 1993; 6-6-2 architecture; 20000 iterations; P_C = 77.33%)

(RED = NON-TREED; GREEN = TREED)



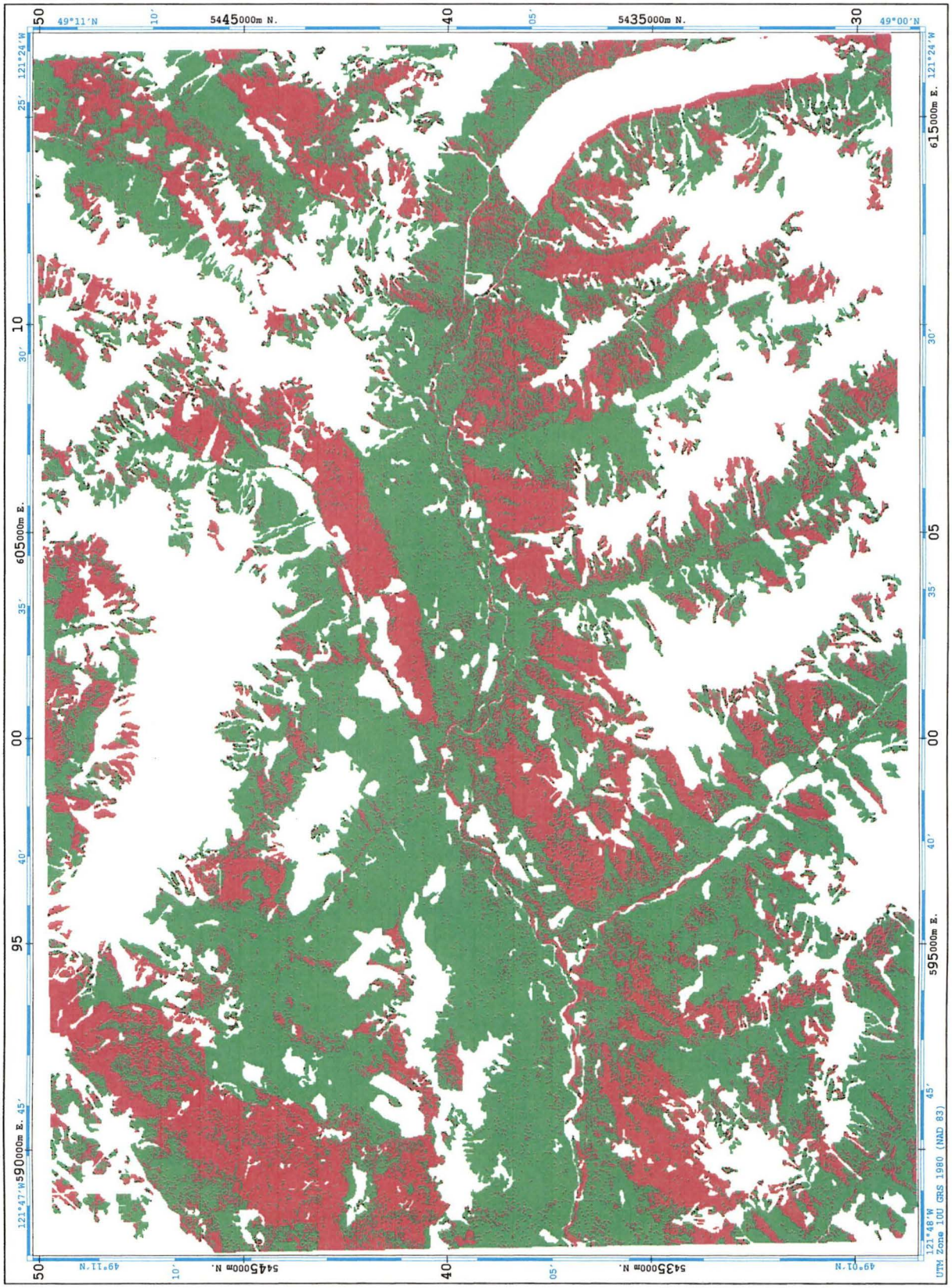
LEVEL TWO: LAND COVER TYPE

NON-TREED, TREED

302

TRIAL 5
(August, 1993; maximum likelihood; P_C = 39.06%)

(RED = NON-TREED; GREEN = TREED)

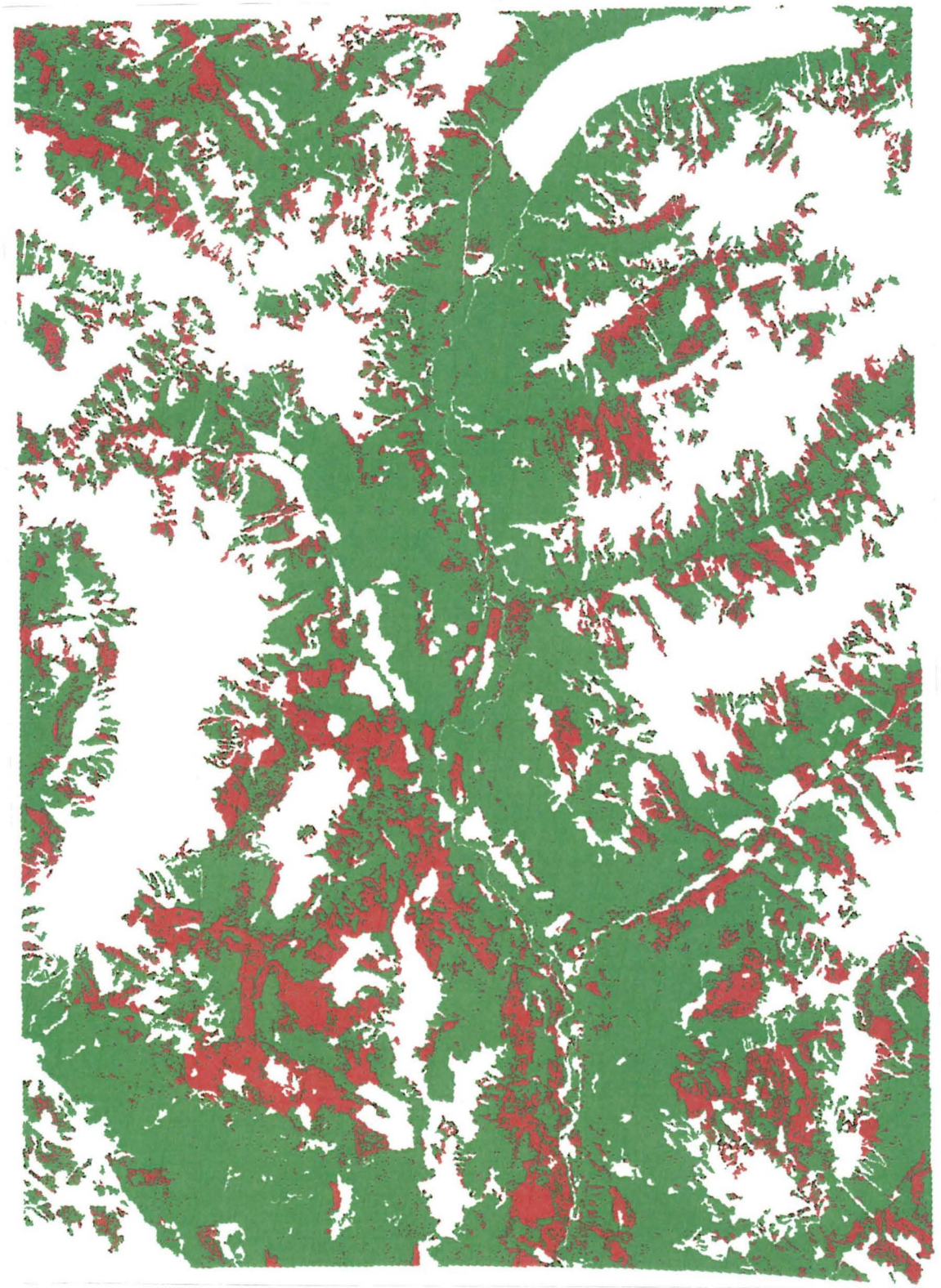


LEVEL TWO: LAND COVER TYPE

NON-TREED, TREED

TRIAL 6
(August, 1993; 6-6-2 architecture; 10000 iterations; 76.97%)

(RED = NON-TREED; GREEN = TREED)



LEVEL TWO: LAND COVER TYPE

NON-TREED, TREED

TRIAL 8
(August, 1993; 6-24-6-2 architecture; 5000 iterations; $P_C = 95.17\%$)

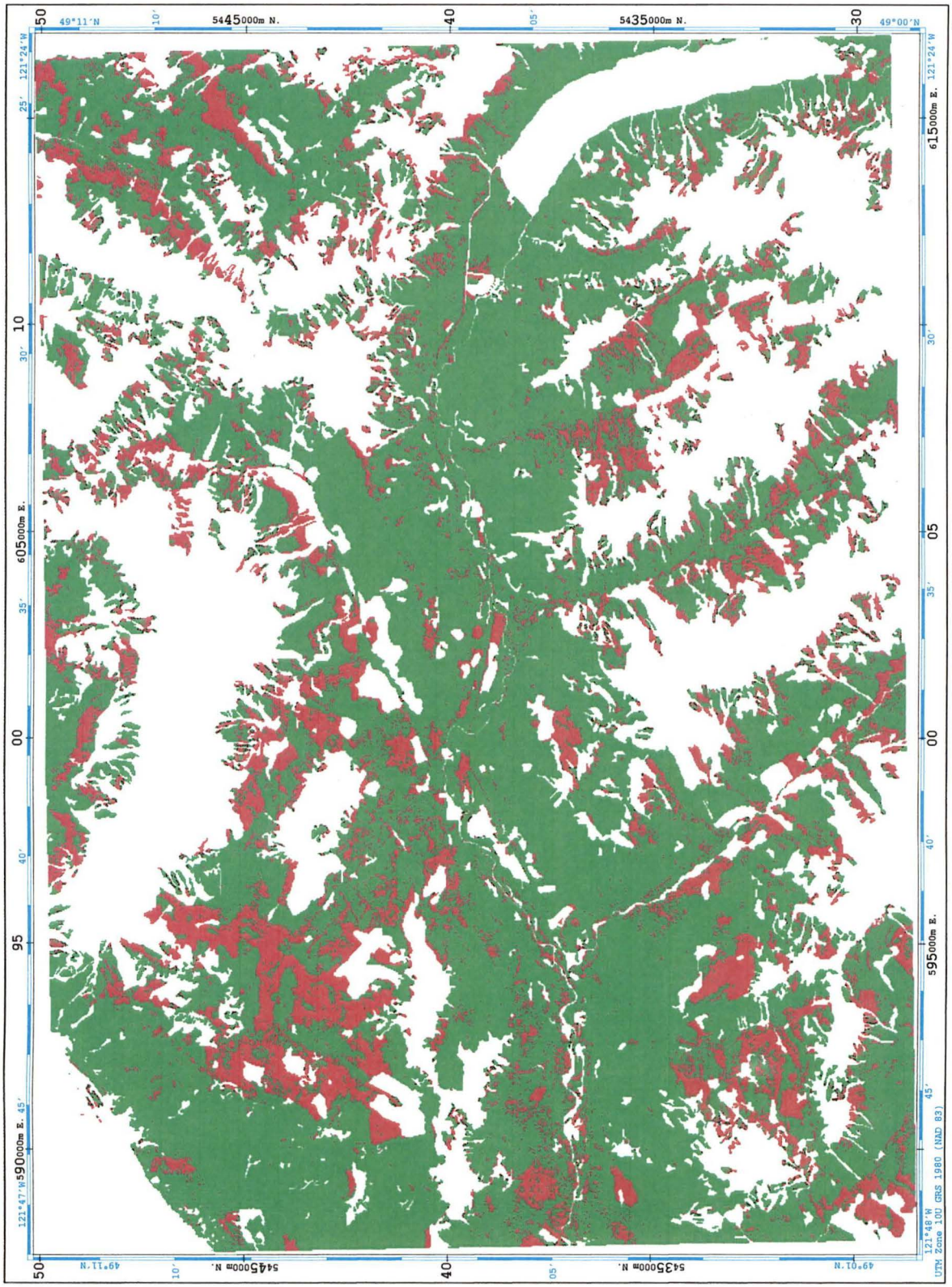
(RED = NON-TREED; GREEN = TREED)

LEVEL TWO: LAND COVER TYPE

NON-TREED, TREED

TRIAL 10
(August, 1993; 6-6-2 architecture; 5000 iterations; P_C = 79.58%)

(RED = NON-TREED; GREEN = TREED)

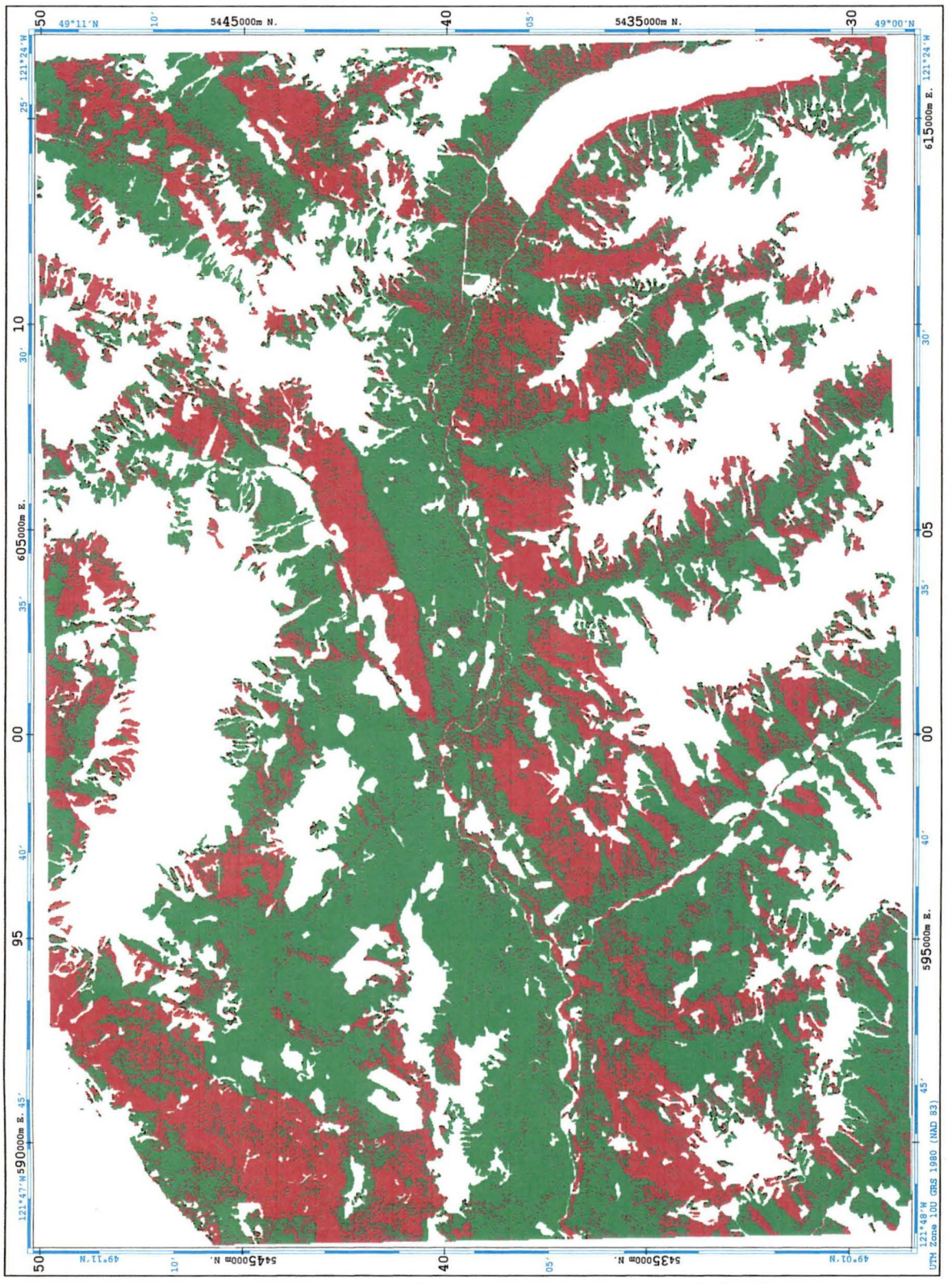


LEVEL TWO: LAND COVER TYPE

NON-TREED, TREED

TRIAL 12
(August, 1993; maximum likelihood; P_C = 41.96%)

(RED = NON-TREED; GREEN = TREED)

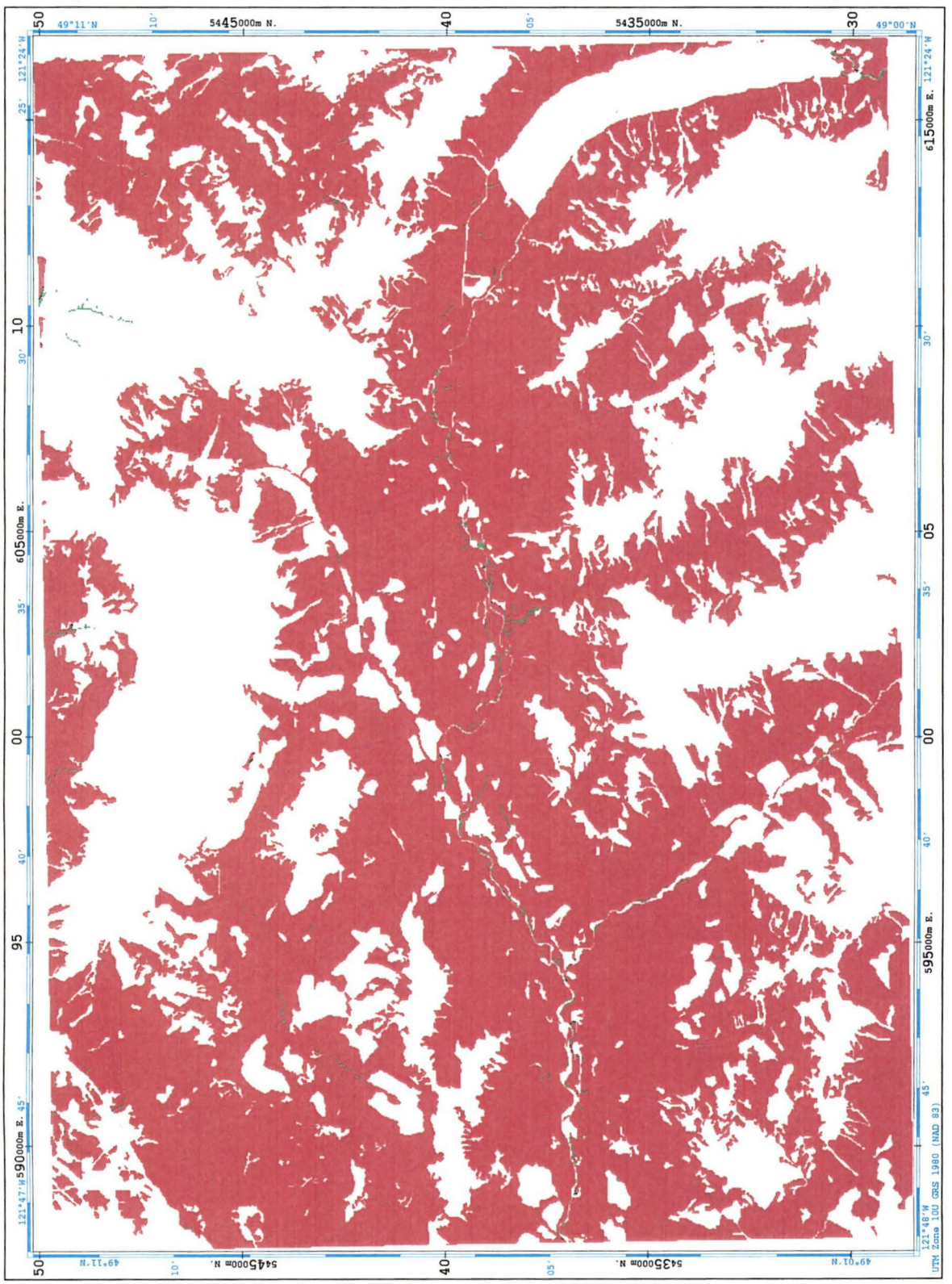


LEVEL THREE: LANDSCAPE POSITION (Vegetated Treed)

WETLAND, UPLAND

TRAINING DATA

(RED = UPLAND; GREEN = WETLAND)

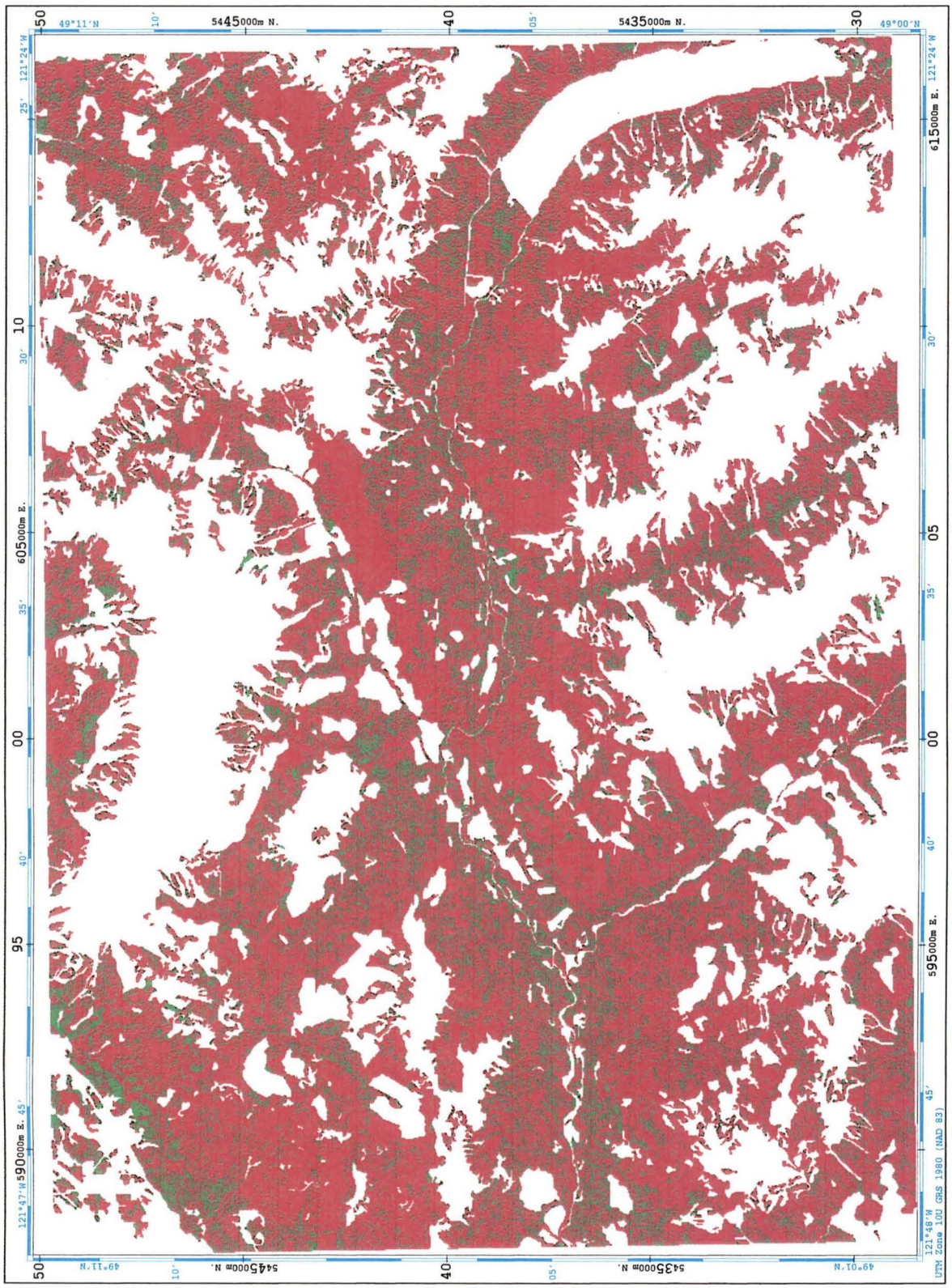


LEVEL THREE: LANDSCAPE POSITION (Vegetated Treed)

WETLAND, UPLAND

TRIAL 1
(August, 1993; 6-6-2 architecture; 10000 iterations; P_C = 83.45%)

(RED = UPLAND; GREEN = WETLAND)

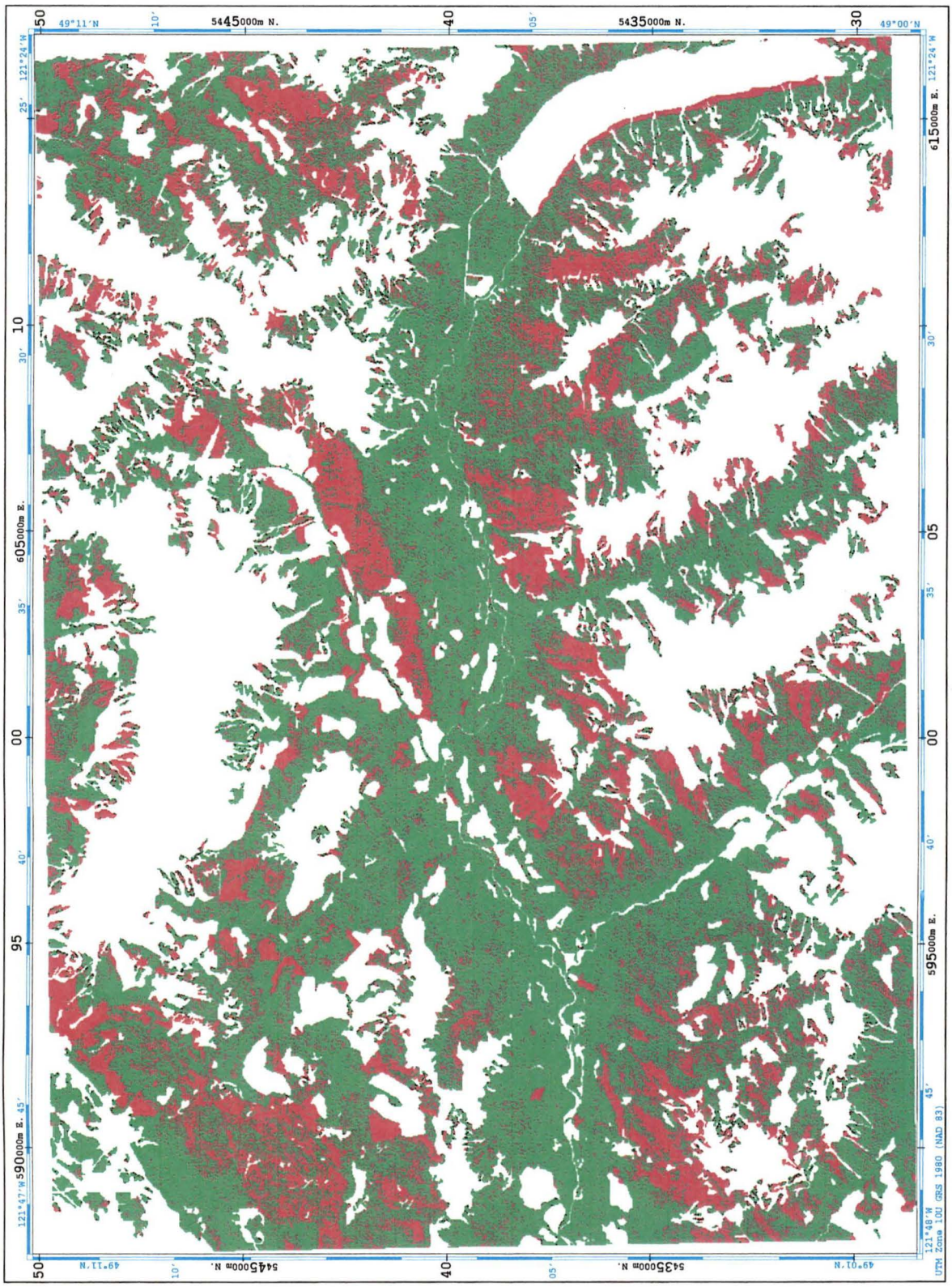


LEVEL THREE: LANDSCAPE POSITION (Vegetated Treed)

WETLAND, UPLAND

TRIAL 5
(August, 1993; maximum likelihood; P_C = 34.87%)

(RED = UPLAND; GREEN = WETLAND)

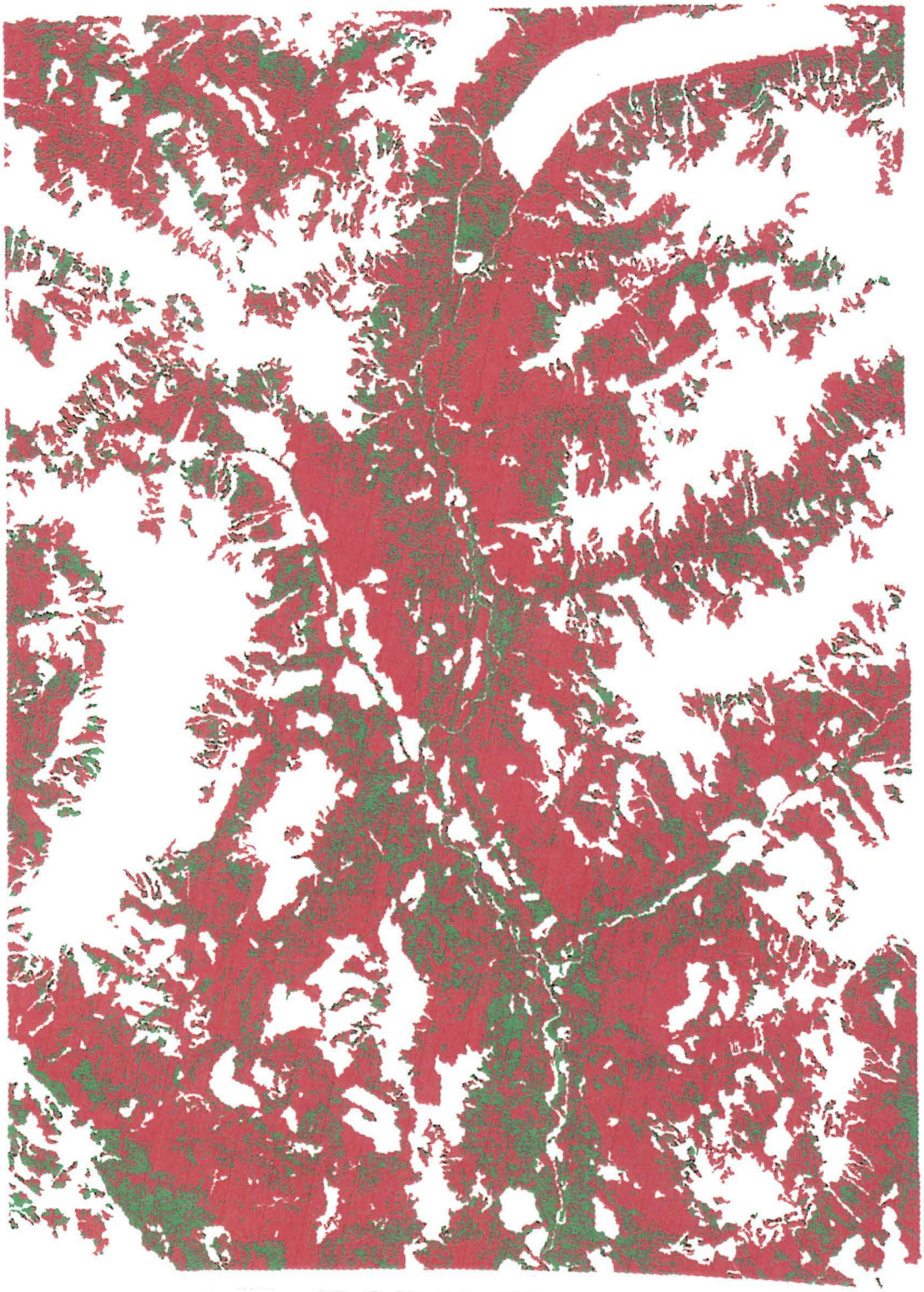


LEVEL THREE: LANDSCAPE POSITION (Vegetated Treed)

WETLAND, UPLAND

TRIAL 6
(August, 1993; 6-6-2 architecture; 10000 iterations; P_C = 77.52%)

(RED = UPLAND; GREEN = WETLAND)

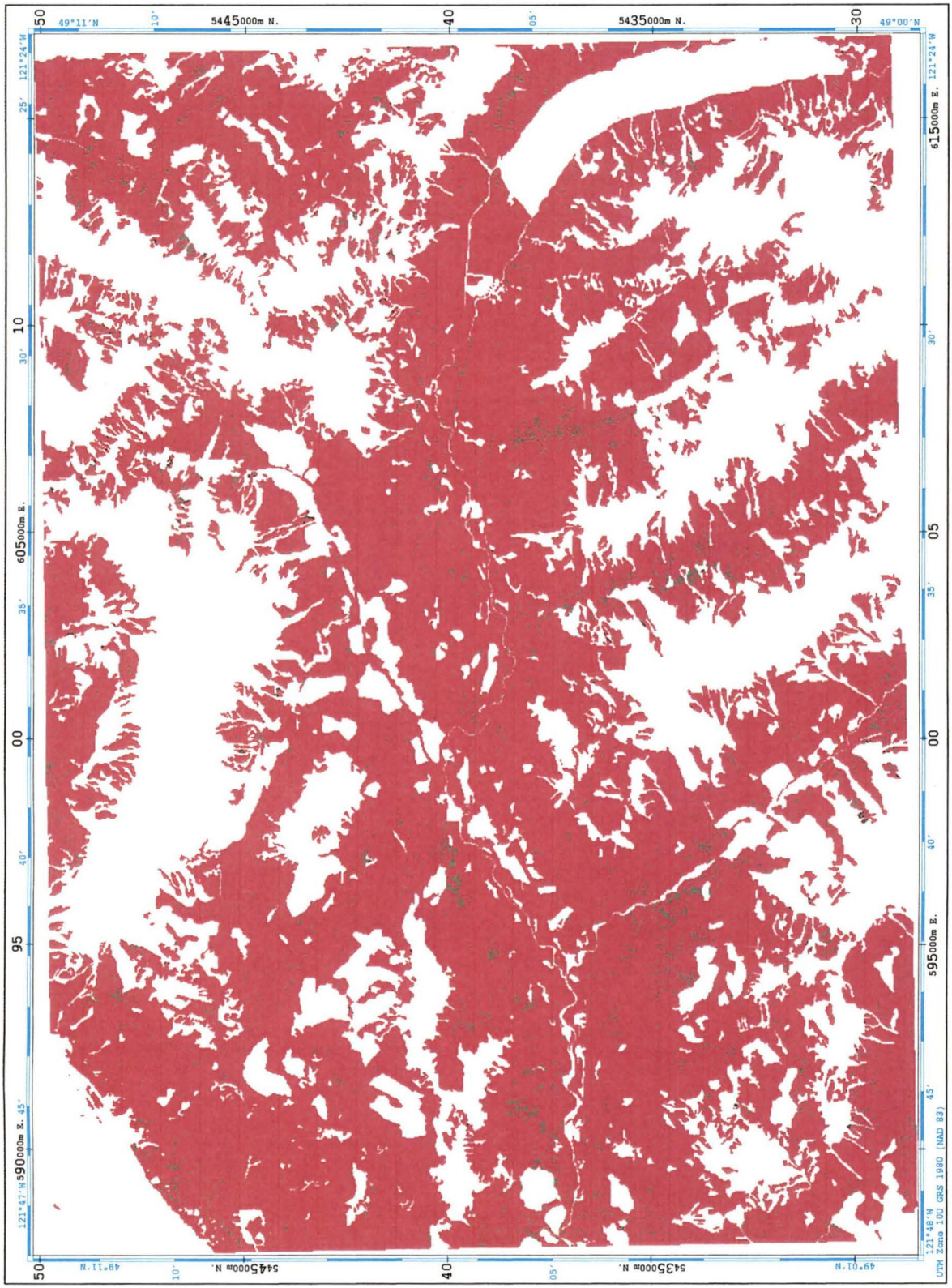


LEVEL THREE: LANDSCAPE POSITION (Vegetated Treed)

WETLAND, UPLAND

TRIAL 9
(August, 1993; 6-24-6-2 architecture; 10000 iterations; P_C=98.93%)

(RED = UPLAND; GREEN = WETLAND)

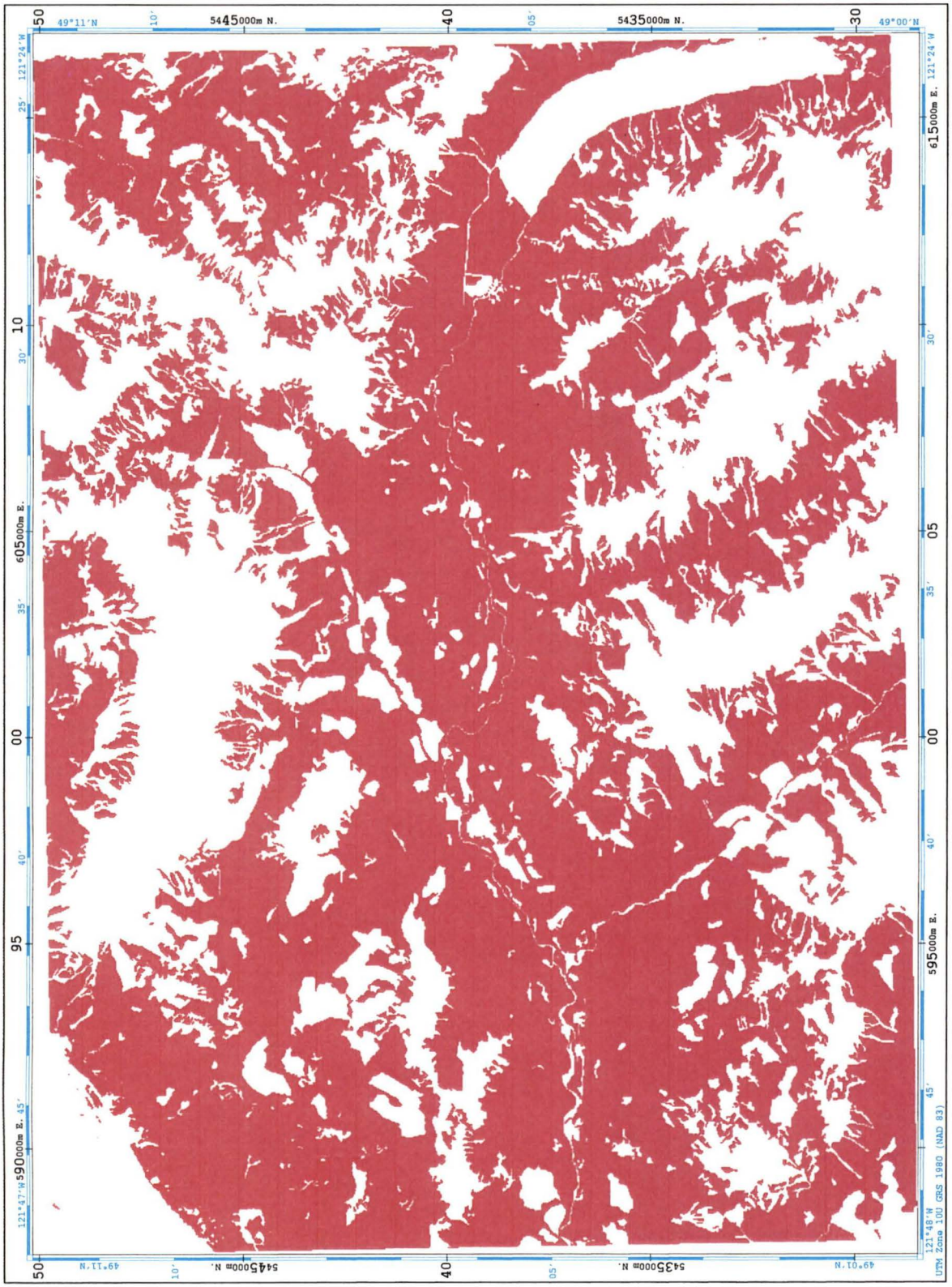


LEVEL THREE: LANDSCAPE POSITION (Vegetated Treed)

WETLAND, UPLAND

TRIAL 12
(August, 1993; 9-36-6-2 architecture; 10000 iterations; P_C=99.46%)

(RED = UPLAND; GREEN = WETLAND)

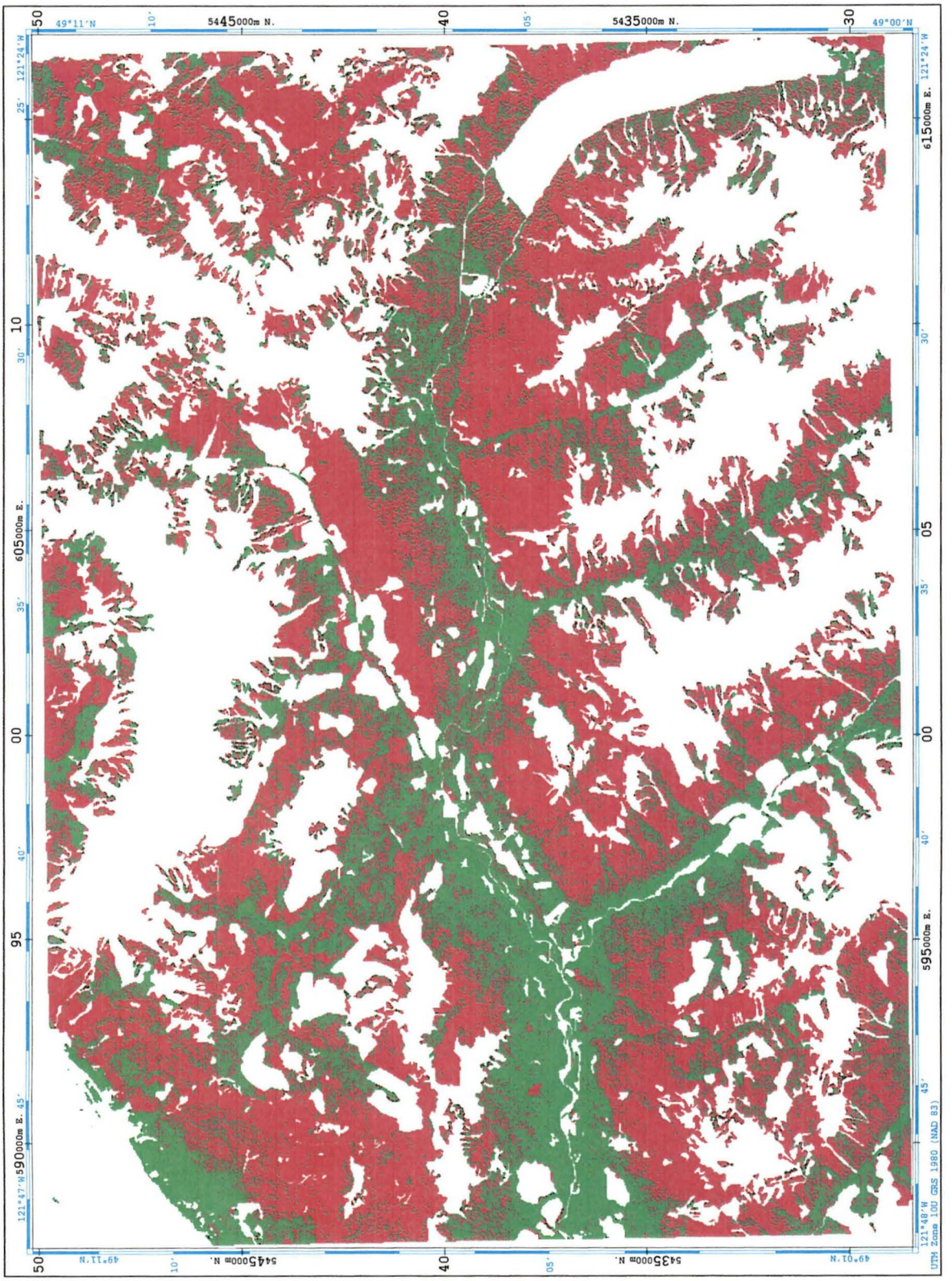


LEVEL THREE: LANDSCAPE POSITION (Vegetated Treed)

WETLAND, UPLAND

TRIAL 14
(August, 1993; 6-6-2 architecture; 10000 iterations; P_C=62.63%)

(RED = UPLAND; GREEN = WETLAND)

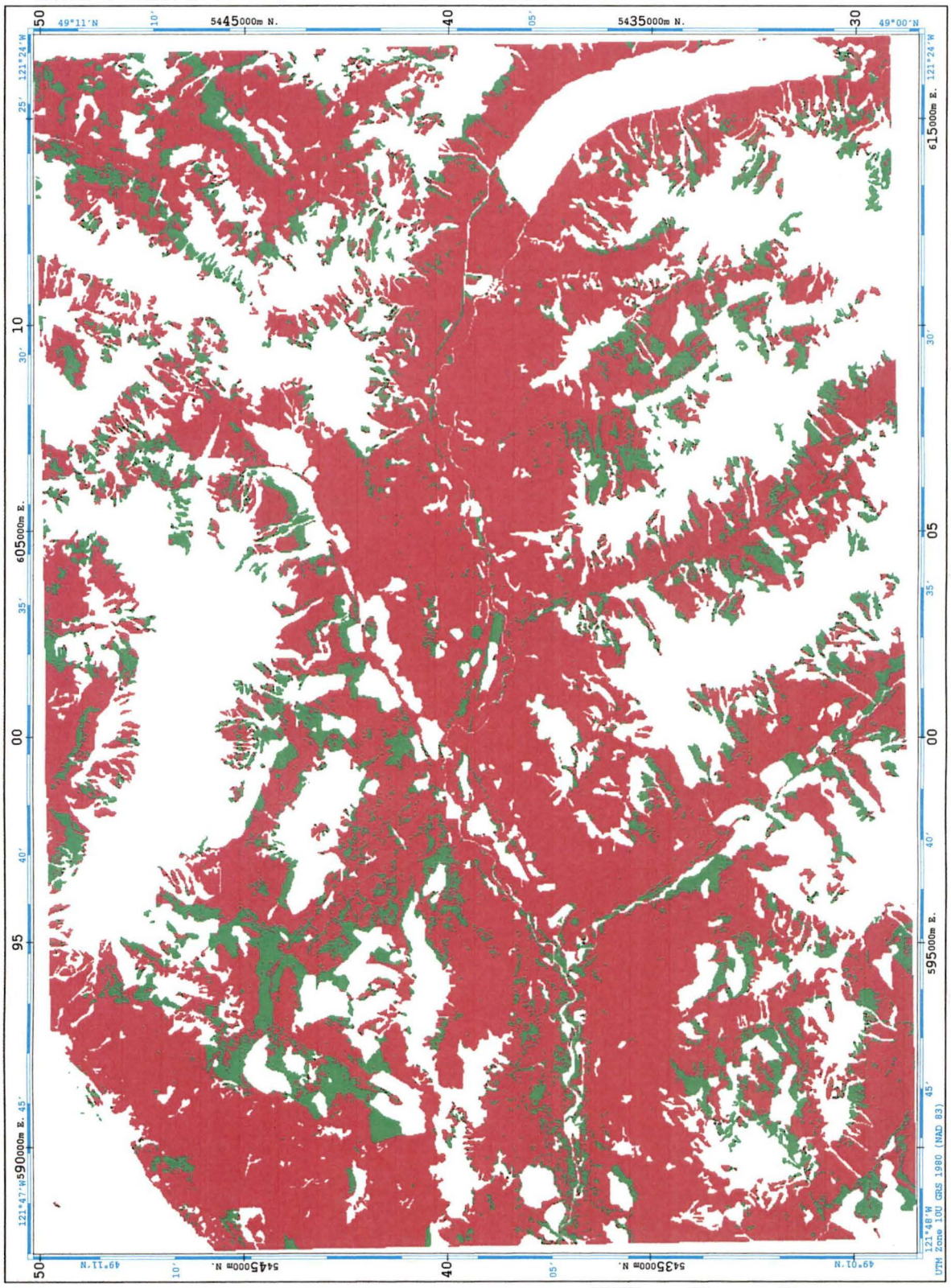


LEVEL THREE: LANDSCAPE POSITION (Vegetated Treed)

WETLAND, UPLAND

TRIAL 16
(August, 1993; 9-9-2 architecture; 10000 iterations; P_C=87.37%)

(RED = UPLAND; GREEN = WETLAND)

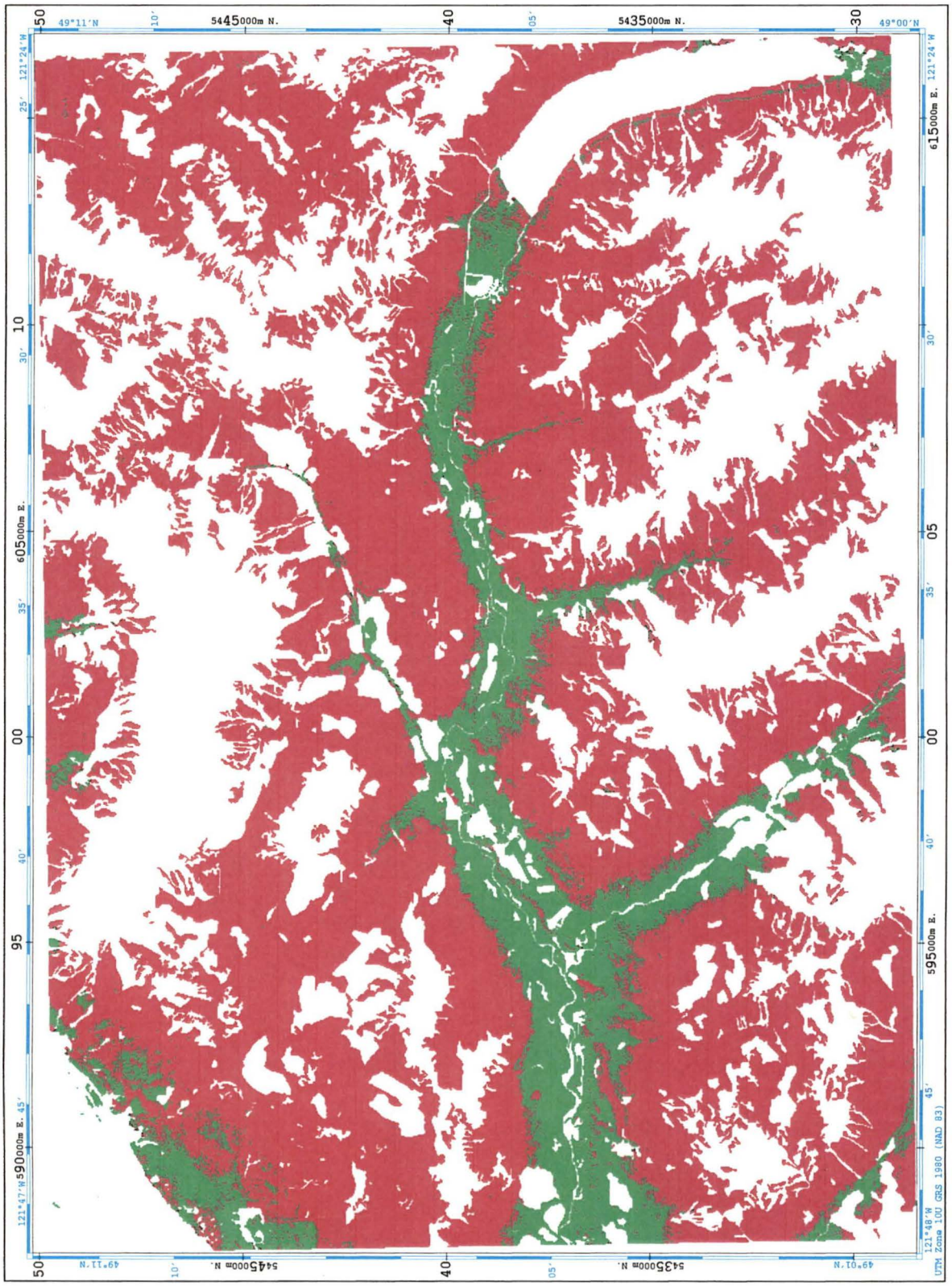


LEVEL THREE: LANDSCAPE POSITION (Vegetated Treed)

WETLAND, UPLAND

TRIAL 18
(August, 1993; 7-7-2 architecture; 10000 iterations; P_C=77.22%)

(RED = UPLAND; GREEN = WETLAND)

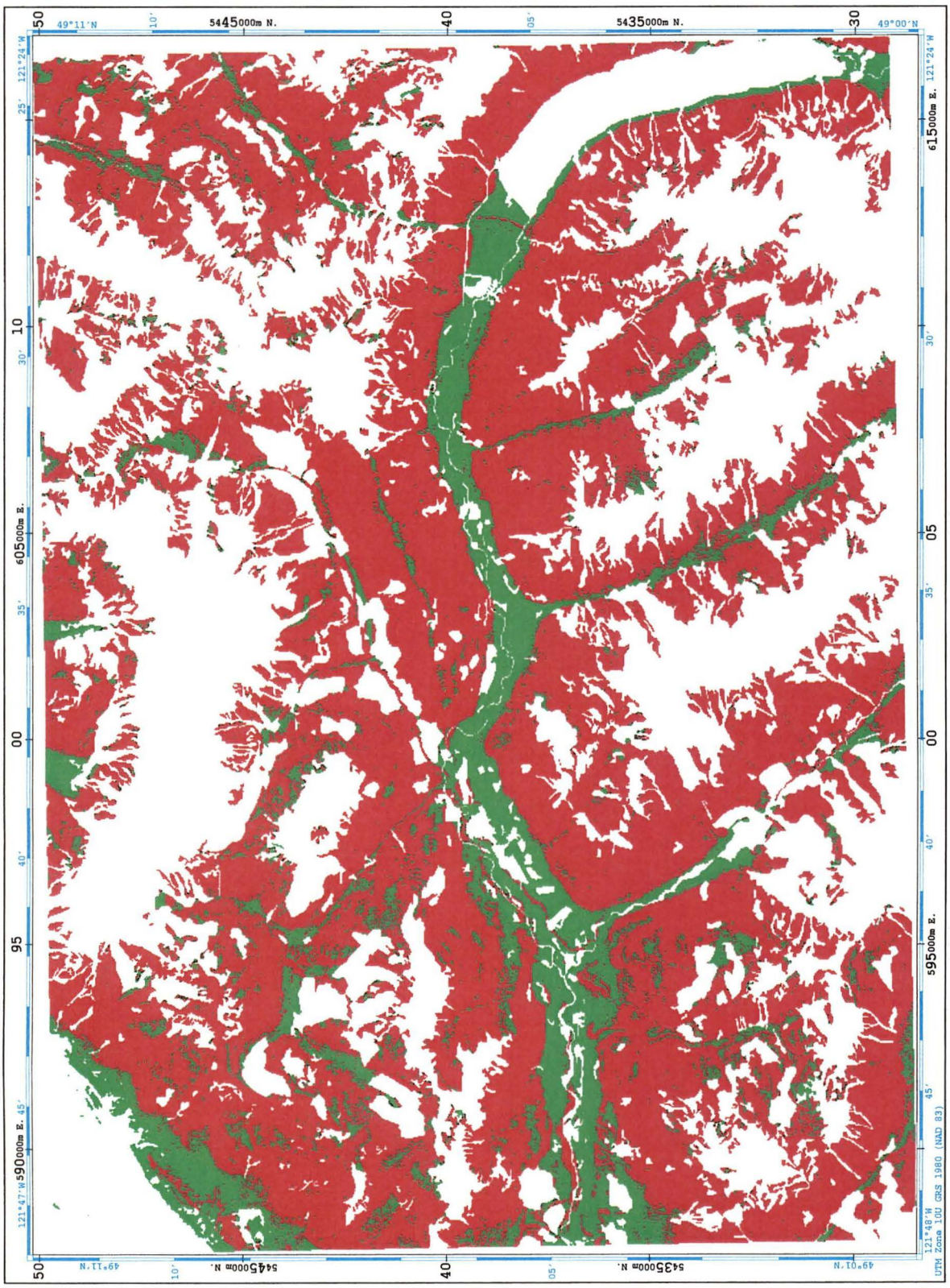


LEVEL THREE: LANDSCAPE POSITION (Vegetated-Treed)

WETLAND, UPLAND

TRIAL 20
(August, 1993; 7-7-2 architecture; 10000 iterations; 80.07%)

(RED = UPLAND; GREEN = WETLAND)

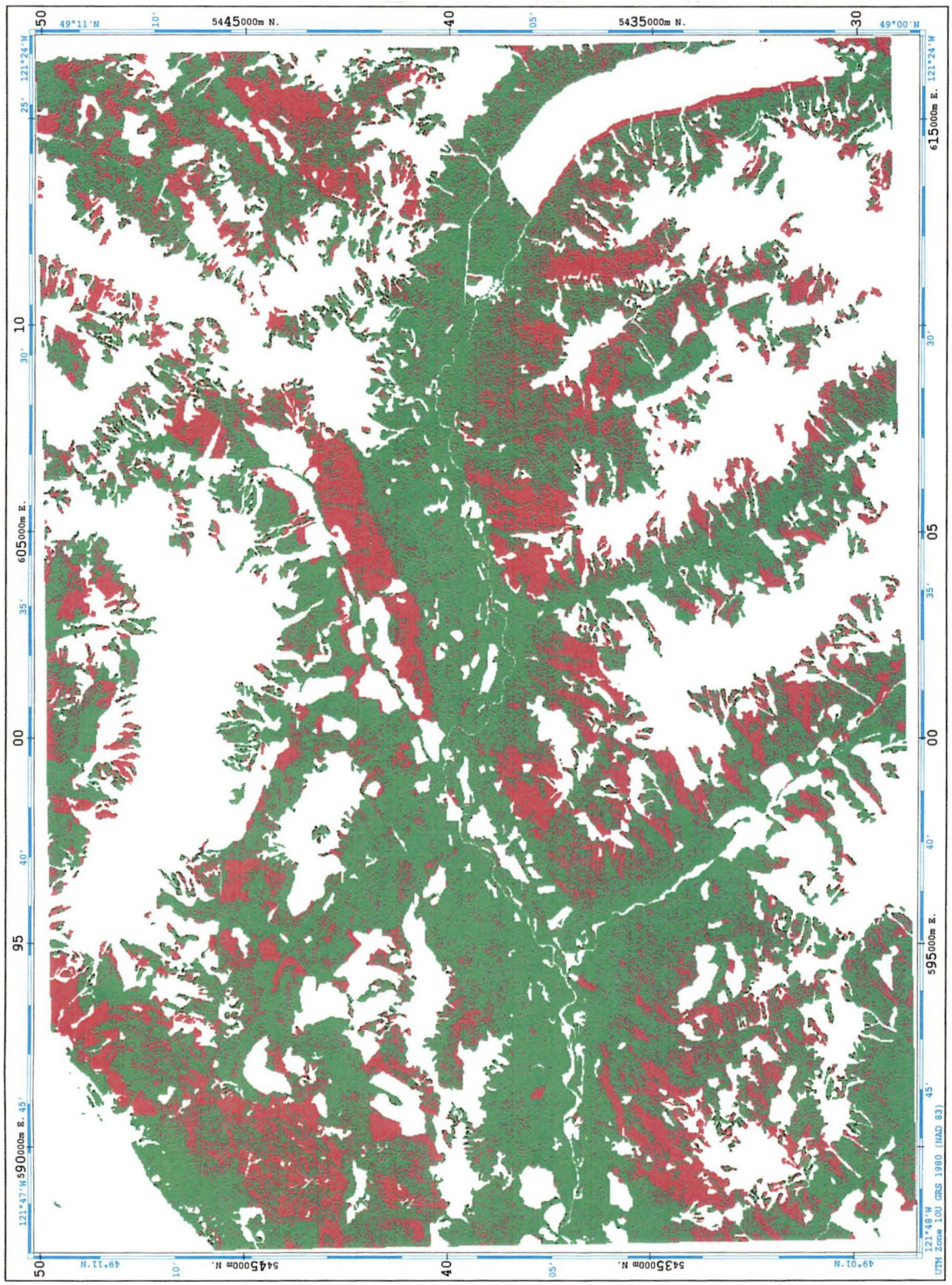


LEVEL THREE: LANDSCAPE POSITION (Vegetated Treed)

WETLAND, UPLAND

TRIAL 21
(August, 1993; maximum likelihood classification; P_C=35.76%)

(RED = UPLAND; GREEN = WETLAND)

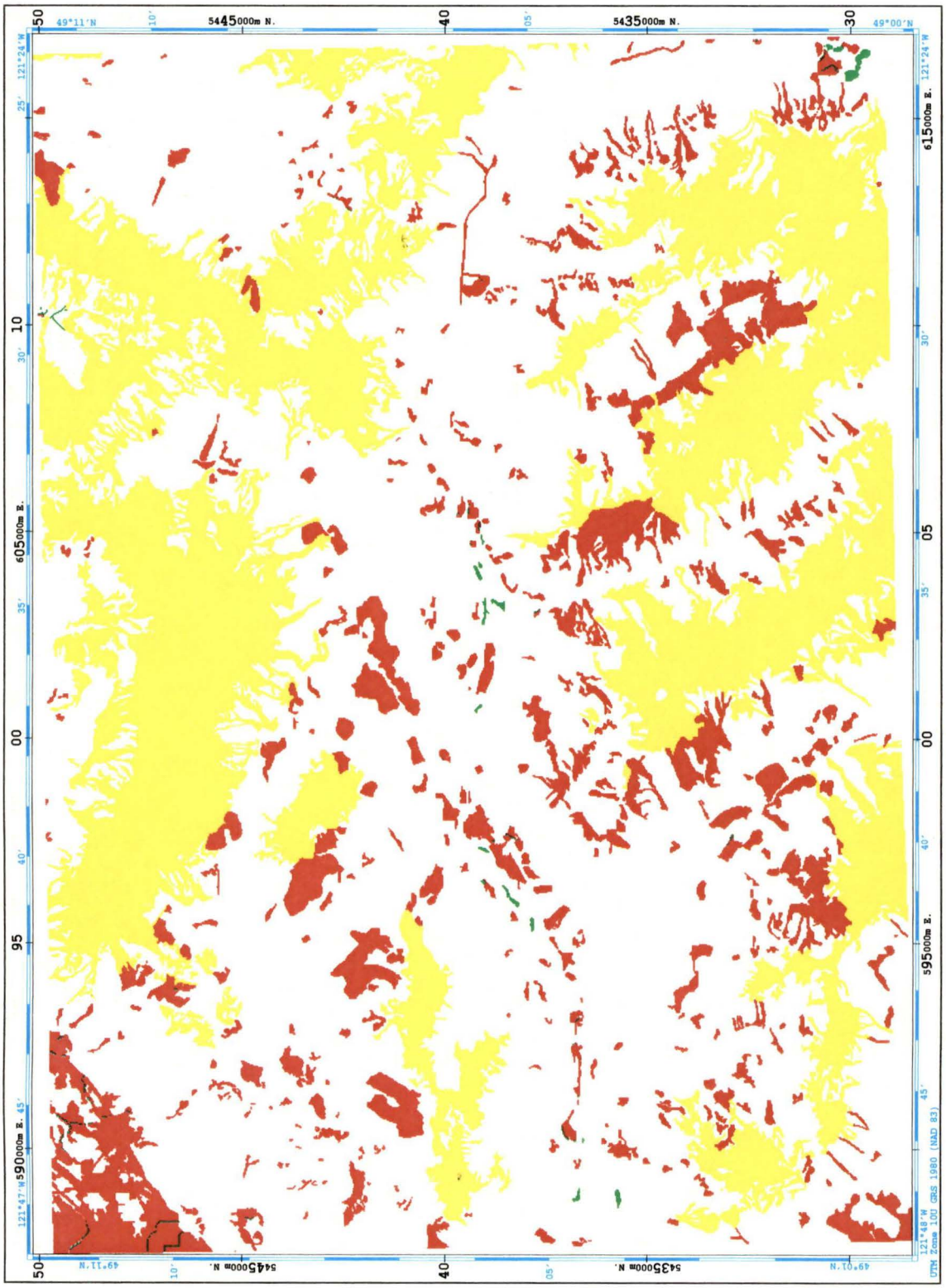


LEVEL THREE: LANDSCAPE POSITION (Non-Vegetated)

WETLAND, UPLAND, ALPINE

TRAINING DATA

(YELLOW = ALPINE; RED = UPLAND; GREEN = WETLAND)

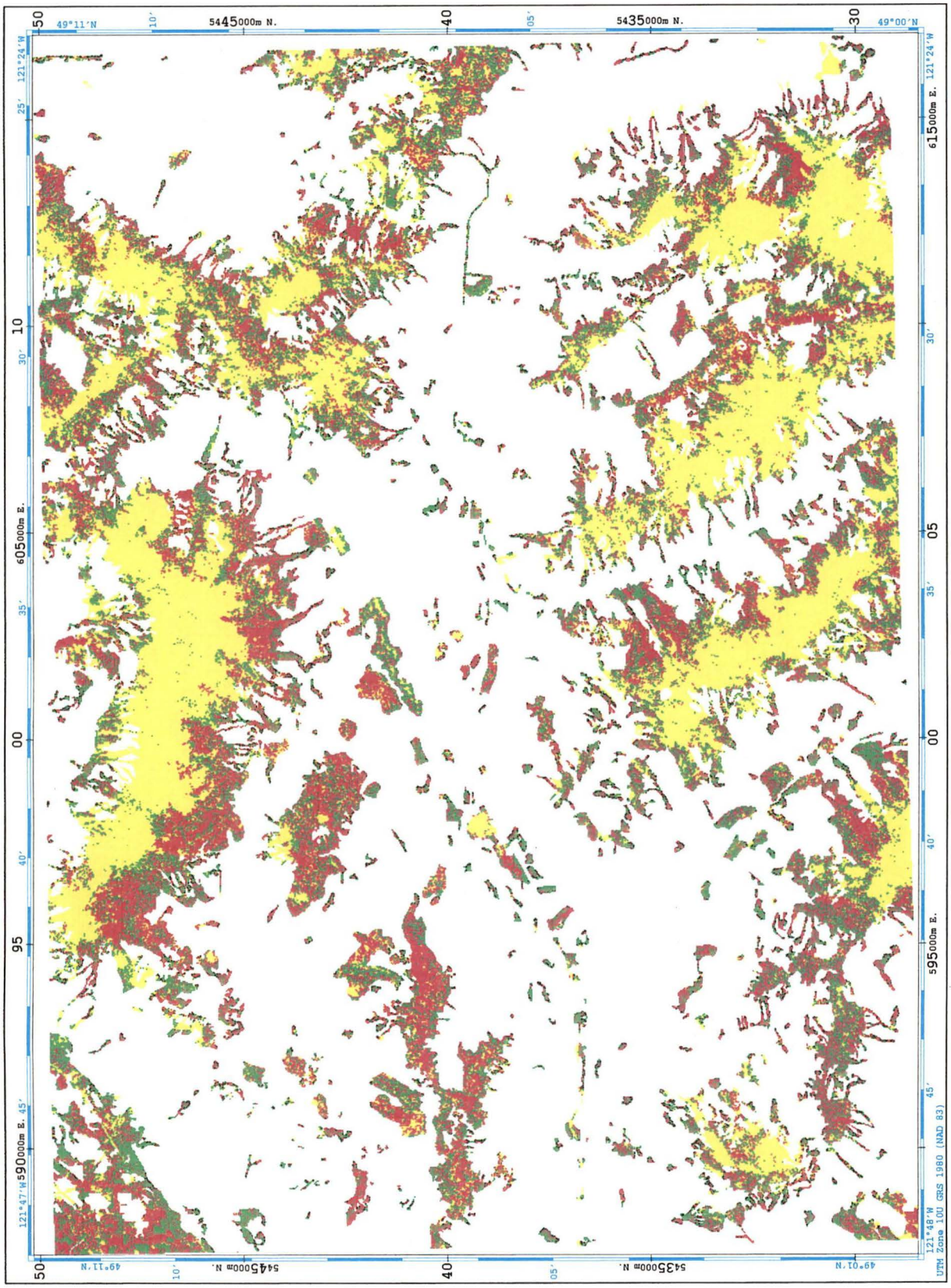


LEVEL THREE: LANDSCAPE POSITION (Non-Vegetated)

WETLAND, UPLAND, ALPINE

TRIAL 1
(August, 1993; 6-6-3 architecture; 10000 iterations; 50.00%)

(YELLOW = ALPINE; RED = UPLAND; GREEN = WETLAND)

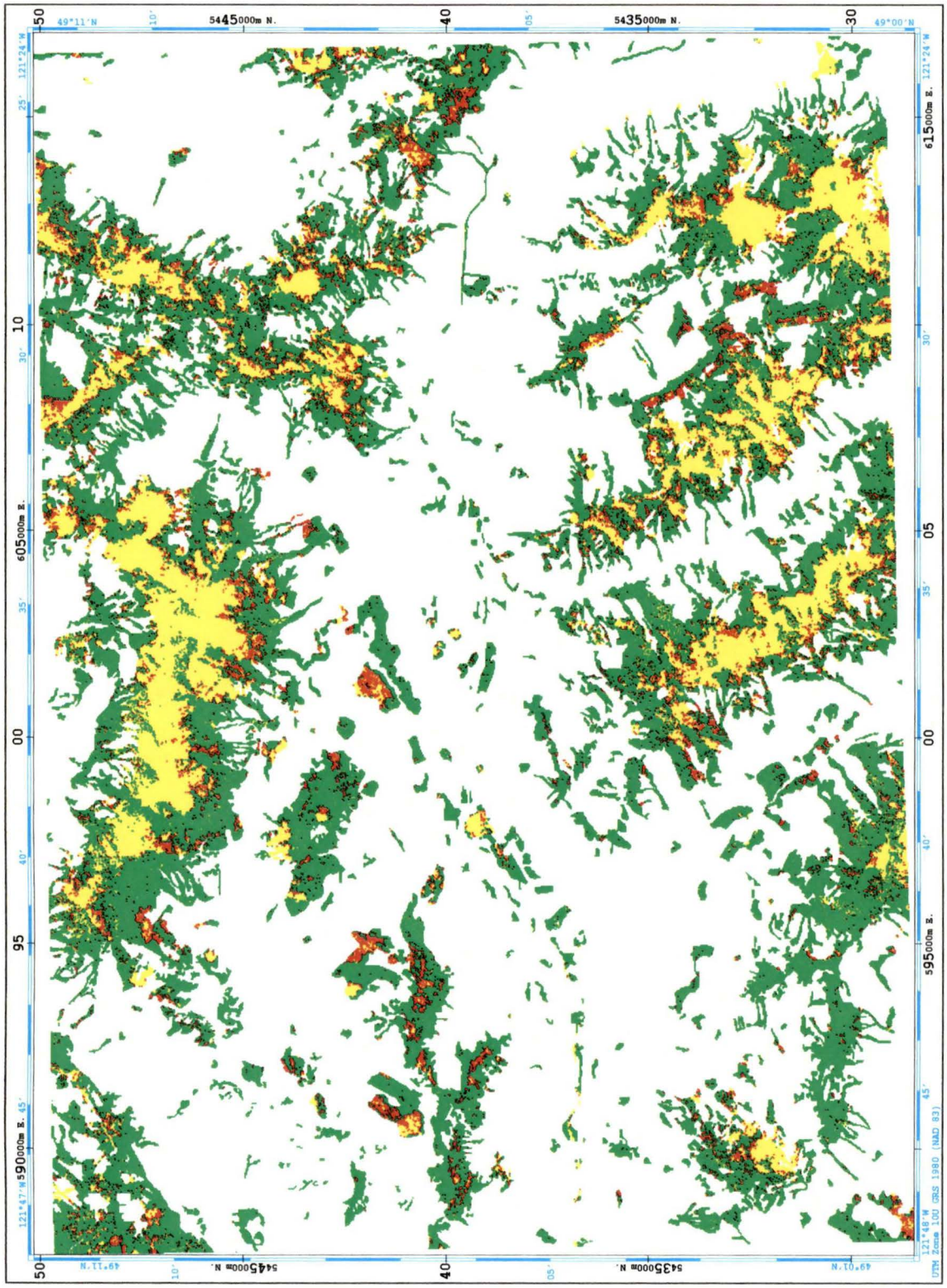


LEVEL THREE: LANDSCAPE POSITION (Non-Vegetated)

WETLAND, UPLAND, ALPINE

TRIAL 5
(August, 1993; maximum likelihood; 22.22%)

(YELLOW = ALPINE; RED = UPLAND; GREEN = WETLAND)

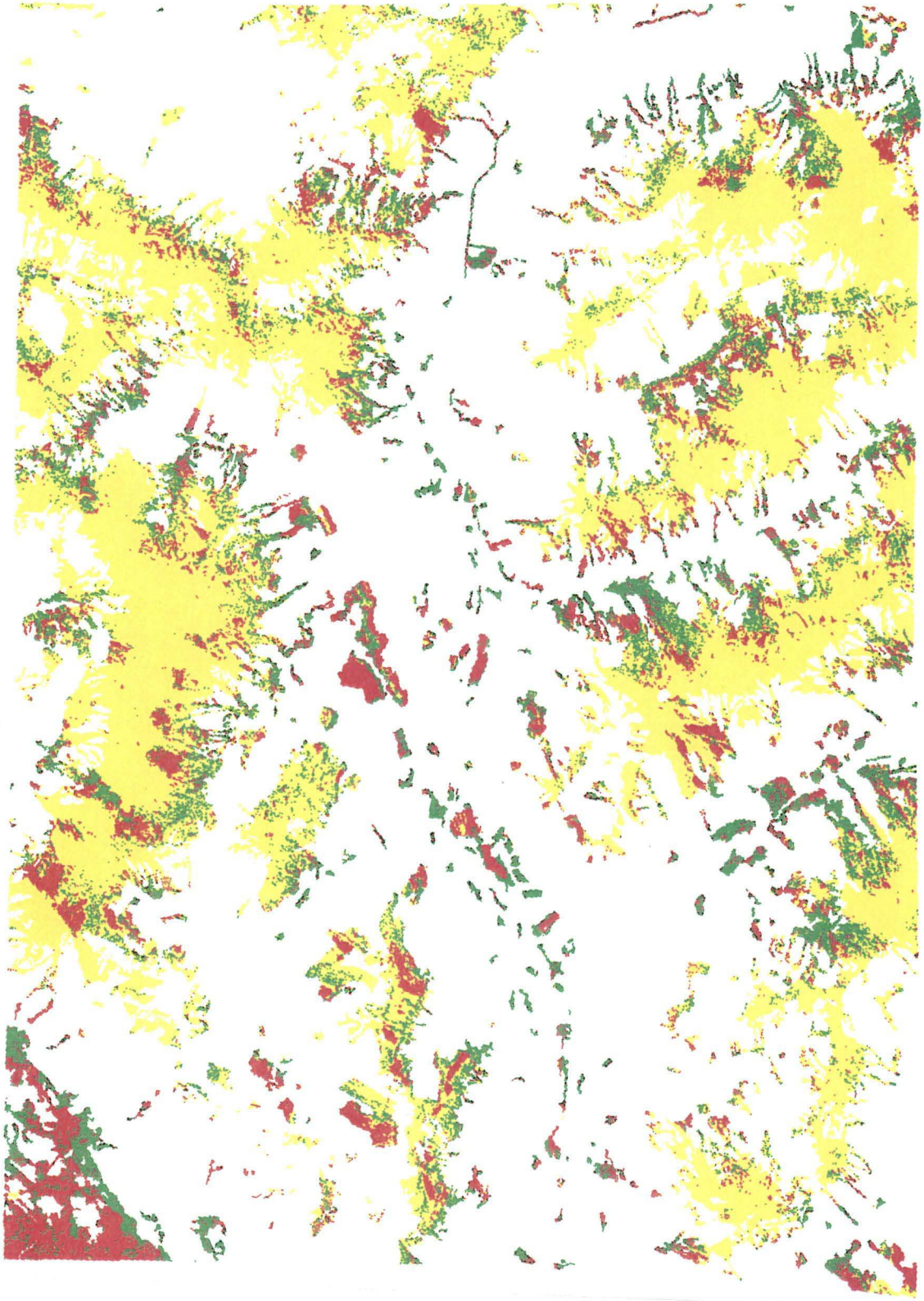


LEVEL THREE: LANDSCAPE POSITION (Non-Vegetated)

WETLAND, UPLAND, ALPINE

TRIAL 6
(August, 1993; 6-6-3 architecture; 10000 iterations; 44.44%)

(YELLOW = ALPINE; RED = UPLAND; GREEN = WETLAND)

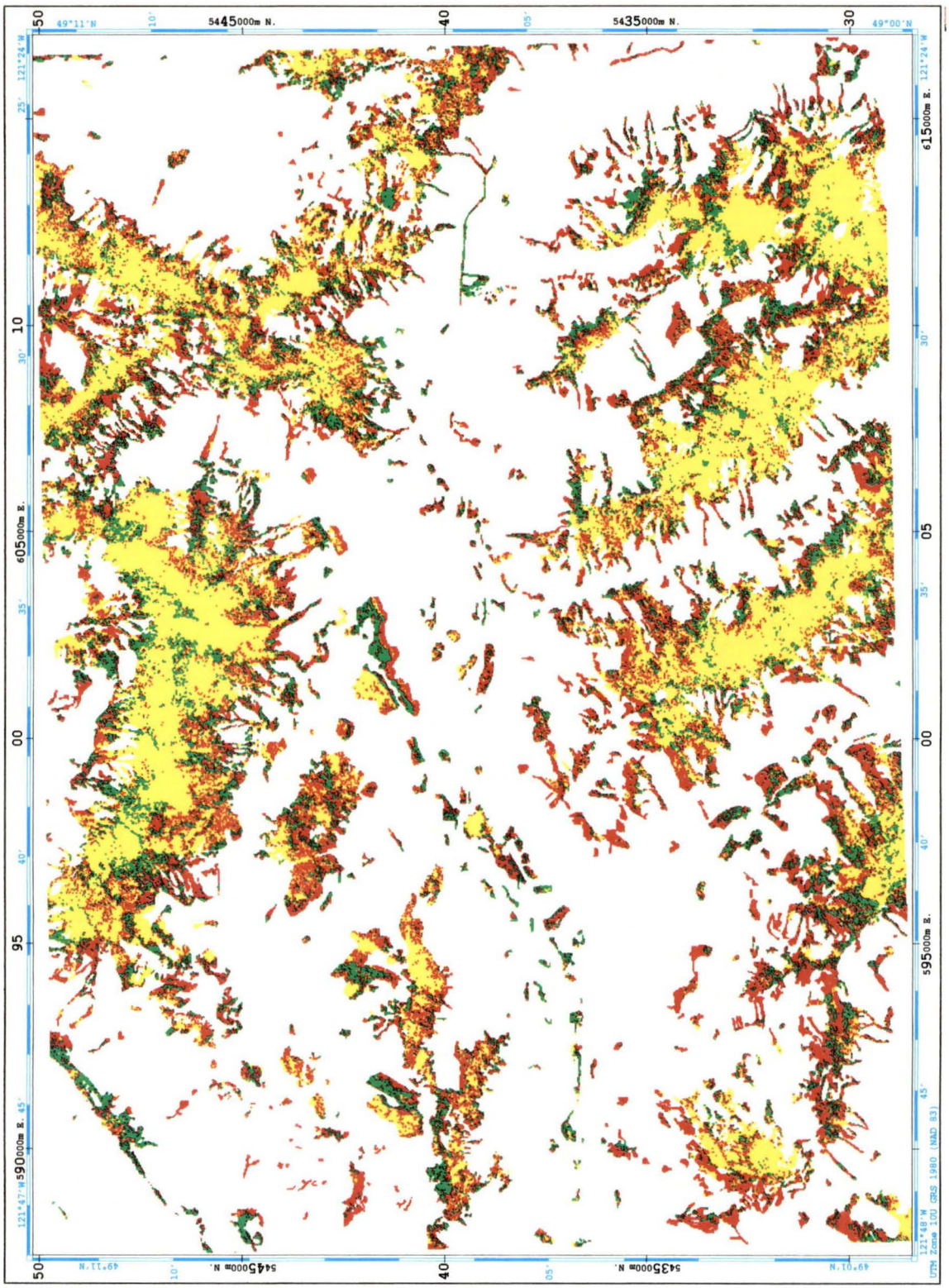


LEVEL THREE: LANDSCAPE POSITION (Non-Vegetated)

WETLAND, UPLAND, ALPINE

TRIAL 9
(August, 1993; 6-24-9-3 architecture; 10000 iterations; 55.55%)

(YELLOW = ALPINE; RED = UPLAND; GREEN = WETLAND)

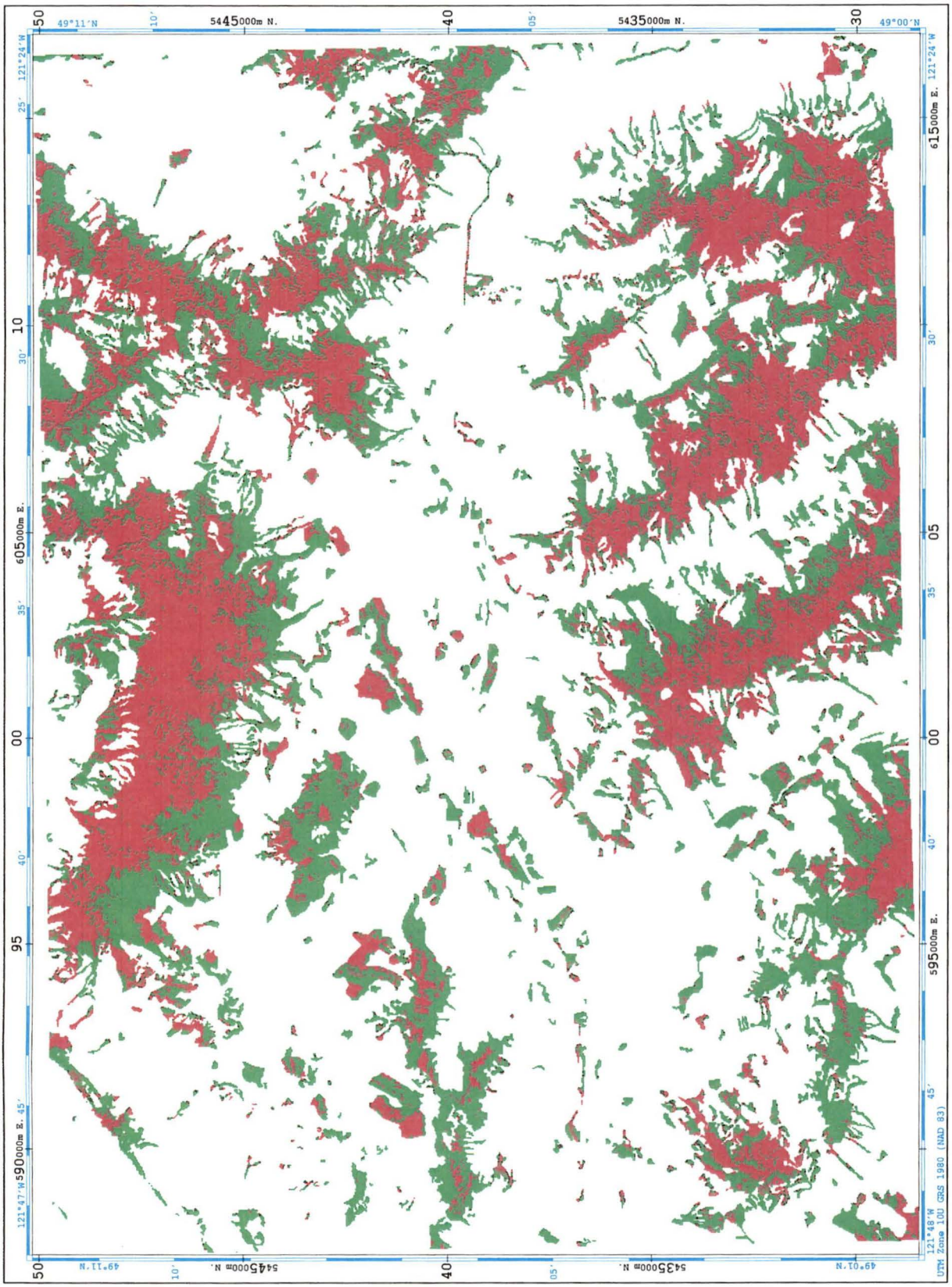


LEVEL THREE: LANDSCAPE POSITION (Non-Vegetated)

WETLAND, UPLAND, ALPINE

TRIAL 12
(August, 1993; 9-36-9-3 architecture; 10000 iterations; 44.44%)

(YELLOW = ALPINE; RED = UPLAND; GREEN = WETLAND)

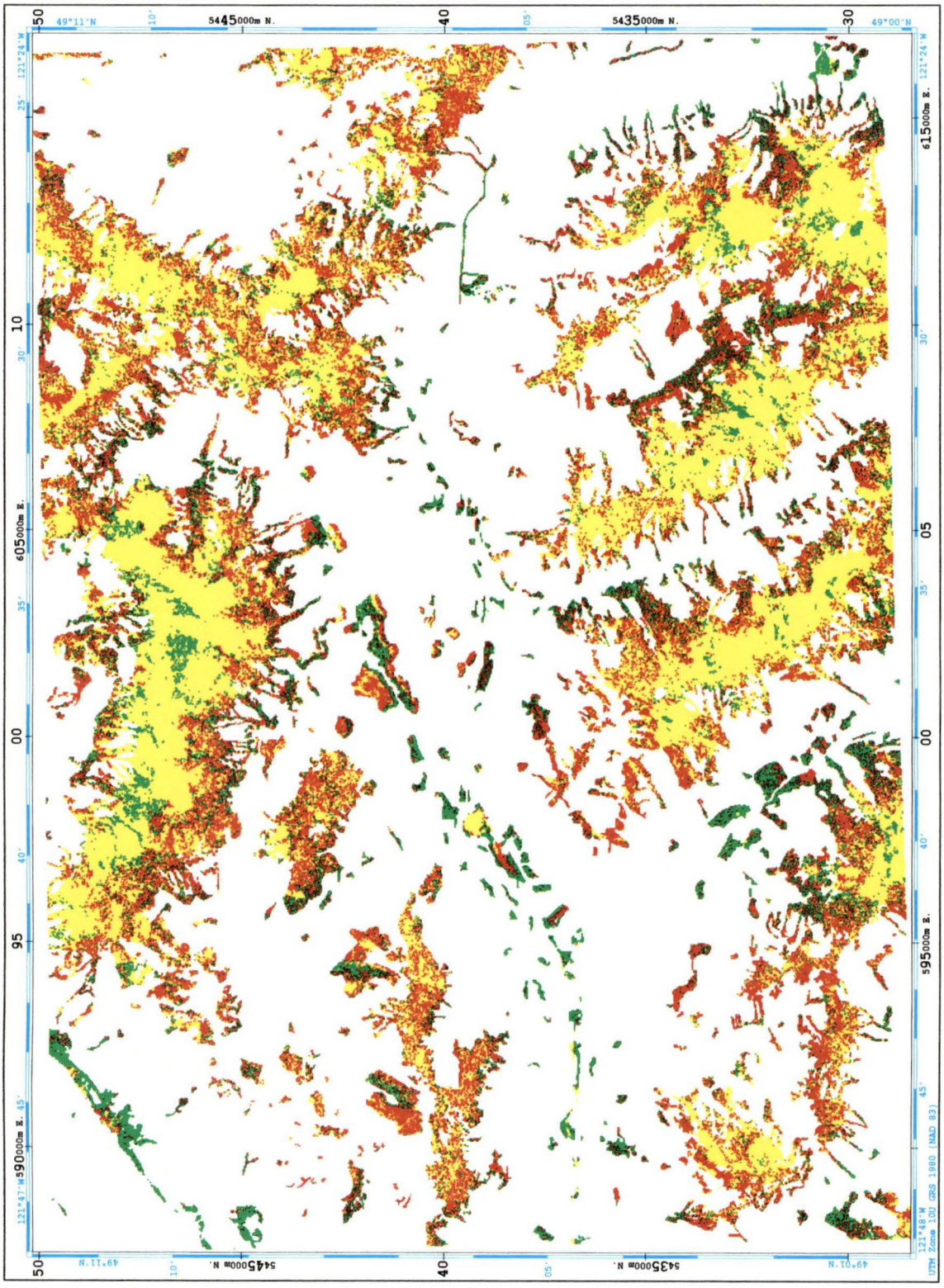


LEVEL THREE: LANDSCAPE POSITION (Non-Vegetated)

WETLAND, UPLAND, ALPINE

TRIAL 14
(August, 1993; 6-6-3 architecture; 10000 iterations; 38.88%)

(YELLOW = ALPINE; RED = UPLAND; GREEN = WETLAND)

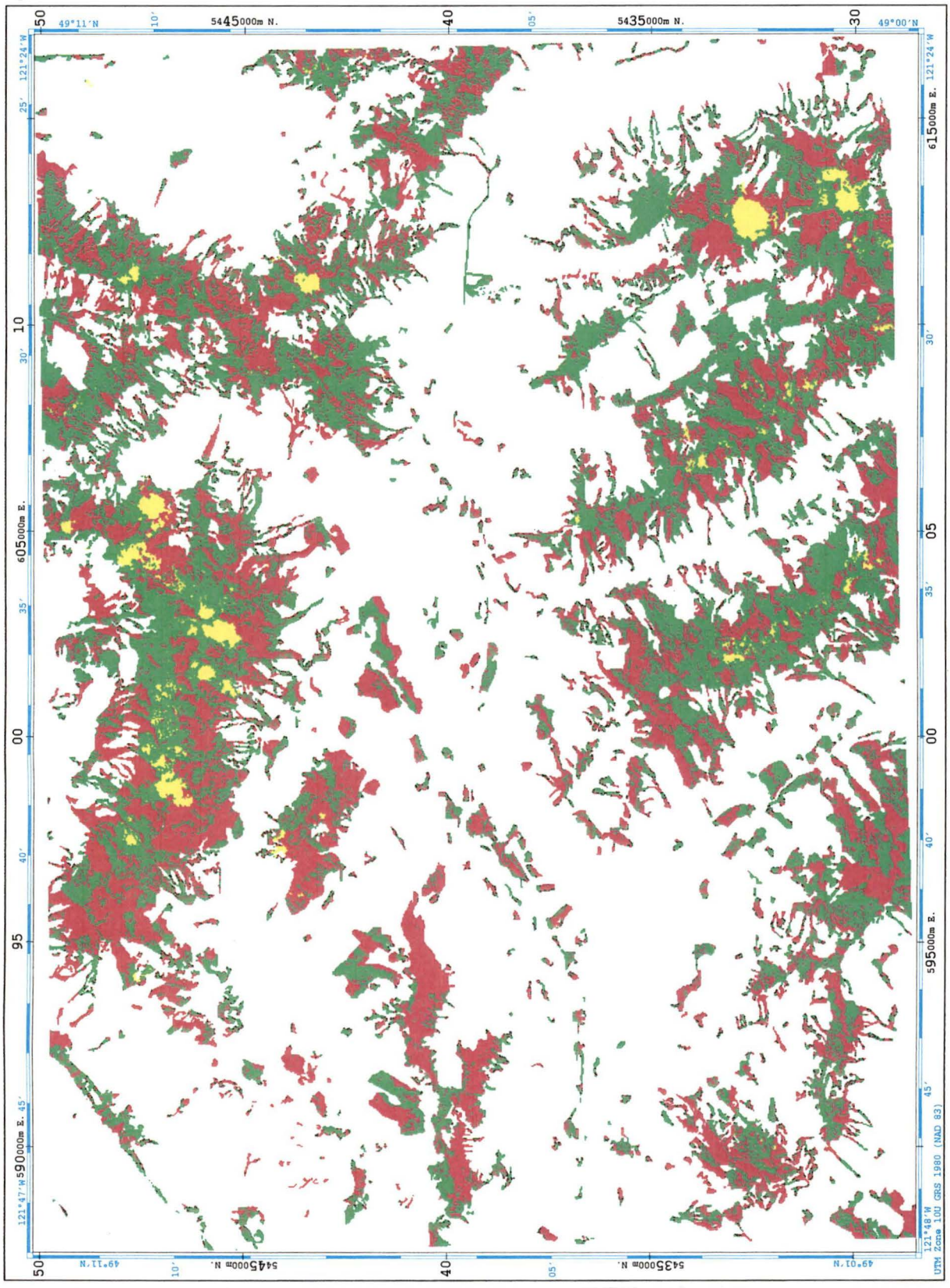


LEVEL THREE: LANDSCAPE POSITION (Non-Vegetated)

WETLAND, UPLAND, ALPINE

TRIAL 16
(August, 1993; 9-9-3 architecture; 10000 iterations; 33.33%)

(YELLOW = ALPINE; RED = UPLAND; GREEN = WETLAND)

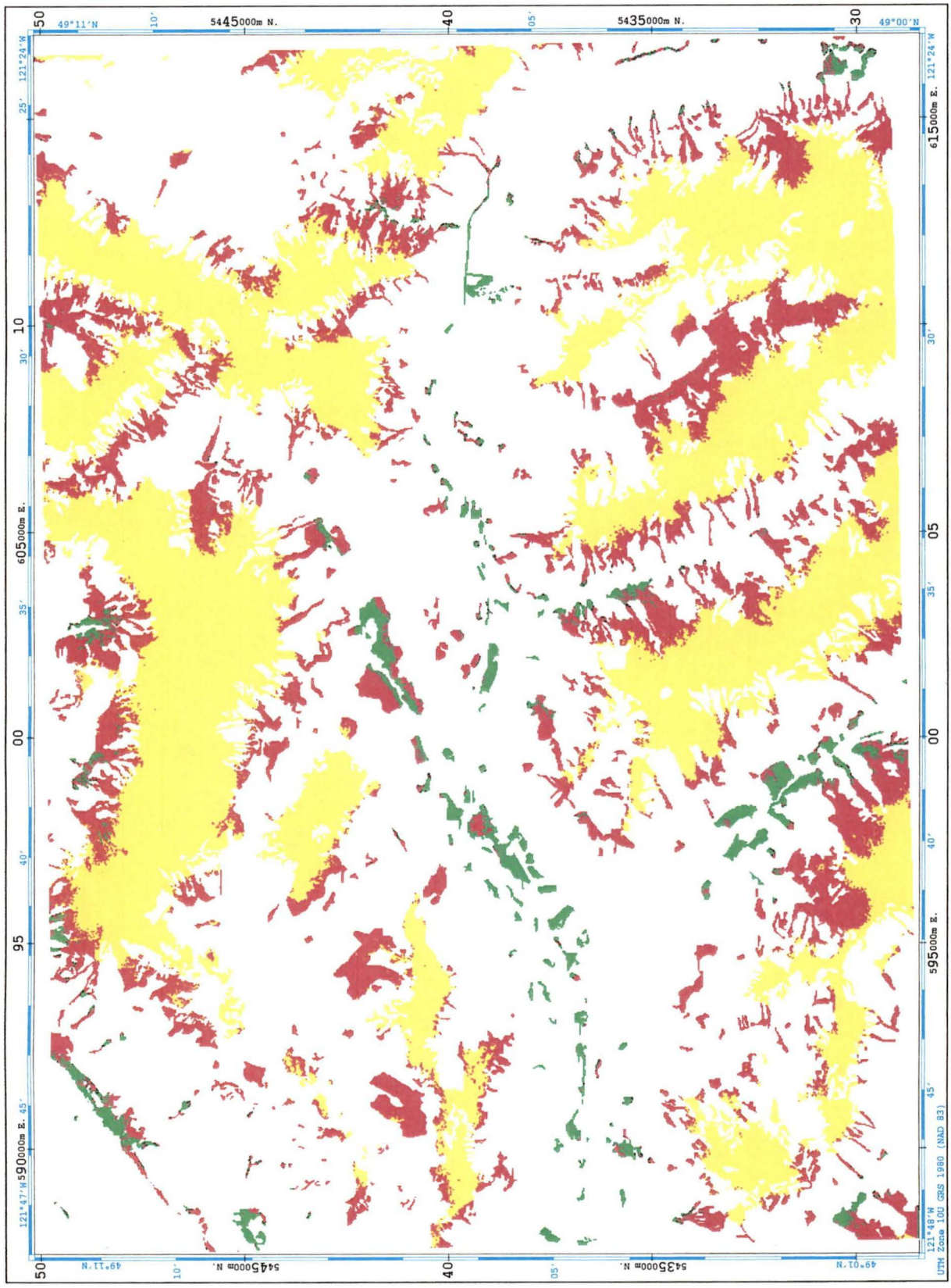


LEVEL THREE: LANDSCAPE POSITION (Non-Vegetated)

WETLAND, UPLAND, ALPINE

TRIAL 17
(August, 1993; 7-7-3 architecture; 5000 iterations; 72.22%)

(YELLOW = ALPINE; RED = UPLAND; GREEN = WETLAND)

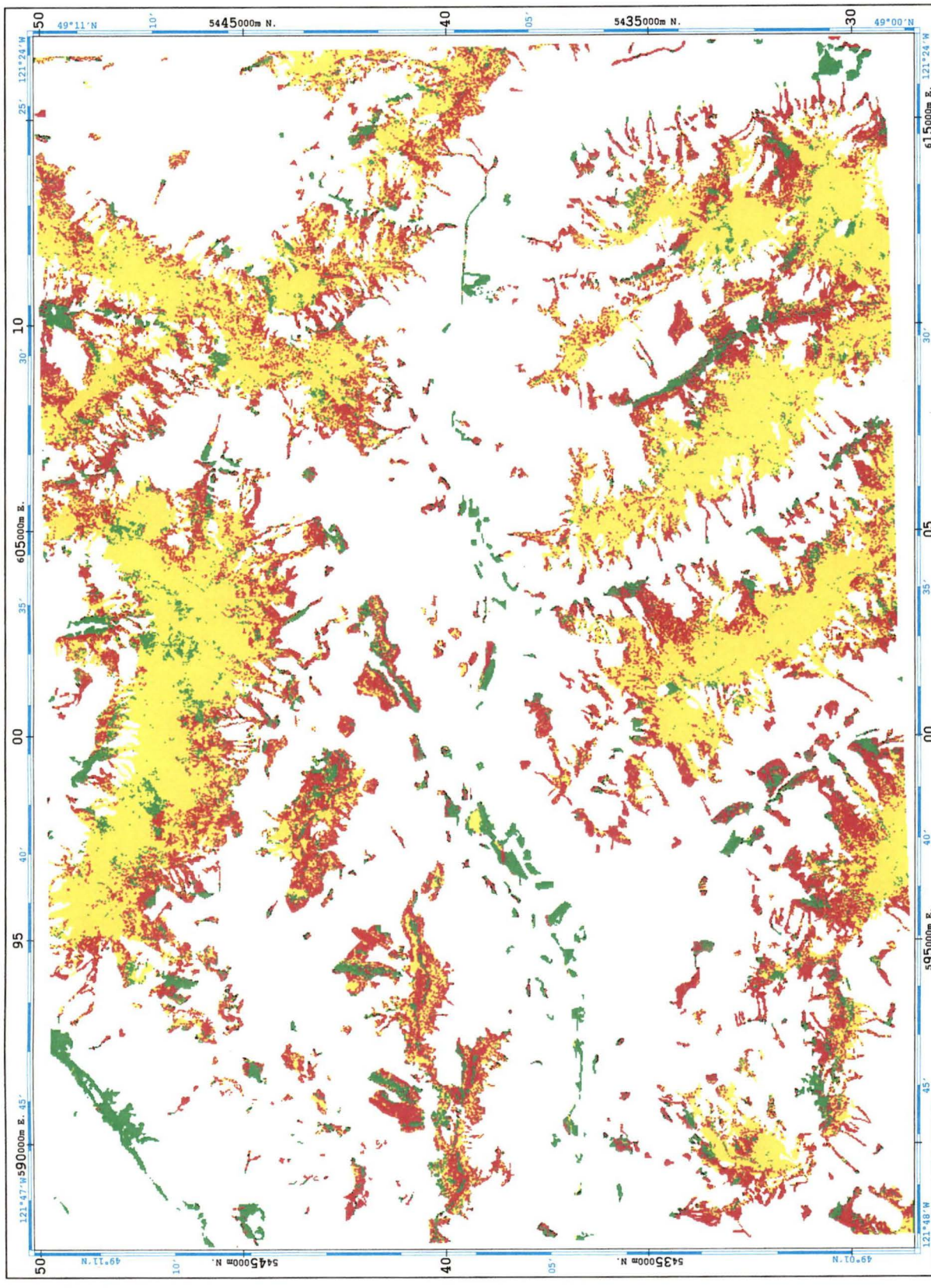


LEVEL THREE: LANDSCAPE POSITION (Non-Vegetated)

WETLAND, UPLAND, ALPINE

TRIAL 20
(August, 1993; 7-7-3 architecture; 10000 iterations; 77.77%)

(YELLOW = ALPINE; RED = UPLAND; GREEN = WETLAND)

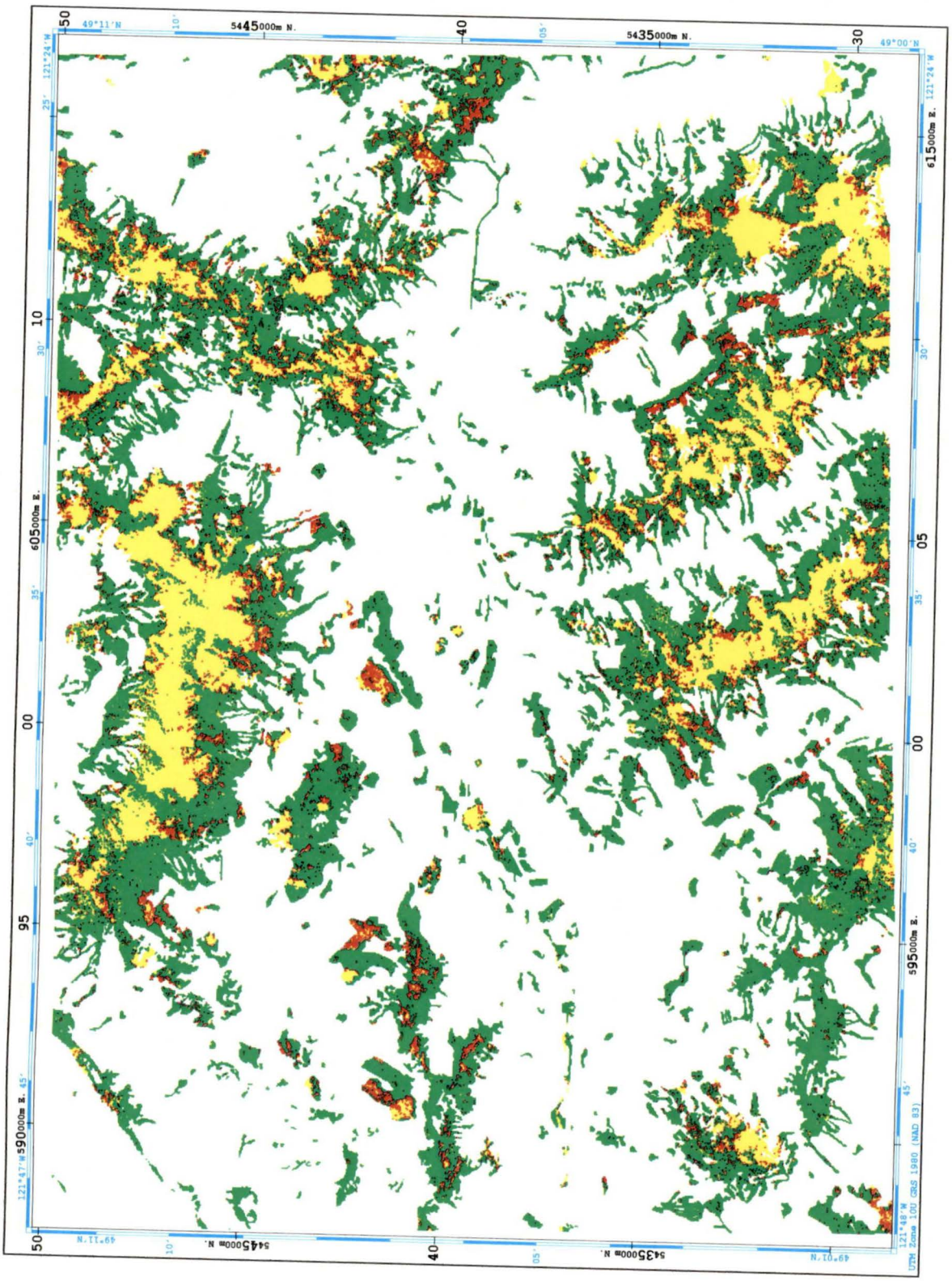


LEVEL THREE: LANDSCAPE POSITION (Non-Vegetated)

WETLAND, UPLAND, ALPINE

TRIAL 21
(August, 1993; maximum likelihood classification; 27.78%)

(YELLOW = ALPINE; RED = UPLAND; GREEN = WETLAND)

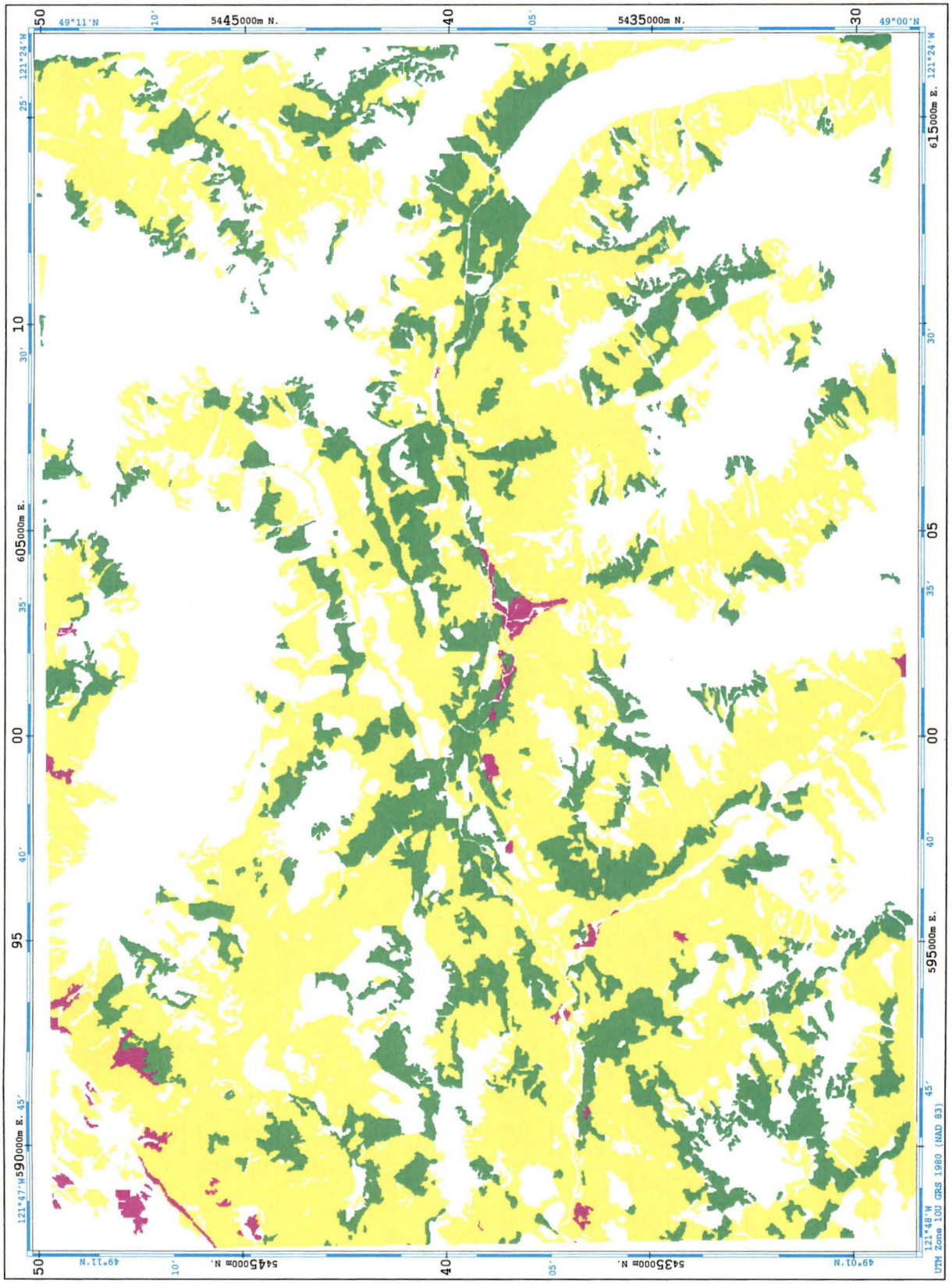


LEVEL FOUR: VEGETATION TYPE

CONIFEROUS, BROADLEAF, MIXED

TRAINING DATA

(YELLOW = MIXED; MAGENTA = BROADLEAF; GREEN = CONIFEROUS)

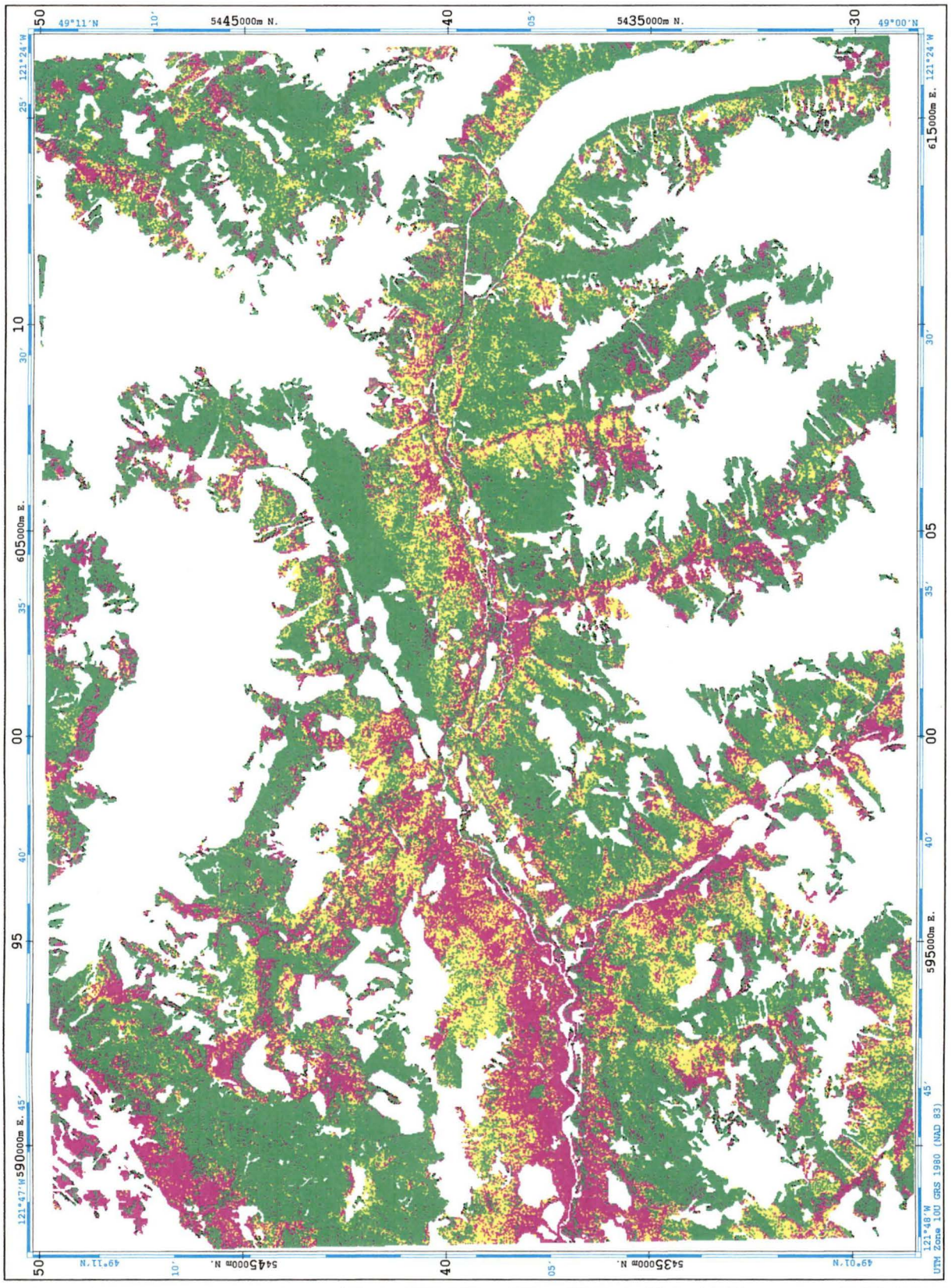


LEVEL FOUR: VEGETATION TYPE

CONIFEROUS, BROADLEAF, MIXED

TRIAL 1
(August, 1993; 6-6-3 architecture; 10000 iterations; 31.13%)

(YELLOW = MIXED; MAGENTA = BROADLEAF; GREEN = CONIFEROUS)

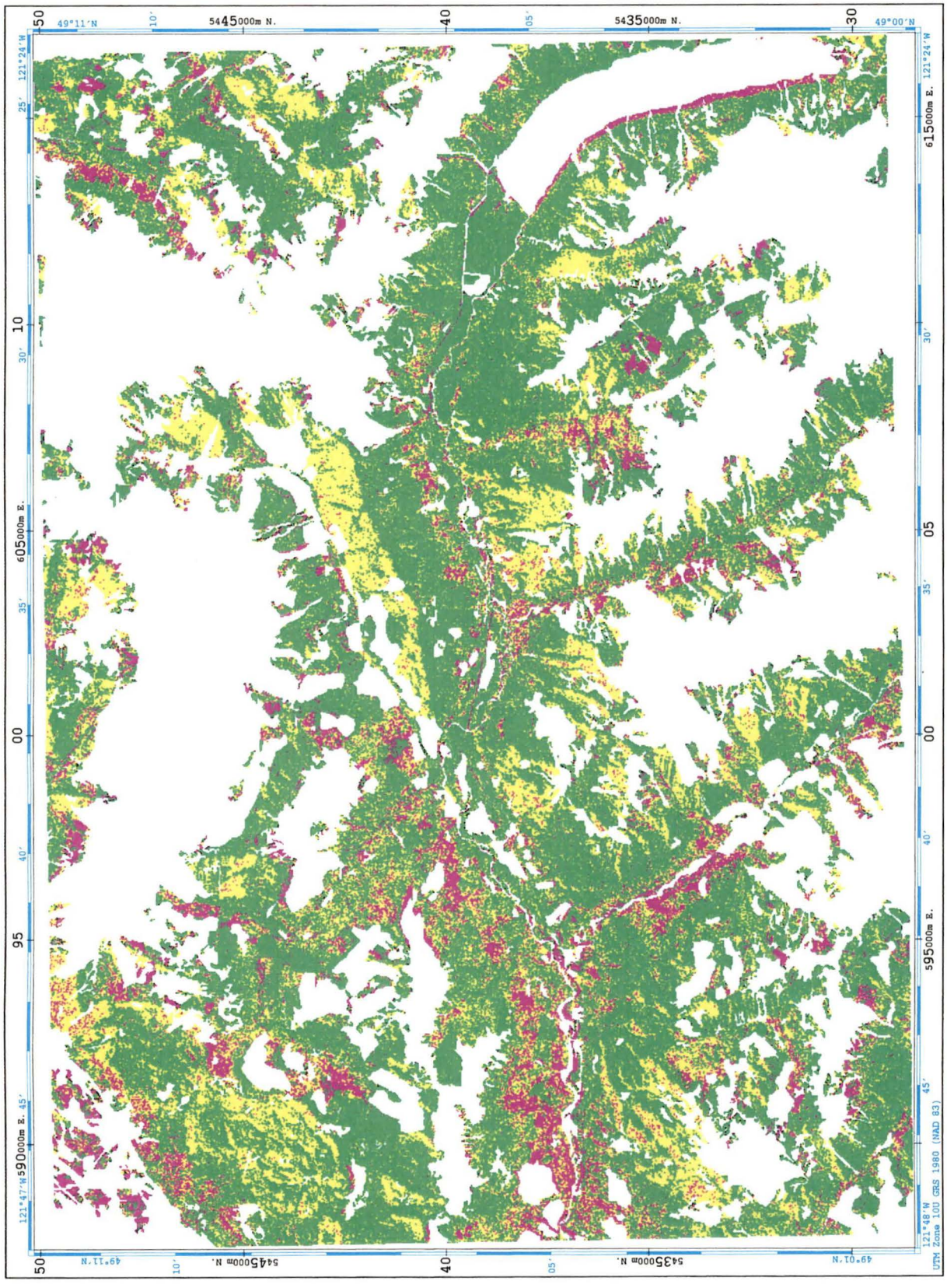


LEVEL FOUR: VEGETATION TYPE

CONIFEROUS, BROADLEAF, MIXED

TRIAL 5
(August, 1993; maximum likelihood classification; 36.99%)

(YELLOW = MIXED; MAGENTA = BROADLEAF; GREEN = CONIFEROUS)



LEVEL FOUR: VEGETATION TYPE

CONIFEROUS, BROADLEAF, MIXED

TRIAL 6
(August, 1993; 6-6-3 architecture; 10000 iterations; 29.89%)

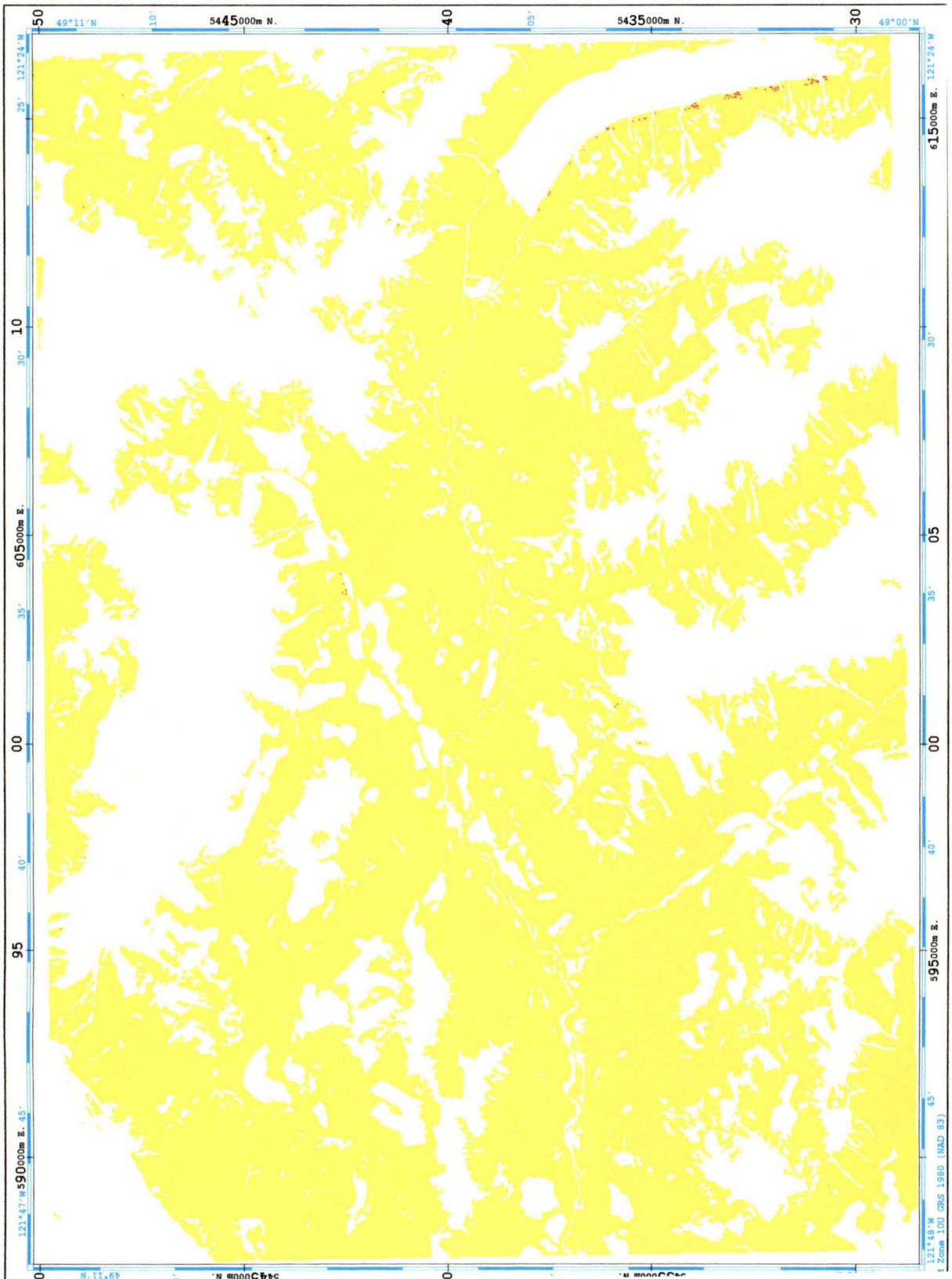
(YELLOW = MIXED; MAGENTA = BROADLEAF; GREEN = CONIFEROUS)

LEVEL FOUR: VEGETATION TYPE

CONIFEROUS, BROADLEAF, MIXED

TRIAL 9
(August, 1993; 6-24-9-3 architecture; 10000 iterations; 72.34%)

(YELLOW = MIXED; MAGENTA = BROADLEAF; GREEN = CONIFEROUS)

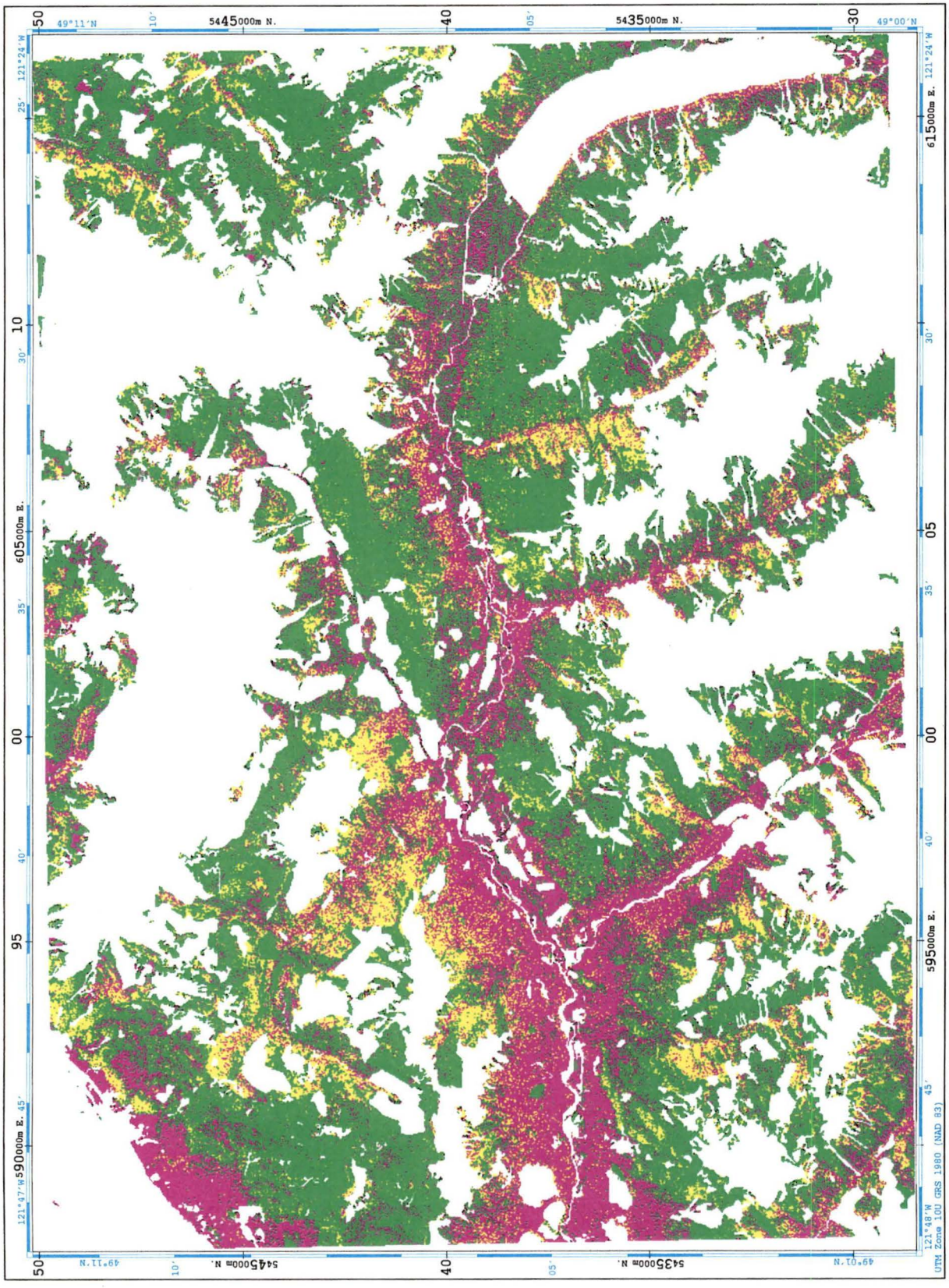


LEVEL FOUR: VEGETATION TYPE

CONIFEROUS, BROADLEAF, MIXED

TRIAL 11
(August, 1993; 6-6-3 architecture; 10000 iterations; 25.82%)

(YELLOW = MIXED; MAGENTA = BROADLEAF; GREEN = CONIFEROUS)

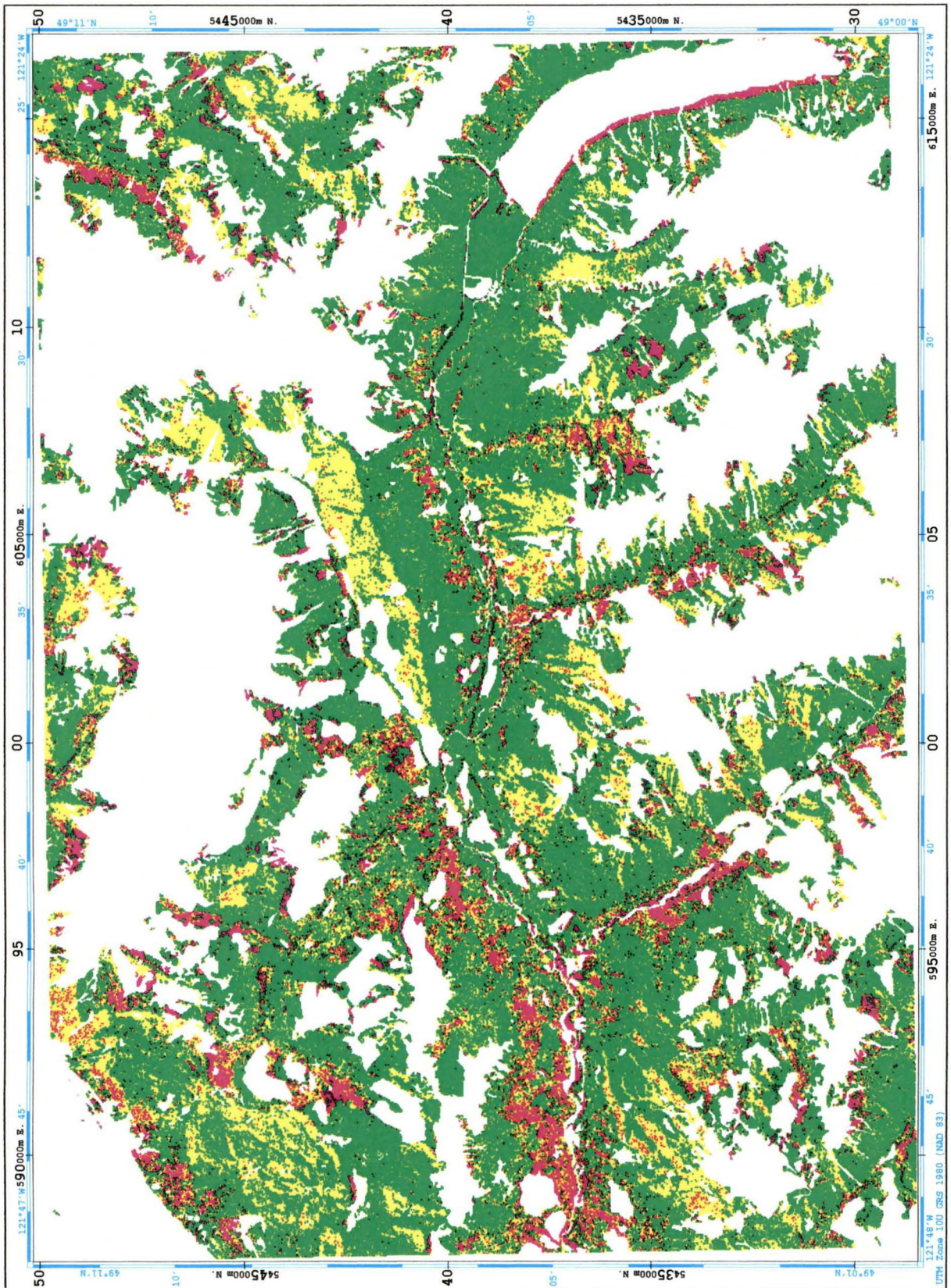


LEVEL FOUR: VEGETATION TYPE

CONIFEROUS, BROADLEAF, MIXED

TRIAL 12
(August, 1993; maximum likelihood classification; 37.17%)

(YELLOW = MIXED; MAGENTA = BROADLEAF; GREEN = CONIFEROUS)

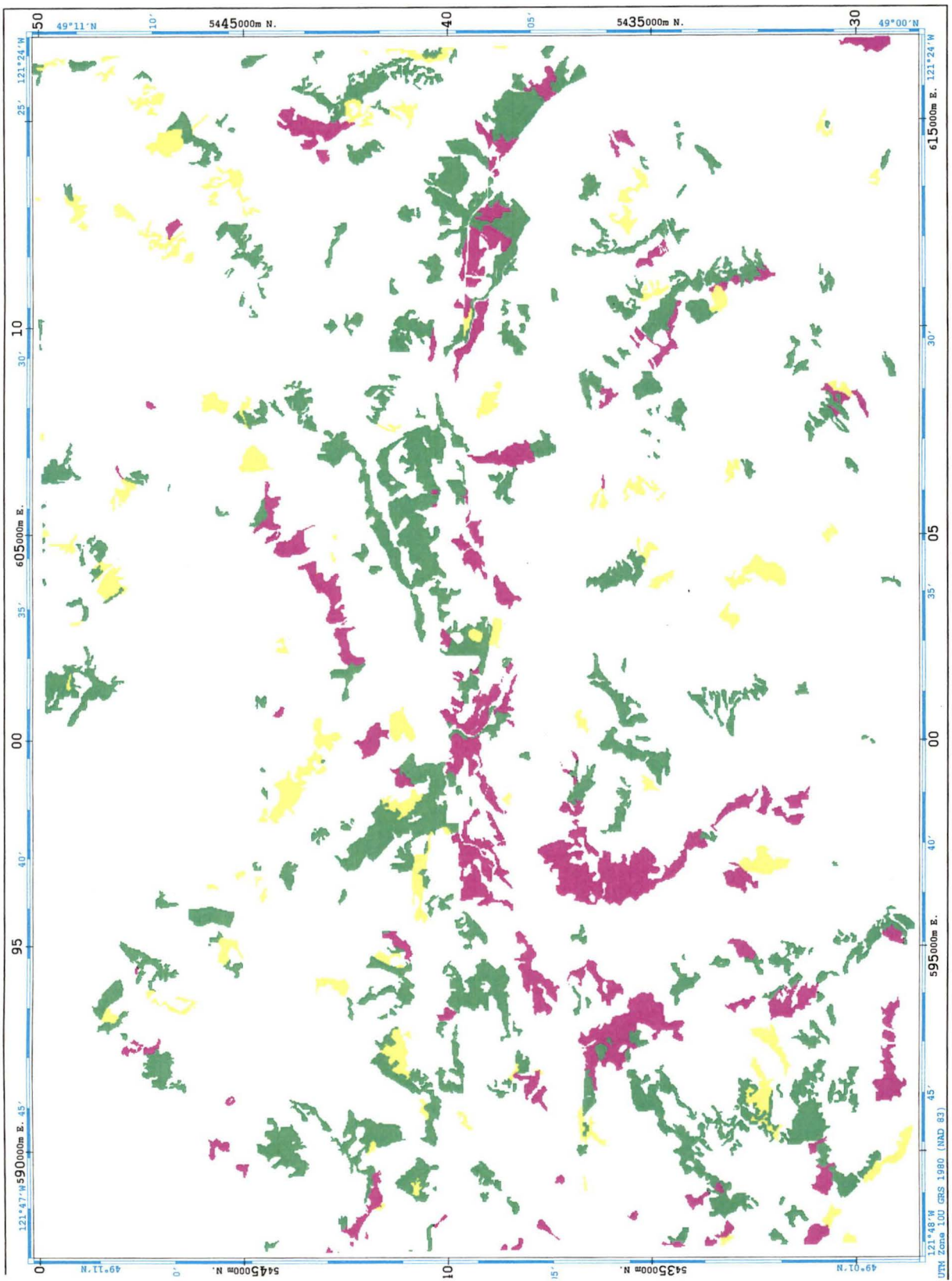


LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Coniferous)

DENSE, OPEN, SPARSE

TRAINING DATA

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)

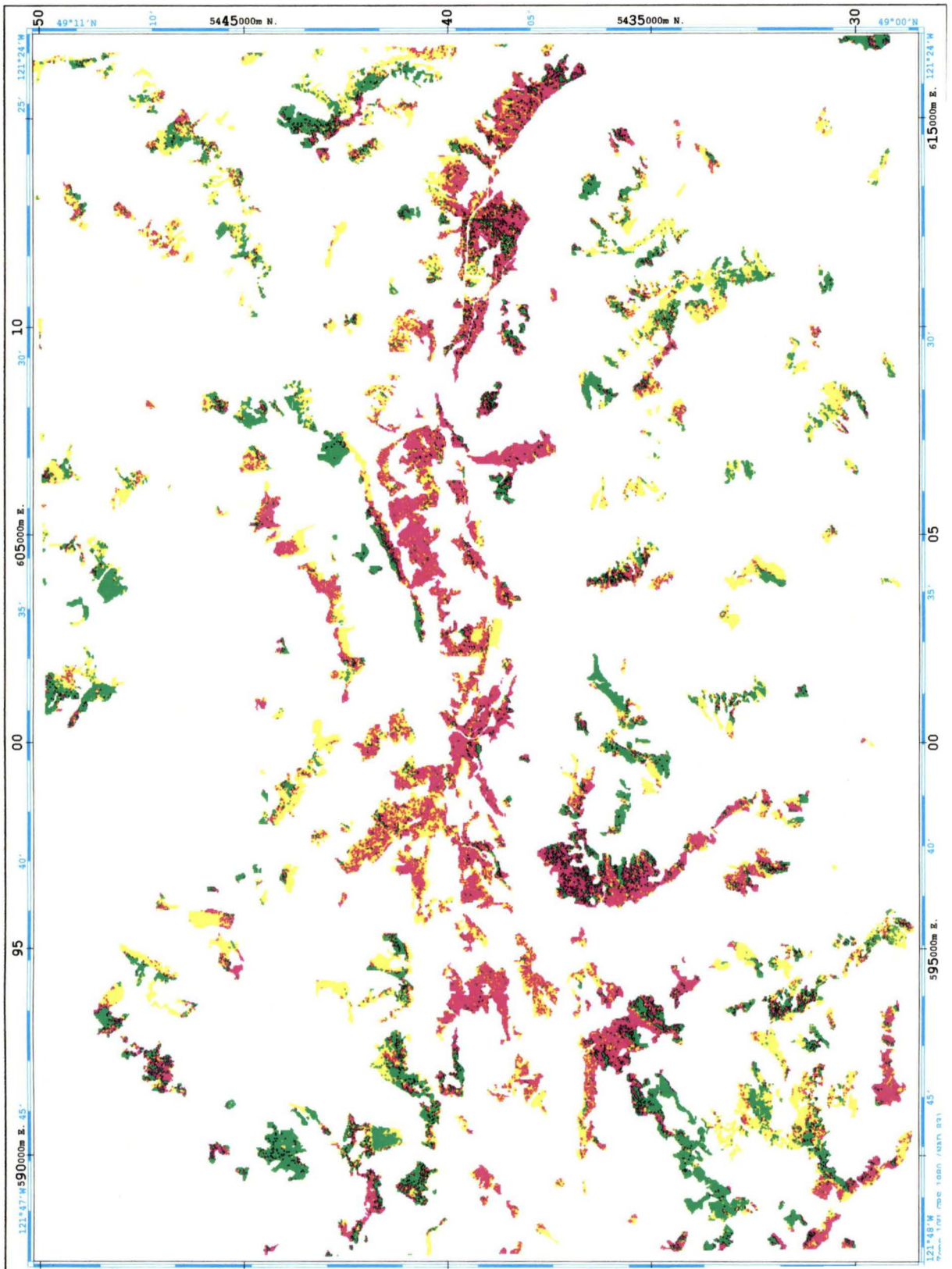


LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Coniferous)

DENSE, OPEN, SPARSE

TRIAL 1
(August, 1993; 6-6-3 architecture; 10000 iterations; 34.00%)

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)



LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Coniferous)

DENSE, OPEN, SPARSE

TRIAL 5
(August, 1993; maximum likelihood classification; 44.00%)

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)

LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Coniferous)

DENSE, OPEN, SPARSE

TRIAL 6
(August, 1993; 6-6-3 architecture; 10000 iterations; 39.58%)

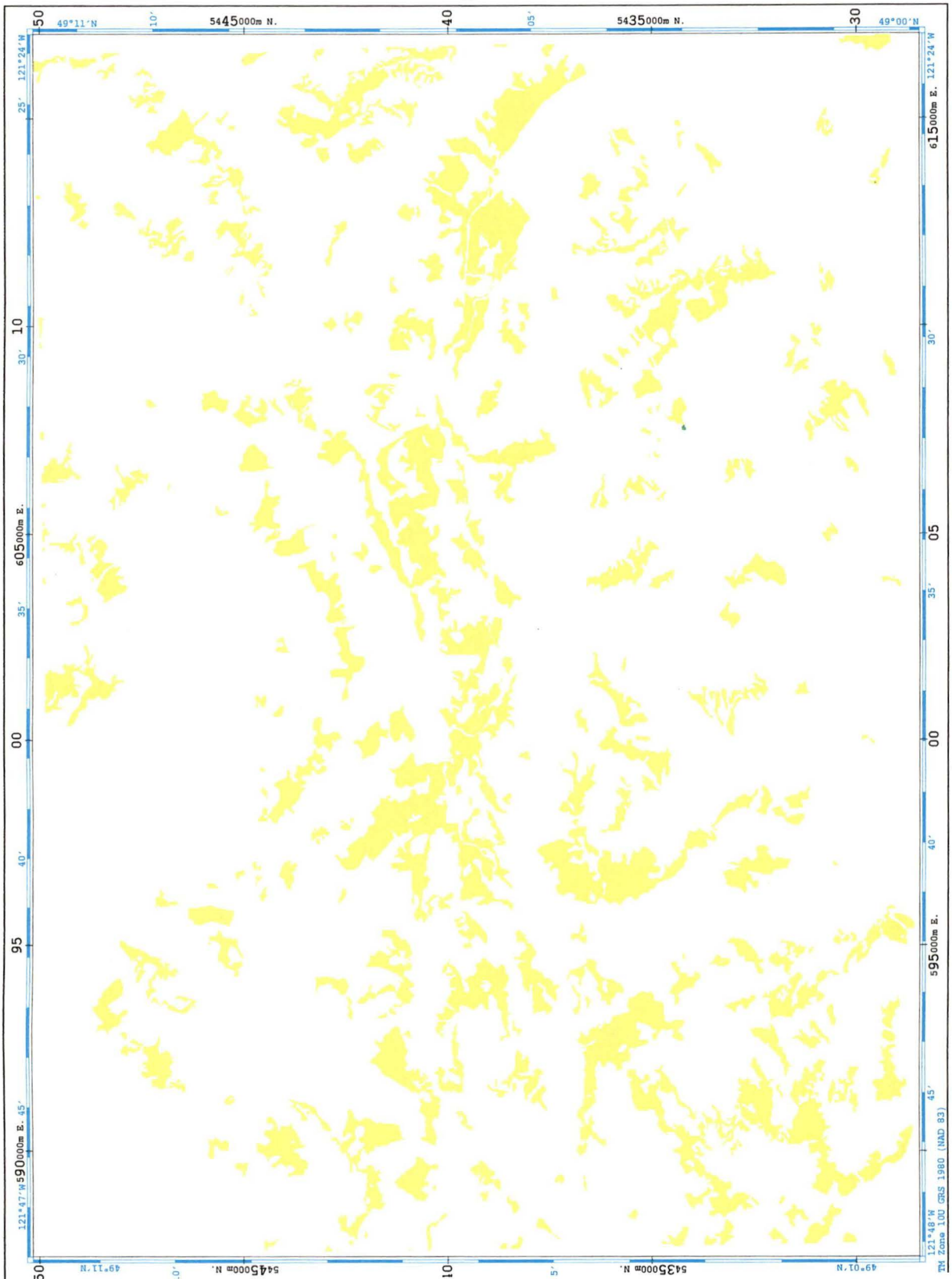
(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)

LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Coniferous)

DENSE, OPEN, SPARSE

TRIAL 9
(August, 1993; 6-24-9-3 architecture; 10000 iterations; 7.06%)

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)

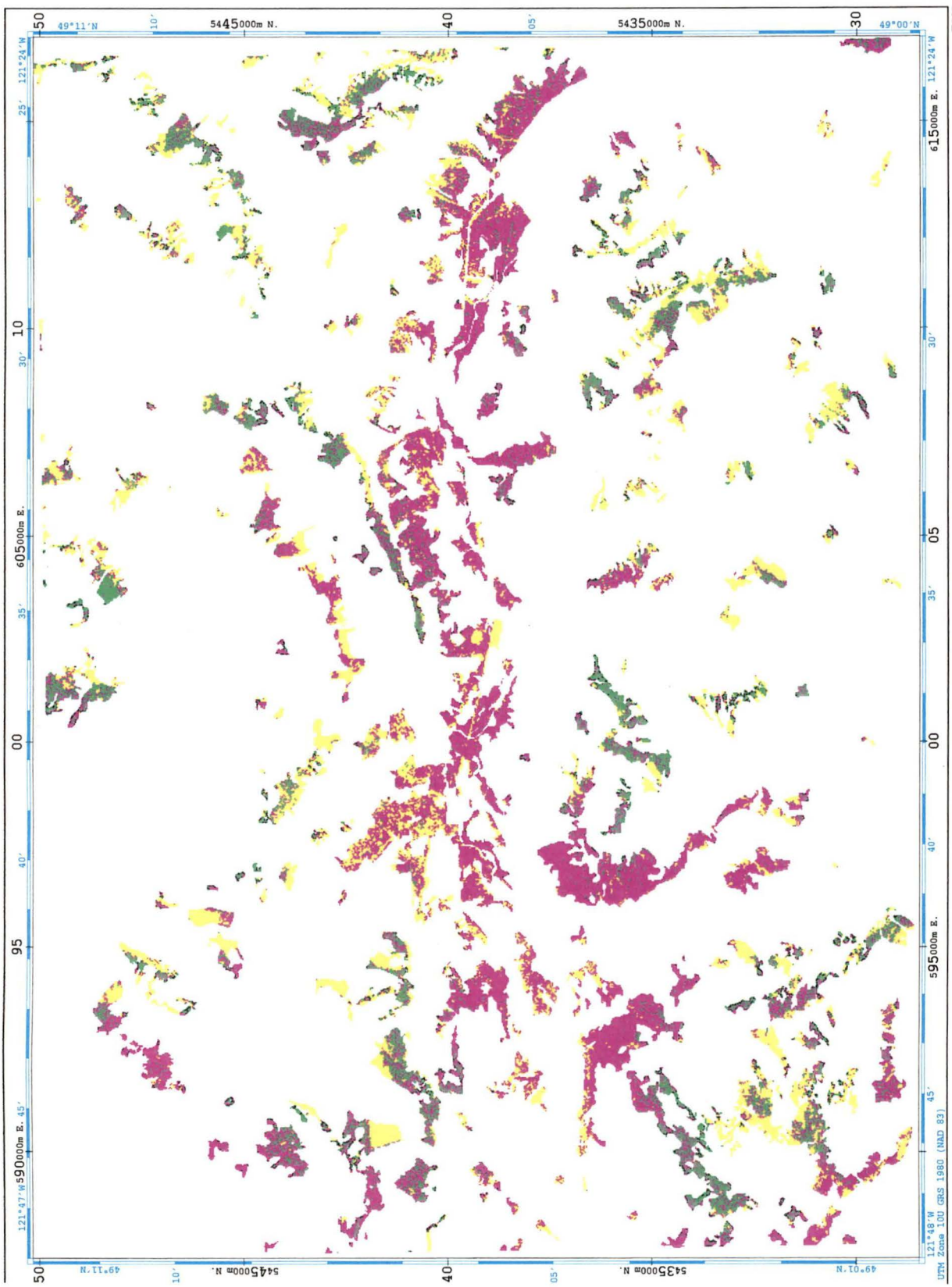


LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Coniferous)

DENSE, OPEN, SPARSE

TRIAL 11
(August, 1993; 6-6-3 architecture; 10000 iterations; 42.00%)

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)



LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Coniferous)

DENSE, OPEN, SPARSE

TRIAL 12
(August, 1993; maximum likelihood classification; 39.32%)

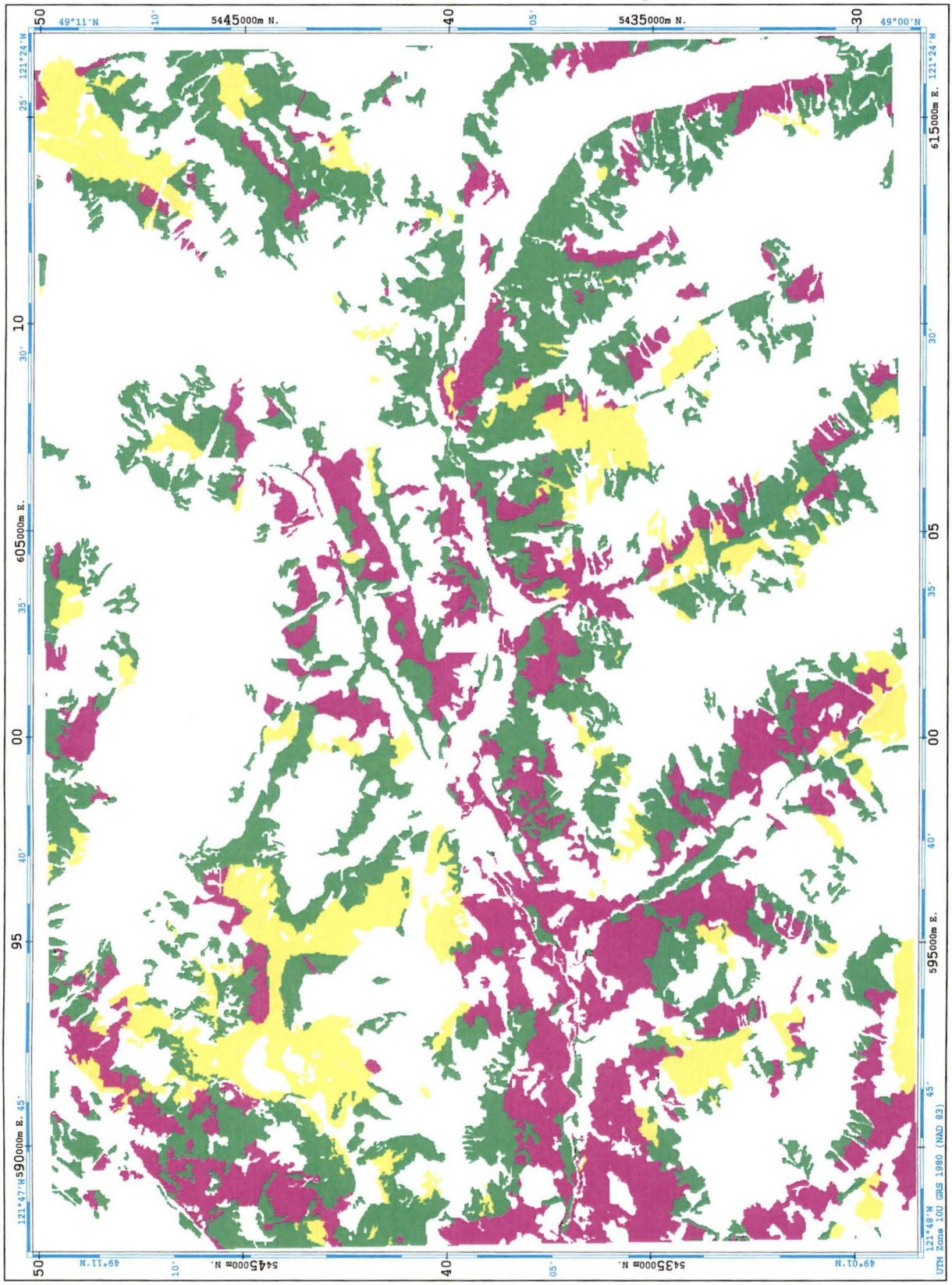
(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)

LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Mixed)

DENSE, OPEN, SPARSE

TRAINING DATA

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)

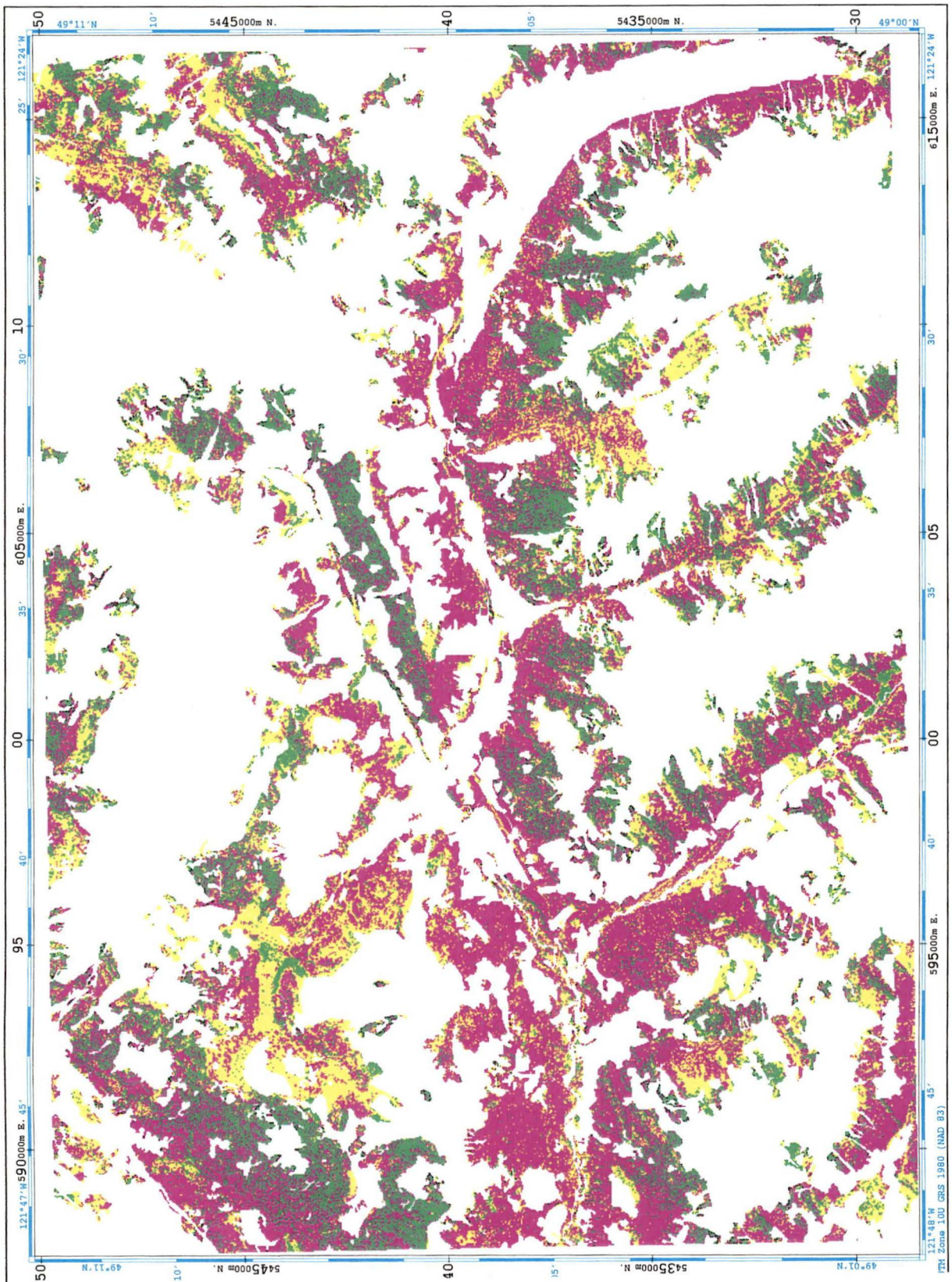


LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Mixed)

DENSE, OPEN, SPARSE

TRIAL 1
(August, 1993; 6-6-3 architecture; 10000 iterations; 42.03%)

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)

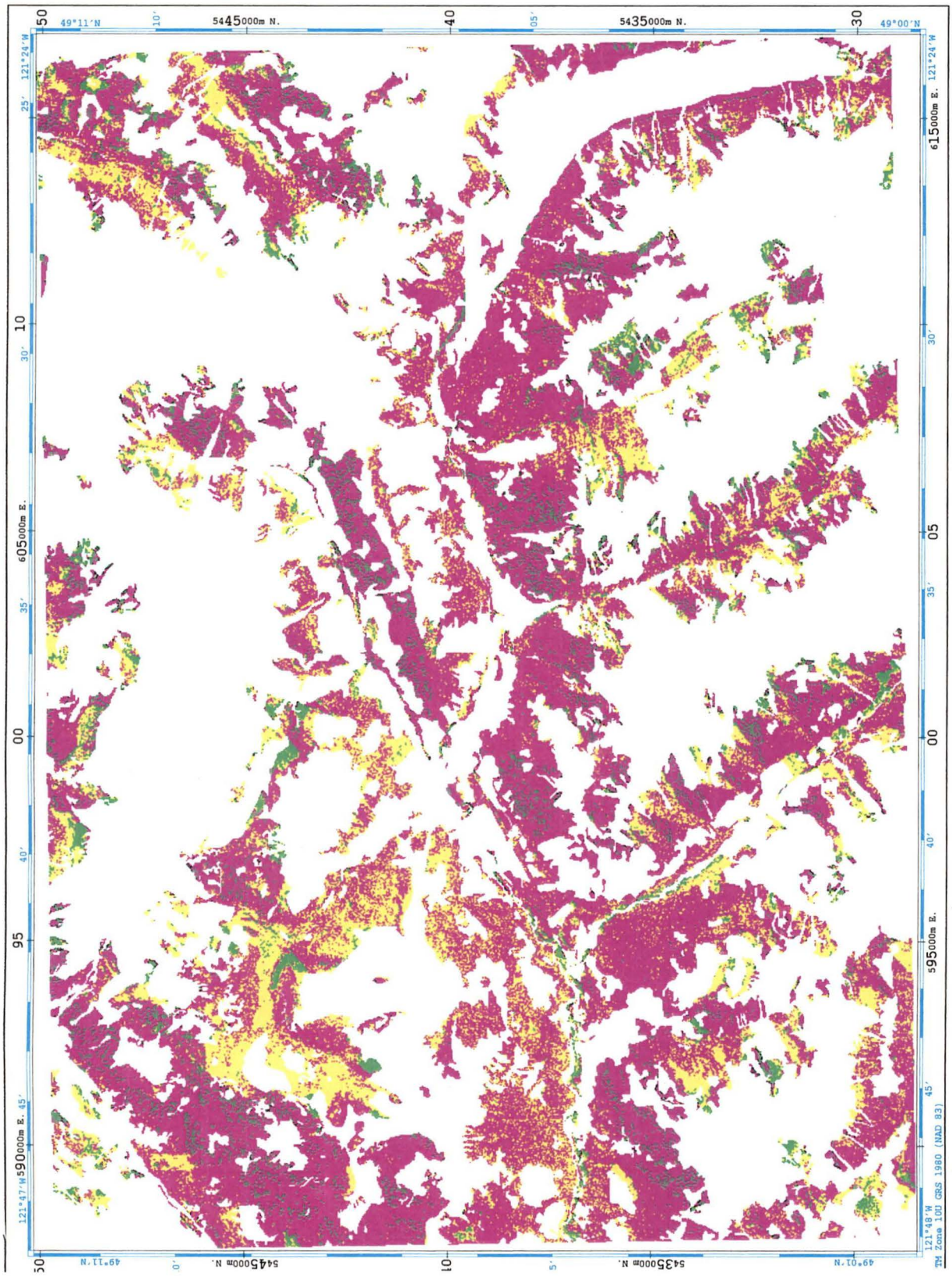


LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Mixed)

DENSE, OPEN, SPARSE

TRIAL 5
(August, 1993; maximum likelihood classification; 48.42%)

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)

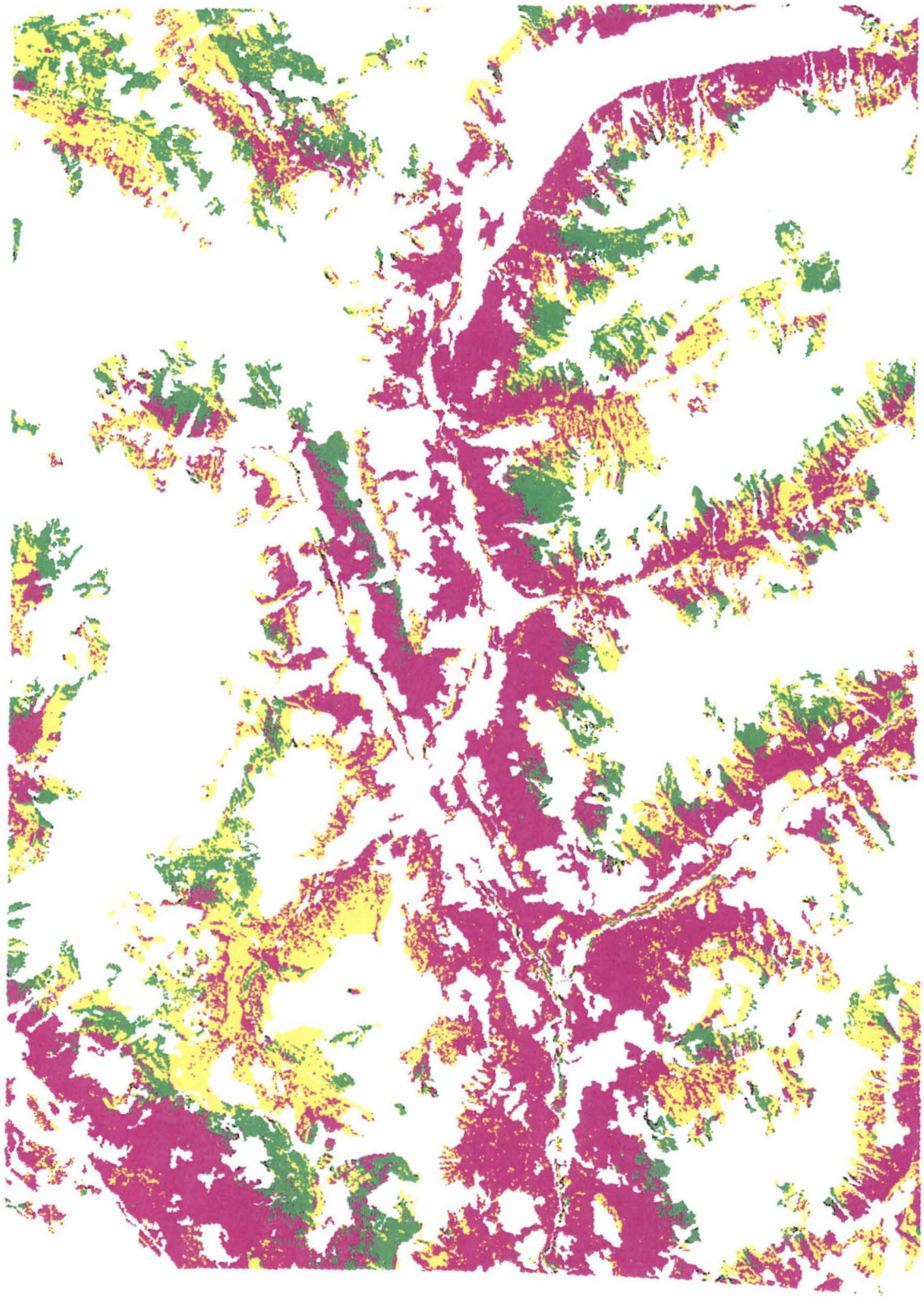


LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Mixed)

DENSE, OPEN, SPARSE

TRIAL 6
(August, 1993; 6-6-3 architecture; 10000 iterations; 77.51%)

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)

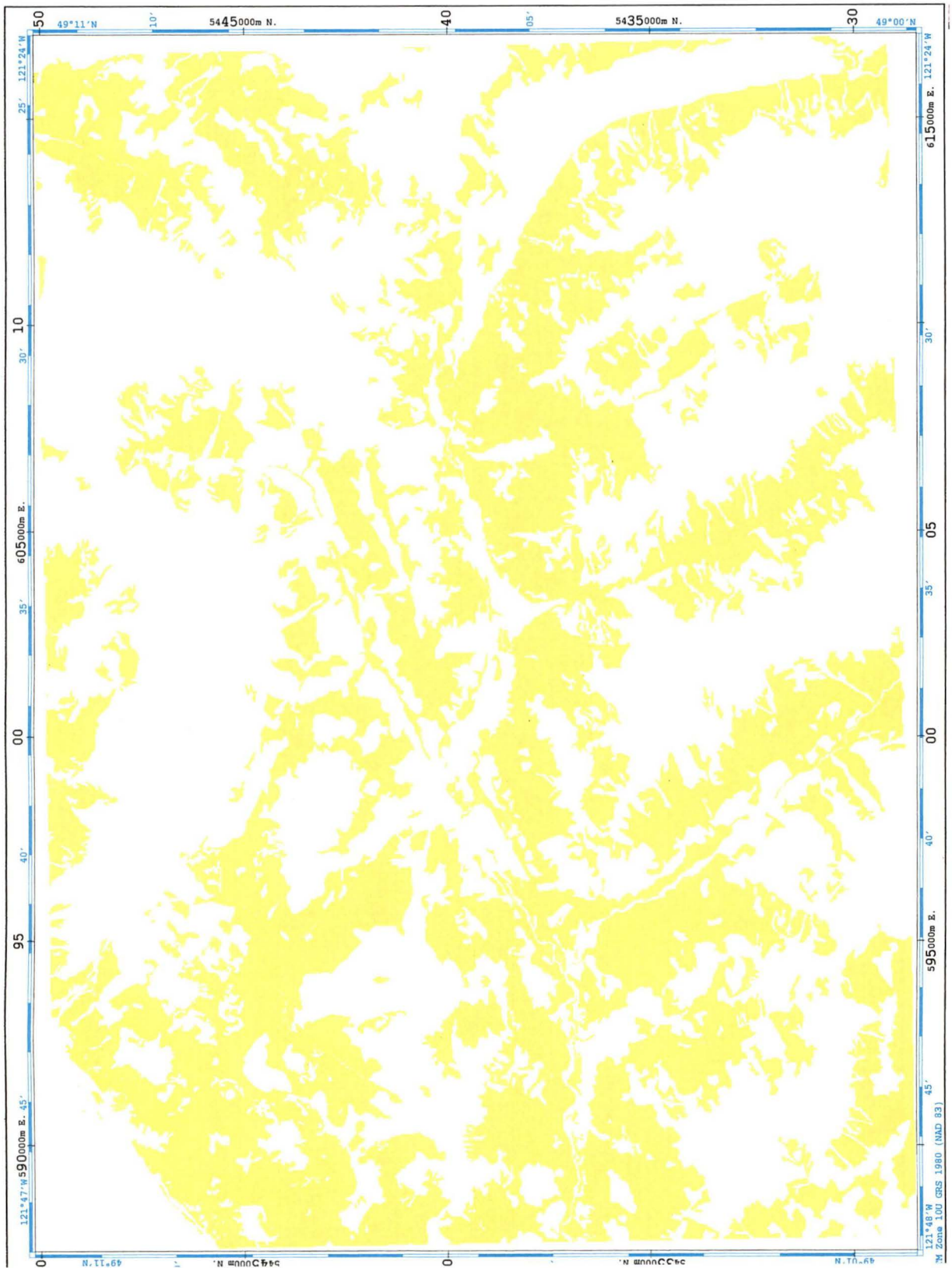


LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Mixed)

DENSE, OPEN, SPARSE

TRIAL 9
(August, 1993; 6-24-9-3 architecture; 10000 iterations; 16.94%)

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)

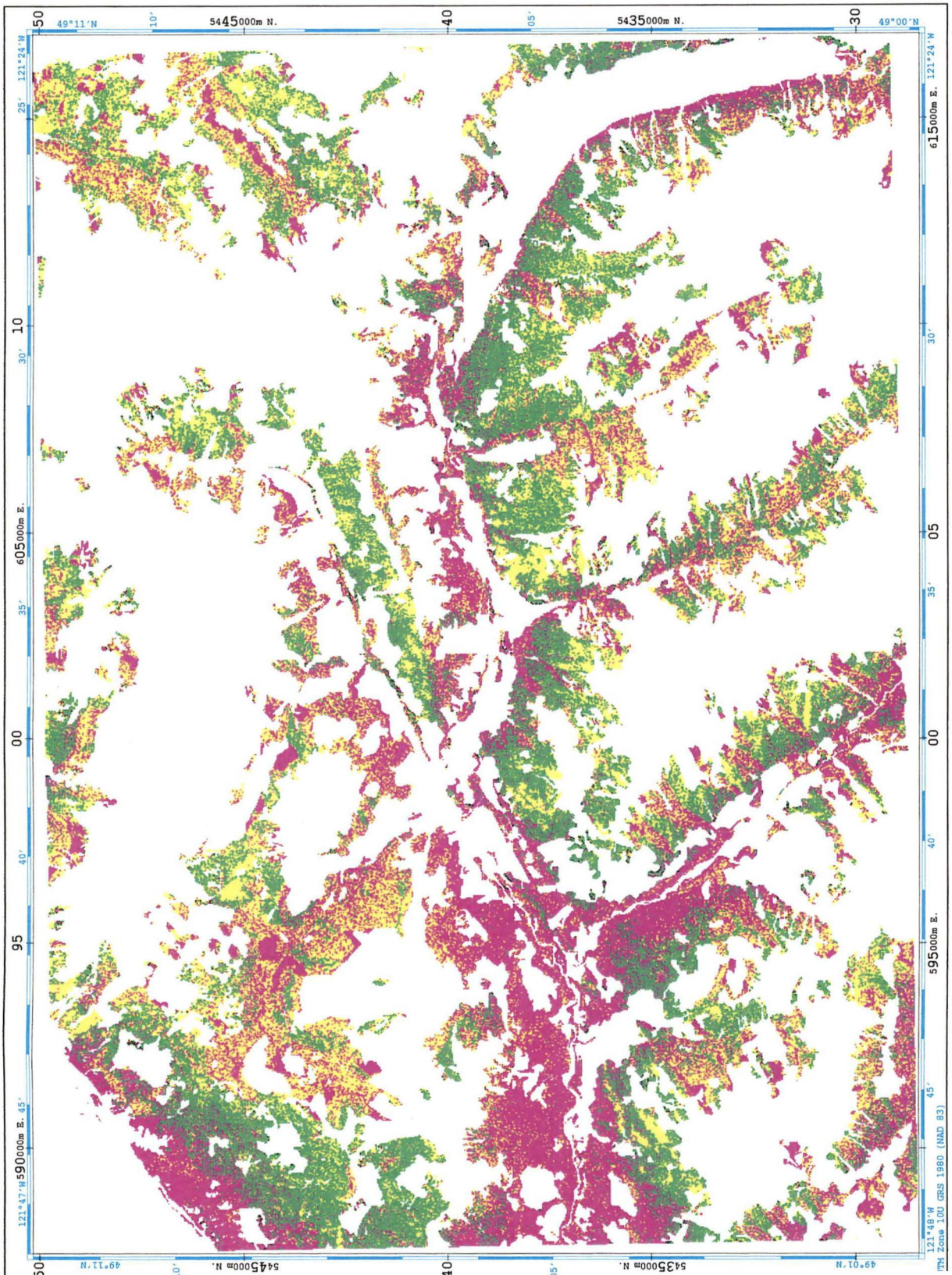


LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Mixed)

DENSE, OPEN, SPARSE

TRIAL 11
(August, 1993; 6-6-3 architecture; 10000 iterations; 39.66%)

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)

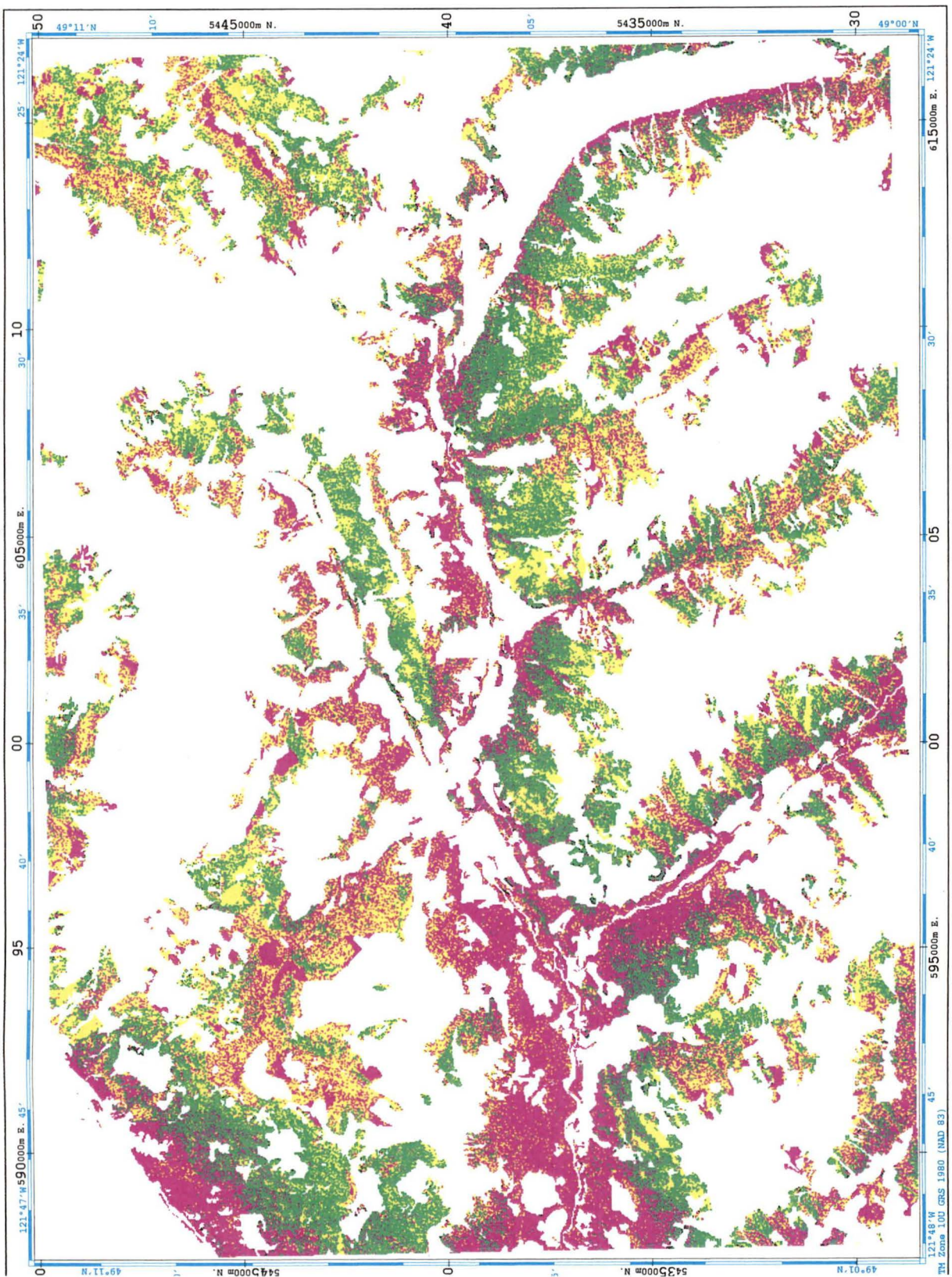


LEVEL FIVE: DENSITY CLASSES
(Vegetated Treed Upland Mixed)

DENSE, OPEN, SPARSE

TRIAL 12
(August, 1993; maximum likelihood classification; 39.32%)

(MAGENTA = DENSE; GREEN = OPEN; YELLOW = SPARSE)



VITA

Surname: White Given Names: Joanne Cheryl

Place of Birth: Moose Jaw, Saskatchewan, Canada

Educational Institutions Attended:

University of Victoria	1989-1994
University of Victoria	1995-1998

Degrees Awarded:

B.Sc. (Co-op)	University of Victoria	1994

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis:

An Investigation Into the Use of Artificial Neural Networks and Landsat Thematic Mapper Imagery for Vegetation Classification in Southwestern British Columbia

Author: [redacted]

Joanne Cheryl White
August, 9, 1998