THE UNIVERSITY OF ALBERTA

DEM Drainage As Ancillary Data to Enhance Digital LANDSAT Classification Accuracies

BY

Knut Olaf Niemann

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY.

DEPARTMENT OF GEOGRAPHY

SPRING, 1988

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

ISBN 0-315-42839-2

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Knut Olaf Niemann

TITLE OF THESIS: DEM Drainage as Ancillary Data to Enhance Digital LANDSAT Classification Accuracies.

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: Spring, 1988

....

#301, 10745-78 Ave
Edmonton, Alberta
Canada
T6E 1P7

Date: November 12, 1987

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for the acceptance, a thesis entitled "DEM Drainage as Ancillary Data to Enhance Digital LANDSAT Classification Accuracies" submitted by Knut Olaf Niemann in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

....
(Supervisor)

....

....

....

....

Date: November 12, 1987

Abstract

Studies integrating digital elevation models (DEMs) with multispectral digital satellite data have typically concentrated on geographic areas characterized by moderate to high topographic relief. Variables such as elevation, slope gradient and aspect contribute most significantly to the zonation of vegetation in these environments. In areas where relief is low, vegetation zonation is based not on individual form elements but rather on physical processes.

The purpose of this research was to integrate multispectral and ancillary process data in such a low relief environment. For this a study area was chosen in the Boreal forest of west central Alberta where the zonation of vegetation is based, to a large extent, on landscape drainage. An initial classification of forest cover based on LANDSAT multispectral data yielded overall classification accuracies of 58% with six clusters representing closed coniferous, deciduous, mixed coniferous/deciduous cover, and open soil. Séparation of wetland from upland areas was impossible.

A DEM was developed from a digitized 1:50000 topographic map sheet. The differential geometry of the DEM was mapped as a series of coverages: slope, aspect, and directional curvatures (down - and across slope). Two additional coverages, relief and a flow path model, were also developed and mapped.

A data set was created by which landscape drainage could be evaluated. Sixty sites, 20 for each of three drainage classes, were interpreted from aerial photographs and digitized from a 1:50000 map sheet. The surface form data at each site were extracted from the digital data base. A univariate analysis of drainage using the form variables resulted in a 45% to 47% explanation of the observed variation. Multivariate analysis combining slope gradient, across and down slope curvatures, relief, and flow paths increased the explanation to 68%.

The initial six MSS based clusters were reclustered inputting the surface form variables resulting in 21 clusters. These were labelled using the spectral values as well as the form data defined by the drainage modelling. The resulting classification accuracy was significantly increased from 58% to 73% with two additional forest cover classes being represented - muskeg/fen and mixed coniferous. This increase in accuracy was also significant in that it showed that DEM information from an area of moderate to low relief could be used to improve digital classifications.

Acknowledgments

I would like to express my gratitude to Dr.J.R. Eyton who provided direction and supervision throughout the past.

Dr. P.Crown, Mr. A.E. Peterson, Dr. A.H. Laycock contributed through discussions and suggestions throughout the project. Dr. P. Howarth also read and commented on the dissertation. I would also like to thank Dr. J.C. Muller for providing funding during the initial year of the degree and for his initial supervision.

I am indebted to the staff of the Resource Evaluation and Planning Division of Alberta Forestry, Lands and Wildlife, especially Mr. D. Thompson and Mr. R. Shelford for making available the various natural resource data bases which have been collated for the study area.

I would like to thank Mr. G. Lester and his staff for their help with all of my cartographic needs.

A special thanks to Martin Scott for all of his help and friendship during these past, and at times trying, years. Also I am grateful to my fellow graduate students for their input throughout the duration of the degree.

Finally I would like to thank Linda who has provided so much support and enthusiasm over the past three years.

Table of Contents

Chapter 1

Introduction

Chapter 2

Study Area and Existing Data Sources

Chapter 3

Conventional Classification of Forest Cover Types Using LANDSAT Multispectral Data

Chapter 4

Digital Elevation Modelling and the Extraction of Spatial Derivatives

Chapter 5

An Examination of Interpretive Map Data as a Potential Ancillary Data Source

Chapter 6

Landscape Drainage Modelling

Chapter 7

Integration of Ancillary and Multispectral Scanner Data

Chapter 8

Conclusions and Future Work

List of Tables

List of Figures

Figure		Page

CHAPTER 1

INTRODUCTION

1.1 Background

The need for the automated classification of digital remotely sensed data has long been recognized. Satellite data have proven to be of benefit in a number of applications in resource management. Hegyi and Quenet (1982) and Hegg and Driscoll (1982), for example, list areas of application where the classification of digital satellite data has been of benefit in the maintenance of forest resource inventories. The sheer volume of data which is collected by the various sensors renders the use of traditional manual interpretive techniques unfeasible. The only viable approach in dealing with such large volumes of data is to use computer - based classification procedures. In addition to the large data volumes, the growing importance of geographic information systems has introduced the need to access information in a digital format based on a spatial resolution unit which is of relevance to the user.

The utility of computer based classification using digital satellite data, however, has often been offset by poor accuracy and the inadequacy of the data for differentiation of cover classes. There are a number of reasons for these problems. The first reason involves the spatial and the spectral resolution of the sensor system and the resulting data. If the feature to be mapped is smaller than the spatial resolution of the data it will not be identified. A second, more fundamental reason can be seen by a comparison of approaches to automated classification and visual (manual) inspection. Over a period of time a methodology for the visual interpretation of imagery has been developed which uses multiple interpretative elements. These photo elements include tone, texture, size, shape, shadow, pattern, and site. Each of these elements contributes to the overall understanding of the observed feature. The conventional approach to the automated classification of remotely sensed data takes advantage of only one of these elements: spectral signature, which is analogous to the photo element tone. Some attempts have been made to incorporate the other interpretive elements into the classification process. The three most commonly attempted integration procedures involve the use of texture, pattern, and site. Spectral variability has been used to define the element texture. This can be accomplished through the calculation of the standard deviation of the spectral values within a designated size of neighbourhood (Goodenough et al., 1986). The resulting standard deviation image is then used as an additional data set in the classification procedure. Contextual classifiers dealing with spectral patterns have also been developed (see for example Tilton, 1983 and Harris, 1985). The improvements in classification accuracy achieved through the use of textural and/or contextual information in addition to satellite derived multispectral data has been limited. Harris (1985), for example, reported a 3.4% reduction in classification error with the introduction of contextual information.

The technique which shows the greatest potential for improvement is the use of site information. The position of a pixel in the landscape can be dealt with in a variety of ways depending on the nature of the classification. Soil characteristics have been digitally overlaid with other multispectral data to improve classifications. Similarly, topographic information has been incorporated with multispectral data. It is this final element, site, which has the greatest potential to improve the classification accuracy and is investigated in this study.

A number of studies have been reported in the literature which have integrated ancillary and LANDSAT data. Boresjo (1984) used a digital template to exclude non-wetland areas. This mask was constructed by digitizing areas delineated as wetland on topographic maps. Only the data falling within the wetland designated areas were interpreted.

Saterwhite et al. (1984) recognized that different vegetation communities, with similar spectral characteristics, were associated with different geomorphic areas. Separation of these vegetative communities was possible through the use of a geomorphic mask. An interpretation of the major geomorphic units: basin, alluvial fan, and mountainous areas was made. Digital masks, outlining each of these units were constructed and the vegetation cover was classified for each separately. Loveland and Johnston (1983) incorporated soil survey data, in addition to a slope coverage, with the LANDSAT derived cover types to develop sophisticated crop analysis models. Soil survey and slope data were used to determine the irrigation potential of an area.

LANDSAT derived cover classes were incorporated into this model to inventory crop type, determine water requirements and predict potential energy costs.

Niemann and Langford (1984) used slope gradient and soil survey data to improve wildland classification accuracies. Effects of both slope and aspect were corrected for through the use of digital terrain data. Wetland areas were stratified through the use of soil maps. Areas which had poorly drained organic soils were classified separately from the well drained uplands. Walsh (1980) found that by incorporating data from digital elevation models, a sufficiently detailed forest classification could be developed which was useful for mapping coniferous classes. Strahler et al. (1978) noted that the low classification accuracies of forest cover classes resulting from low spatial and spectral resolution could be improved by exploiting the ecological preferences of vegetation species. By introducing slope gradient and aspect and elevation into their classification process, they concluded that elevation contributed the most to improve the classification accuracy, while slope and aspect contributed the least.

Franklin and LeDrew (1984) and Franklin (1987) incorporated digital terrain with LANDSAT data to facilitate the digital mapping of land classification units. It was concluded, as was the case with Strahler et al. (1978), that of the terrain variables used, elevation contributed the most to improve the MSS- based classification. Mapping accuracy increased from 46 to 65 percent by incorporating elevations with the MSS data, while slope and aspect only contributed an additional 1% to improving the final classification. A common feature in these studies using digital elevation models was that they were located in moderate to high relief environments where either zonation of vegetation occurred by elevation or on the bases of slope gradient or aspect. Low relief areas may have similar terrain/vegetation relationships, but the interaction between the landform elements and vegetation cover may not be that simplistic.

1.2 Research Objectives and Hypothesis

The objective in this research project was to investigate the possibility of improving classification results obtained through the analysis and interpretation of digital data by the integration of ancillary data sources. There were fundamental differences between this study and previous studies which have attempted to integrate the various data sources. While earlier studies have addressed problems of ancillary data integration in areas of moderate to high relief, where the effects of individual terrain components such as slope and aspect are pronounced, this study investigated the potential for integration in an area of moderate to low relief where the individual terrain parameters were of less importance. A review of literature, included in Chapter 6, indicated that the influence of terrain on vegetation in areas of low relief was much more complex than was the case where elevation differences were extreme over short distances. Therefore, the initial questions to be addressed were:

1) **What types of ancillary data are the most useful for integration in low relief terrain?**

2) How may these data be integrated with the remotely sensed data?

An analysis of the literature demonstrated that a relationship existed between forest cover characteristics and landscape drainage. Based on this, the working hypothesis for the project was that an improvement in forest cover mapping accuracy could be achieved through the integration of landscape drainage information. The term landscape drainage applies to the process by which water is removed from the landscape. This topic will be dealt with in detail in Chapter 6.

The remainder of this dissertation will deal exclusively with LANDSAT multispectral data. It was recognized that other higher resolution sources of digital remotely sensed data do exist, including Thematic Mapper and SPOT satellite and airborne data for example. LANDSAT multispectral data were chosen for use in this study because of the large number of scenes which have been archived since the initial LANDSAT 1 satellite was launched in 1972. The high degree of accessibility renders this data source substantially more attractive than other types of remotely sensed data mentioned above.

1.3 Organization of the Dissertation

The dissertation is divided into eight chapters. Chapter 1 is an introduction to the study and the objectives and working hypotheses are stated. The geographic area chosen to test the working hypothesis and the existing data bases available from outside sources are described in Chapter 2. These data include sources which will be used to generate the data base necessary for the forest cover classification as well as information sources required to evaluate the results of the analysis. A conventional classification of forest cover from LANDSAT data is presented and discussed in Chapter 3. Included in this third chapter is a discussion of geometric corrections and the effects of terrain geometry (slope, aspect, and relief) on the spectral reflectance recorded by the LANDSAT Multispectral Scanner Subsystem (MSS), followed by a description of the classification procedure and an analysis of the results. Digital terrain models and the modelling process are addressed in Chapter 4. A discussion of surfacing techniques is presented in this chapter followed by the procedure used to surface the study area. A description of the spatial derivatives generated and used in this study is also presented. A comparison between two types of thematic coverages is made in Chapter 5. These data sources represent viable options for integration with multispectral data in classifying forest cover. An evaluation of their suitability as ancillary data products for integration with satellite multispectral data was appropriate for this research project. Landscape drainage modelling is discussed in Chapter 6. In this chapter the literature related to two topics is reviewed. The first topic is the relationship between forest stand composition and soil drainage. The second topic is the effect of terrain geometry on landscape drainage. The conventional digital landform elements described in Chapter 4, as well as less conventional elements used in the drainage analysis are described. A discussion of the modelling processes follows. The integration of the LANDSAT data and ancillary landscape drainage data is described in Chapter 7. This includes a review of methods previously used to achieve this integration. The procedure adopted for this study and an analysis of the results through an error analysis is presented. A comparison with the results obtained through the use of the MSS data alone is made. Chapter 8 is a summary of the findings presented in the dissertation and offers directions for further research. The methodology used in this research project is summarized in the form of a flow diagram in Appendix 1. This diagram has been included as a referencing aid to the reader.

CHAPTER 2

Study Area and Existing Data Sources

2.1 Introduction

Two criteria were used as the basis for the selection of the study area. The first criterion was that the area chosen contains a wide variety of landscape types and forest cover classes. The second criterion was the existence of a large data base for the area with relevant terrain and forest cover information. A site in west central Alberta was chosen to fit these criteria. The area is located in the Coal Branch area to the south of Hinton, Alberta (Figure 2.1).

2.2 Study Area

2.2.1 Physiography

The study area is situated to the east of the Eastern Slopes of the Rocky Mountains. In the eastern portion the physiographic unit is the Alberta Plateau of the High Interior Plains (Bentz et al., 1985). The western boundary is the eastern flank of the Rocky Mountain Foothills physiographic unit. The elevation of the area ranges from a minimum of 1000 metres above sea level in the northeast to 1600 metres in the southwest. The relief is gentle with a low, undulating topography in the eastern two thirds of the area. A ridge bisects the area from northwest to southeast

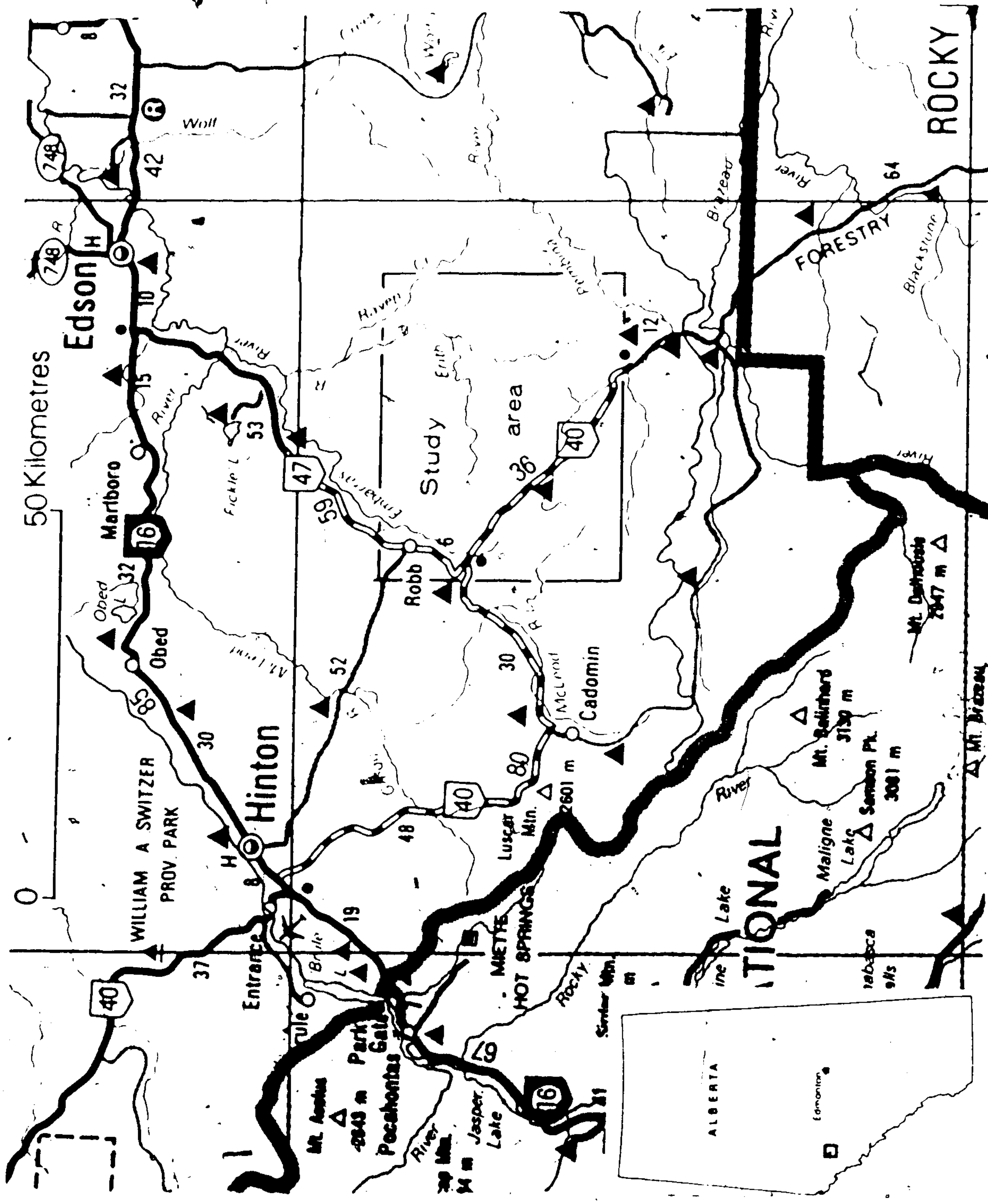


Figure 2.1. Location map

and accounts for most of the topographic variation. This ridge also represents the eastern fringe of the Rocky Mountain Foothills physiographic unit (Bentz et al. 1985).

2.2.2 **Surficial Deposits**

The surficial deposits are composed primarily of glacial till (Dumanski et al., 1972). The nature of the till has been strongly affected by the local bedrock geology, which is primarily sandstone, conglomerates and shale. The thickness of these deposits is variable, depending on the topography. Some small deposits of glaciofluvial and ice contact materials are scattered throughout. Modern alluvial deposits are found in the river valleys.

2.2.3 **Vegetation**

The most common forest stand type in the area is composed of coniferous vegetation. The most dominant conifer specie is lodgepole pine (*Pinus contorta*) which is found over a large proportion of the study area. Both white spruce (*Picea glauca*) and black spruce (*Picea mariana*), are located in the test area, but are not nearly as dominant as pine. Poplar (*Populus spp.*) is the dominant deciduous tree species. Balsam poplar (*Populus balsamifera*) is found in some of the poorly drained areas while aspen poplar (*Populus tremuloides*) is found in the well drained sites.

The wetland sites, gleyed and peat soils, have a predominant black spruce and tamarack (*Larix laricina*) coniferous tree cover. This grades to a deciduous shrub and a herbaceous cover with a decrease in soil drainage condition. A more detailed discussion of the vegetation distribution is presented in Chapter 6.

2.4 Data Sources

A large data base with relevant natural resource information exists for the selected test area. These data can be divided into two types. The first includes data used to classify the forest types. For this study these included LANDSAT images and topographic map data. A second type of data was used to verify the classifications based on the first. A LANDSAT 2 scene from August 12, 1981 was selected, because of the high quality of the imagery due to favourable atmospheric conditions. In addition, the midsummer date of the imagery should maximize the separation between deciduous and coniferous cover, as well as vegetative versus nonvegetative areas. The false colour composite of the satellite scene of the study area is presented in Figure 2.2. The topographic data used to develop the digital terrain model and the map derivative products were taken from a topographic map (NTS 83F2) at a scale of 1:50000 (Canada Department of Energy Mines and Resources, 1982).

A recent forest inventory compiled by the Alberta Forest Service, Alberta Department of Forests, Lands and Wildlife, was used for vegetation identification and verification. These data were available in a paper map format (the Phase 3 forest inventory maps) and in a statistical data base (the AFORISM data

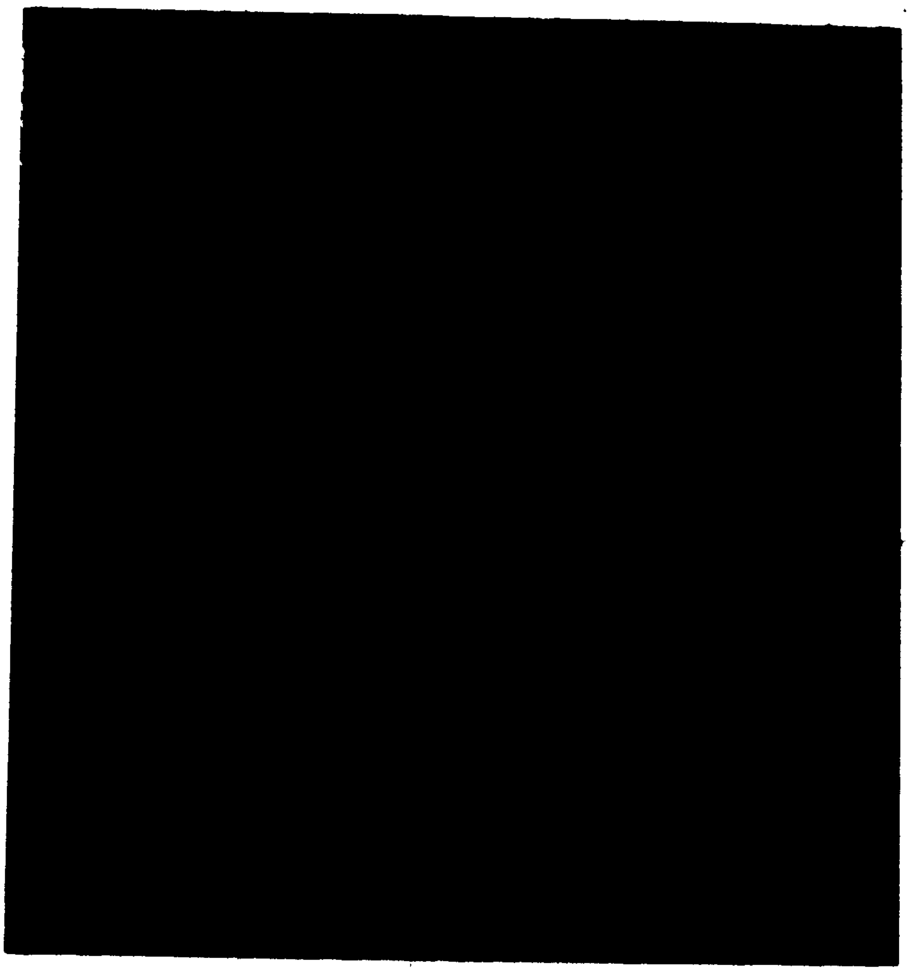


Figure 2.2 A false colour composite (bands 4,5, and 7) LANDSAT image of the study area. The image represents a 512 by 512 pixel window.

base). The Phase 3 maps used were produced at a scale of 1:50000 and displayed information on stand type, crown density and understorey composition. Additional information on commercial viability of each stand was also included. For the purposes of this study, only the stand composition and crown closure class were used. The AFORISM digital data base contains the forestry statistics complementing the Phase 3 inventory maps. The data are compiled and stored on a township basis. Only the percentages of crown closure were used in this study.

A physical land classification (PLC) compiled by Resource Evaluation and Planning Division, Alberta Forestry, Lands and Wildlife, was available for the test area. The terrain characteristics mapped as part of this evaluation included soil parameters, parent material genesis, drainage type, and classed slope gradient. The maps were produced at scales of 1:50000 and 1:150000.

2.4 Summary

To satisfy the objectives of this research, as stated in Chapter 1, a test area was selected. This area had to fulfill two fundamental requirements. The first requirement was variability in forest cover and physiography- geomorphology. The second requirement was that a large existing natural resource data base which included forest inventory and land classification information be available.

CHAPTER 3

Conventional Classification of Forest Cover Types Using LANDSAT Multispectral Data.

3.1 Introduction

The classification of MSS data and assessment of the resulting accuracy with respect to the mapping of large areas is discussed in this chapter. The classification of digital MSS data can be divided into four steps:

1) development of a classification key or legend,
2) selection of training sites and generation of training statistics,
3) classification,
4) accuracy analysis.

These four steps are discussed, first, with respect to multispectral classifications in general, and second with respect to their application to the classification of the forest cover for the test area used in the current study. In addition, two factors which could affect the classification results are examined. The first factor was the geometric correction of the multispectral data, and the second factor was the effect of the terrain geometry on the radiance values recorded by the multispectral scanner system.

3.2 LANDSAT MSS Data in Forest Cover Mapping

The extent to which MSS data can be used to map wildland forest cover has been reported in the literature. Beaubien (1979) discussed the results obtained in the mapping of large heterogeneous wildland areas using LANDSAT MSS data. In the first of two studies Beaubien (1979) noted that forest class definition and associated classification accuracies were high. This high degree of classification accuracy was due to the relatively simple landform units resulting in uniform, non - complex forest stands. The class labels that Beaubien (1979) generated, however, reflected the problem inherent in separating some types of forest stands. For example, peat bogs were grouped together with areas which had been deforested through clear cutting or disease. A second study conducted by Beaubien (1979), however, did not yield the same success rates as both class definition and classification accuracy decreased substantially in an area where the terrain and vegetation assemblages were more complex. He noted that the principal causes of reflectance differences between similar forest types are due to the slope gradient and aspect. Also, mixed hardwood/softwood stands did not yield substantially different signatures from stands of decadent conifers, and wetlands such as peat bogs were confused with mixedwood stands. Beaubien stated that "...reflectance expresses a balancing of factors so that a single class will contain forest types with more or less varying characteristics" (Beaubien, 1979, p.1142).

Bryant et al. (1980) attempted to define a standard forest inventory map legend and map a test area using a computer classification of satellite MSS data. Their goal was to match, as closely as possible, the computer classification to inventory legends commonly used in forest inventory studies. The classes mapped were softwood, hardwood, mixedwood, nonforested and water. The classification results indicated that a degree of confusion existed between the forest classes. Both commission and omission error rates were in the order of 30%. They found that wetland areas, or bogs, were consistently misclassified. A 92% classification error was found to be associated with this class. The bogs, they noted, were confused consistently with upland forests, and did not separate as a unique cover class.

Pettinger (1982), in the mapping of forest lands, was able to delineate major cover classes including conifer and sagebrush. Spectral overlap prevented the consistent separation of aspen classes, agricultural land and perennial grasses. The agricultural and grassland classes were particularly difficult to separate. Various stands of aspen, differentiated by the proximity to water and site drainage conditions, were also not distinguishable through the spectral data.

Hame (1984), in a study in Finland, observed consistently low classification accuracies of forest types. Although the amount of confusion between the various classes was not specified, the separation of the various cover classes, as evidenced by the feature space plot provided, was poor. Niemann and Langford (1984), working in the Boreal Mixedwood region of Alberta, concluded that the classification of conifers, based on spectral data alone, resulted in very high accuracies. This broad breakdown is not useful for most resource planning applications. Attempts to map the conifer classes at the species level reduced the accuracies significantly. Confusion between mixed conifer, open conifer and pine classes was high. The spruce dominant class had a low classification accuracy and exhibited a substantial amount of confusion between the mixed conifer and pine dominant classes.

An important feature of the majority of LANDSAT inventory studies is the poor discrimination of the various peatland and other wetland cover classes. Peatlands have been defined as areas where the production of organic materials exceeds the rate of degradation (Moore and Bellamy, 1974). More importantly, however, is that these features may be considered landforms, formed by an interaction between vegetation cover, drainage, and soils through time (Ivanov, 1981). The resulting floristic component which covers the peatlands is secondary in defining that landform (Wilde et al., 1954). As a result, peatlands do not yield a distinctly different vegetation cover from the surrounding uplands, at least with respect to the spectral characteristics recorded by the MSS.

Palylyk (1985) discussed the potential of differentiating peatlands and discriminating these areas from upland forest stands with multispectral data. Using data from the Boreal Mixedwood forest in Alberta, obtained for two dates during the growing season, Palylyk demonstrated that peatlands were spectrally distinct from upland areas with May imagery, whereas the wetland and upland cover types were not spectrally distinct when using imagery obtained later in the growing season. Although the muskeg/fens were separable with the May data, upland cover types were not. This lack of discrimination severely limits the potential usefulness of the imagery. The time window during which this discrimination can be accomplished is narrow, so that the relevant imagery may not be available. The timing of this window also varies from year to year depending on the prevailing climatic conditions.

3.3 Multispectral Scanner Data

Prior to discussing the definition of cover classes through the use of MSS data and the resulting accuracies, it is necessary to identify some of the potential sources of error which are of particular significance to this study. Two topics are dealt with in this section. Because this research requires two or more registered data sources based on different cartographic projections, the first topic is the geometric correction of the MSS data. The second topic is the variation in spectral reflectance caused solely by terrain geometry as discussed by Beaubien (1979), which, if substantial, can significantly affect the classification results.

3.3.1 Geometric Corrections

The geometric correction of LANDSAT data must be performed to correct for a number of errors. These errors, as outlined by Bernstein and Ferneyhough (1975), result from earth rotation, satellite velocity variations, altitude and attitude changes of the satellite, panoramic distortions and variations in mirror velocity. All of the above influences, with the exception of the altitude and attitude variations and panoramic distortions, are systematic and are easily corrected. The altitude and attitude variations and panoramic distortions, however, must be evaluated and dealt with for each scene individually. The magnitude of these perturbations can be determined through the use of ground control points (GCP). Once the errors have been modelled, a geometrically consistent map projection can be related to the error model. A second step assigns the appropriate digital number (DN), or spectral value, to the geometrically corrected grid cell.

The accuracy of the error model derived from the GCP's is determined by both the number and distribution of points. Butlin et al. (1978) and Shlien (1979) noted that data geometrically corrected at the Canada Centre for Remote Sensing (CCRS), using the Digital Image Correction System (DICS), used approximately 320 GCP's per LANDSAT frame, or 20 per 1:50000 map sheet, resulting in registration accuracies to within 50 metres. This accuracy claim depended on the number and distribution of the GCP's, as outlined by Orti (1983). He noted that GCP's should, in the first instance, be placed around the periphery of the scene. In the more populated areas, where potential points are abundant, the accuracy claim by Shlien (1979) could easily be met with a minimum number of points. In the less well developed areas, however, where accurate GCP's are less abundant, the optimum positioning of points may not always be possible. In these cases a greater number of points are used to compensate for the lower degree of positional accuracy.

The second part of the geometric correction process for MSS data, termed resampling, entails the assignment of a DN to a specific grid cell. There are a number of ways this can be accomplished. The simplest is the nearest neighbour method. In this approach the Dn of the geometrically - corrected pixel is assumed to be the same as that of the geographically closest pixel from the uncorrected data set. This method is the fastest computationally as no calculations involving the original reflectance data are performed.

A second approach to resampling is the convolution method, which uses mean DN values of a neighbourhood of pixels centered on the desired position. The cubic convolution interpolator applies a third order polynomial to a series of four points in the X and Y directions separately and calculates a mean grey level value (GLV) from the two lines for each pixel on the resampled data set (Shlien, 1979). This method creates a smoother image than is achieved with the nearest neighbour approach and avoids replication errors or missing pixels altogether. It is, however, computationally more demanding, requiring twenty times the computing time (Schowengerdt, 1983).

The radiometric accuracy of the spectral data after geometric correction is difficult to measure (Shlien, 1979). A variety of factors influence the radiometric values recorded by the LANDSAT MSS sensors. Some of these factors are independent of the sensors and not systematic, making compensation extremely difficult. For example, the resampling of pixels using the approaches to resampling which interpolate, or derive the new pixel GLV through the influence of neighbouring pixels, will integrate the values of the surrounding pixels so that if a very bright or dark grid cell is within the neighbourhood of the interpolator, the extreme values will influence the resampled data number. Dana (1982) stated that for very extreme variations in reflectance the effects were noted up to 500 metres away.

For the purpose of this study the CCRS DICS data were used. A cubic convolution resampling approach was selected over the nearest neighbour method. Although the number of GCP's used to correct the subscene was not reported by CCRS, the "guaranteed" positional accuracy as reported by Butlin et al. (1978) and Shlien (1979) is 50 metres.

3.3.2 **Effect of Terrain Geometry on Spectral Reflectance**

A point which is often overlooked when classifying digital MSS data is the effect of the terrain geometry on the radiance values which are recorded by the sensor. In a study describing the potential for correcting MSS data based on slope and aspect, Teillet et al. (1982) evaluated a number of previously published calibration methods. These methods included the Lambertian reflectance model, a view foreshortening model, the Minnaert reflectance model, a cosine model, and a diffuse and specular model. Each of the methods was used to calibrate LANDSAT MSS data in high relief zones. The calibrated LANDSAT data were then compared to the uncalibrated data with respect to effects on the resulting classification accuracy. Teillet et al. (1982) concluded that for overall classification results there were no improvements with the use of calibrated data. In some cases a substantial decrease in the classification accuracy was experienced. Some individual classes did yield quite different results using the various calibration models and Teillet et al. (1982) concluded that "...the behaviour of the LANDSAT data as a function of incidence angle is class dependent...there is no gain in correct forest classification with slope corrected MSS data even when atmospheric parameters are included..."(Teillet et al., 1982, p.99). In a separate, earlier study by Smith et al. (1980), the Minnaert-derived radiance model was used to address the effect of slope and aspect on the LANDSAT MSS data. They stated that in areas of relatively low relief, with corresponding low slope angles, variations in the scene radiance are the result of variations in surface cover.

Although Teillet et al. (1982) suggested that there is little to be gained from correcting the data for slope effects on the classification as a whole, it was necessary to investigate the influence of the terrain to evaluate the effects on individual classes for the present study. To evaluate the effects of terrain, three considerations were made: the effect of aspect, the effect of slope gradient, and the effect of shadows.

To investigate the considerations outlined above, sites identified as pine from Phase 3 forest inventory maps, verified from aerial photographs were selected for use. The pine class was used because this class represented the dominant cover class for the study area. In addition, earlier studies by Smith et al. (1980) and Teillet et al. (1984) used this cover in their analyses. From distribution of the histogram of the slopes * presented in Figure 4.8, it was decided that the slope gradients for 6 % up to 15% would be used. The 15% slope gradient represented the 85 percentile of the study area. A range of slope gradients was used so that, when combined with the direction of solar azimuth the size of the sample extracted for analysis would be sufficiently large. The aspects were linearized for the solar azimuth at the time of image acquisition, so that those aspects at 144° were assigned 0° and those at 324° were assigned 180°. All other aspects were assigned the appropriate values between 0° and 180°. To test the effect of aspect on radiance values recorded by the sensor samples with linearized aspects between 0° and 22° and 158° and 180° were extracted. This range was used so as to yield a sufficiently large sample for analysis. The descriptive statistics for the samples are presented in Table 3.1. The difference in spectral values introduced by the slope aspect are less than the variation in spectral values contained within the pine class for a particular aspect. Based on the observed differences in reflectance values in Table 3.1, aspect is not considered a dominating influence on the MSS data for the test area. As can be seen in Figure 3.1, the vast majority of

---

* The slope gradients and aspects used in this section were extracted from the data base developed for this study. The development of this data base is dealt with in detail in Chapter 4.

Table 3.1

Summary of Grey Level Values for Pine Stands on Slopes Ranging from 6% to 14% for Sun Facing and Sun Opposing Aspects

Aspect = 144° (Sun Facing)

Band	Minimum	Maximum	Mean	Standard Deviation
4	26	30	27.8	1.1
5	26	28	27.3	0.7
6	46	55	51.0	2.9
7	46	59	53.2	4.8

Aspect = 324° (Sun Opposing)

Band	Minimum	Maximum	Mean	Standard Deviation
4	25	29	27.4	1.2
5	24	29	26.4	1.3
6	41	57	47.4	4.7
7	43	58	49.0	4.9

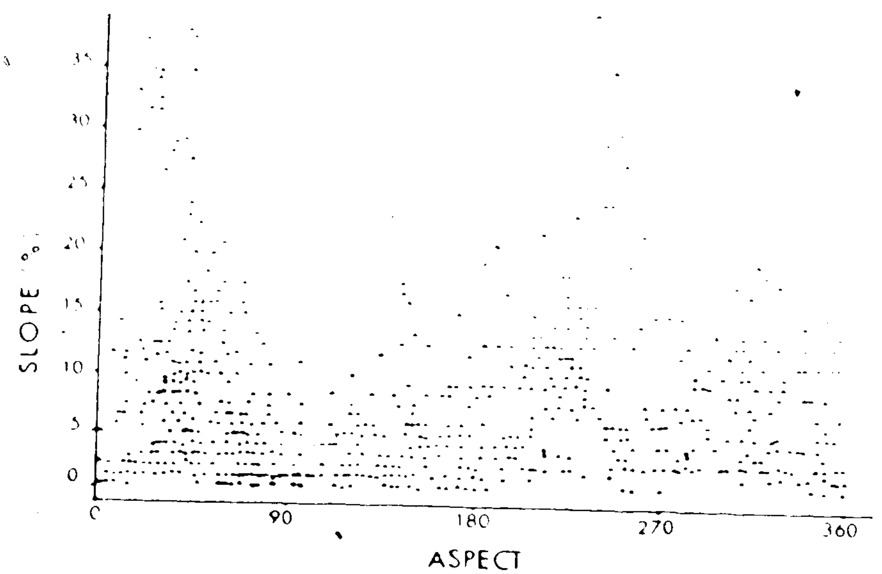


Figure 3.4. Relationship of slope and aspect for the study area.

the area is perpendicular to the 144° - 324° axis. The overall effect of aspect, from the distribution illustrated in Figure 3.1, is that the variation in spectral value is less than the extreme effect for the major part of the study area. This would further reduce the misclassification resulting from the effects of aspect.

The second area of consideration was the effect of slope gradient on the values recorded by the LANDSAT MSS sensor. Smith et al. (1980) calculated the critical slope angles and solar elevations necessary to detect reflectance differences caused by surface cover and not topographic variations. The graph presented by Smith et al. (1980) was used to determine the critical slope angle necessary to detect reflectance differences of 10% for the current study area. Using the solar elevation value at the time of imaging, 45°, a critical slope angle of 9° was found. In other words for that part of the study area with slopes less than 9°, GLV variations of 10%, or more, may be attributed to differences in the surface cover and not slope gradient. This slope gradient translates to approximately 15%. From Table 3.1 it is evident that the within class GLV variation for pine is greater than 10% for both the sun - facing and sun - opposing aspects.

The third area of concern, with respect to the effects of terrain, is that of shadow. The problem of shadow is of particular concern in areas of very high relief. To evaluate the potential shadow effects, the relief variable, which is discussed in section 6.5.1.1, was used as a measure of topographic variation. The maximum relief value for the study area, was indicated in Figure 6.4, was 20 metres. The length of a potential shadow was calculated using the trigonometric relationship:

$$b = a * \tan \alpha$$

where b is the length of the shadow, a is the relief, α is the solar elevation. As mentioned earlier the solar elevation was 45 o and the maximum relief 20 metres. The relationship assumes that the obstacle causing the shadow is at a right angle to the ground surface. The maximum length of a potential shadow is 20 metres which was less than one half of a pixel.

The conclusions for the study area based on these analyses suggest that the perturbations caused by the terrain geometry are not of such a magnitude to contribute to major classification error. The use of a single classification algorithm for the entire test area, in this instance, is therefore feasible. It follows that the majority of classification errors which may be encountered by using this algorithm are caused by factors outside the influence of the terrain geometry - radiance relationship.

3.4 **Classification of Digital Multispectral Scanner Data**

The conventional classification of digital MSS data into classes representing cover types involves a number of steps.

1) to develop a classification key or legend.

2) to choose sites which contain the information by which statistics can be generated and the classifier trained.

3) to examine the spectral characteristics of an individual pixel and assign it to a cover class.

This assignment is based on the properties of the training sites used in the second step.

4) to evaluate the accuracy of the classification.

Each of these steps will be discussed in turn as they apply to the test area.

3.4.1 Development of a Classification Key and Choice of Training Sites

There are two approaches to defining the training statistics which are to be used in the classification process. These are supervised and unsupervised approaches. A supervised approach to class definition assumes **a priori** knowledge regarding the surface cover. Sites are selected and labelled to reflect either the entire range of existing cover classes, or specific inventory classes. In most cases, the training sites are small and composed of a uniform distribution of cover. The resulting clusters of spectral data are assigned the appropriate class label based on an interpretation of the training sites. Unsupervised class definition, on the other hand, makes no assumption regarding the existing cover types. Relatively large training sites which contain a number of cover types are usually defined . Spectral clusters occurring within the training sites are identified and labels attached which correspond to the cover types. The labeling process is carried out after the classification of the image in order that the various classes can be identified and labelled.

For this study a hybrid approach to training site selection was implemented using components of both the supervised and unsupervised approaches. This approach was taken because a large amount of published information existed regarding the distribution of the cover classes present for the test area. An unsupervised approach was also beneficial because the forest cover stands in the study area were generally small and heterogeneous. A classification key for the area was developed using existing forest cover maps (Phase 3 Forest Inventory, Alberta Forestry, Lands and Wildlife). Two constraints were recognized in developing the key. The first constraint was the nature of the cover classification. In other words, a differentiation of conifer species was desirable rather than merely separating conifer from deciduous stands. A second constraint was the nature of the MSS data and the recognition of the limitations of the multispectral classification procedure. Given the relatively low spatial and the coarse spectral resolution of the LANDSAT MSS data, it was not possible to assign class labels developed in previous surveys to the MSS based classes. An example of this was the use of crown closure in the class definition. The classification key had to reflect a balance between the desired classification level and what could realistically have been expected using the LANDSAT MSS data.

Based on these considerations, an initial classification key was developed (Table 3.2). This key reflected the nature of the cover distribution of the study area and the desired mapping categories. Also recognized was the basic limitations of the spectral data such as the degree of crown closure.

Several stands representing each of these classes were identified throughout the test area using aerial photographs and Phase 3 forest inventory maps. A point location within each cover class stand was selected and plotted on a 1:50000 scale map sheet. These points were then digitized and the UTM location for each translated into a pixel address. Based on these addresses, data values at the digitized location, as well as a 3 X 3 pixel neighbourhood surrounding the digitized point, were extracted. This size neighbourhood was used because the size of some of the cover stands were sufficiently small so that a 5 X 5 pixel (250 metre X 250 metre) neighbourhood would sample pixels belonging to other cover stands. A total of 120 stands were selected to define the training areas. With the surrounding 3 X 3 neighbourhoods, this yielded 1080 points which were used to develop the training statistic for the classification.

3.4.2 Classification Procedure

The assigning of a pixel to a particular cover class based on the spectral characteristics may take a number of different approaches. The classification methods, although operationally different, are all similar in that they divide the spectral feature space into regions. Each of the regions defines a class. The methods used to define these regions are the base of the various classifiers. The simplest approach to classification is the box or parallelepiped classifier. In this approach the

Table 3.2

Classification Key

Upland conifer
Mixed conifer (spruce and pine)
Mixed conifer and deciduous (spruce and aspen)
Upland deciduous
Muskeg/fen (conifer)
Muskeg/fen (deciduous)
Open Soil

classes are defined through the use of straight lines which outline the bounds of the spectral clusters which define the class. The simplest approach to isolate and define the clusters only allows rectangular boxes to be constructed. A more sophisticated parallelepiped classifier allows for stepped rectangular boundaries to be defined. An advantage in using the parallelepiped classifier is that it is computationally simple. A second benefit is that problem areas which contribute to misclassification and confusion can easily be dealt with through examination of the feature space and appropriate adjustment of the box boundary.

A second approach to subdividing the feature space calculates a probability that a pixel will belong to a specific class based on the characteristics of the spectral values for a particular training area. An example of this approach is the discriminant classifier. A disadvantage is that an assumption is made that every cover type has been defined and that every pixel to be classified belongs to one of the training classes, which may not always be the case.

For the purposes of this study a discriminant classifier was chosen. Classification by discriminant analyses uses a classification function to decide whether a case belongs to a group (Klecka, 1980). This function is a linear combination of discriminating variables (eg. spectral bands). The decision as to the appropriate class is made by maximizing between group differences while minimizing within group variation. The function has the form:

$$h_k = b_{k0} + b_{ki} X_4 + b_{k2} X_5 + b_{k3} X_6 + b_{k4} X_7$$

where h_k is the score for class k, b_{k0} is a constant, b_{ki} is a coefficient for discriminant variable i and class i, and X_i represents the spectral values for bands 4 through 7. The value of h is calculated for each pixel times the number of classes (k). A pixel is grouped into the class with the highest score (h). The reason for using this approach was principally due to the large number of data sets which would potentially be used when incorporating the ancillary data into the classification. Any more than two data sets makes the use of a parallelepiped classifier unwieldy.

Although the classes to be mapped, as defined in Table 3.2, represented rigid boundaries categorizing the cover types, there was a continuum between cover types. Training sites which were defined as pine stands, for example, may have in fact ranged from closed stands to open pine with a considerable amount of understorey reflection contributing to the spectral values recorded by the satellite MSS sensor. It was recognized, therefore, that the isolation of spectral clusters occurring within a cover class which corresponded to other cover classes was necessary. This reduced the degree of misclassification by not including in the training data those sites which were labelled as one class but had the spectral characteristics of another.

The use of a clustering approach entails a very specific problem; that is, determining the choice of the appropriate number of clusters. This problem can be reduced in magnitude through the use of the Cubic Clustering Criterion (CCC) discussed by Sarle (1983). By performing a series of cluster analyses and progressively increasing the number of clusters, a set of CCC values are generated. When the CCC values are plotted against the corresponding number of clusters, an interpretation of the potential number of clusters present in the data set can be made. Sarle (1983) states that a major break in slope following a steep upward rise in the line tracing the plotted points suggests a possible candidate number of clusters. Clusters with differing forms result in differing shaped CCC curves. Spherical clusters yield a distinctive peak following a sharp rise in the CCC value. Elliptical clusters result in a rapid rise in the appropriate number of clusters followed by a break in the slope of the curve followed by a more gentle increase in the CCC value rising to a peak, after which there is a decrease in the CCC value. Care must be exercised in interpreting these plots, because multiple peaks may represent subdivision of the clusters into smaller, and sometimes meaningless, subclusters.

With this in mind, the training sites selected from the test area were subjected to a series of cluster analyses using a k-means nearest centroid clustering procedure (SAS, 1985). The clustering process uses seed points which act as the centroids of the initial clusters. Each observation is assigned to the nearest seed point based on the Euclidean distance. Once all of the observations have been assigned to a seed then the mean is calculated for each of the clusters. This cluster mean replaces the seed values and the process is repeated with the observation changing clusters if the Euclidean distance to a different cluster mean is less than to the original one. Once there is no change in the cluster means, the process stops.

The CCC values were plotted against the appropriate number of clusters. The resulting plot (Figure 3.2) was interpreted to mean that potentially 3, 6, and 12 clusters were present in the data set. All of these represented elliptical clusters according to Sarle (1983).

The first of these runs, containing three clusters, was interpreted as corresponding to the very basic cover characteristics, based on the summary of mean spectral values for each cluster (Table 3.3); that is coniferous forest, deciduous forest and a mixture between the two. The clusters were cross tabulated with the original training site labels (Table 3.4).

The training data were then grouped according to the 6 and 12 cluster runs. The data points were cross tabulated based on original training site labels and cluster numbers. The mean spectral values and crosstabulation results are presented in Tables 3.5 to 3.8. The crosstabulation of the 6 clusters with the cover labels (Table 3.6) shows that four distinctive classes can be interpreted. Clusters 3 and 4 are labelled as closed conifer. Clusters 2 and 5 are designated as deciduous, and cluster 6 as a mixture of deciduous and coniferous vegetation. Cluster 1 represents open soil. There is not a unique cluster which describes the wetland categories.

It can be seen from Table 3.8 that for the 12 cluster run, 4

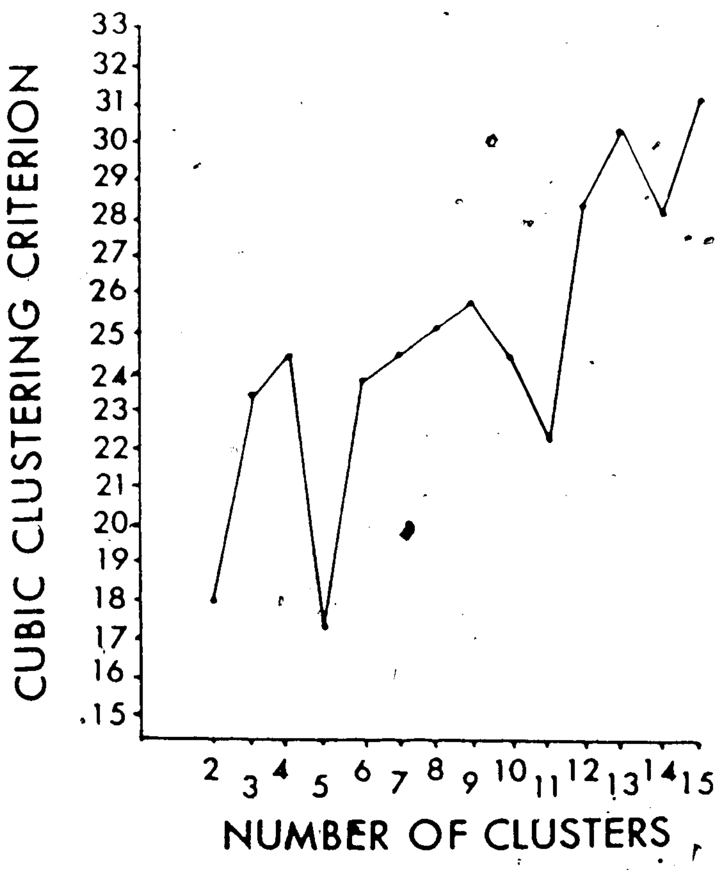


Figure 3.2. Plot of cubic clustering criterion versus number of clusters

Table 3.3

Statistical Summary of Grey Level Values for Three Clusters

Cluster 1

Band	Mean	Standard Deviation
4	30.2	3.0
5	30.4	5.1
6	61.8	4.1
7	68.0	5.4

Cluster 2

Band	Mean	Standard Deviation
4	31.2	2.4
5	30.9	4.0
6	76.4	6.5
7	88.4	9.8

Cluster 3

Band	Mean	Standard Deviation
4	28.5	2.0
5	28.0	3.4
6	50.0	3.9
7	53.0	4.9

Table 3.4

Crosstabulation Results for Three Clusters and Cover Classes

COVER CLASS	CLUSTER # 1	2	3
Upland conifer	94	21	263
Mixed conifer	20	0	124
Mixed conifer/ deciduous	81	16	83
Deciduous	108	10	8
Muskeg (treed)	62	25	57
Muskeg (deciduous)	34	35	30
Open soil	7	0	2

Table 3.5

Statistical Summary of Grey Level Values for Six Clusters

Cluster 1

Band	Mean	Standard Deviation
4	37.6	3.8
5	45.4	5.9
6	57.3	5.8
7	58.2	6.6

Cluster 2

Band	Mean	Standard Deviation
4	31.0	2.5
5	30.8	4.3
6	73.7	3.7
7	84.3	5.1

Cluster 3

Band	Mean	Standard Deviation
4	28.0	1.2
5	27.3	1.4
6	47.8	2.8
7	49.1	3.4

Cluster 4

Band	Mean	Standard Deviation
4	28.9	1.5
5	28.1	1.8
6	55.5	2.7
7	59.6	3.4

Cluster 5

Band	Mean	Standard Deviation
4	31.6	1.7
5	30.8	2.3
6	88.3	7.1
7	107.1	9.4

Cluster 6

Band	Mean	Standard Deviation
4	30.2	2.3
5	30.1	3.4
6	63.7	3.0
7	70.8	3.9

Table 3.6

Crosstabulation Results for Six Clusters and Cover Types

COVER CLASS	CLUSTER # 1	2	3	4	5	6
Upland conifer	[illegible]	25	168	132	0	46
Mixed conifer	0	0	87	45	0	12
Mixed conifer/ deciduous	0	13	55	65	0	47
Deciduous	0	46	19	9	14	38
Muskeg (treed)	21	22	25	41	1	34
Muskeg (deciduous)	3	15	21	28	0	32
Open soil	6	0	0	0	0	3

Table 3.7

Statistical Summary of Grey Level Values for Twelve Clusters

Cluster 1

Band	Mean	Standard Deviation
4	40.0	5.2
5	50.3	5.6
6	63.4	1.8
7	63.8	2.9

Cluster 2

Band	Mean	Standard Deviation
4	29.8	1.8
5	29.1	2.2
6	64.7	2.3
7	72.7	2.4

Cluster 3

Band	Mean	Standard Deviation
4	29.5	2.0
5	29.2	2.2
6	59.7	2.3
7	65.1	2.6

Cluster 4

Band	Mean	Standard Deviation
4	27.7	1.2
5	26.9	1.1
6	45.8	2.2
7	46.4	2.5

Cluster 5

Band	Mean	Standard Deviation
4	31.8	1.2
5	30.7	1.5
6	97.2	3.4
7	119.2	5.5

Cluster 6

Band	Mean	Standard Deviation
4	31.1	2.2
5	30.8	3.6
6	76.9	2.3
7	88.9	3.6

Cluster 7

Band	Mean	Standard Deviation
4	30.1	1.8
5	29.2	2.3
6	70.5	2.6
7	80.6	2.5

Cluster 8

Band	Mean	Standard Deviation
4	31.5	1.9
5	30.8	2.7
6	84.2	3.7
7	101.5	3.8

Cluster 9

Band	Mean	Standard Deviation
4	28.4	1.2
5	27.6	1.6
6	50.0	2.0
7	51.9	2.1

Cluster 10

Band	Mean	Standard Deviation
4	28.8	1.3
5	27.9	1.8
6	54.9	2.2
7	58.6	2.3

Cluster 11

Band	Mean	Standard Deviation
4	35.0	1.9
5	39.2	3.2
6	68.2	4.2
7	74.2	4.3

Cluster 12

Band	Mean	Standard Deviation
4	37.2	2.7
5	44.1	4.5
6	52.7	3.4
7	53.2	4.0

Table 3.8

Crosstabulation of Twelve Clusters and Cover Types

COVER TYPE	CLUSTER # 1	2	3	4	5	6	7	8	9	10	11	12
Upland conifer	6	17	42	62	0	4	7	0	128	99	12	0
Mixed conifer	0	0	10	46	0	0	0	0	56	30	2	0
Deciduous	0	38	10	14	10	43	47	8	5	5	0	0
Mixed conifer/ deciduous	0	24	32	20	0	8	9	4	10	15	4	0
Muskeg (treed)	3	16	24	14	0	11	14	1	15	23	7	16
Muskeg (deciduous)	0	15	29	11	5	6	8	8	0	10	5	2
Open soil	3	1	0	0	0	0	0	0	0	0	3	2

of the classes defined in the classification key presented earlier in the chapter are represented. These four classes were the same as the 6 clusters described above. For example, the upland conifer (pine) category was represented by cluster numbers 4, 9 and 10. The distribution of original cover types within these three clusters was consistent between the three. Cluster 4 showed a somewhat higher number of deciduous labelled sites. The mixed deciduous/coniferous and open spruce classes were both represented by cluster numbers 2 and 3, with cluster 2 having a greater percentage deciduous points and cluster 3, herbaceous points. The deciduous class was defined by cluster numbers 5, 6 and 7. The wetland categories, again, were not defined as unique clusters, but had sites scattered in all of the clusters. The conclusion which can be drawn from this is that the 12 groups did not represent clusters of spectral data which could be related to unique cover classes. Several of the larger clusters encountered in the 6 cluster run were separated into smaller groups, but with no consistent pattern to the groupings.

An analysis of the separation of the clusters for the three different clustering runs was performed. Tables 3.9a-c summarize the results of the three analyses. The three clusters (Table 3.9a) attained a separation of 98.4%. This separation increased table 3.9a slightly to 98.7% using the six clusters (Table 3.9b), but decreased to 97% for the twelve clusters (Table 3.9c). The results indicated that there was little confusion between the clusters in all three of the analyses. Using these results, as well as the crosstabulations discussed above it was concluded that six clusters would be the most appropriate in defining the

Table 3.9a

Separation of Three Clusters

	Classified into Cluster			
From Cluster	1	2	3	total
1	363 (97.84)	8	0	371
2	0	126 (100)	0	126
3	9	0	574 (98.46)	583
total	372	134	574	1080

Overall separation = 98.4%

Table 3.9b

Separation of Six Clusters [a]

Classified into Cluster:

From Cluster	1	2	3	4	5	6	total
1	36 (97.3)	0	0	1	0	0	37
2	2	117 (96.69)	0	0	0	2	121
3	0	0	372 (99.73)	1	0	0	373
4	0	0	2	315 (99.06)	0	1	318
5	0	0	0	0	19 (100)	0	19
6	4	0	0	1	0	207 (97.64)	212
total	42	117	374	318	19	210	1080

Overall separation = 98.7%

Table 3.9c

Separation of Twelve Clusters.

From Cluster	Classified into Cluster 1	2	3	4	5	6	7	8	9	10	11	12	total
1	11 (91.67)	0	0	0	0	0	0	0	0	0	1	0	12
2	0	102 (96.23)	0	0	0	0	4	0	0	0	0	0	105
3	0	5	159 (95.21)	0	0	0	0	0	0	3	0	0	167
4	0	0	0	175 (98.87)	0	0	0	0	2	0	0	0	177
5	0	0	0	0	8 (100)	0	0	0	0	0	0	0	8
6	0	0	0	0	0	51 (98.08)	0	1	0	0	0	0	52
7	0	0	0	0	0	2	56 (95.65)	0	1	0	0	0	59
8	0	0	0	0	0	0	0	13 (100)	0	0	0	0	13
9	0	0	0	6	0	0	0	0	214 (96.83)	1	0	0	221
10	0	1	0	0	0	0	0	0	1	204 (99.03)	0	0	206
11	0	0	0	0	0	0	0	0	0	0	31 (100)	0	31
12	0	0	0	0	0	0	0	0	0	0	0	20 (100)	20
total	11	107	150	181	6	53	70	14	217	208	33	20	1080

Overall separation = 97%

distribution of the multispectral data for the study area. The vegetation cover of the study area was classified into six clusters representing four cover classes, using linear discriminant analyses. The resulting image is displayed in Figure 3.3 (see back pocket).

3.5 **Error in Classification of Digital Remotely Sensed Data**

The assessment of classification accuracies has received a great deal of attention in the literature. Many issues have been raised pertaining to the accuracy assessment of classified data, including the nature of classification error, sample design for error assessment (including the size of the sample necessary to yield a credible summary of the classification accuracy), and the assessment of the interpretation as a whole versus individual classes.

Todd et al. (1980) discussed three sources of error inherent in LANDSAT MSS - based classifications. The first error source was that introduced through the LANDSAT system itself, including geometric errors as described earlier in this chapter. The effect of any residual geometric errors, after geometric correction, could be reduced by choosing training sites, and sampling for classification accuracy, in the centre of the vegetation stand.

A second source of error, according to Todd et al. (1980), was the overestimation of the potential of the sensor and data to detect and discriminate cover classes. They noted that the definition of classes with a separability approaching the noise level of the data accounted for almost one half of the recorded error in the classification. Some of this noise may be attributable to atmospheric disturbances. Other sources of error may be attributable to bidirectional reflectance of the vegetation or to effects from neighbouring high reflectors. A third source of noise may be attributed to topographic variations. By attempting to generate classes which are spectrally very close, the chance of misclassification becomes increasingly significant.

A somewhat related error source is the choice of appropriate class labels. This has been neglected to a large degree in the literature, but is of prime importance in the generation of classification error. Rosenfield and Fitzpatrick-Lins (1983) addressed the question of appropriate class definition by stating that within a multispectral classification, error was either caused by sensor related effects or class labeling. Chrisman (1982), in a discussion of classification accuracy, stated that the usefulness of a multispectral classification depends on the validity of the class labels chosen. This is a crucial point which has eluded many who were developing classification keys. One of the most common errors is to confuse land use with cover type. In many of the forest resource applications, the labels of clear cut, for example, have been included with cover type labels such as deciduous and coniferous vegetation. Clearly, the cover represented by the land uses (or conditions) may not be unique and the areas on which clearcuts have taken place may be in varying stages of regeneration. A second important area of confusion is the inclusion of landforms within classifications based on spectral data. Todd et al. (1980) in their ten class key, included the label "cliff". Similarly the landform(s) which comprise "wetlands", including fens, muskegs, and bogs, have been included in numerous multispectral classifications (see for example Kalensky et al., 1981). The inappropriate selection of class labels for MSS based classifications will, therefore, lead to a significant reduction in the mapping accuracy.

The third area of error as identified by Todd et al. (1980) is due to inappropriate decisions made by the analyst. This may include some of the problems discussed above with regard to mislabeling of classes. It may also include the incorrect interpretation of ground cover from aerial photographs, or field checks. An insufficient number of training sites may also lead to class boundaries not being properly defined. These errors, according to Todd et al. (1980), accounted for over one half of the total classification errors.

Given these potential sources of error, any classification using multispectral data must also include an objective evaluation of the classification accuracy. In designing a procedure for assessing the amount of error within a map, some pertinent issues must be addressed. The first issue is sample design. The second is the assessment of the classification as a whole versus individual categories. Finally, the third issue is the problem encountered in the comparison of multiple classifications.

A number of different strategies for designing a sampling programme have been put forward to adequately represent a region.

Van Genderen and Lock (1978), Fitzpatrick Lins (1980) suggested the use of a stratified random method. The stratification was approached in two ways. The first approach was thematically based. In other words, a sufficiently large sample was extracted from each of the mapped classes. The second approach was to stratify the region based on area. This assured that the entire area of interest was represented by the sample design. The first approach is not always feasible because it implies **a priori** knowledge of the relative percentage of the total area that a class occupies. An alternative approach is to randomly choose the sample points without stratification. The use of this latter approach ensures that the existing cover classes are represented according to their distribution and that a particular cover class will not be over sampled with respect to the distribution. This will allow an accurate understanding of the classification accuracy.

Associated with the sample design is the number of points which must be sampled to satisfy statistical assumptions. Van Genderen and Lock (1978) and Rosenfield et al. (1981) suggested that a minimum of 20 sample points be required per class to assure an adequate representation of the cover class. The use of this number suggests that for a ten - class map, the total number of points sampled should be 200, which is an absolute minimum assuming that classes would not be over sampled. Krejicie and Morgan (1970) determined that the minimum required sample size should be determined by the total size of the population that was being sampled. They noted that the rate of increase in sample size is not linear so that the required number of samples

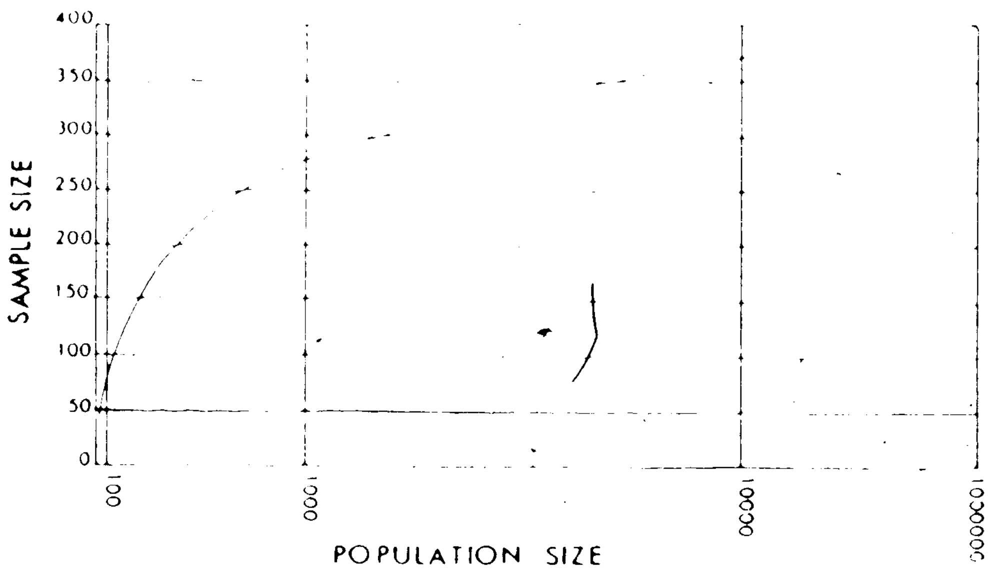


Figure 3.4 Plot showing the relationship between population and sample size (adapted from Krejcie and Morgan, 1970)

with a small population is high compared to the sample size required with larger populations. A plot showing the relationship between population and sample size is presented in Figure 3.4. From this plot it can be seen that the sample size required to present populations consisting of 9,000 members is 370 points. This increases to 380 points for a population of 400,000.

For the present study the assessment of classification errors and the evaluation of accuracy was essential. The overall accuracy of the classification, as well as for the individual classes had to be addressed and evaluated in some detail. To achieve this goal, a random sampling network was designed. This sampling strategy was used rather than a stratified one for the reason stated earlier; to arrive at an understanding of the classification accuracy as a whole, the individual, particularly smaller, classes should not be over sampled. A random sample ensured that the individual classes were sampled according to their distribution. The total number of points used for this design was 400. This surpassed the number of points suggested by Krejicie and Morgan (1980) as the minimum for a population of this size. To calculate the population size the number of rows was multiplied by the number of columns in the test area matrix. The coordinates of the sample points were generated through the use of a random number table. The boundaries for the coordinates were based on the original size of the elevation matrix. With the generation of the various landform coverages the size of this matrix became smaller with the boundary rows and columns being eliminated. Thus, the final number of sample points was reduced to 372 from the original 400. Given the minimum size restrictions stipulated by Van Genderen and Lock (1978) and Rosenfield et al. (1980) all but one of the classes specified in the original classification key were adequately sampled. The open soil class had a small number of sample points because it was restricted to a small geographic area.

To address the final error source discussed by Todd et al. (1980), that of misinterpretation of cover types used in the training site selection, as well as for the accuracy analysis, an evaluation of the Phase 3 map data was performed. The accuracy of these map products is normally evaluated by selecting a set of sites and performing independent interpretations. The original map class and the independent assessment is recorded on a log sheet. These log sheets were examined for the test area. A tabulation of the results of this evaluation is presented in Table 3.10. The results of the analyses did not represent the overall map accuracy of the Phase 3 product, since only the stand composition information was used. The resulting interpretation accuracy of the Phase 3 map information used in this study was 82%.

An error matrix for the MSS classification is presented in Table 3.11. A number of points must be noted regarding this matrix. The first is that the values represented by the verified samples are found in the rows and the MSS classification in the columns. Second, the omission and commission percentages have been calculated for only the diagonal.

As discussed earlier in this chapter, it was not possible to subdivide the conifer classes into smaller, meaningful clusters.

Table 3.10

Summary of Accuracy Evaluation of Phase 3 Forest Inventory for Study Area.

VERIFICATION

PHASE 3	Closed pine	Open spruce	Closed spruce	Deciduous	Mixed deciduous/ coniferous	Mixed conifer
Closed pine	36 (67)	2	6	0	4	7
Open spruce	0	6 (67)	0	2	0	0
Closed spruce	4	0	11 (55)	2	0	0
Deciduous	4	0	0	130 (93)	20	0
Mixed deciduos/ coniferous	4	0	0	6	88 (77)	4
Mixed coniferous	6	1	3	0	2	49 (82)

number
(% correct)

Total number of samples= 397
Total correct= 320
Map accuracy=81%

Table 3.11

Accuracy Analysis for MSS Based Classification.

Aerial Photographs/ Phase 3

	closed conifer	mixed conifer	mixed deciduous coniferous	deciduous	muskeg/ fen	open soil
closed conifer (cluster 3 and 4)	116 (48) (69)	22 (32) (81)	36	11	18	0
mixed coniferous/ deciduous (cluster 6)	24	1	37 (50) (42)	6	6	0
deciduous (cluster 2 and 5)	26	4	15	37 (59) (69)	7	0
open soil (cluster1)	1	0	0	0	0	5 (17) (100)

number
(% commission errors)
(% correct)

Total number of samples=372
Number accurate=217
Mapping accuracy=58%

The result, therefore, was that the pine stands and the mixed conifer classes were not mapped separately based on MSS data alone. The mapping accuracy for the upland conifer class, which was reported in this portion of the analysis, was based solely on the one class. An examination of the individual classes yielded some very interesting patterns in the misclassifications which did occur. The conifer class yielded a moderate to low classification accuracy with a 32% commission error and 29% omission error rate. A high degree of confusion occurred with the mixed coniferous/deciduous, deciduous and wetland classes. The pine stands confused with the mixed coniferous/deciduous class while the mixed coniferous stands did not. The reason for this confusion was that the crown closure for the pine stands, especially with the older stands, was less than the mixed conifers, where the spruce trees formed the understorey, thereby eliminating a substantial amount of the herbaceous and deciduous understorey which would have affected the reflectance detected by the MSS scanner. An 18% commission error with mixed coniferous/deciduous class occurred. This error rate was high but easily explained given the arbitrary boundary imposed by the classification scheme which separated the conifer class from the mixed coniferous/deciduous class. The arbitrary boundary positioning also explained some of the omission errors reported below. It was obvious from the accuracy rates reported for the mixed coniferous/deciduous class (42% correct and 50% commission error) that this was a significant problem. The confusion with the deciduous class was 5%, while the wetland class was almost 9%. Similar trends in the omission rates also occurred, with a

13% confusion rate with mixed coniferous/deciduous class, and 15% for the deciduous class. The deciduous class displayed a commission error of 31% and an accuracy of 59%. There was a great deal of confusion with all of the other classes, with the possible exception of the mixed coniferous class for the reasons discussed.

One significant problem with this classification, as indicated earlier, was the inadequacy of the MSS data to define areas of poorly drained, organic soil which comprises the muskegs and fens. As is evident from Table 3.11, the wetland class included samples from all of the clusters representing the vegetation. This point was also observed earlier in this chapter during the cluster analyses of the training site information.

3.6 **Summary**

From the foregoing analysis and discussion it can be concluded that for forest cover classifications the use of MSS data alone is not sufficient to yield satisfactory results both in terms of the breadth or accuracy of the classes. Additional data must be considered to improve the classification. The most successful approach has in the past been the inclusion of information analogous to the photo element site. The use of these data is investigated in a subsequent chapter. Prior to integration, however, an investigation of the nature of these potential data is made.

CHAPTER 4

Digital Elevation Modelling and the Extraction of Spatial Derivatives

4.1 Introduction

One potential source of ancillary information, which may be useful when integrated with LANDSAT MSS data, defines the geometry of the earth's surface. This geometric information is used to describe a variety of surface form characteristics which pertain to the landscape. One data source used to relate to a point in an absolute sense is elevation above sea level. Sources which relate to a point within a designated area are slope, both magnitude and direction, and the rate of change of the slope, or curvature. All of these sources will be discussed in this chapter. The derivation of the various products which were used in this study are described in detail.

4.2 Digital Terrain Models and Digital Elevation Models

A number of different terms have been used to describe the mathematical representation of the earth's surface. Two different terms will be used in this dissertation to describe the various types of models. The definitions used follow the terminology suggested by several authors (for example Monmonnier (1982) and Yoeli (1984)). A digital tèrrain model (DTM) is used to describe a set of irregularly - spaced points representing the earth's surface. Information regarding the geographic positioning of the data points, as well as the elevation, are recorded in the model, at each point (X,Y,Z triplet). From the DTM a regularly spaced grid of elevation values where the geographic location information is implicit may be produced. The process is known as surfacing and the resulting grid is termed a digital elevation model (DEM). As the points are regularly spaced, information regarding the geographic positioning is implicit with only the location of an origin being necessary. From this DEM a large number of products can be derived which characterize the geometric properties of the landscape at a point, or which model processes based on the geometric properties of the landform. Those products termed DEM derivatives in this dissertation, will be restricted to those created through the process which will be described later in this chapter. A second term, landscape coverages, will also be used which denotes those products not produced through differential geometry but by other means.

4.3 Digital Elevation Modelling

A variety of methods may be used to create a DEM. The first method is the direct generation of an elevation matrix through the use of a photomapper (see for example Franklin, 1985), or a photogrammetric analytical plotter. A second method is through the construction of elevation profiles from contours (for example Maxwell and Turpin, 1968). Third, an elevation matrix may be surfaced from a DTM (for example Maxwell and Turpin, 1968; Evans, 1972; Hardy, 1972; Sampson, 1978). Surfacing from a DTM can be accomplished through the use of two basic approaches. The first is a numerical approach which creates the matrix without creating an intermediate mathematical function. This approach interpolates the elevation for each cell in the elevation grid from surrounding known elevation points. An example of this is the inverse weighted distance method. The distances from a candidate point (grid cell) to its nearest neighbours are calculated. The inverse distance of this point to its neighbours determines the contribution of each of the points to the final elevation value assigned to the candidate point. This approach is computationally simple so that a large area may be surfaced within an acceptable period of time. The main disadvantage of this is that interpolations do result in a noisy representation of the actual surface.

Another approach to surfacing generates a continuous mathematical surface. This is known as an analytical method. There are a number of advantages in using this approach to surfacing. The main one is that the mathematical solution may derive a surface which passes through all of the points in the DTM exactly. There is also none of the noise caused by interpolation as in the case of the numerical approach.

In his discussion of topographic surfacing Hardy (1971) reviewed the more common analytical approaches used. These included the Fourier and the polynomial series. He rejected both of these approaches as being unsuitable for use in topographic surfacing, because of their inability to represent topographic variation adequately without the use of very large data sets. He stated that both of the series noted above "with relatively few data points...(are)...unmanageable in representing the sometimes rapid and sharp variations in real topographic surfaces"(Hardy, 1971, p1906). To avoid the problems encountered through the use of the two surfacing series mentioned above Hardy (1971) discussed the use of a multiquadric solution to the surface. The equation to solve for elevation was given by Hardy as:

$$Z = \sum_{j=1}^{n} C_j \left[(X - X_j)^2 * (Y - Y_j)^2 \right]^{0.5}$$

This approach yields a cone at each of the X-Y points. The shape of the cone, that is its peakedness, is controlled by the coefficient C. As well, Hardy stated that the shape of the individual cones is controlled by the characteristics of all of the other cones in the area being surfaced. He noted that "...strong influences from sharp cones farther away may override the weak or flat cones...however every cone will cause some significant change in the slope of the multiquadric surface regardless of its flatness" (Hardy, 1971, p1908). It was demonstrated by Eyton (1974) that the multiquadric solution fitted to a series of topographic points is suitable for surfacing large areas.

The main disadvantage of the multiquadric approach is the computational complexity and the computer memory requirements necessary to surface an area. Solving the multiquadric function using a matrix approach (Eyton, 1974) requires a matrix with a size equaling the number of X-Y-Z triplets in the DTM squared. If a DTM with 500 triplets is used the matrix created will have

250,000 entries. When using 4 byte data this matrix would require 1 megabyte of memory or 2 megabytes if 8 byte data are used. To lessen this problem and allow for large DTM's to be surfaced, the grid being created is sectioned into smaller neighbourhoods. One of two strategies may be adopted to achieve this. The first is to step the surfacing by processing adjacent neighbourhoods within the larger matrix. The second strategy also uses stepped smaller neighbourhoods, but each neighbourhood overlaps the adjacent ones. Only a central core of each of these overlapping neighbourhoods is sampled and retained. This is termed a roving approach. The roving methodology is computationally more demanding than the stepped approach because of the larger number of data points needed to solve each neighbourhood, but is more accurate at the boundaries. The shape of a surface which is calculated by an analytical approach is very sensitive to the distribution of DTM triplets within the area being surfaced. If there is no overlap between adjacent neighbourhoods then the boundaries may not match. This is especially the case in areas of low relief where the density of points describing the surface is low, resulting in a small number of points describing each neighbourhood. In some cases entire neighbourhoods may have no points contained within them. The roving method is, therefore, preferable to the stepped approach to surfacing large areas. Eyton (1974) used a stepwise approach in solving the problems created by a large DTM data set. He noted that some minor boundary problems resulted from the use of this approach. However, these were not significant. Subsequently, the stepwise approach was replaced by a roving solution (Eyton, 1987).

To generate a DTM for this study a 1:50000 NTS topographic map was digitized (Figure 4.1). The contours on the map sheet were traced using point mode digitizing. A data point was digitized at each break point along the contour line. The geographic position and elevation were recorded for each point. To save processing time at a later stage, the contour lines were thinned so that only those points which had a spacing of greater than 50 metres apart (1 millimetre on the source document) were retained. This generalization tended to reduce the information content in high relief areas while not affecting areas of low relief amplitude. Using this thinning approach the file size was reduced from 29,000 to 22,500 points.

To deal with the large DTM data set, the roving approach to surfacing discussed above was adopted. For each of the neighbourhoods a subset of the original DTM was sampled so as to be centered on, or as near as possible to, the neighbourhood. The subset size was chosen so as to cover a neighbourhood which was substantially larger than that which was being surfaced. In this way there was an overlap between the DEM subset for adjacent neighbourhoods. The choice of the appropriate neighbourhood and DTM subset size varied depending on the complexity of the terrain. In areas of great detail the neighbourhoods were small with a large triplet subset. In flat areas, the subset remained large but the neighbourhood size was increased. If the DTM subset did not have sufficient overlap with the adjacent areas, or the subset did not extend beyond the boundaries of the neighbourhood

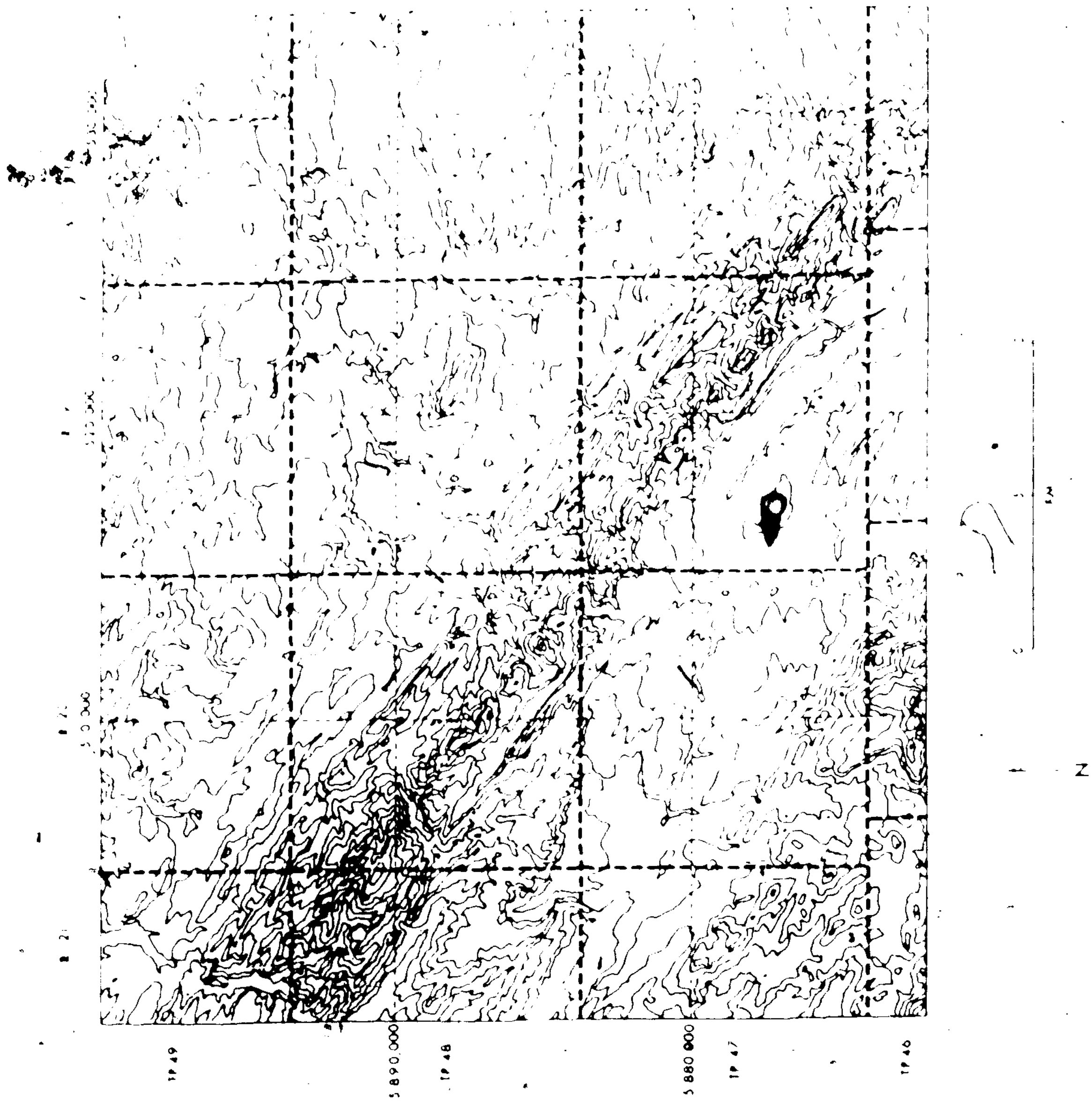


Figure 4.1. Topographic map of the study area. The contour interval is 30 metres (100 feet).

then visible seams were produced at the boundaries. In terrain with variable relief the choice of an appropriate neighbourhood size was difficult. A compromise was made because processing time increased substantially with an increase in either the size of the DTM subsample or a decrease in the neighbourhood size.

Another concern was that care had to be exercised when surfacing from digitized contours since the elevation triplets were concentrated along the contours. This required that a relatively large number of points was chosen for the triplet subset. If the subset size was too small then the points chosen would all fall on the same contour line.

Given the relatively limited processing capabilities of the VAX 11-730, a large neighbourhood size (11 X 11 cells) and a large triplet subset (200 points) were chosen for the surfacing of the DTM for this project. Several surfacing runs were made on the DTM to arrive at this combination. The larger DTM subsets reduced the number of seams between the neighbourhoods in the lower relief areas, as discussed above. The areas of higher relief, however, were best suited for smaller neighbourhoods due to the density of DTM points. Given the substantially larger proportion of the test area occupied by the low relief as well as the time requirements for the processing a combination of subset and neighbourhood size mentioned above was employed. The time required to generate a surface of 520 X 640 grid cells ranged from 60 CPU hours for an 11 X 11 neighbourhood and a triplet size of 75 points to 210 CPU hours for a 7 X 7 matrix and 100 points. The final run to create the DEM with the input parametres mentioned above, 11 X 11 neighbourhood and 200 points, required approximately 110 CPU hours to process.

The combination chosen was found to be suitable for all but a very small portion of the test area. The resulting perspective plot, shown in Figure 4.2, is a reconstruction of the DEM. A small number of seams are detectable in this representation which are not as visible in the grey tone image from the DEM (Figure 4.3).

The histogram of the DEM is shown in Figure 4.4. As is evident, the majority of the area is contained within a narrow elevation range, from approximately 1000 metres above sea level (MASL) to 1150 MASL. A secondary peak at 1450 MASL represents the northwest to southeast trending ridge. The maximum elevation is approximately 1600 MASL.

4.4 Geometric Derivatives

The derivative products from the DEM, which are described below, have been generated through the use of the methodology developed by Eyton (1987) and a FORTRAN programme developed by Eyton. The following a summary of that given by Eyton (1987). The methodology is based on the finite difference approximation of the first and second derivatives of a 3 X 3 neighbourhood of elevation values from the DEM. This neighbourhood and the associated notation which will be used in subsequent portions of this chapter are illustrated in Figure 4.5.

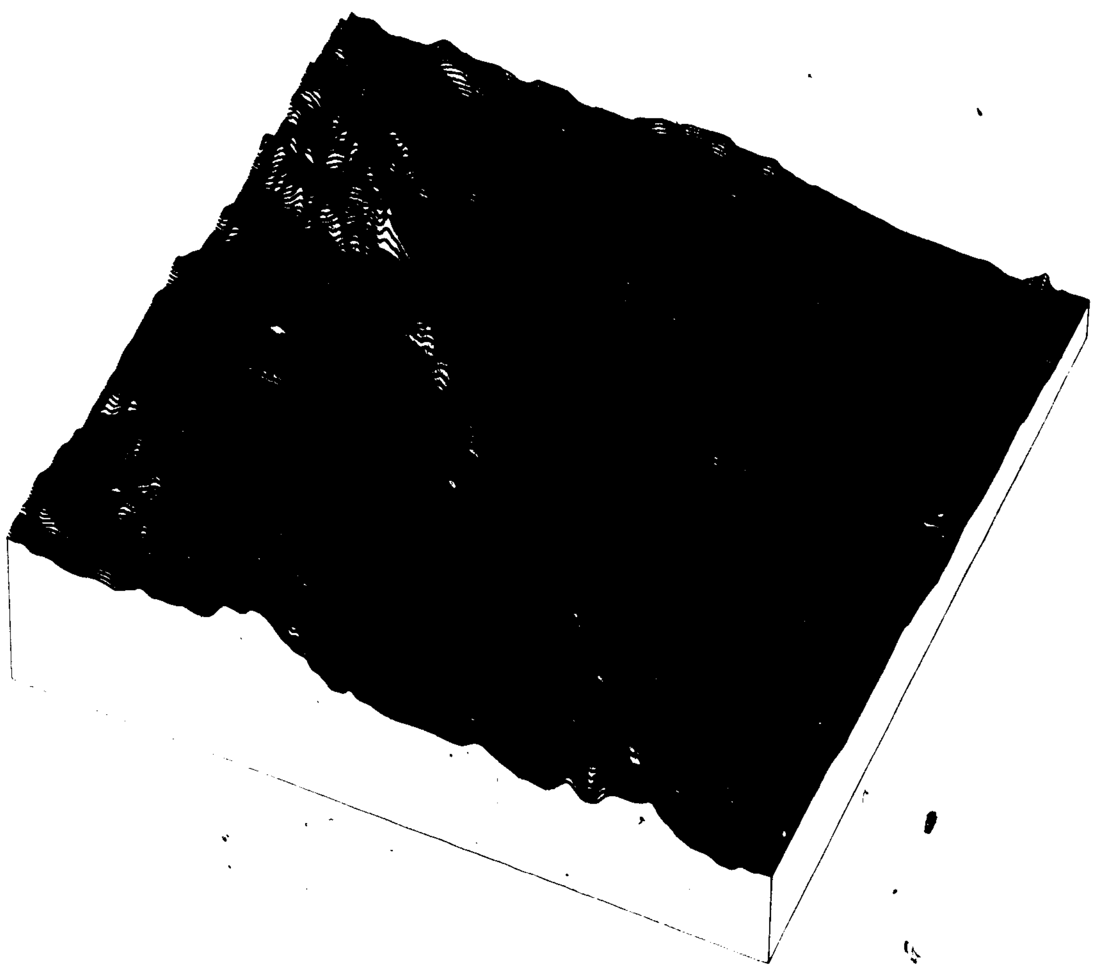


Figure 4.2. A perspective plot of the study area based on the digital elevation model derived from the digitized topographic map (vertical exaggeration 8x).

Figure [illegible].3. Grey tone image of the digital elevation model developed from the digitized topographic map

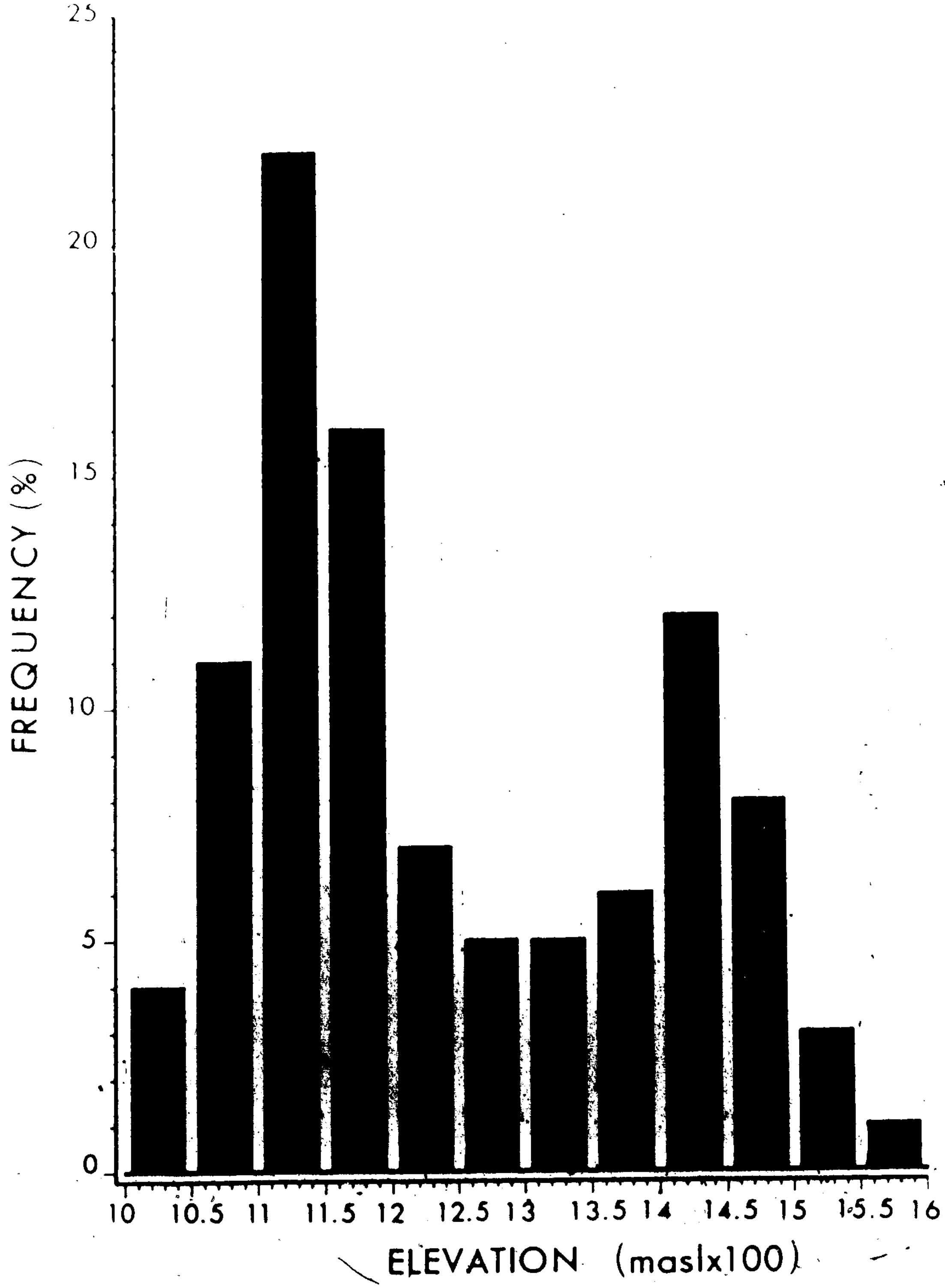


Figure 4.4. A histogram of the distribution of elevation values for the study area.

1,1	1,2	1,3
2,1	2,2	2,3
3,1	3,2	3,3

Figure 4.5. Neighbourhood notation used in this study.

4.4.1 First Derivative

The first derivative of elevation describes the rate of change of elevation, or the slope of the terrain. This slope refers to the maximum slope of a tangent plane at a point on the slope surface. The methodology used to calculate the slope is two stepped:

1) Calculation of the slope in the X and Y directions.

2) Calculation of the maximum slope.

Based on the neighbourhood notation outlined in Figure 4.5, this is calculated by:

1) $$\text{SLOPE X} = \frac{Z(2,2) - Z(2,1)}{D} + \frac{Z(2,3) - Z(2,2)}{D} = \frac{Z(2,3) - Z(2,1)}{2D}$$

$$\text{SLOPE Y} = \frac{Z(2,2) - Z(1,2)}{D} + \frac{Z(3,2) - Z(2,2)}{D} = \frac{Z(3,2) - Z(1,2)}{2D}$$

where D is the grid cell spacing. Positive slope values are those resulting from a decrease in the elevation towards the origin, while negative slope values indicate an increase in elevation away from the origin. This is illustrated in Figure 4.6.

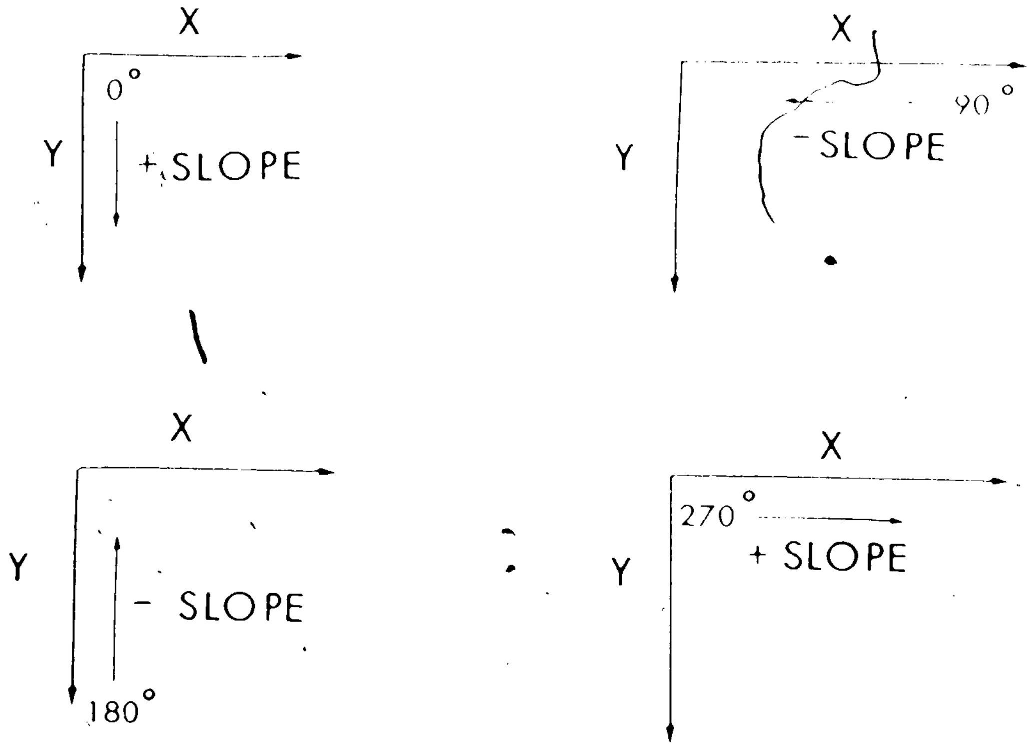
2) The maximum slope $= [(\text{SLOPE X})^2 + (\text{SLOPE Y})^2]^{0.5}$

The resulting slope values are expressed as a tangent.

To calculate the direction of maximum slope, or the aspect, the angle between the maximum slope and the slope in the X direction is determined:

$$O_s = \cos^{-1}(\text{SLOPE X}/[(\text{SLOPE X})^2 + (\text{SLOPE Y})^2]^{0.5})$$

A second step converts the local angle, O_s to an azimuth.



(from Eyton, 1987)

Figure 4.6. Signing convention for the calculation of aspect.

Conversion is based on the signs of the slopes in the X and Y directions. Table 4.1 summarizes this conversion.

The slope image is shown in Figure 4.7 for the test area. Figure 4.8 is a histogram showing the distribution of the slope values in Figure 4.7. A large proportion of the test area had very low slopes. Slightly more than 85% of the study area was located on slopes with gradients of less than 15%, and that 99% was on slopes less than 35%.

A compass rose of the distribution of the aspect values for the study area is shown in Figure 4.9. The highest percentage of area had a northeast and southwest orientation. This reflects the influence of the northwest to southeast trending topography of the Rocky Mountain Foothills.

4.4.2 Second Derivative

The second derivative measure obtained for this study was a directional curvature, giving measures for the shape of the landscape in both the down slope and across slope directions. These forms of curvature were used instead of other measures such as plan curvature because when both the across slope and down slope curvatures are considered together they represent a three dimensional perspective of the landscape. These curvature measures indicate the rate of change of the slope along the axis of maximum slope and an axis perpendicular to this line. These axes do not have to conform to the X and Y axis of the neighbourhood matrix used, but may occur on a diagonal. The

Table 4.1

Conversion of Local Angle to Slope Azimuth

Slope Sign: Slope X	Slope Sign: Slope Y	Slope Azimuth
+	+	$270^\circ + \theta_s$
+	−	$270^\circ - \theta_s$
−	−	$90^\circ + \theta_s$
−	+	$90^\circ - \theta_s$

(from Eyton, 1987)

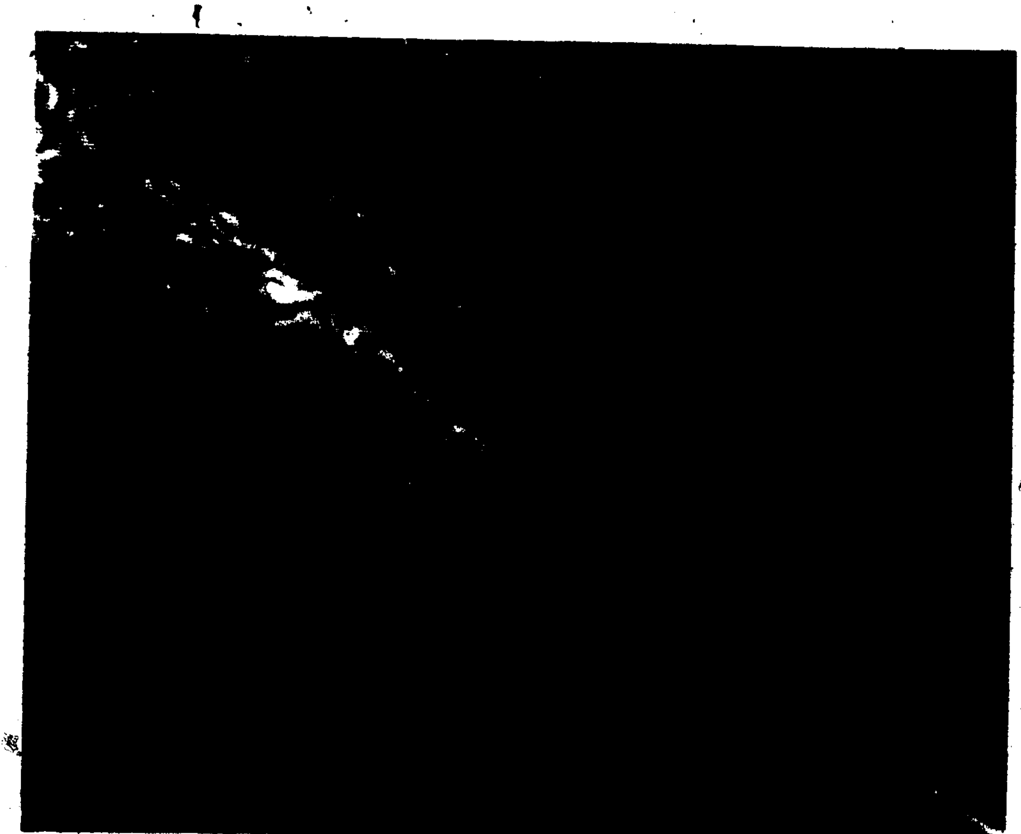


Figure 4.7. Grey tone representation of the slope gradient values of the study area.

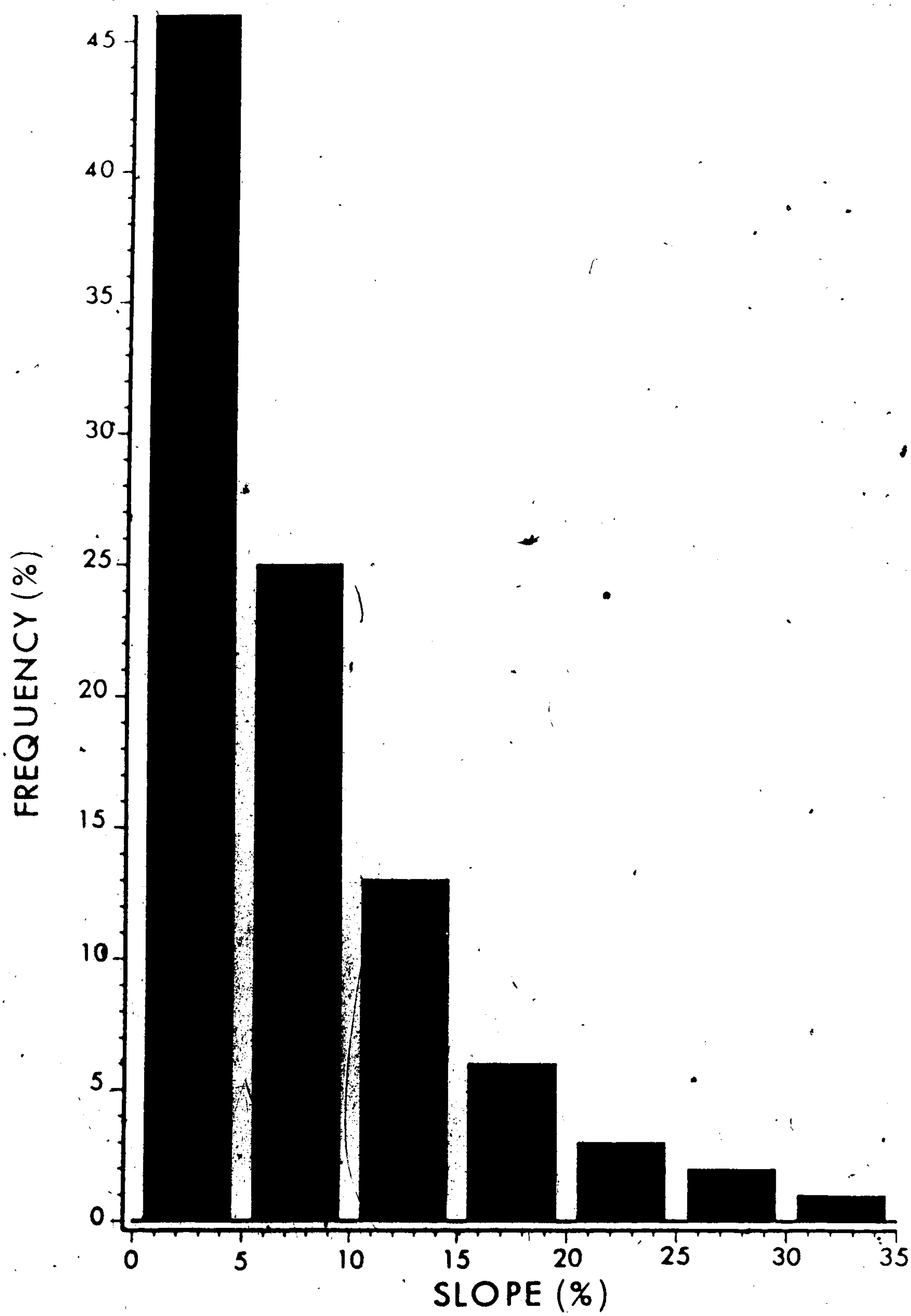


Figure 4.8. Histogram of the slope values for the study area.

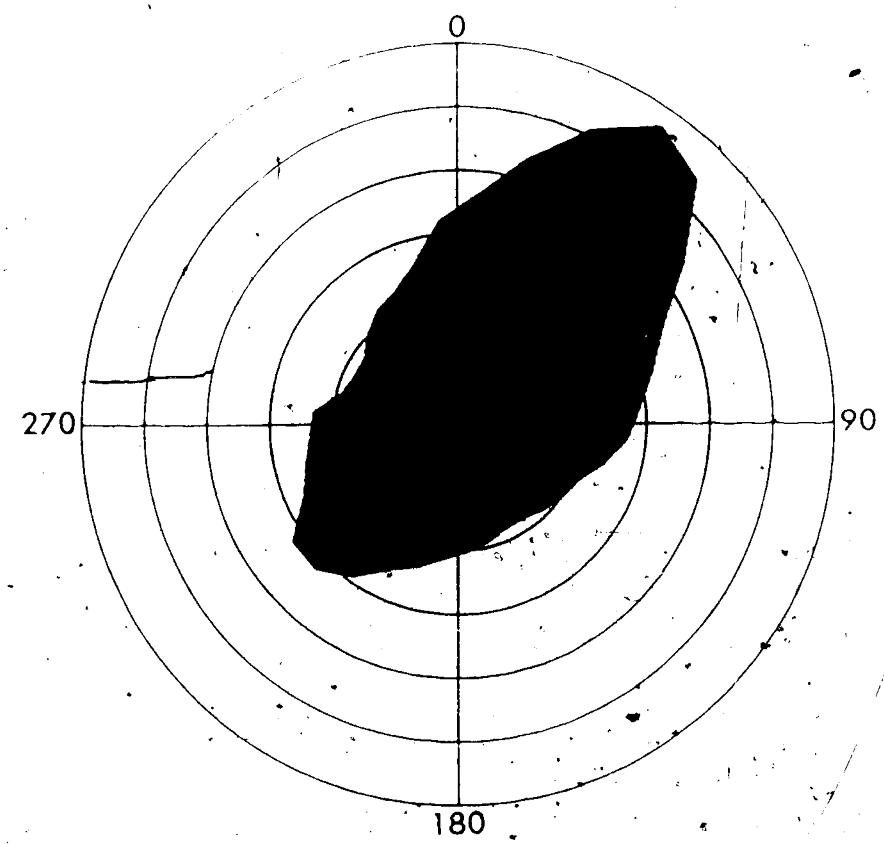


Figure 4.9. Rose diagram of the aspects in the study area.

Figure 4.12. Grey tone of across slope values for the study area.

procedure for calculating these curvature measures was described by Eyton (1987) and is paraphrased as follows:

1) The curvature along an axis from the points a and b may be calculated by the use of the following relation:

$$\text{Curvature} = \frac{2Z(2,2) - Za - Zb}{2D}$$

where Za is the elevation at point a and Zb is the elevation at point b; D is the grid spacing and Z(2,2) refers to the elevation at the center of the 3 by 3 cell neighbourhood.

2) Estimation of the values at Za and Zb requires the down slope azimuth O_s as defined earlier. As the azimuth distribution is circular it is necessary to linearize it by changing the values to a 0-180 degree distribution (O_1).

To determine the Cartesian coordinates for the points a and b:

O_1 from 0 to 90 degrees

$Xa=2.0+SIN\ O_1$

$Ya=2.0-COS\ O_1$

$Xb=2.0-SIN\ O_1$

$Yb=2.0+COS\ O_1$

O_1 from 90 to 180 degrees

$Xa=2.0+COS\ O_1$

$Ya=2.0+SIN\ O_1$

$Xb=2.0-COS\ O_1$

$Yb=2.0-SIN\ O_1$

Calculation of the elevation values for points a and b:

For θ_s from 0 to 90 degrees,

$$Za = (2.0-Ya)\,[Z(1,3)(Xa-2.0) + Z(1,2)(3.0-Xa)] + (Ya-1.0)\,[Z(2,3)(Xa-2.0) + Z(2,2)(3.0-Xa)]$$

$$Zb = (3.0-Yb)\,[Z(2,2)(Xb-1.0) + Z(2,1)(2.0-Xb)] + (Yb-2.0)\,[Z(3,2)(Xb-1.0) + Z(3,1)(2.0-Xb)]$$

For θ_s from 90 to 180 degrees,

$$Za = (3.0-Ya)\,[Z(2,3)(Xa-2.0) + Z(2,2)(3.0-Xa)] + (Ya-2.0)\,[Z(3,3)(Xa-2.0) + Z(3,2)(3.0-Xa)]$$

$$Zb = (2.0-Yb)\,[Z(1,2)(Xb-1.0) + Z(1,1)(2.0-Xa)] + (Yb-1.0)\,[Z(2,2)(Xb-1.0) + Z(2,1)(2.0-Xb)]$$

To calculate the across slope curvature the θ_1 value equals θ_s + 90, and the calculation of Za and Zb is repeated with the new values of θ_1.

Figure 4.10 is a grey tone image of the down slope curvatures for the test area. Figure 4.11 is a histogram of the distribution of values from this data set. Figure 4.12 and 4.13 represent the same for the across slope curvatures. What is evident from both of the histograms is that the curvatures calculated from the test area DEM were quite low. This reflects the generally low slopes contained within the study area.

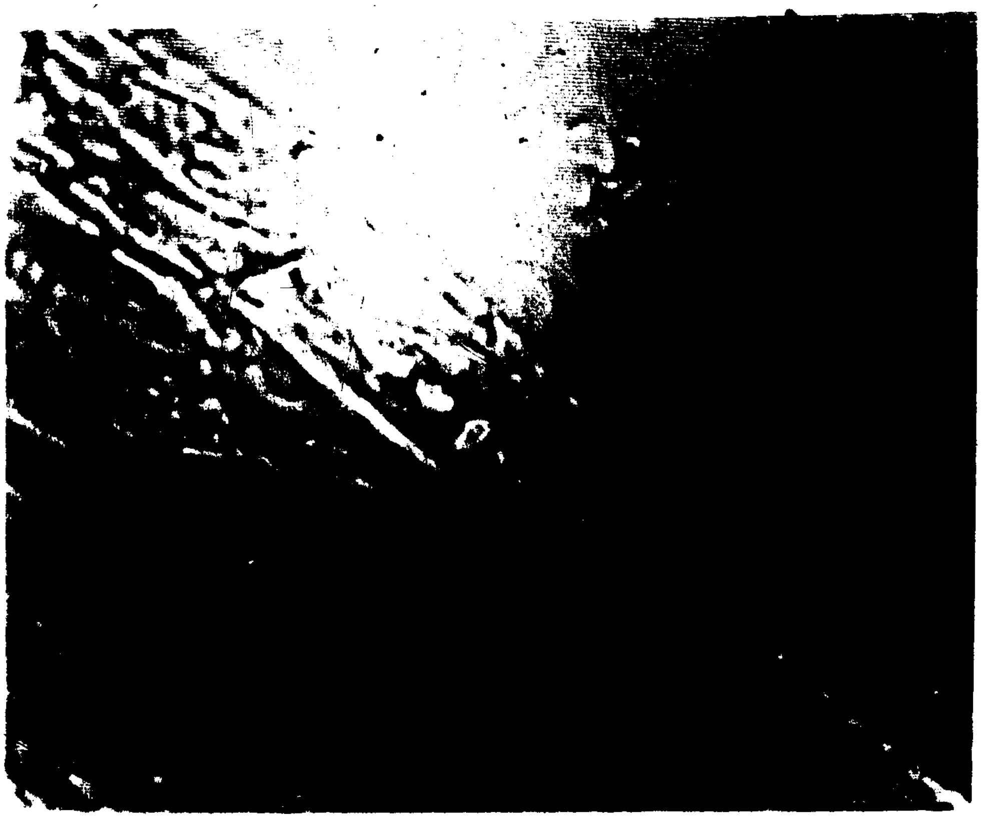


Figure 4.10. Grey tone image of down slope curvature coverage.

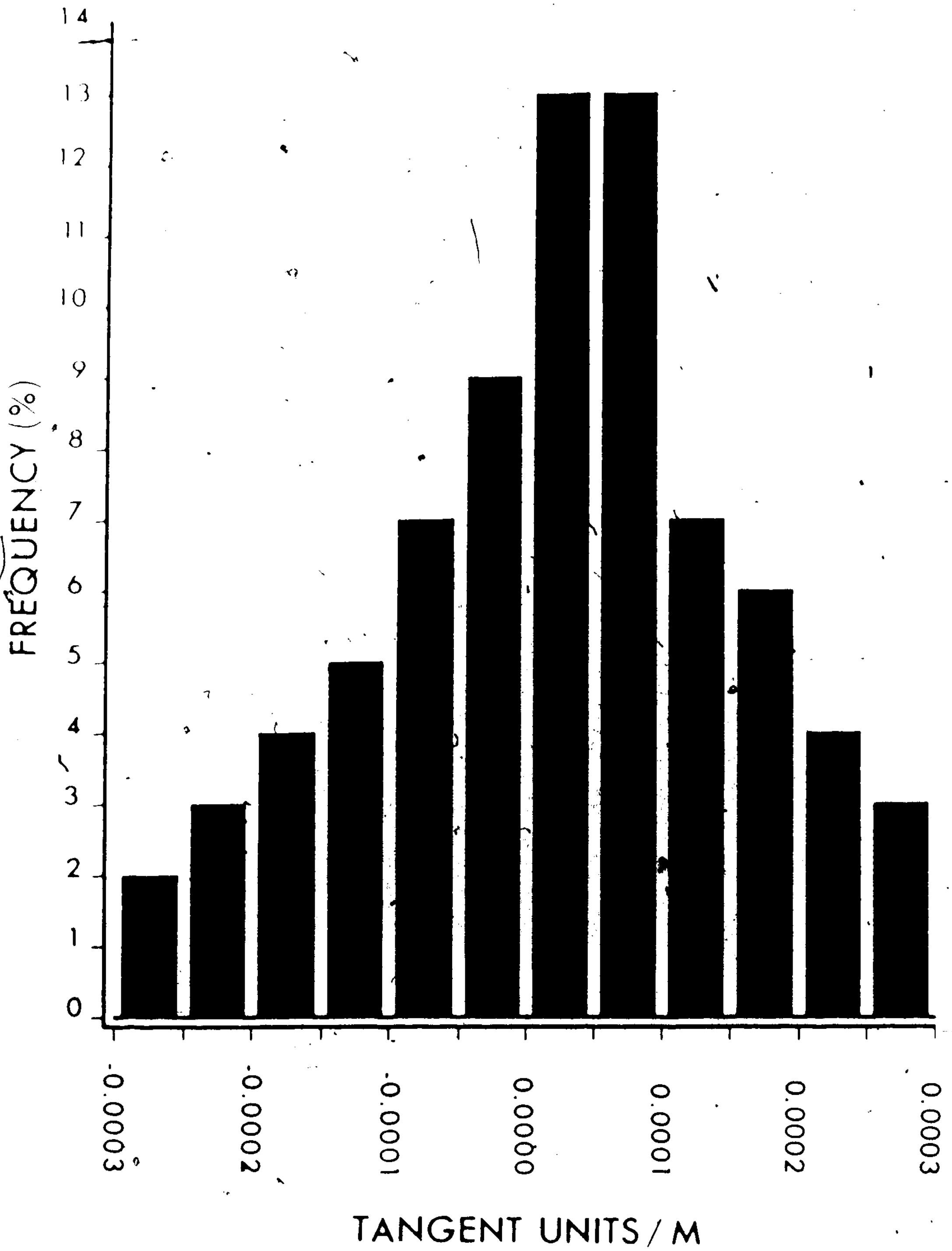


Figure 4.11. Histogram of down slope values.

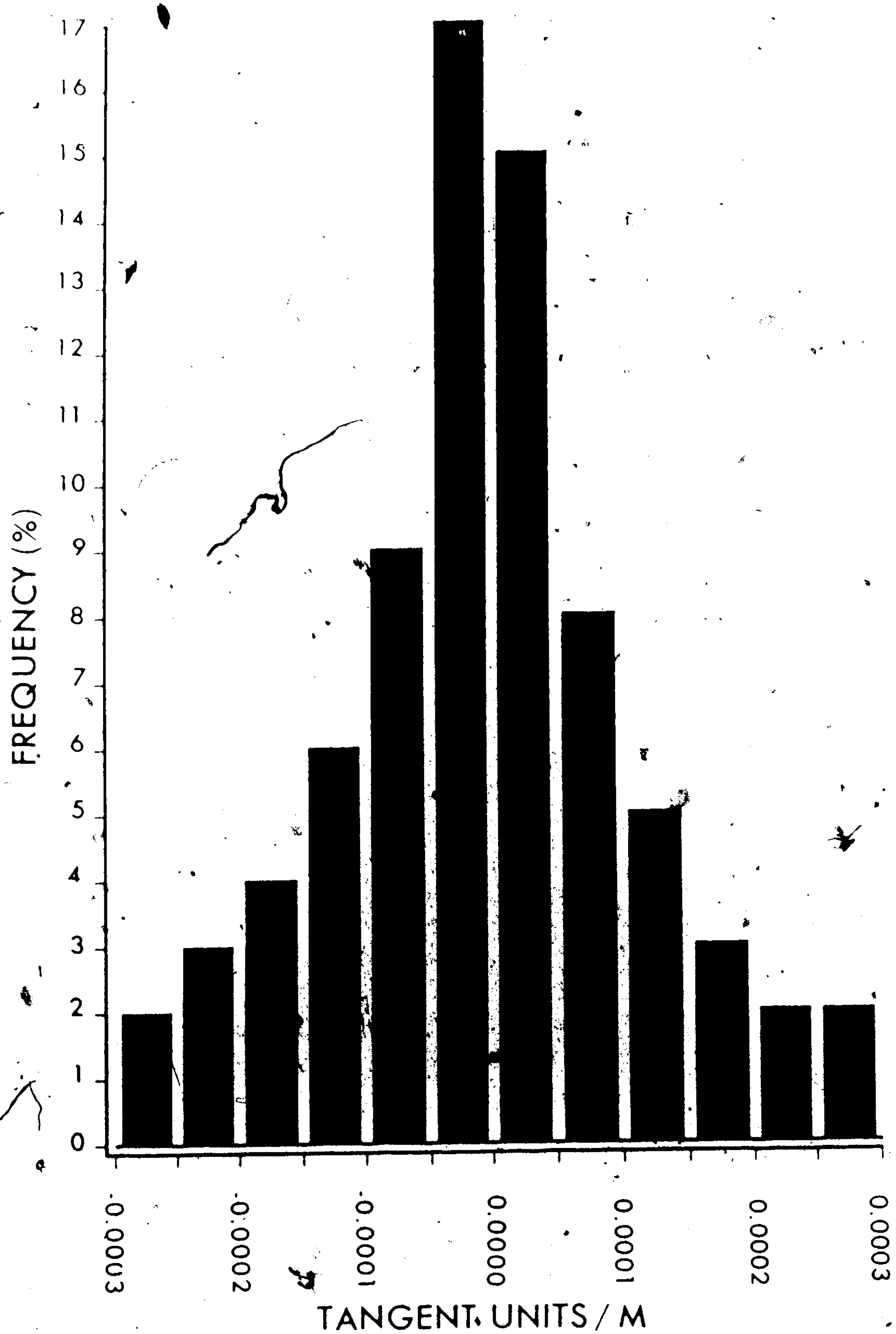


Figure 4.13. Histogram of the across curvature values for the study area

4.5 Summary

This chapter is an examination of the development of a data base which can be used as ancillary data for the MSS integration. Attention was given to the surfacing approach used to generate the DEM. The importance of this product cannot be overestimated as it formed the basis of the subsequent products which describe the geometric properties of the land surface. The first derivatives generated were those which described the slope gradient and the direction of maximum slope gradient. The second derivative measure described the form of the landscape both in terms of the across slope and down slope curvatures.

CHAPTER 5

An Examination of Interpretive Map Data as a Potential Ancillary Data Source

5.1 Introduction

This chapter deals with a comparison of two different sources of ancillary information which could potentially be used with MSS data to enhance forest cover classifications. Each of these data products have unique advantages and disadvantages as ancillary data sources. The first data source is the conventional interpreted polygon map. The advantage to using this type of map is that it was produced using the range of interpretive elements that were discussed in Chapter 1. The second data source is the derived map, described in the preceeding chapter. There are specific reasons for comparing these two information sources. The first is that numerous attempts at ancillary data integration (discussed in Chapter 7), have involved a pre or post classification sorting process based on a mapped theme. The second reason for this comparison is that the interpreted polygon map is the most common map coverage contained in Geographic Information Systems. This type of map is, therefore, the most readily available for integration.

The use of interpreted landform, or landscape, data in modelling may introduce problems. The first, and most significant problem, is one of generalization. Simplification and data reduction tend to occur at a number of stages throughout the mapping process. Unless source documents allow for resolutions which are greater than, or equal to, the LANDSAT data, a generalization in the remotely sensed data must be made to compensate for the lower level of detail. A second problem is that of error introduced by the ancillary data coverages. Errors contained in map sheets are multiplicative when the map sheets are overlaid. If an accuracy of 80% is achieved for a soils map, for example, and an accuracy of 70% is attained for the LANDSAT classification, then the accuracy of the resulting classification, or model, is only 56%. The implications for the incorporation of multiple coverages is evident. It is for these reasons that it was decided that a comparison between the two ancillary data sources was necessary.

5.2 Interpreted Map

The interpreted theme to be compared was derived from a 1:50000 scale physical land classification (PLC) map (Figure 5.1) prepared by Alberta Forests, Lands and Wildlife, Resource Evaluation Division. The mapped polygons represent the spatial distribution of soil properties, landscape - forming processes, soil moisture conditions and landform geometry (slope gradient and aspect). The information mapped was based primarily on the interpretation of aerial photographs with supporting field verification.

The PLC map corresponding to the study area was digitized. A matrix of slope values was produced by gridding the digitized

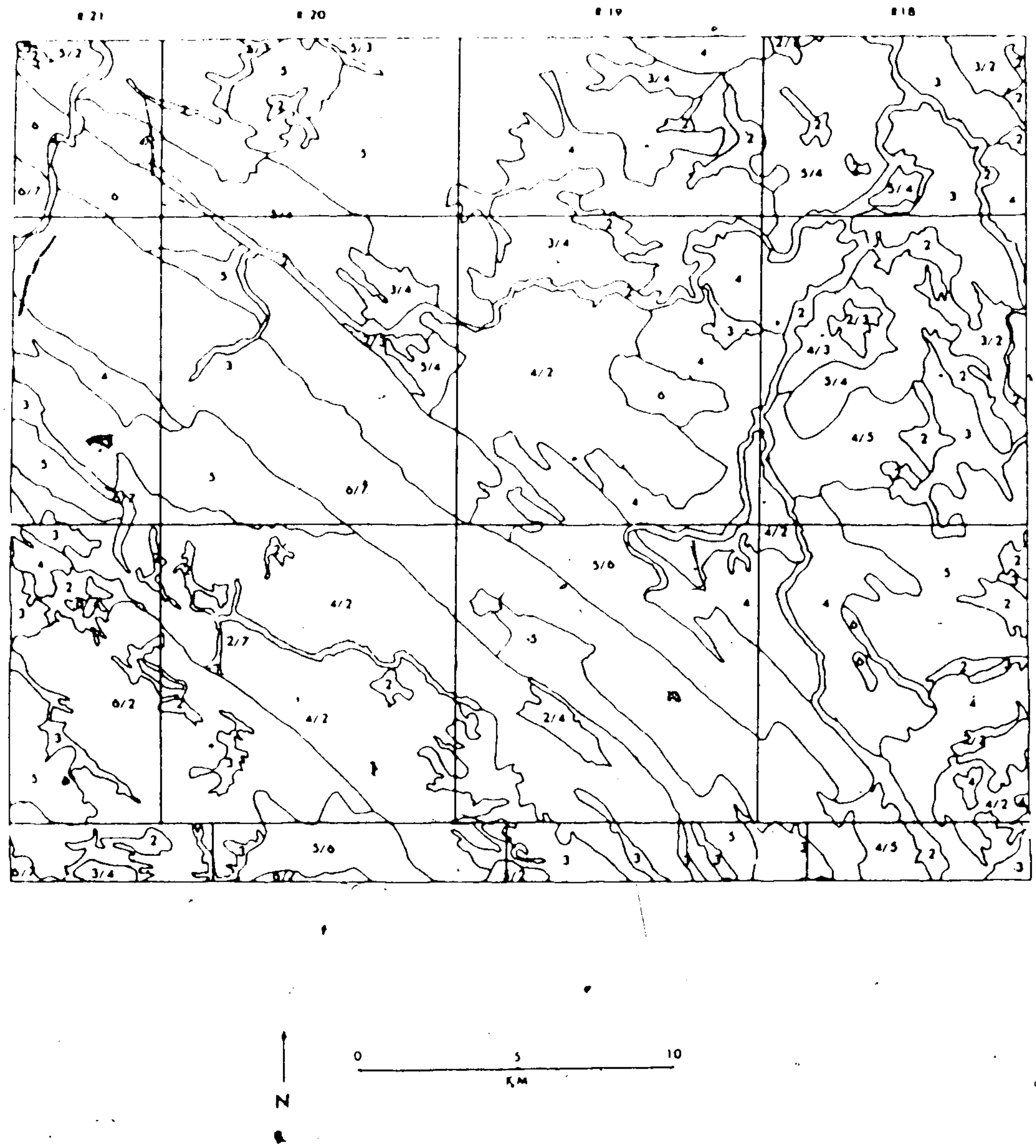


Figure 5.1. Interpreted slope map used in the analysis. (Polygon labels represent slope classes which are defined in Figures 5.2 and 5.3)

polygon coverage. The resulting matrix consisted of 520 rows and 640 columns of 50 metre grid cells. Each of these pixels contained a unique theme.

5.3 Comparison of Two Map Coverages

To evaluate the two map products (PLC map and slope distribution derived from the DTM/DEM) as to which was the most appropriate for LANDSAT classification improvement, a theme common to both was selected. This theme was slope gradient. In the case of the PLC map the mapped theme consisted of polygons with simple or complex slope designations. The simple class polygons were characterized by a uniform slope throughout. The complex class polygons may have had the entire specified range of slope classes occurring within its boundaries.

The two slope gradient coverages were spatially registered to coincide exactly with each other. This was a relatively simple procedure since the NTS map sheet used as the source for the DTM had also been used as a base for the PLC map. The origins of the digital coverages (the maximum and minimum northings and eastings) were known for each so that corresponding windows could be extracted.

The two maps were digitally overlaid to facilitate comparison. The results of this overlay are illustrated in Figures 5.2 and 5.3. The distribution of DTM/DEM slopes within the simple-classed PLC polygons are illustrated in Figure 5.2. For the lower classed PLC polygons (classes 2, 3 and 4) the distributions were similar. They were unimodal with the mode at 2% slope for all three

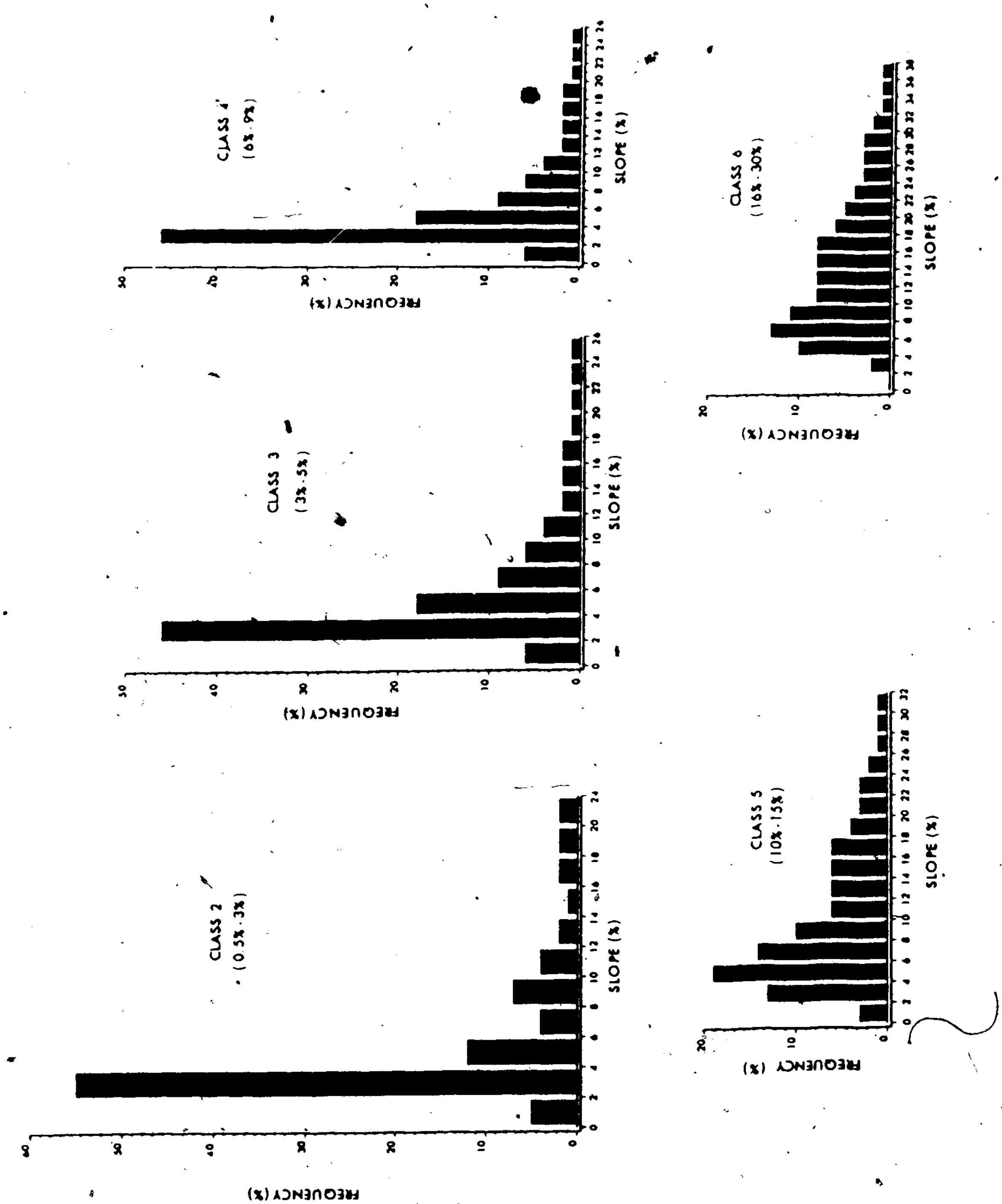


Figure 5.2. Distribution of DEM slope values within the simple class interpreted polygons

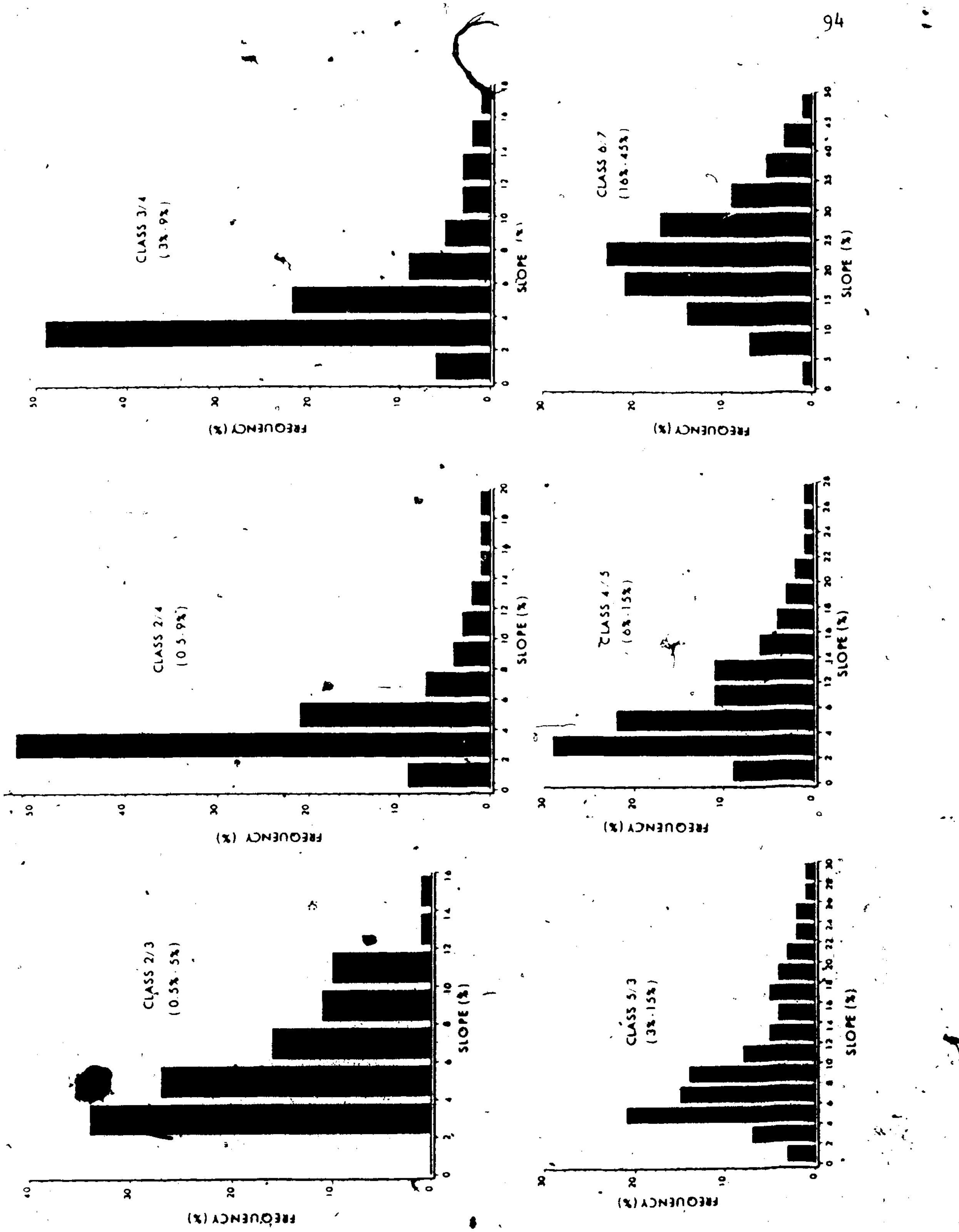


Figure 5.3. Distribution of DEM slope values within the multiple classed interpreted polygons.

classes. There was however an increasing positive skewness as the PLC gradients increased. The distribution changed for the higher slope classes (classes 5 and 6). The ranges of the slopes contained within the PLC polygons were much greater than the lower classed units and the modes less well defined.

The histograms for the distribution of the complex PLC polygons are shown in Figure 5.3. The distribution of the slope values shown in the histograms are similar to the pattern of the simple classed units. The modes of the lower classes are still low with the distributions positively skewed. The higher classed polygons with greater ranges in slope classes have a corresponding larger range in DTM/DEM values.

5.4 Discussion

The comparison of these two map products highlighted the high degree of generalization associated with the interpreted map. This generalization may have been the result of three factors. The first factor, as shown by the polygons representing the higher slope gradients, was that a large degree of cartographic generalization occurred during the various stages of map production. This generalization led to an apparent misclassification of areas within the polygons. For example, in the polygons with slope class 6, only 33% of the area fell within the range specified. The generalization may have been caused by several factors. The first factor may have been the method of survey, which may not have used slopes as the primary reason for delineating the polygon boundary. Rather, the primary reason for the delineation may have been uniform soil properties, with the slopes of only secondary importance. A second factor contributing to the generalization may have been that the regional slope within the polygon was not considered, but rather, the description of the PLC unit refers to a complex slope assemblage. The polygons, therefore, may have been interpreted and delineated with a smaller mapping scale in mind. In other words, although cartographically produced at a scale of 1:50000, the interpretation may have been carried out for mapping at scales of 1:150000. This point was rejected for two reasons. The first was that some of the polygons mapped at a 1:50000 scale were too small to be reproducible at smaller scales. The second reason was that the 1:150000 series was not a generalized version of the 1:50000 PLC map, but rather a different classification product which integrated vegetation information with the terrain interpretation. The third factor contributing to the apparent misclassification, especially important for low slope zones, was that dense forest cover of varying ages, densities, and species, tended to confuse the interpretation of actual slope breaks, thereby leading to erroneous interpretations.

In favour of the polygon map was that the contour interval of the original topographic map used to generate the DTM may have been too great to account for some of the variations in slope. This was especially important when dealing with the areas within the low slope polygons, where the variations in slope were relatively subtle. In the higher slope zones, where a change in class resulted in a variation of over 10% to 15% rather than 2%

as in the lower zones, the large contour interval was not as significant.

The rationale for the two slope coverages not coinciding was secondary to the fact that they did not match and that the spatial information content of the PLC map was substantially less than that contained within the DTM/DEM coverage. (This point was reinforced by the findings of a previous study by Nix et al. (1984), although the mismatch of their map products was attributed to differences in cartographic scale.) The conclusion of this analysis was that, for the purposes of integration with MSS data, the high degree of generalization of the polygonal map would introduce a substantial generalization in the final product.

5.5 Summary

Based on the results of this comparison it was concluded that the generalization of the information contained within the interpreted, polygonal maps greatly reduced their usefulness. The interpretations for other themes likely have similar characteristics as the slope gradient theme. The high degree to which the information content had been generalized placed serious limitations on the potential usefulness of this type of data source for use with LANDSAT MSS data in the classification of forest cover.

CHAPTER 6

Landscape Drainage Modelling

6.1 Introduction

Three points can be made from the discussions in the dissertation to this point. The first point is that the spectral resolution of the LANDSAT MSS data is not sufficient to discriminate the forest cover in sufficient detail to yield a classification which is useful beyond defining very broad classes such as coniferous, deciduous, and open soil. The second point is that the topographic variation in the area chosen for this study is small with 85%, of the terrain being located on slope gradients of less than 15%. The third point is that the generalization which is inherent in the interpreted polygon map renders this source of information of limited value for integration with MSS data.

Based on these three points three tasks were deemed necessary to fulfill the objectives of this project. They were:

1) To determine, through the use of previous studies reported in the literature, what characteristics affect the natural distribution of forest vegetation cover.

2) To investigate, through the use of published literature, if these characteristics have previously been modelled using DEM derivative products.

3) To investigate the relationship of the DEM derivatives developed for this study to the terrain characteristics which contribute to forest vegetation cover.

These three tasks are dealt with in this chapter.

6.2 Relationship of Terrain Characteristics and Stand Composition

Numerous studies have investigated the relationship between natural forest site characteristics and the quality and composition of the natural vegetation cover. Coile (1952) found that the quality of the site with respect to vegetation growth is largely determined by edaphic properties, terrain characteristics, climate, and the inherent rooting habits of different tree species.

Wilde et al. (1954), working in the Algoma area of northern Ontario, related soil profile characteristics including texture, drainage, permeability, pH, organic content, and soil depth to the distribution of natural vegetation cover. The descriptions provided for these soil/vegetation relationships were very general and did not discuss the influence of various soil characteristics on the vegetation types in detail. A correlation was made, however, between soil drainage and vegetation distribution. The more rapidly drained soils which had developed on the colluvial and coarser glaciofluvial materials were covered with stands of conifer composed primarily of pine with some spruce. The soils with a high clay content, thereby less well drained, were covered by conifer stands composed of fir and spruce. The least well drained soils, including the muck and peat deposits, supported predominantly black spruce, sedges, mosses and deciduous shrubs. Site quality was also implied by noting that stands which had developed on deeper, well drained soils are faster growing than those developed on the more poorly drained lowland or shallow, skeletal soils.

Daubenmire (1968), working in Northern Idaho, stated that the soil moisture regime during the growing season not only played a significant role in determining the quality of the site for a particular species, but also influenced the forest composition at that point. He noted that as the soils became less well drained there was a change in the predominant forest species from pine to fir to black spruce.

Matcock and Curtis (1960) found that of the environmental factors most commonly cited, soil moisture regime was the one which most commonly differentiated forest species composition. They stated that the species which were most frequently associated with coarse textured, well drained sandy soils were markedly distinct from these species characteristic of poorly drained clay-rich and organic types.

Jeglum (1974) reported a gradation from closed black spruce stands found on mineral soils to open stunted stands on organic soils and peatlands. The quality of a site for black spruce decreased with an increase in moisture and a decrease in aeration. Lowry (1975) stated that black spruce stands exhibit the best growth on mesic sites. The influences of microclimate as reflected by slope, aspect, relief and exposure, he noted, exert a lesser influence. The relationship of site quality of black spruce to drainage was also discussed by Dryness and Grigal (1979). They noted that well drained and very poorly drained sites offered the poorest environments for black spruce growth.

Zobel et al. (1976) stated that "...factors related to temperature, moisture, nutrients, and mechanical stress often correlate with the observed vegetation pattern" (Zobel et al., 1976, p.47). They noted that soil moisture availability has been recognized as the most influential factor in determining the distribution of forest vegetation by acting as a limiting factor. A second limiting factor, temperature, became dominant when higher precipitation and lower evapotranspiration rates prevailed. This zonation, based on temperature regimes was relevant when large geographic areas or areas with a large elevation range, were being investigated.

Spurr and Barnes (1980) stated that in addition to the edaphic characteristics of a site, the topography was important in controlling the moisture conditions. They noted that the slope gradient and aspect related to the quality of a site for growth, as did the relative position of a site on the slope. This latter point was discussed by a number of investigators. Richards and Stone (1964) found that the internal drainage of the soil profile was related to the slope gradient as well as the slope shape. Graney and Ferguson (1971) studied site quality relationships for pine stands and reported that the topographic variables used in their study all contributed in some way to the degree of soil drainage. Spurr and Barnes (1980) also found that site quality could be equated to the more common physiographic elements including aspect, slope gradient and elevation. They concluded that other topographic factors, not usually considered, such as slope convexity, concavity and the position relative to the crest of the hill slope, were of importance. Bowersox and Ward (1972) found that site quality for oak was influenced by slope aspect, gradient and slope position. The variable slope position, described an integration of slope form as well as the distance from the top of the hill slope as was the case with Graney and Ferguson (1971). Using multiple regression analysis, they found that of the topographic variables, the slope position was the most significant. In their study, Spies and Barnes (1985) used a stepwise discriminant analysis to investigate the contribution of 55 physiographic and soil variables in describing the nature of forest ecosystems. From the results they concluded that the three physiographic variables which were most significant included slope gradient, topographic position and the distance to a body of water. The topographic position used a qualitative classification of the slope components, for example, upper slope, mid slope, and lower slope. The ecosystems, they found, could be classified into two basic groups, upland and wetland ecosystems. The upland ecosystems were distinguished from the wetland ones by the physiographic variables as well as soil drainage and in some cases soil texture.

Another factor influencing vegetation composition and site quality is the depth to the water table. A shallow water table will act in much the same way as any other restriction, by not allowing the establishment of deep rooting species (Spurr and

Barnes, 1980). In the case of black spruce, the rooting depth is quite shallow so that a high water table in mineral soils is not a factor in their establishment.

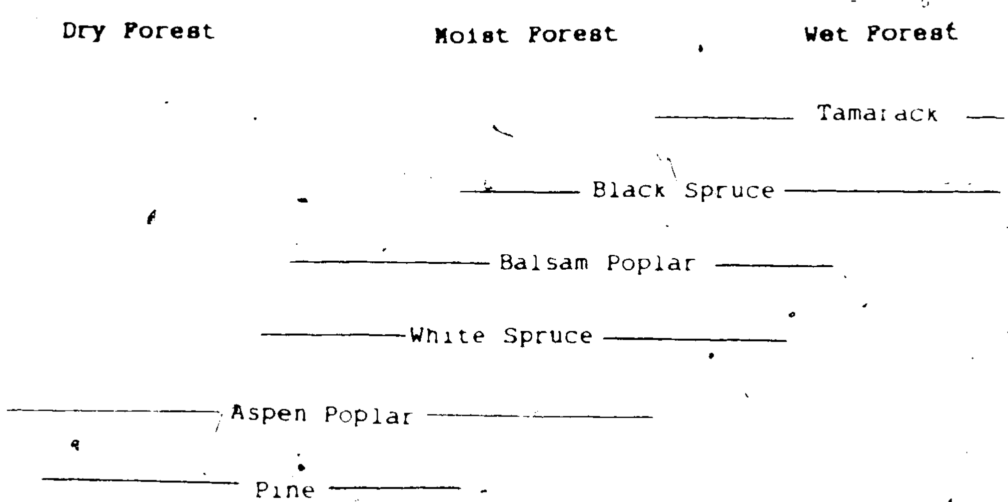
A common conclusion of the studies cited above was that although a number of factors influenced the type of vegetative cover and the quality of conditions for growth, the recurring limiting factor was moisture. This limiting factor may have been on the side of excess as in the case of pine or fens, however, conditions may also have been excessively xeric which would have affected the distribution of black spruce. The predominant factor was therefore the degree to which the soil profile was droughty or saturated during the growing season.

6.3 Vegetation Characteristics of the Boreal Forest with Special Reference to West Central Alberta

A very general zonation of Boreal forest vegetation, based on soils drainage, is illustrated in Figure 6.1. The zoning is not well defined, and there is some overlap between some of the species represented. A few of the species such as birch (Betula spp.) and white spruce (Picea glauca) do exist within a wide range of moisture conditions.

Some general conclusions regarding the zonation were made by Corns (1983) who discussed the distribution of forest community types in the Boreal forest of west central Alberta. He noted that the distribution of the forest species depended on elevation as well as the soil moisture regime. He found that pine (<u>Pinus contorta</u>) had a very high elevation range, but was generally restricted to well drained and to moderately well drained sites. Some instances of imperfectly drained sites were also recorded. At the other moisture extreme was black spruce (Picea mariana) which was found in imperfectly to very poorly drained sites. White spruce (Picea glauca) was found bridging these two extremes. Subalpine fir (Abies lasiocarpa) was located in the higher elevations (greater than 1600 metres) on well drained sites.

For the present study area, an evaluation of the vegetation cover was carried out by Bentz et al. (1985). Figures 6.2 a-d represent four typified cross sections relating terrain characteristics to the vegetation pattern. Figures 6.2 a and b represent the vegetation cross sections found on the undulating morainal deposits located in the eastern sections of the study area. The first of these cross sections, Figure 6.2 a, is from an area of fluted and hummocky moraine. The areas which this figure represents are undulating with linear ridges and hummocks separated by depressional zones. The predominant forest cover for these areas is pine on the higher, well drained sections. This cover grades to black spruce on the lower, less well drained portions of the slopes and along the peripheries of the depressional areas. The poorly drained depressional areas are covered by deciduous shrub and herbaceous cover. Figure 6.2b represents areas with fluted and hummocky moraines but which are covered with a thin veneer of glaciofluvial deposits. The upland areas are well drained, with the drainage decreasing towards the base of the slopes, as was the case above in the previous cross



(from Larson, 1980)

Figure 6.1. Relationship of soil moisture to forest vegetation cover.

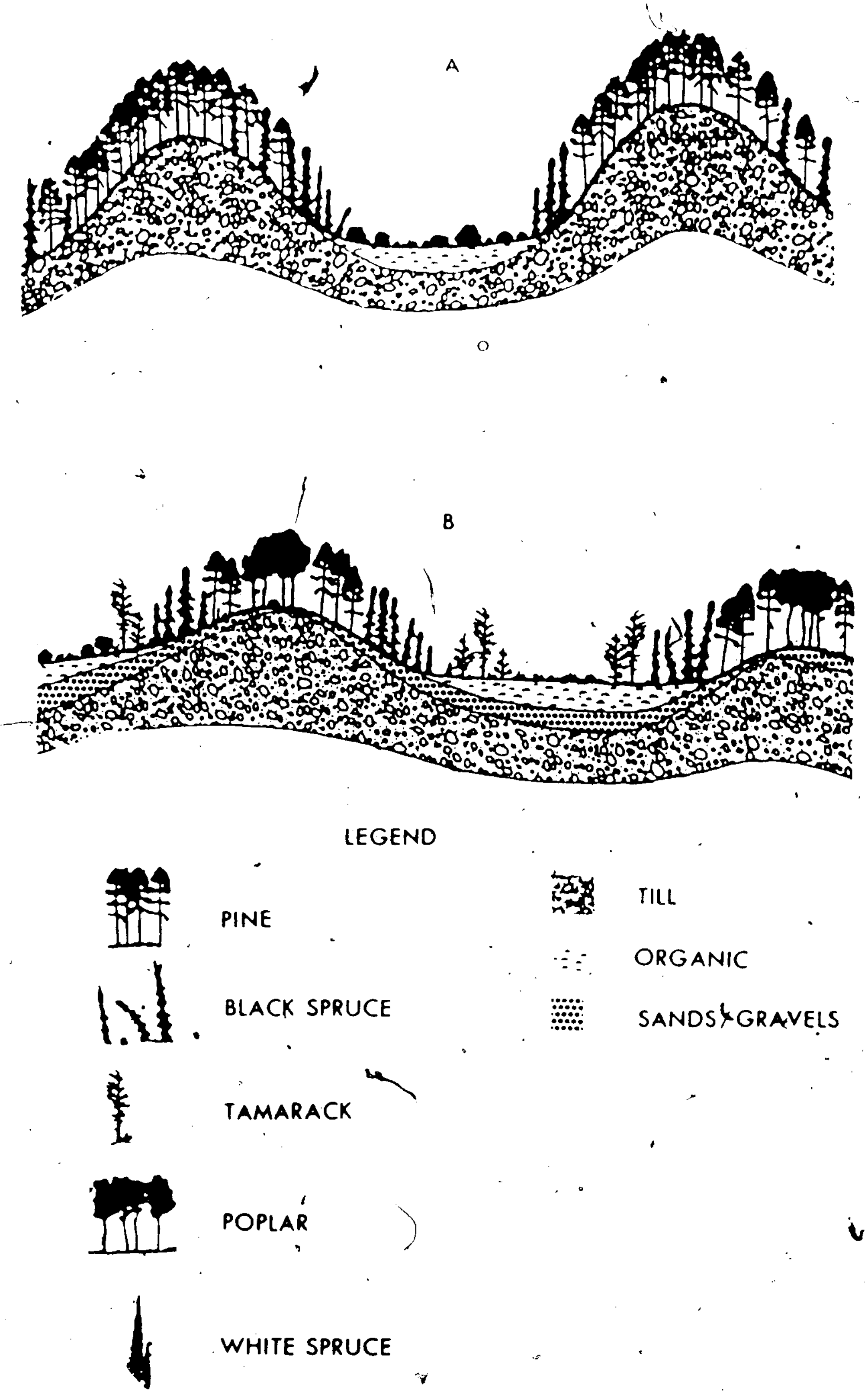


Figure 6.2. Cross sections showing typified vegetation topographic relationships (based on Bentz et al., 1985). See text for discussion.

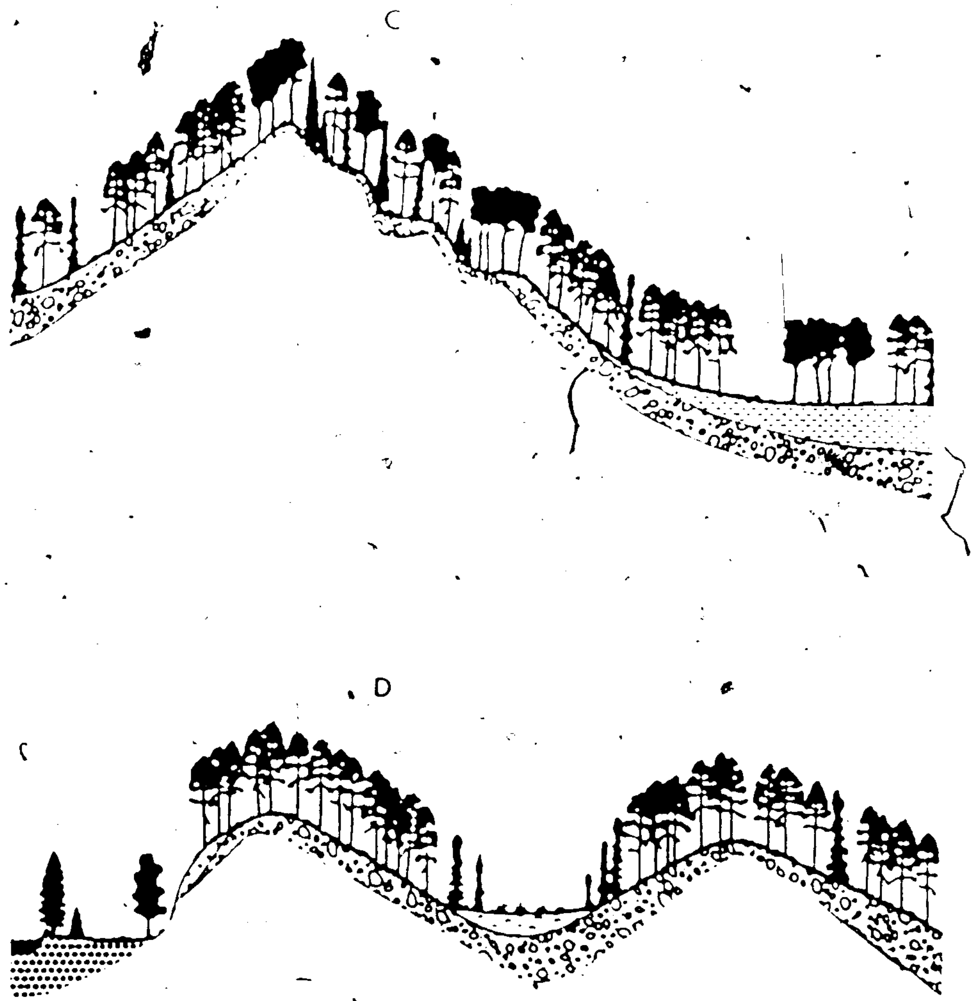


Figure 6.2 cont'd

section (Figure 6.2a). The difference between the areas portrayed in cross sections 6.2a and 6.2b is that the glaciofluvial deposits, predominantly sands and gravels, promote a greater degree of drainage. The pine stands may, therefore, contain significant amounts of aspen poplar, especially at the summits of the hummocks. Away from the hummocks the aspen stands are replaced by stands of pure pine. The cover grades into black spruce towards the base of the slopes and to black spruce and tamarack around the periphery of the depressional areas, which are covered by deciduous shrubs (for example birch) and graminoids. Although this cross section (Figure 6.2b) has been included in this discussion, it does not represent a large proportion of the study area. The first cross section is representative of the majority of the study area occupied by the upper plains.

The third vegetation cross section (Figure 6.2c) represents the northwest to southeast trending, bedrock - controlled ridge located in the centre of the study area. The soil drainage ranges from excessive at the crest of the ridge to poor in the depressions. The excessively drained crests of the ridges are covered by stands of pure aspen poplar and mixtures of pine and aspen poplar. As the drainage decreases down slope, the aspen is replaced by white spruce and further down slope by black spruce. The base of the slopes are covered by mixed stands of spruce, pine and balsam poplar.

The final cross section (Figure 6.2 d) is representative of the vegetation assemblages in the western - most portions of the study area. The topography of these areas is characterized by bedrock controlled ridges which are covered by a morainal veneer or blanket. The depressions separating the ridges contain accumulations of organic soils. The ridges are well drained and covered by closed coniferous forests composed of lodgepole pine, subalpine fir, and black spruce. The lower slopes have black spruce and pine cover, while the depressions are covered by black spruce and tamarack at the edges and birch and graminoids in the centres.

Based on the foregoing descriptions it is evident that the forest vegetation assemblages are influenced to a large extent by their position on the slope and the associated moisture conditions. The vegetation cover corresponding to soil drainage, therefore, ranges from aspen and lodgepole pine on the best drained sites, to lodgepole pine and white spruce on the sites with moderate drainage, to lodgepole pine, black spruce and balsam poplar on less well drained sites, to black spruce and tamarack on imperfectly to poorly drained sites, and to deciduous shrubs and graminoids on the very poorly drained sites.

6.4 Terrain Geometry and Moisture Drainage

From the previous discussions a relationship between drainage characteristics at a site and the location of the site in the landscape can be established. The relationship of the drainage characteristics and terrain geometry have been extensively studied both in the field as well as in the laboratory. Troeh (1964), in an examination of soils maps and contour maps, concluded that the curvature of the contour lines (that is convexity and concavity) were significantly related to landscape drainage. Areas of concave slopes were related to areas of poor drainage while the better drained sites were associated with convex sites. Examination of the geometric properties of the landscape by Troeh (1964) indicated that slope gradient correlated the highest with drainage, followed by curvature and the radius of slope curvature.

Young and Mutchler (1969) noted that there was a substantial effect of slope shape on moisture drainage. Moisture drained faster from convex slopes than from either straight or concave ones. Calve, Kirkby and Weyman (1972) found that drainage was most affected by three topographic elements: the length of the slope; the shape of the slope; and the gradient of the slope. The effect of lower slope gradients was to decrease response times which led to the accumulation of moisture. Slope shape concentrated the flow so that depressions and hollows became saturated while ridges would dessicate. Finally, as the distance increased from the crest of the slope, the amount of antecedent moisture increased which resulted in a more rapid saturation at the slope base. All of these components have combinatorial effects so that a complex-shaped hillslope behaved differently to hydrologic events than a simple-shaped one. Whipkey and Kirkby (1978) concluded similarly that the soil moisture regime was controlled to a large extent by the shape of the hillslope. They noted that all other things being equal, there was a uniform increase in moisture with an increase in distance from the divide and that the degree of saturation was inversely related to the slope gradient. They also noted that the coarse textured soils were usually associated with the steeper slopes which also enhanced the drainage of moisture.

Beven (1978) found that differences between the location within the drainage basin and amount of soil moisture could be explained by the geometric properties of the landscape. The headwaters, he noted, normally were broadly convex with high slopes. These areas, therefore, drained rapidly. The lower portions of the basin were generally concave and lower in gradient so that these areas were more poorly drained. This supported earlier work by Weyman (1973) and Dunne et al. (1975) who found that areas adjacent to drainage channels and at the base of drainage basins were the first to become saturated followed by depressions, or swales, and the foot slopes.

In summary, a number of physical conditions which affect the hydrologic response of an area were listed by Fleming (1975). These conditions included contributing (or drainage) area, elevation, landslope, length of flow path, vegetation cover, soil characteristics, bedrock and surficial geology, and landuse.

6.5 Drainage Modelling

Fleming (1975) discussed two possible approaches to modelling of earth surface processes. The first modelling approach treated the area in question as a homogeneous surface, with uniformly distributed physical characteristics, such as soil texture or slope gradient. The second subdivided the area of interest into a number of finite elements and calculated the process response for each of these elements. The advantage of the first, lumped, modelling strategy was that only a very generalized knowledge of the physical characteristics was necessary. The problem was that similarly general results are derived, so that only broad conclusions could be made regarding the process. The second, or distributed, modelling approach treated the basin as being composed of spatially discrete elements. As these spatial entities were treated as independent elements, it was imperative that detailed information be used regarding the distribution of the physical properties. This latter approach is less widely used, because of the level of detail required. It does, however, yield results which are suited to gridded, or raster, data bases such as the one used in this study.

6.5.1 **Digital Products**

Certain DEM and derivative coverages such as elevation, slope gradient, slope aspect* and curvature are the most commonly used terrain variables for modelling and have been introduced earlier. Given the preceeding discussion, however, it was apparent that other digital landscape coverages would also have to be used. These will be discussed below.

6.5.1.1 **Relief**

A measure of the relief of the terrain was developed for this study. This coverage described the variation in elevation in a

---

* See Appendix 2 regarding the use of aspect in this analysis.

Figure 6.3. Grey tone image of relief coverage.

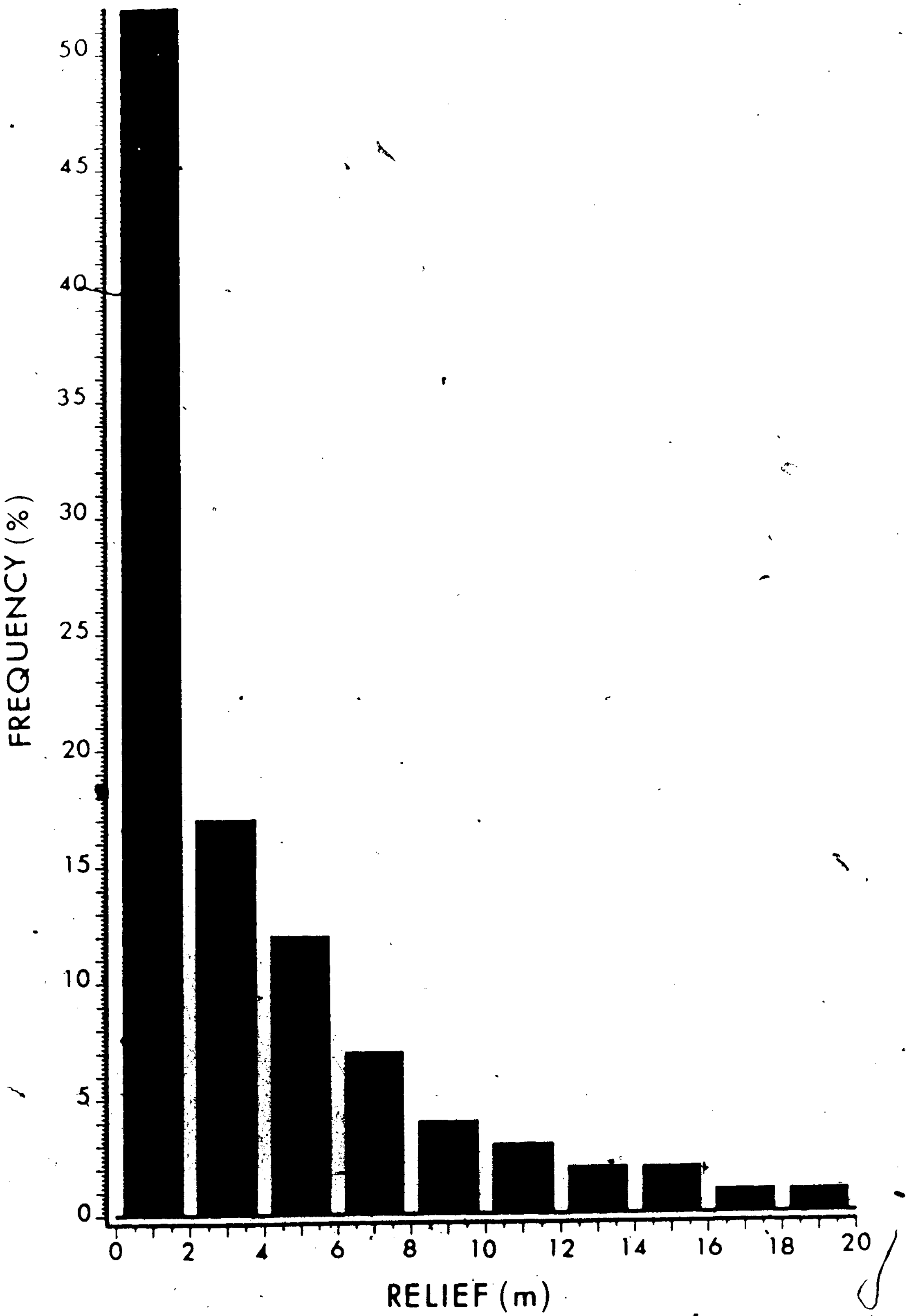


Figure 6.4. Histogram of distribution of relief values.

neighbourhood of pixels. This variable was described by Evans (1972) and more recently by Franklin (1985) as the standard deviation of elevation within the pixel neighbourhood. For the purposes of this study the neighbourhood size chosen was 5 X 5 pixels (Franklin, 1985). This resulted in a sample which was sufficiently large to yield valid standard deviation results without using an area which was too large. In the case of a 5 X 5 neighbourhood a 250 metre by 250 metre grid is used. A grey tone image of this coverage is presented in Figure 6.3. A histogram of the distribution of the relief values is presented in Figure 6.4. The predominantly low relief values indicated in the histogram reflect the low slope values in the study area as shown in Figure 4.8.

6.5.1.2 **Flow Paths**

All of the DEM derivative coverages discussed so far have dealt with only the land form geometry at a point or within the context of a 3 X 3 (or 5 X 5 cell) neighbourhood. These variables do not reflect the influences of the geometric properties up slope, away from the sampling point or neighbourhood. A measure of slope length, or the relative position of the sample site on the slope, has been identified by a number of workers as being fundamental in determining site quality for many forest species. This is not represented by any of the point measures. To account for the relative position, a flow path measure was adopted.

Past studies have reported on the generation of drainage networks from DEMs. O'Callaghan and Mark (1984) presented a method by which major drainage paths could be extracted using DEMs. Their approach was to derive a drainage accumulation matrix (DAM) by examining a drainage direction matrix. This latter data set was a representation of the direction of maximum slope. The resulting DAM was a summary of the total drainage area which contributed to that element. Threshold values were chosen by which channelized flow could be developed. Zones with values greater than the threshold were retained to represent the channels. An approach similar to the derivation of the DAM was adopted by Eyton (1987). In a DEM from Pennsylvania, however, he chose to separate channelized flow from nonchannel, or sheet, flow. Yuan and Vanderpool (1986) in their approach to simulating drainage also considered the effect of depressions on the drainage pattern. With the exception of Eyton, the studies mentioned above chose to ignore the lower flow path values and concentrated solely on the definition of channelized flow. Eyton's inclusion of sheet flow represents an important extension to the previous uses.

For this study the flow path number, as defined by Eyton (1987), produced a viable representation of the slope length above a particular point. The algorithm used was the one developed and presented by Eyton (1987). The flow path modelling procedure is outlined in Figure 6.5 a-c. A grid is initialized with values of 1 (or any other value preferred). This grid is illustrated in Figure 6.5 a. A 3 X 3 neighbourhood starting at the upper left of the DEM (Figure 6.5 b) is searched for its lowest value relative to the center of the neighbourhood. A pointer is positioned at this lowest point and the equivalent a: **Initialized Answer Matrix**

1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1

b: **Gridded Surface**

140	135	130	125	120	130	137	140	146	146
130	125	110	100	100	110	125	125	125	130
105	100	90	80	85	95	100	100	105	110
85	75	70	65	67	72	75	80	87	90
60	65	55	51	53	55	57	69	62	65
40	43	40	38	39	43	50	57	60	53

c: **Result of First Pass**

1	1	1	1	1	1	1	1	1	1
1	2	1	1	1	1	1	1	1	1
1	1	3	1	1	1	1	1	1	1
1	1	1	4	1	1	1	1	1	1
1	1	1	5	1	1	1	1	1	1
1	1	1	6	1	1	1	1	1	1

Figure 6.5. Outline of flow path generation. See text for explanation.

pixel is incremented in the answer matrix (Figure 6.5c). This process is repeated until the path which is being followed reaches the edge of the DEM or a closed depression is encountered

3 neighbourhood of the DEM (row 2, column 3) and the flow path is traced down slope. The process is repeated until the entire DEM has been evaluated to row n-1, column n-1. The flow path number resulting from the algorithm used, reflects not the linear distance from the crest of the slope to the sampling point, but the sum of all of the grid cells which contribute to the amount of flow through that point. This may incorporate the linear distance down slope of a single path as well as a large number of tributary pixels. A word of caution in using this flow path number is that it does not relate to an absolute value. Rather, it is a statistical variable which reflects the length of flow paths within a basin.

The grey tone image for the flow path is shown in Figure 6.6. The dark areas represent crests of hills and ridges. The grey values increase with an increase in distance down slope. The small linear features represent areas of concentrated drainage, possibly channelized flow. The frequency of the flow paths shown in Figure 6.6 is shown in Figure 6.7.

6.5.2 Drainage Model

From the literature presented earlier in this chapter, it is evident that there are a number of topographic elements which contribute to the movement of moisture through the landscape.

Figure 6.6. Grey tone of flow path coverage.

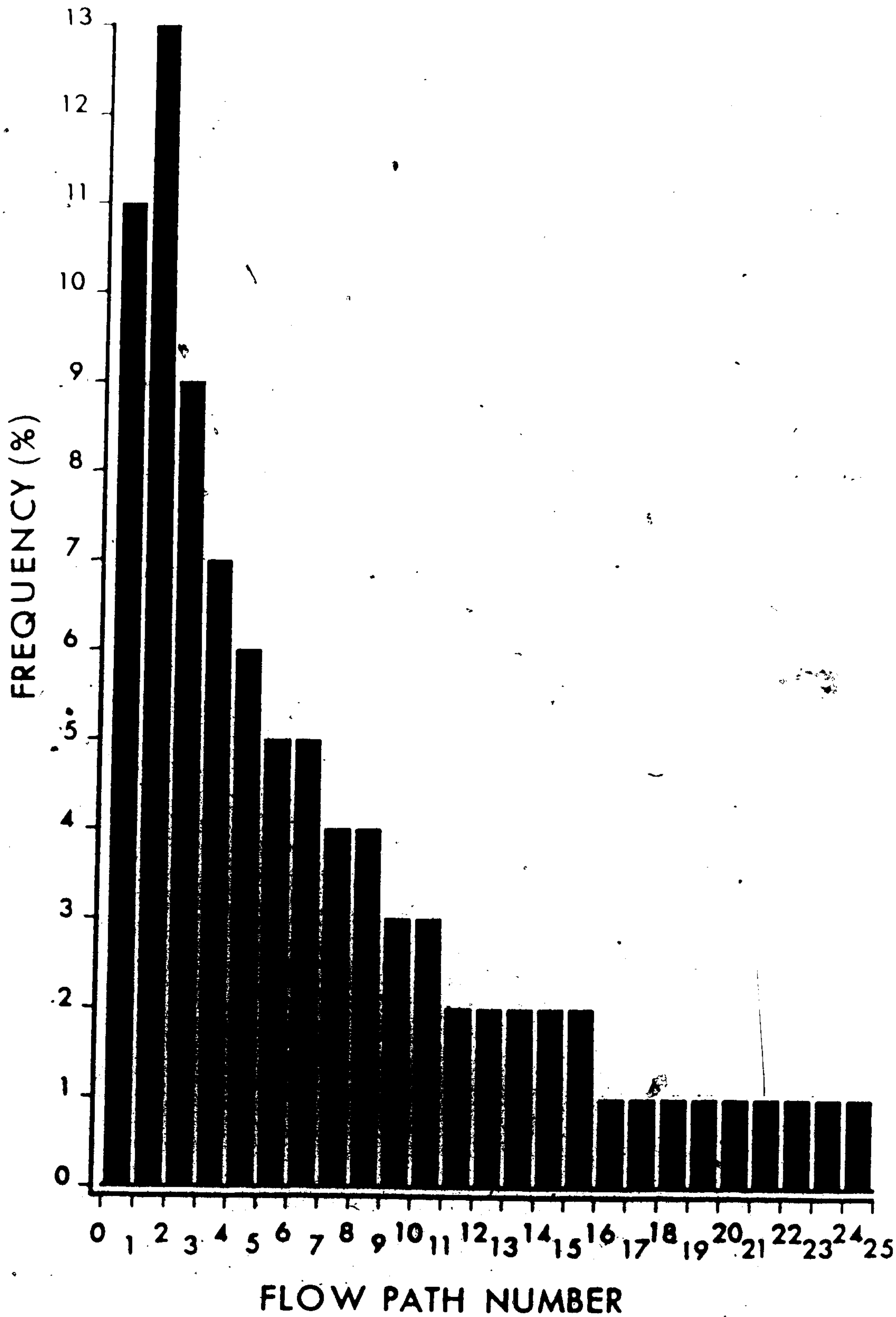


Figure 6.7. Histogram of flow path distribution.

These elements do not act in a simple, linear fashion, but rather, in a complex manner together with a variety of soil variables such as texture, structure, and parent material.

To develop an understanding of the contribution of the derivative and landscape coverages developed in defining landscape drainage, a number of sites were photointerpreted into three classes: well drained, moderately well drained, and poorly drained. This concept of landscape drainage is analogous to soil drainage as was defined by the the National Soil Survey Committee of Canada (1974) as being determined by the:

1) Actual soil moisture content of the soil in excess of the field moisture capacity
2) The extent of the period during which such excess water is present.

It was noted by the National Soil Survey Committee that the main diagnostic field criterion was the degree of mottling of the soil profile. The concept of landscape drainage views the landform as a three - dimensional entity and therefore extends beyond the soil profile. As such, other important diagnostic criteria in addition to the soil characteristics, including the topographic position and the vegetation cover are also used.

Based on descriptions of the drainage, vegetation and land form relationships, discussed earlier in this chapter as well as unpublished data supplied by Alberta Forestry, Lands and Wildlife, the three landscape drainage classes for the study area were defined as follows:

1) Well Drained: The well drained sites in the study area are found on the upper sections of slopes. They

predominate on the steeper slopes but may also be found on lower gradients in areas associated with coarse textured parent materials, such as those of fluvial and glaciofluvial origin. As soils developed on these materials are restricted to a small portion of the study area, the majority of the well drained sites are located on the upper sections of slopes. The vegetation characteristic of these sites is predominantly lodgepole pine of varying crown closure densities. There are also occurrences of pure aspen stands.

2) Moderately Well Drained: The moderately well drained sites are found on the mid- to lower- portions of long slopes and predominate on the shorter slope environments such as in the hummocky and ground moraine environments. The predominant vegetation type associated with this drainage class is lodgepole pine, white spruce and aspen mixtures occurring as closed stands. In transitional zones with poorly drained sites there may be mixtures of lodgepole pine and black spruce.

3) Poorly Drained: The poorly drained areas are located at the base of slopes and in depressional sites. These sites are characterized by organic soils and are the result of a large influx of mineral rich groundwater from the surrounding upland areas. The vegetation cover varies throughout

the area ranging from dense black spruce to black spruce/tamarack to nontreed shrub and herbaceous fens.

Based on these descriptions a total of 60 sites, 20 sites for each class, were chosen for detailed analysis using the landscape and derivative coverages. The drainage class at each site was interpreted from aerial photographs (scale 1:30000 approximately). The location of each of these sites was plotted on a 1:50000 topographic map and digitized. The values at the sites for each of the derivative and landscape coverages were extracted from the digital data base. The location of an additional 18 field - checked sites were digitized separately. The data for each of these sites was also extracted from the data base. These sites were used to evaluate the initial interpretation and subsequent modelling of the original 60 sites.

6.5.3 **Results**

The distributions of the derivative and landscape values for the 60 landscape drainage sites are illustrated in Figures 6.8 to 6.11. The slope distributions for the sample sites (Figure 6.8) indicated that the best drained sites were located on the highest sloping surfaces, as would be expected given the definition of this drainage class. The slope gradients associated with the other two drainage classes were low, predominantly below 5%. The distributions of the flow path numbers are illustrated in Figure 6.9. As expected, well- and moderately- drained sites yielded low values whereas those values associated with poorly drained sites

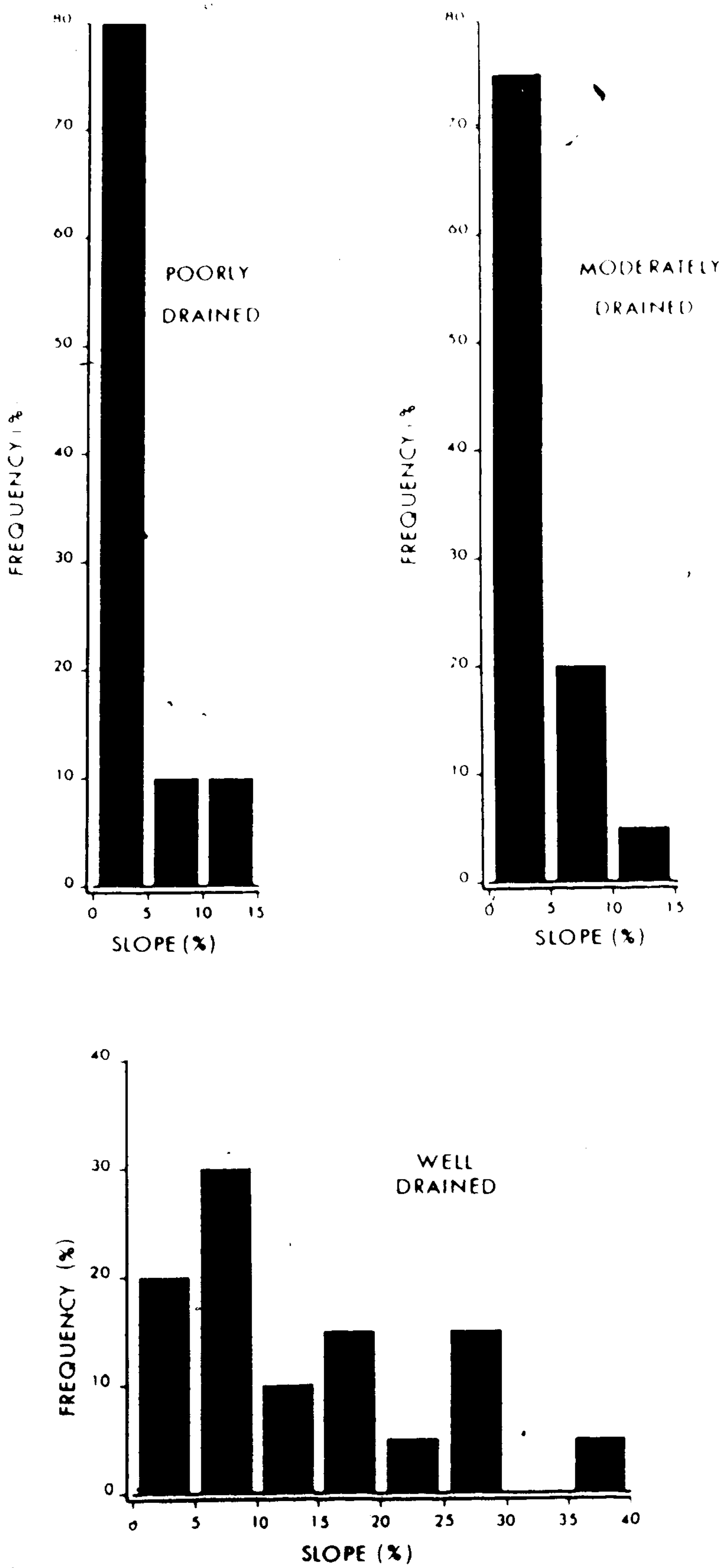


Figure 6.8. Histograms of the slope distributions for landscape drainage sites.

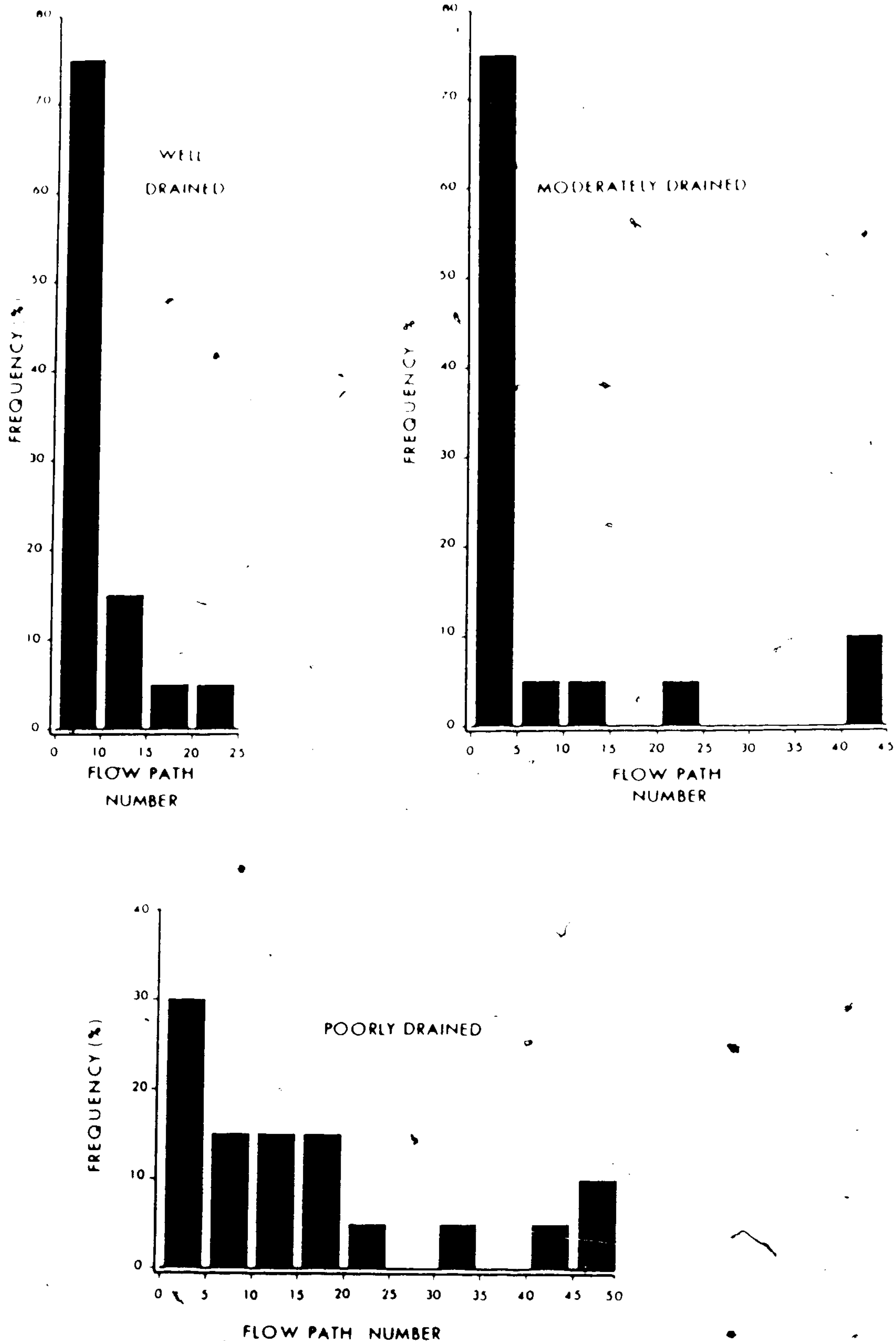


Figure 6.9. Histograms of the distribution of flow path values for the landscape drainage sites.

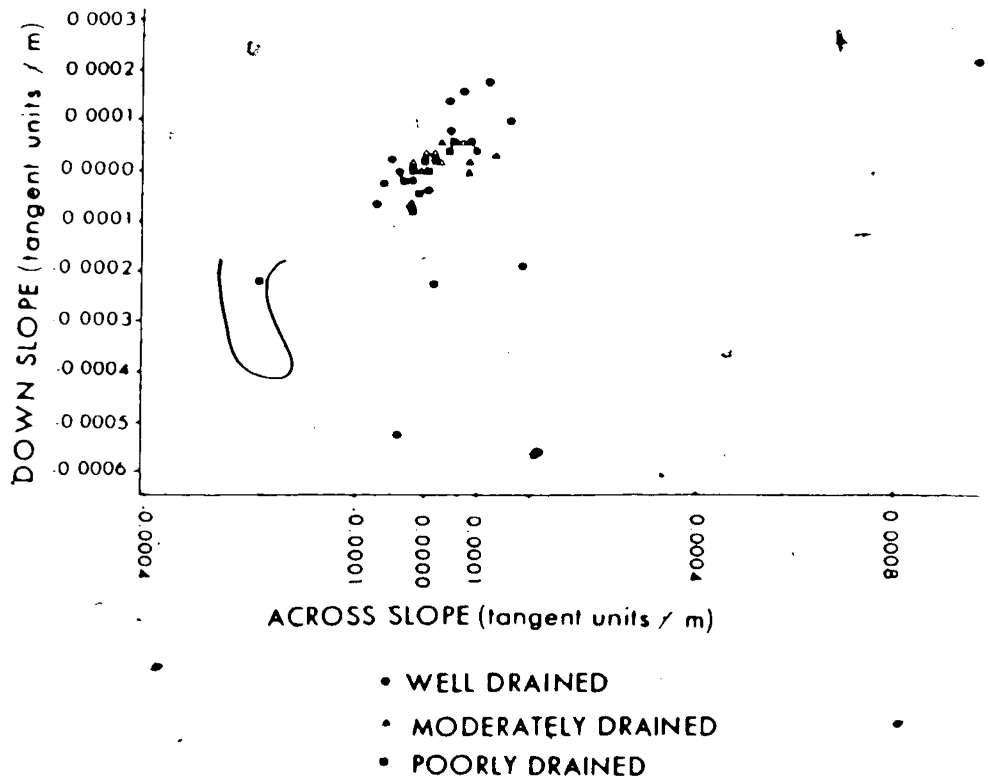


Figure 6.10. Distribution of down slope and across slope curvatures for the landscape drainage sites.

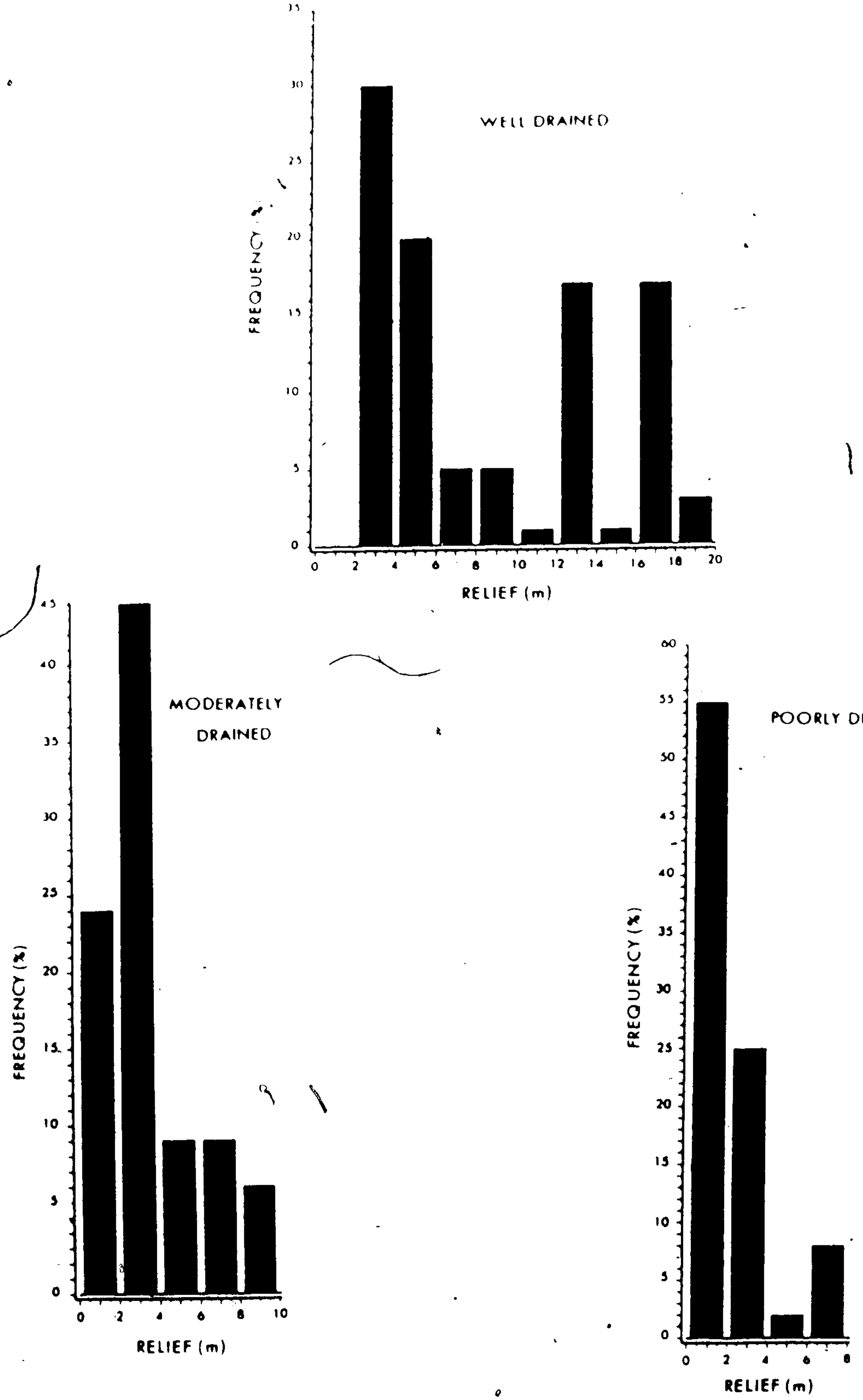


Figure 6.11. Histograms of the distribution of relief values for the landscape drainage sites.

were considerably higher. Figure 6.10 is a scatterplot of the two curvatures (across and down slope). These were plotted as a bivariate plot because, as discussed in Chapter 4, both were needed to define the landform shape. As can be seen by this scatterplot, there was a substantial degree of overlap. The tendencies were, however, for the better drained sites to be convex while the more poorly drained sites were concave. The distribution of the relief variable is shown in Figure 6.11. The well - drained sites had a very broad range of relief values from 2 metres to 20 metres. This range decreased, as did the overall values with a decrease in the landscape drainage.

To further evaluate the potential of each of the derivative and landscape variables to describe the landscape drainage properties at a site, linear discriminant analysis was used. In addition to the classification role of discriminant analyses, described in Chapter 3, this procedure can also be used in an interpretive capacity (Klecka, 1980). Discriminant analysis is useful, in this latter capacity, in studying the differences between several groups, or classes, with respect to one or more variables. In the case of the drainage classes, discriminant analysis was used first to investigate the differences as defined by the DEM derivative and landscape variables separately and second with all of the variables combined. The results of the analyses are displayed as confusion matrices (Tables 6.1 to 6.5) which summarize the degree to which the variables explain the between group variation. The curvature matrix summarizes both the across and down slope variables combined. An examination of the confusion matrices, indicates that a single variable did not

Table 6.1

Confusion Matrix for Slope Variable.

		Into Class			
		well drained	moderately drained	poorly drained	total
From Class	well drained	10	8	2	20
	moderately drained	3	4	13	20
	poorly drained	2	4	14	20

Overall Accuracy = 28/60 = 47%

Table 6.2.

Confusion Matrix for Across and Down Slope Curvatures.

		Into Class			
		well drained	moderately drained	poorly drained	total
From Class	well drained	5	8	7	20
	moderately drained	3	13	4	20
	poorly drained	0	11	9	20

Overall Accuracy = 27/60 = 45%

Table 6.3

Confusion Matrix for Relief Variable.

From Class	Into Class: well drained	moderately drained	poorly drained	total
well drained	10	7	3	20
moderately drained	3	4	13	20
poorly drained	2	4	14	20

Overall Accuracy = 28/60 = 47%

Table 6.4

Confusion Matrix for Flow Path Variable.

		Into Class			
		well drained	moderately drained	poorly drained	total
From Class	well drained	15	4	1	20
	moderately drained	16	2	2	20
	poorly drained	5	7	8	20

Overall Accuracy = 25/60 = 45%

Table 6.5

Confusion Matrix for all DEM Derivative and Landscape Variables.

From Class		Into Class			
		well drained	moderately drained	poorly drained	total
	well drained	15	4	1	20
	moderately drained	5	12	3	20
	poorly drained	2	4	14	20

Overall Accuracy = 41/60 = 68%.

adequately describe the variation in landscape drainage. All of the variables individually described approximately 45% to 47% of the variation in the landscape drainage sites.

Table 6.1 is the confusion matrix using only slope values for the three drainage classes. A great deal of confusion existed between all three of the drainage classes. The class with the fewest omission errors was the one with the poorest drainage. The highest commission rate was with the moderately drained sites. The well - drained sites are poorly classified based solely on slope values with a great deal of confusion existing with the sites classed as moderately drained. The mean values for the slope variable for the three classes were 13.8% for well drained, 4.4% for moderately drained, and 3.2% for poorly drained sites. The standard deviations for the three classes were high, with 10% for the well drained class, 3.9% for the moderately drained and 3.2% for the poorest drained sites.

The confusion matrix for the curvature variables is presented in Table 6.2. Overall, classification accuracies were similar to those obtained for the slope variable. A higher degree of explanation was attained for the moderately drained sites than with slope gradients. The degree of explanation for the poorly drained sites was very low, with only a 45% accuracy. The well drained sites were poorly discriminated with the use of curvatures only. The mean curvature values were all in the category that Young (1972) termed as straight slope shapes. For the moderately and poorly drained sites, this was not surprising as the slopes in these categories were very low so that any change in the slope values would be small over a grid cell distance. The mean values for the poorly drained sites indicated that a slight concavity did occur both in the up slope as well as the down slope directions. This is indicative of the broad shallow basins into which the moisture tends to drain. The moderately drained sites, also found in predominately low slope environments, were characterized by convex slope profiles. The well drained sites were characterized as being straight in the down slope direction and convex in the across slope direction. The high standard deviation of the well drained sites could be attributed to the greater definition of the topography in the high slope environments.

The landscape variables defined by relief and flow paths yielded overall classification accuracies of 47% and 45% respectively. The relief variable provided a classification matrix (Table 6.3) which was identical to that yielded by slope. The mean values were 9.2 metres for well drained sites, 3.1 metres for the moderately drained sites and 2.3 metres for the poorly drained sites. The standard deviations varied similarly, with values of 6.1 metres, 2.3 metres and 1.9 metres, respectively. The final variable, flow path, resulted in a slightly different confusion matrix (Table 6.4) than the other variables described above. The matrix indicates that the well drained sites were described quite well by this variable while the other two classes are not. The mean values for the flow paths indicated that the well drained sites had a value of 4.6 flow units, with a standard deviation of 4.5 flow units. This is to be expected as most of the well drained sites were located close to the crest of the slopes. The moderately drained sites yielded a mean flow path value of 9.0 flow units, but with a high standard deviation of 19.8 flow units. The poorly drained sites had a mean flow path value of 23.6 flow units and a standard deviation of 26.8 flow units. The flow path distribution had a high degree of positive skewness (Figure 6.7), which increased with the decrease in the landscape drainage (Figure 6.9), hence the high standard deviation.

The preceeding analysis indicates that the degree of explanation of drainage obtained from the examination of the individual variables was low. Therefore a multivariate approach, combining all of the variables, was performed. All five of the derivative and landscape products were combined in a discriminant analysis. The resulting confusion matrix is displayed in Table 6.5. The overall percentage of drainage sites which were correctly classified using all of the variables increased to 68%. A Chi-square test (Place, 1985) was used to evaluate whether the change in classification accuracy between the two data sets was statistically significant (Table 6.6). The increase in classification accuracy of the multivariate analysis compared to the univariate approach was found to be statistically significant based on this test. A summary of the mean and standard deviation values for the derivative and landscape variables is displayed in Table 6.7.

Table 6.6

Summary of Test for Statistical Significance of Change in Classification Accuracy Between Single Variable and Multivariate Analysis.

H_1 = There is a statistically significant difference between the results of the univariate and multivariate analysis.

H_0 = The difference in accuracy between the two classifications is not statistically significant.

χ^2 = 3.84 $(p<0.05)$ df=1

	success	fail	total
single variable	28	32	60
multiple variable	41	19	60
total	69	51	N=120

$$\chi^2 = \frac{120(|(28*19)-(41*32)|-(120/2))**2)}{69*51*60*60}$$

$= 4.91$

H_0 is rejected (4.91>3.84), therefore there is a statistically significant difference between the two classifications.

Table 6.7

Summary of Drainage Class Statistics

	Mean	Standard Deviation
Well Drained		
Slope	13.[illegible]%	10.0%
Relief	9.[illegible] metres	6.1 metres
Curvature(a)	0.00006	0.00016
Curvature(d)	0.000007	0.00018
Flow Path	4.6	4.[illegible]
Moderately Drained		
Slope	4.3%	3.2%
Relief	3.1 metres	2.3 metres
Curvature(a)	0.000015	0.000031
Curvature(d)	0.000015	0.000028
Flow Path	9.0	19.9
Poorly Drained		
Slope	3.4%	3.2%
Relief	2.3 metres	1.9 metres
Curvature(a)	-0.000018	0.000054
Curvature(d)	-0.000025	0.000070
Flow Path	23.5	26.8

6.5.4 Discussion

The 68% overall classification accuracy obtained using the digital variables must be viewed as being of significance when considered with other factors. The first of these factors is that landscape drainage is not only controlled by the geometry of the landscape, but also influenced by edaphic factors such as soil texture and structure. The second factor is the scale of the landforms present in relation to the resolution of the source document for the DTM. This point was discussed in Chapter 5 which presented the comparison of the two data sources. The contour interval of the source document was approximately 30 metres (100 feet). Some of the land forms in the study area had an elevation amplitude which was less than 30 metres, hence the processes that were influenced by these small forms would not be represented by the DEM and its derivative products.

The trends in the groupings of derivative and landscape variables, shown in Table 6.7, summarize numerically the qualitative criteria used in the photo interpretation. This was significant in that a categorization of landform drainage was in fact possible based on landform geometry. In articles by Heerdegen and Beran (1982) and Burt and Butcher (1985) it was concluded that the spatial distribution of landscape moisture did not relate well to single variables, such as curvature or slope, or to indices based on two variables such as slope and basin area. This point was in agreement with the findings of this study. This research unlike the two studies cited above, however, recognized that a multivariate approach must be taken in describing the processes operating. It was not reasonable to assume that a single variable, such as slope, could adequately account for the variations in a process across the entire spectrum of landscapes. In some of the lower slope areas, drainage variations were caused, as suggested by the findings of this study, by the shape of the landscape or the position on the slope. A multivariate approach therefore was necessary to account for the changing influence of the landforms' geometric properties.

The addition of the flow path information to the modelling contributed substantially to the explanatory power of the model. As mentioned earlier, flow paths have been traditionally used to define drainage channels. Only a few studies have recognized that the flow path values may be useful in other ways (Eyton, 1987). In this study, the flow path values were used as a measure of the extent of the up slope region that affects a particular site. This concept of up slope area differed from that described by previous workers (see, for example, Burt and Butcher, 1985) in that it recognized that the entire drainage basin up slope of the site did not contribute to the process operating at a site. The use of the flow path variable recognizes that there was an information transfer down slope and that a process operating at a site could not be adequately described by using only point measures. The failure of past studies to incorporate this variable may account in part, along with the use of a univariate approach, for the low success rates in defining processes through the use of DEM derivative products.

6.5.5 **Comparison With Field Checked Samples**

To establish the validity of the modelling process and results discussed above it was necessary to relate the drainage sites which were interpreted from aerial photographs to sites which had been interpreted based on field observations. A total of 18 sites were checked in the field. The location of each of these field sites was plotted on a 1:50000 map sheet and digitized. The same process used with the photo interpreted sites was carried out on this second set. The location of each point was converted into a pixel address and the relevant information extracted from the digital data base. To compare these sites with the photo interpreted ones, the drainage classes were summarized into three classes corresponding to the categories previously used. The poorly and imperfectly drained sites were assigned a Class 3, rating while the excessively and well drained sites Class 1. All of the remaining sites were grouped into Class 2.

Discriminant analysis was performed using slope, relief, flow path and curvature variables. The results are presented in Table 6.8. The overall classification accuracy of the control sites was tested against the accuracy obtained for the 60 photo interpreted sites using a standard Chi-square analysis (Place, 1985). The null hypothesis for this test was that there was no significant difference in results obtained for the two sets of samples (Table 6.9). With the acceptance of the null hypothesis two major conclusions could be made. The first was that the photo interpreted sites were valid with respect to landscape drainage, and second that the DEM derivatives and landscape variables

Table 6.8

Confusion Matrix for Field Checked Landscape Drainage Sites

From Class	Into Class: well drained	moderately drained	poorly drained	total
well drained	5	0	0	5
moderately drained	0	8	2	10
poorly drained	0	1	2	3
total	5	9	4	18

Table 6.9

Summary of Test for Statistical Significance of Results Comparing Photointerpreted Sites and Field Checked Sites.

H_1 - There is a difference between the results of the field surveyed sites and the interpretation from aerial photographs.

H_0 - There is no statistically significant difference.

P = 6.61 ($p<0.01$) df=1

	success	fail	total
field checked	15	3	18
photo-interpreted	41	19	60
total	56	22	N=78

$$\chi^2 = \frac{78((|(15*19)-(41*3)|-(78/2))**2)}{19*56*22*60}$$

$$= 0.83$$

H_0 is rejected (0.83<6.61), therefore there is no difference between the classification results.

described, to a large degree, the landscape moisture distribution. The results obtained from this analysis indicated that the null hypothesis could be accepted ($p < 0.01$) and that there was no statistically significant difference between the two sets of samples.

6.6 Summary

From the preceeding analysis and discussions it was concluded that a description of landscape drainage conditions at a site can be approximated using only the geometric properties of the landscape surface. These variables include traditional ones such as slope, as well as less frequently used variables such as directional curvature measures and relief. These variables all refer to the geometric conditions at a point, while the flow path data set was used to reflect the conditions up slope from the sample point.

It must be emphasized that the landscape drainage conditions are only approximated by the geometric variables. As mentioned previously, edaphic factors such as soil texture and structure also play important roles in determining the moisture conditions at a site. A further influence on landscape moisture distribution is that of bedrock geology. Toth (1962) described the process by which groundwater was introduced into the landscape through seepage. In areas where this type of process occurs, poorly drained sites can be found on sections above the base of the slope. A final influence on the amount of moisture in the soil profile which has not been considered in this analysis is that of the removal of water by the vegetation itself through evapotranspiration.

CHAPTER 7

Integration of Ancillary and Multispectral Scanner Data

7.1 Introduction

Previous chapters in this dissertation have established that; first, information from MSS data alone are insufficient to adequately define forest cover types in the Boreal ecozone, second, that ancillary landscape drainage information may be useful to improve the MSS based classification, and third, that landscape drainage can be defined, to a large degree, by the geometric properties of the terrain. The integration of the ancillary landscape drainage information with the MSS data and evaluation of the resulting classification accuracy changes are discussed in this chapter.

7.2 Approaches to Integrating Ancillary and Remotely Sensed Data

Three approaches to the integration of remotely sensed and ancillary data have been suggested (Hutchinson, 1982). These approaches include preclassification sorting of the MSS data, direct integration of data and postclassification sorting. The first and third approaches are typically those associated with Geographic Information Systems (GIS) which contain coverages of classified data. The second approach, direct integration of data, is most commonly used in conjunction with image processing systems. This latter procedure usually involves the incorporation of additional bands, or channels, of information with the remotely sensed data during the classification process. Most statistical classification techniques such as maximum likelihood or linear discriminant analysis assume that the input data sets are ratio or interval scaled. This precludes the use of previously classed data sets such as soil or slope classes which are nominally and ordinally scaled. As a result, studies using these data types as ancillary data rely on the pre- or postclassification sorting routines.

In preclassification sorting, the remotely sensed data are stratified according to some criteria. Boresjo (1984) outlined areas of poorly drained terrain and classified only those areas contained within these polygons. Satterwhite et al. (1984) interpreted the vegetation cover stratified on the basis of geomorphic units.

The majority of studies have used postclassification sorting to enhance remotely sensed classifications. Pettinger (1982) found that discrimination of wet and dry meadows was not possible through the use of MSS data alone. The meadow class, originally based on MSS data, was reclassified based on the location within upland or wetland zones. The studies by Niemann, Langford, and More (1984) and Niemann and Langford (1984) also used a postclassification sorting approach.

As discussed in Chapter 5 of this dissertation significant problems caused by generalization may be associated with the introduction of preclassified data into a classification of MSS data. To overcome these problems, ancillary data can be introduced directly into the classification process. This has only been attempted in a limited number of cases as it requires that the ancillary data be continuous and not nominally or ordinally scaled as is usually the case with GIS-type data bases.

Strahler (1980) and Strahler et al. (1980) found that classifier modification through the use of prior probabilities is an appropriate technique for integrating MSS and ancillary data. Strahler (1980) suggested that the posterior probability rule of a Bayesian-type classifier may be altered to reflect the probability of a sample point belonging to a class based on the ancillary data (for example, slope, aspect, elevation). The weighting factor assigned to each pixel was calculated based on the ancillary information. For example, there is an elevation range in which subalpine fir will grow. Areas outside of this range will have a progressively lower probability of encountering this species. This type of probability stratification was used by Strahler to help separate some of the spectrally indistinct species. The use of this probability modification approach has a definite advantage according to Strahler in that discrete, or ordinal, data may be incorporated directly into a classifier which would normally require a continuous data set.

An approach which reflects a more direct integration of the data sources is the use of all of the data sets in the classifier. Tom and Miller (1980) modelled forest productivity integrating LANDSAT MSS, slope, aspect, elevation and solar radiation data, through a linear discriminant analysis, to classify forest site quality. By using a stepwise approach they produced a classification which was based only on those variables which contributed significantly to the final results. They noted that site quality based on MSS data alone accounted for classification accuracies of 43%. Combining landscape and MSS variables, accuracies were increased to over 95%. Franklin (1985) used a similar stepwise linear discriminant analysis to study the effects of adding topographic variables to LANDSAT data to improve classification accuracies. Through a stepwise approach overall classification accuracies were increased from 46% to 66%.

7.3 Integration of Ancillary Data

The integration of ancillary information with the MSS data sets directly into the classification process was chosen as the most appropriate for use in this study. The use of pre- or postclassification sorting assumes **a priori** knowledge of the relationships between the ground cover and the MSS and ancillary data. Without this understanding of these relationships it is more useful to investigate the existence of naturally occurring clusters. This can better be accomplished with the use the direct integration of data sets, where the data have not been previously classed. The next decision was whether or not to integrate the data sets using traditional methods, as used by Tom and Miller (1980) and Franklin (1985). The classification they adopted used a Bayesian classifier, such as maximum likelihood or discriminant analysis, integrating all of the data sets into predefined classes. There was no allowance for the definition of naturally occurring clusters of data. In many instances the nature of the association of the data variables may not be known and, therefore, the exact number of naturally occurring clusters unknown. It is not reasonable to assume that the naturally occurring clusters correspond to the classes defined by another classification procedure. The desired approach, therefore, should allow for the analysis to define the classes while the classifier assigns the pixels to the most appropriate class.

7.4 Procedure

Although the results of the initial MSS - based classification were not satisfactory with respect to the desired classes, the six clusters generated did represent naturally occurring groupings of spectral data. The main problem encountered in the MSS based classification was that the six initial clusters resulted in too few distinct cover classes. A second problem was that of misclassification based on spectral similarity. The aim of the addition of the ancillary data based on the results of the drainage analyses was to subdivide these original clusters into smaller subclusters which represented the desired cover types defined in the original classification key.

With this aim in mind, the training sites within each of the initial MSS clusters were reclustered using the five DEM derivative and landscape variables used to define the drainage classes in Chapter 6. Clusters 1 and 5 were omitted from this reclustering because they were considered too small, with 19 and 36 sample points, respectively, to yield meaningful results. Sarle (1983) suggested that the maximum number of clusters generated from a data set be 10% of the original number of input data points. The remaining four MSS based groups were clustered using the same methodology as used for the MSS data. The CCC values were plotted against the number of clusters and the candidate number of clusters interpreted. These plots are presented in Figures 7.1a through d. Figure 7.1a represents the plot for cluster number [illegible] Nine subclusters were chosen to represent this cluster. This number was chosen over seven subclusters, because it was the point where the line showed a sharp break in the slope following a rise. For cluster number 3 (Figure 7.1b), three subclusters were chosen. In this case, the five subclusters were not chosen, because the line between three and five subclusters was almost straight. Only at 3 subclusters was there any indication of a peak in the line. Ten subclusters were not chosen because, although there was a substantial decrease prior to a rise there was no peak in the pattern. Rather, the curve was almost flat, with a slight rise following that point. Three subclusters were chosen from MSS cluster number 4 (Figure 7.1c) based on the same reasoning as for the second cluster. For MSS cluster 6 (Figure 7.1d) four subclusters were chosen. This point represented the highest point on the curve, in addition to being preceeded by a rise and followed by a fall in the CCC values. The peak was, however, not pronounced.

The reclustering, in addition to the other two clusters which were left intact, resulted in a total of 21 subclusters. A

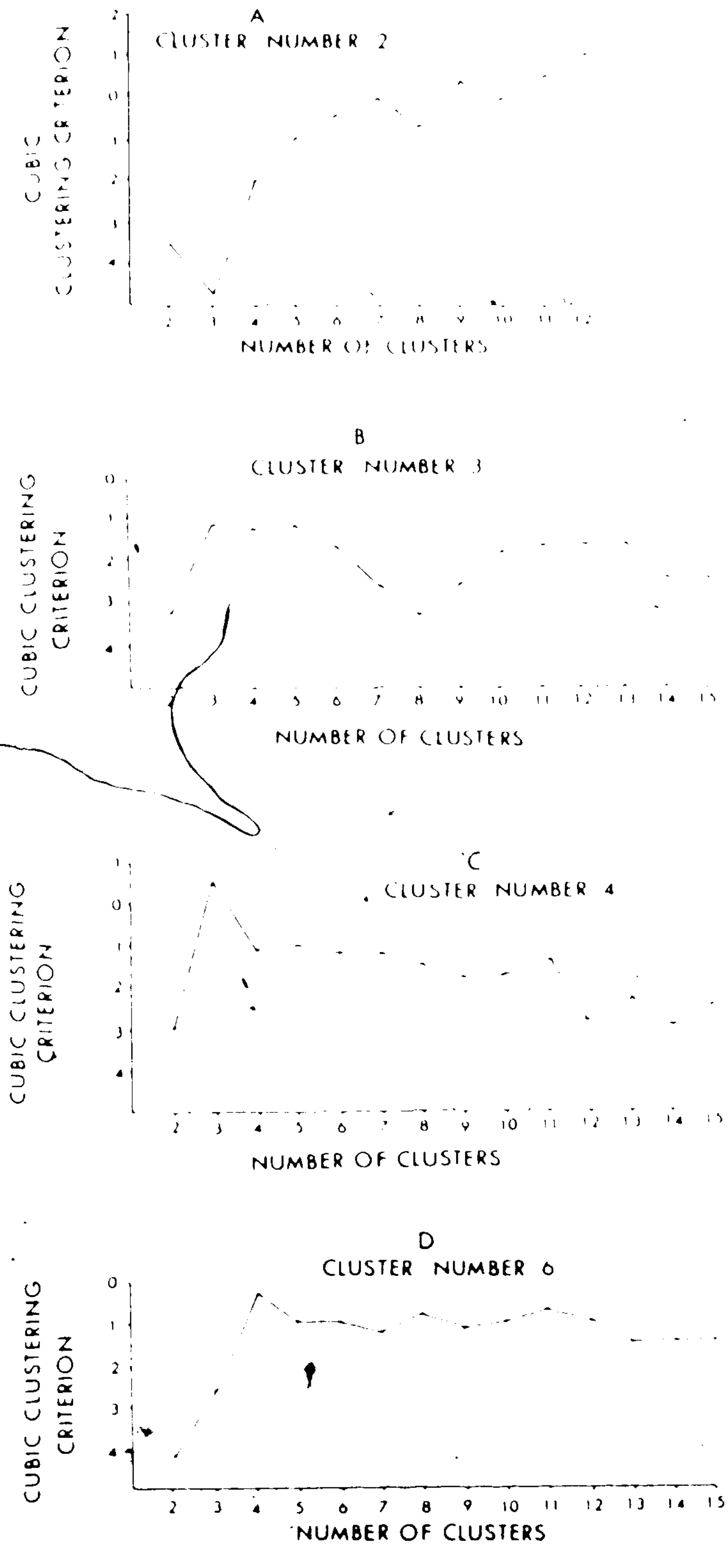


Figure 7.1. Cubic clustering criteria versus number of clusters for MSS defined clusters.

crosstabulation of the original cover class labels and the new cluster number was performed (Table 7.1). An initial labeling of the clusters was made based on the distribution of the original cover classes within each of the new groups. A second labeling pass was made based on the clustering of the ancillary data coverages. The drainage interpretation discussed in Chapter 6 and summarized in Table 6.7 was used to label the classes. A summary of the characteristics of each class is presented in Table 7.2. Those clusters with high slopes, high relief, low flow path numbers, and straight to convex slope shapes were considered well drained (numbers 2, 7, 9, 13, 16, 17 and 19). Clusters which had low slope gradients, low relief, high flow path numbers and straight to concave slope shapes were considered poorly drained (numbers 4, 5, 8, 10, 12, 15 and 21). The moderately drained sites had generally low slope gradients, low relief , and low to moderate flow path numbers. The remaining clusters were grouped into this category. The only exception was cluster number 1 which was initially interpreted as being open soil and so was not restricted to a particular drainage class. The final labeling of the clusters is summarized in Table 7.3.

The separation of the 21 clusters was analyzed using discriminant analysis. The confusion matrix indicating the degree of separation is shown in Table 7.4. The overall separation of the 21 clusters is 90%. Several of the 21 clusters were labelled similarly. When the clusters which were similarly labelled were merged, confusion dropped so that the resulting separation rose to 92%.

Table 7.1

Results of Cross Tabulation Between Cluster Number and Cover Class for Training Sites.

Cluster Number

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Upland Conifer	8	6	4	10	3	4	0	0	32	0	122	0	49	92	9	5	0	25	0	19	0
Mixed Conifer	0	0	0	0	0	0	0	0	5	0	105	0	10	19	0	0	2	3	0	0	0
Mixed Conifer/ Deciduous	1	3	9	0	0	0	9	0	0	23	6	32	0	18	16	8	4	34	5	18	0
Deciduous	0	2	5	0	19	19	11	0	1	0	13	0	5	2	0	2	19	10	5	10	0
Treed Muskeg	21	0	21	0	1	0	0	0	1	17	20	0	8	8	9	0	1	35	0	0	0
Deciduous Muskeg	4	0	2	5	10	10	0	13	0	0	6	13	0	0	4	0	10	5	0	10	7
Open Soil	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0

Table 7.2

Summary of the Descriptive Statistics for the 21 Clusters.

Variable	Mean	Standard Deviation
Slope (tan units)	0.091	0.086
Flow Path	6.6	39.5
Curvature(D)	-0.00004892	0.00012123
Curvature (A)	-0.00000477	0.00006454
Relief (m)	6.07	5.30
B4	37.4	3.9
B5	45.0	6.2
B6	58.1	6.8
B7	59.3	8.3

Cluster 2

Slope (tan units)	0.149	0.033
Flow Path	4.4	1.7
Curvature (D)	0.00002500	0.00009443
Curvature (A)	0.00003827	0.00003097
Relief (m)	9.2	1.2
B4	32.0	3.1
B5	33.8	7.3
B6	73.4	3.4
B7	83.5	4.8

Cluster 3

Slope (tan units)	0.026	0.019
Flow Path	5.4	3.3
Curvature (D)	0.00000834	0.00002831
Curvatreu (A)	0.00000366	0.00002491
Relief (m)	1.8	1.0
B4	31.5	2.0
B5	32.1	2.8
B6	73.6	4.0
B7	83.3	5.1

Cluster 4

Slope (tan units)	0.036	0.008
Flow Path	200.0	1.5
Curvature (D)	0.00001620	0.00002699
Curvature (A)	0.00002400	0.00002763
Relief (m)	2.82	0.43
B4	34.2	1.1
B5	34.6	2.8
B6	75.6	3.8
B7	84.2	4.2

Cluster 5

Slope (tan units)	0.059	0.021
Flow Path	32.2	8.7
Curvature (D)	-0.00002015	0.00004657
Curvature (A)	-0.00002808	0.00006343
Relief (m)	4.06	1.08
B4	30.0	2.3
B5	29.1	2.9
B6	74.8	4.3
B7	86.2	4.1

Cluster 6

Slope (tan units)	0.078	0.034
Flow Path	8.7	8.2
Curvature (D)	-0.00004389	0.00017294
Curvature (A)	-0.00001400	0.00008095
Relief (m)	5.81	1.02
B4	30.4	2.6
B5	29.0	4.5
B6	72.8	3.2
B7	84.4	5.5

Cluster 7

Slope (tan units)	0.263	0.035
Flow Path	2.5	0.5
Curvature (D)	0.00005200	0.00006224
Curvature (A)	0.00002050	0.00007197
Relief (m)	18.1	1.5
B4	27.8	0.4
B5	27.8	0.9
B6	74.3	3.1
B7	86.1	3.8

Cluster 8

Slope (tan units)	0.024	0.004
Flow Path	134.1	15.4
Curvature (D)	-0.00000700	0.00003455
Curvature (A)	-0.00000175	0.00001549
Relief (m)	1.5	0.1
B4	31.3	1.0
B5	30.4	2.9
B6	74.8	2.5
B7	84.6	5.1

Cluster 9

Slope (tan units)	0.125	0.065
Flow Path	4.3	4.3
Curvature (D)	0.00003430	0.00020181
Curvature (A)	0.00001571	0.00013679
Relief (m)	8.9	2.8
B4	27.6	1.1
B5	26.4	1.2
B6	47.6	2.8
B7	48.8	4.1

Cluster 10

Slope (tan units)	0.019	0.003
Flow Path	156.9	54.4
Curvature (D)	-0.00000567	0.00000598
Curvature (A)	-0.00000753	0.00000400
Relief (m)	1.5	0.2
B4	28.9	1.7
B5	27.9	1.0
B6	48.0	2.3
B7	49.3	2.6

Cluster 11

Slope (tan units)	0.039	0.022
Flow Path	10.8	15.1
Curvature (D)	0.00000067	0.00005734
Curvature (A)	-0.00000253	0.00004848
Relief (m)	2.8	1.3
B4	28.1	1.1
B5	27.4	1.4
B6	47.9	2.7
B7	49.2	3.2

Cluster 12

Slope (tan units)	0.022	0.004
Flow Path	147.7	16.5
Curvature (D)	-0.00000333	0.00000208
Curvature (A)	-0.00001167	0.00001002
Relief (m)	1.63	0.412
B4	30.3	3.2
B5	28.3	0.6
B6	57.0	4.0
B7	59.0	2.5

Cluster 13

Slope (tan units)	0.107	0.037
Flow Path	4.9	4.3
Curvature (D)	-0.00000218	0.00013589
Curvature (A)	0.00001577	0.00011206
Relief (m)	7.4	1.4
B4	28.8	1.6
B5	27.9	2.0
B6	55.6	2.5
B7	59.9	3.5

Cluster 14

Slope (tan units)	0.037	0.021
Flow Path	5.0	4.2
Curvature (D)	-0.00000110	0.00004449
Curvature (A)	0.00001202	0.00004907
Relief (m)	2.6	1.2
B4	28.9	1.3
B5	28.3	1.8
B6	55.6	2.8
B7	59.6	3.3

Cluster 15

Slope (tan units)	0.043	0.017
Flow Path	37.8	12.2
Curvature (D)	-0.00002305	0.00005817
Curvature (A)	-0.00003534	0.00006260
Relief (m)	3.2	0.9
B4	29.2	1.4
B5	28.4	1.8
B6	55.6	3.1
B7	59.2	3.6

Cluster 16

Slope (tan units)	0.181	0.075
Flow Path	3.6	2.7
Curvature (D)	0.00003680	0.00028029
Curvature (A)	0.00005293	0.00020924
Relief (m)	13.5	2.1
B4	28.6	1.8
B5	26.3	0.9
B6	54.3	2.8
B7	58.8	3.3

Cluster 17

Slope (tan units)	0.199	0.121
Flow Path	4.0	2.5
Curvature (D)	0.00001237	0.00004266
Curvature (A)	0.00001332	0.00009149
Relief (m)	14.2	8.0
B4	31.6	1.7
B5	30.8	2.3
B6	88.3	7.1
B7	107.1	9.5

Cluster 18

Slope (tan units)	0.026	0.0168
Flow Path	9.6	11.3
Curvature (D)	0.00001059	0.00004103
Curvature (A)	0.00000345	0.00003595
Relief (m)	2.0	1.0
B4	30.4	2.1
B5	30.0	2.8
B6	63.6	3.0
B7	70.0	3.6

Cluster 19

Slope (tan units)	0.189	0.050
Flow Path	4.3	1.5
Curvature (D)	0.00004408	0.00012345
Curvature (A)	0.00006569	0.00006283
Relief (m)	13.0	2.8
B4	29.0	2.3
B5	30.0	4.2
B6	63.3	2.1
B7	71.6	2.4

Cluster 20

Slope (tan units)	0.095	0.032
Flow Path	5.1	5.1
Curvature (D)	0.00000532	0.00010347
Curvature (A)	-0.00000041	0.00008964
Relief (m)	6.5	1.3
B4	29.7	2.4
B5	29.8	4.0
B6	63.8	3.1
B7	71.9	4.2

Cluster 21

Slope (tan units)	0.023	0.008
Flow Path	171.9	25.8
Curvature (D)	0.00000314	0.00001188
Curvature (A)	-0.00000300	0.00003663
Relief (m)	1.9	0.7
B4	33.1	1.5
B5	35.3	1,8
B6	65.6	4.0
B7	73.0	3.5

Table 7.3

Final Labelling of 21 Clusters.

Cluster #	Label
1	Open Soil
2	Mixed Coniferous
3	Mixed Coniferous
4	Muskeg (Deciduous)
5	Deciduous
6	Deciduous
7	Deciduous
8	Muskeg (Deciduous)
9	Upland Conifer
10	Muskeg (Coniferous)
11	Mixed Conifer
12	Muskeg (Conifer)
13	Upland Conifer
14	Mixed Conifer
15	Muskeg (conifer)
16	Mixed Conifer/Deciduous
17	Deciduous
18	Mixed Conifer/Deciduous
19	Mixed Conifer/Deciduous
20	Mixed Conifer/Deciduous
21	Muskeg (Deciduous)

Table 7.4

Confusion matrix showing separation of twenty-one clusters.

Number of Observations Classified into Cluster:

From Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	TOTAL
1	32 82.05	2	0	0	0	0	0	0	0	0	0	[illegible]	0	2	0	0	0	0	0	0	2	35
2	0	6 72.73	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[illegible]
3	0	0	38 92.68	0	1	1	0	0	0	0	0	0	0	0	0	0	0	[illegible]	0	0	0	4
4	0	0	0	5 100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
5	0	0	1	0	11 84.62	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13
6	0	4	0	0	2	30 81.08	0	0	[illegible]	0	0	0	0	0	0	0	0	0	0		0	37
7	0	0	0	0	0	0	6 100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
8	0	0	0	0	0	0	0	8 100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	8
9	0	0	0	0	0	0	0	0	58 92.06	0	0	0	0	0	0	5	0	0	0	0	0	63
10	0	0	0	0	0	0	0	0	0	15 100.00	0	0	0	0	0	0	0	0	0	0	0	[illegible]
11	0	0	0	0	0	0	0	0	3	0	268 91.76	0	2	[illegible]	12	0	0	0	0	0	0	292
12	0	0	0	0	0	0	0	0	0	0	0	3 100.00	0	0	0	0	0	0	0	0	0	3
13	0	0	0	0	0	0	0	0	0	0	0	0	90 96.77	0	0	[illegible]	0	0	[illegible]		0	93
14	0	0	0	0	0	0	0	0	0	0	4	0	10	148 87.57	0	0	0	[illegible]	0	0	0	69
15	0	0	0	0	0	0	0	0	0	0	0	0		0	37 97.37	0	0	0	0	0	0	38
16	0	0	0	0	0	0	0	0	0	0	0	0	[illegible]	0	0	[illegible] 80.00	0	0	0	0	0	15
17	0	0	[illegible]	0	0	3	2	0	0	0	0	0	0	0	0	0	3 66.41	0	0	0	0	5
18	0	0	3	0	1	1	0	0	0	0	0	0	0	4		0	0	[illegible] 85.41	0	3	0	23
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3 [illegible]	0	0	13
20	0	0	0	[illegible]	0		0	0	0	0	0	[illegible]	3	0	0	0	0	0		6 [illegible]	0	69
21	0	0	0	[illegible]	0	0	0	0	0	0	0	0	[illegible]	0	0	0	0	0	0	0	6 [illegible]	[illegible]

The test area was classified into the 21 classes using discriminant analysis. Nine input data coverages: the four spectral bands plus five ancillary data coverages, were used in the classification. The resulting classified image is shown in Figure 7.2 (see back pocket). The poorly drained units (muskeg and fens classes) mapped in Figure 7.2 were grouped into a single class, so that the image does not distinguish between the deciduous shrub and treed muskegs. The areas within the various muskeg classes were sufficiently small to make it difficult to distinguish them separately.

7.5 Error Analysis and Discussion

The identical data set used to test the accuracy of the classification of the MSS data was used to evaluate this second classification. The results of this analysis are presented in Table 7.5. The classes generated through this classification approach correspond much more closely to the desired classes which were discussed earlier in this dissertation in Chapter 3. The legend for this second classification differed substantially from the one obtained for the MSS based classification where only labels describing the very broad cover classes were possible. The results tabulated in the error matrix for the new classification also indicates an increase in overall accuracy. The classification accuracy of this second classification was 73%, an improvement of 15% over the initial MSS based classification. This change in accuracy was statistically

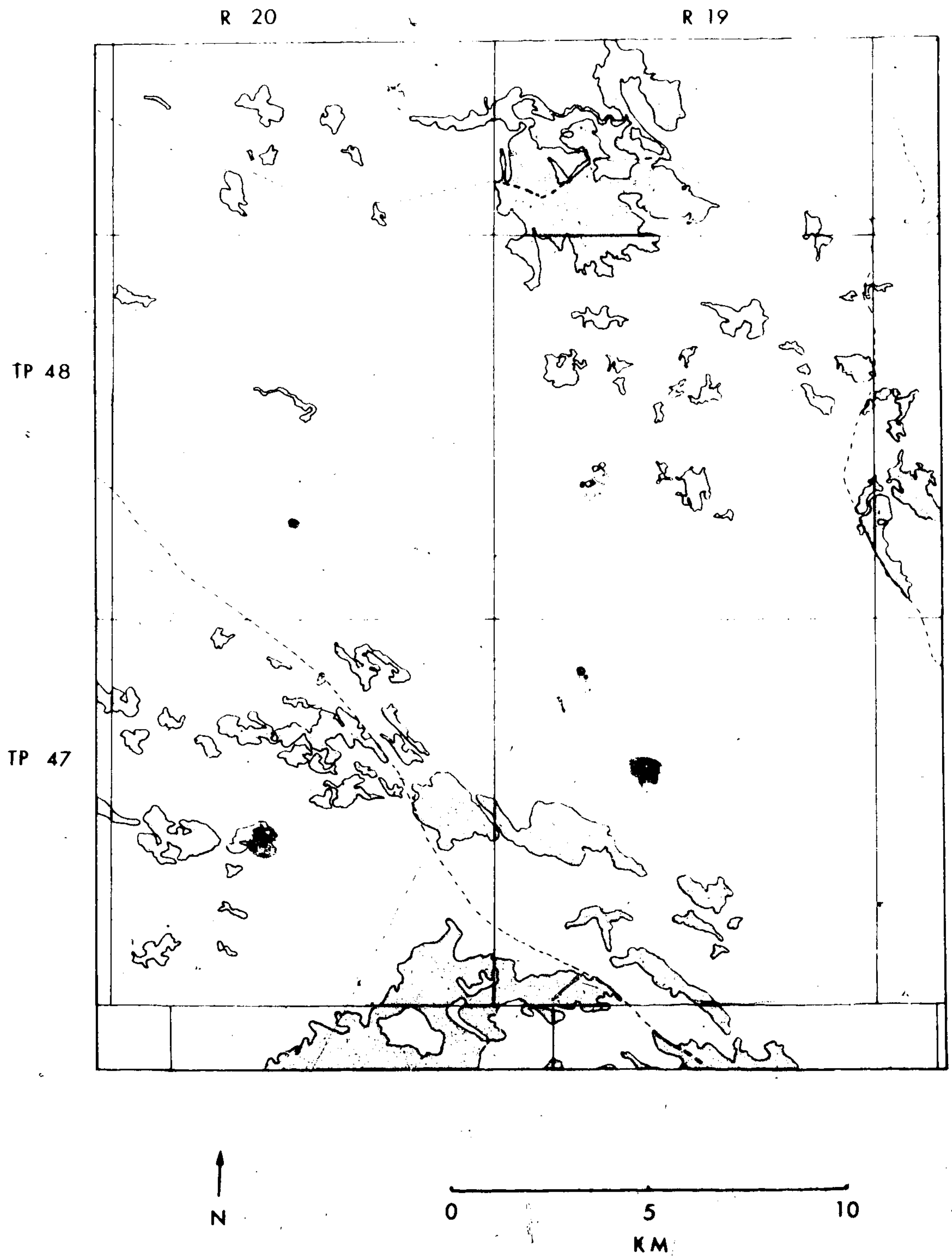


Figure 7.3. Single class map showing the distribution of mixed conifer stands in the study area. (source: Phase 3 Forest Inventory)

Table 7.5

Accuracy Matrix for Classification Based on MSS and Ancillary Data.

AERIAL PHOTOGRAPHS/PHASE 3

CLASSIFICATION	Closed coniferous	Mixed coniferous	Mixed coniferous/ deciduous	Deciduous	Muskeg /fen	Open soil
Closed conifer	120 ([illegible]) ([illegible])	8	4	8	4	0
Mixed conifer	14	13 (52) (4[illegible])	2	[illegible]	0	0
Mixed conifer /decid.	14	3	78 ([illegible]) (89)	4	[illegible]	0
Decid.	15	2	4	41 (35) (73)	1	0
Muskeg/ fen	2	0	2	3	18 (22) (60)	0
Open soil	1	0	0	0	0	5 ([illegible]) (100)

number of points
(%commission errors)
(%correct)

Total number of sample points=372
Total number accurate=276
Percent accuracy=73

significant ($p < 0.01$) based on the Chi-square analyses (Place, 1985) (see Table 7.6). Other figures showing the change in classification accuracy for the individual classes are also presented in Table 7.6.

The closed conifer class in this second analysis has been subdivided into two distinct classes, upland conifer (pine) and mixed conifer. The mapping accuracy of the upland conifer has not increased significantly from the MSS based classification, with an omission accuracy of 71% and a commission accuracy of 83%. The commission error rate dropped from 44% to 17%, which was due mainly to the decrease in the confusion with the mixed conifer/deciduous class. This mixed class, as described earlier, occurred mainly in a moderately to poorly drained environment, although aspen poplar and pine mixtures did occur in some well drained sites. There was also a substantial decrease in the confusion between the muskeg pixels and those representing the upland conifers.

The second cover class to be distinguished through the use of ancillary data is mixed conifers, that is, stands of pine and spruce found in areas of moderate drainage. The commission errors with pine were especially high, being caused by an overlap, not only with respect to the spectral characteristics, but also with regards to the environmental conditions in which these stands were found as pointed out in Chapter 6. This point is particularly evident when comparing the distribution of the mixed conifer class shown in Figure 7.3 and the classified image in Figure 7.2. One possible method of discriminating these stands may be with respect to age, where the spruce would be

Table 7.6

Comparison of Accuracies Obtained for the MSS-based Classification and MSS+Ancillary Based Classification.

Class	MSS Based (%)	MSS+Ancillary Based	Change in Accuracy (%)	χ^2
Closed Conifer	116 (69%)	120 (71%)	+4 (2%)	0.44
Mixed Conifer	NA	13 (48%)	NA	
Mixed Conifer/ Deciduous	37 (42%)	76 (86%)	+36 (44%)	35.6*
Deciduous	37 (69%)	40 (74%)	+3 (5%)	0.18
Muskeg/ Fen	NA	18 (60%)	NA	
Open Soil	5 (100%)	5 (100%)	0%	
Overall Accuracy	217 (58%)	276 (73%)	+59 (15%)	14.46*

* - change is statistically significant (χ^2= 6.61, p<0.01, df=1)

characteristic of a later successional stage than pine.

The mixed deciduous/coniferous class accuracy was notably increased both in terms of omission and commission. The original MSS based classification yielded omission errors of 58% and a commission error rate of 50%. These error rates decreased significantly ($p<0.01$) in the second classification (Table 7.6). The omission accuracy in the second classification was 14%, while the commission rate was 27%. The most notable decrease in confusion was with the pine class, as was discussed above. Of the 50% commission error associated with the original classification 32% were the result of confusion with the pine class. This was reduced to 13% with the addition of the drainage classes. A slight decrease in the confusion was found with the poorly drained and deciduous classes, while the confusion with the mixed conifer classes increased slightly from 1.3% to 2.8%. The omission errors decreased notably with respect to confusion with the conifer classes. The area where confusion was most likely to occur was with regard to the pine class.

There was an improvement in the classification accuracy of the pure deciduous class. However, this increase was not statistically significant. In the case of the MSS based classification, a commission error of 59% and omission error of 31% occurred. A slight reduction in the omission error percentage did take place with the inclusion of the drainage information. The most significant change was in the reduction of the commission error rate to 35%. Varying degrees of reduction in confusion with all of the other classes did occur. The most notable was with respect to the mixed conifer/deciduous class, from 17% to 6%, and the muskeg class, from 8% to 1%.

The final area of change was the muskeg/fen class. Whereas with the MSS based classification this class was not separable, the addition of the drainage information yielded a distinct class. For the sake of this evaluation the two muskeg classes, conifer and deciduous, were grouped into a single group. The reason for this was that the small size of the majority of these areas rendered them unresolvable with respect to distinct cover types. The resulting class had a 40% omission error and a 28% commission error. The greatest confusion with respect to commission was with the mixed deciduous/coniferous class (23%) and pine (13%). Figure 7.4 is a single class map illustrating the distribution of the muskeg areas in the study area. The visual comparison of this map with the classified image supports the statistics presented in Table 7.5 . Also noted through this comparison was that, while the poorly drained areas were quite well represented in the areas with relatively pronounced local and regional relief, they were missed almost entirely in the areas where the relief was low.

The confusion between all of these classes was partly the result of transitional areas where there was an overlap between the various types of forest stands, both in terms of spectral signatures, as well as environmental conditions for growth. More importantly, the contour interval of the DTM source document was not sufficiently small in some areas to define landforms which controlled the drainage process. As discussed earlier in this dissertation, the contour interval may be adequate in defining

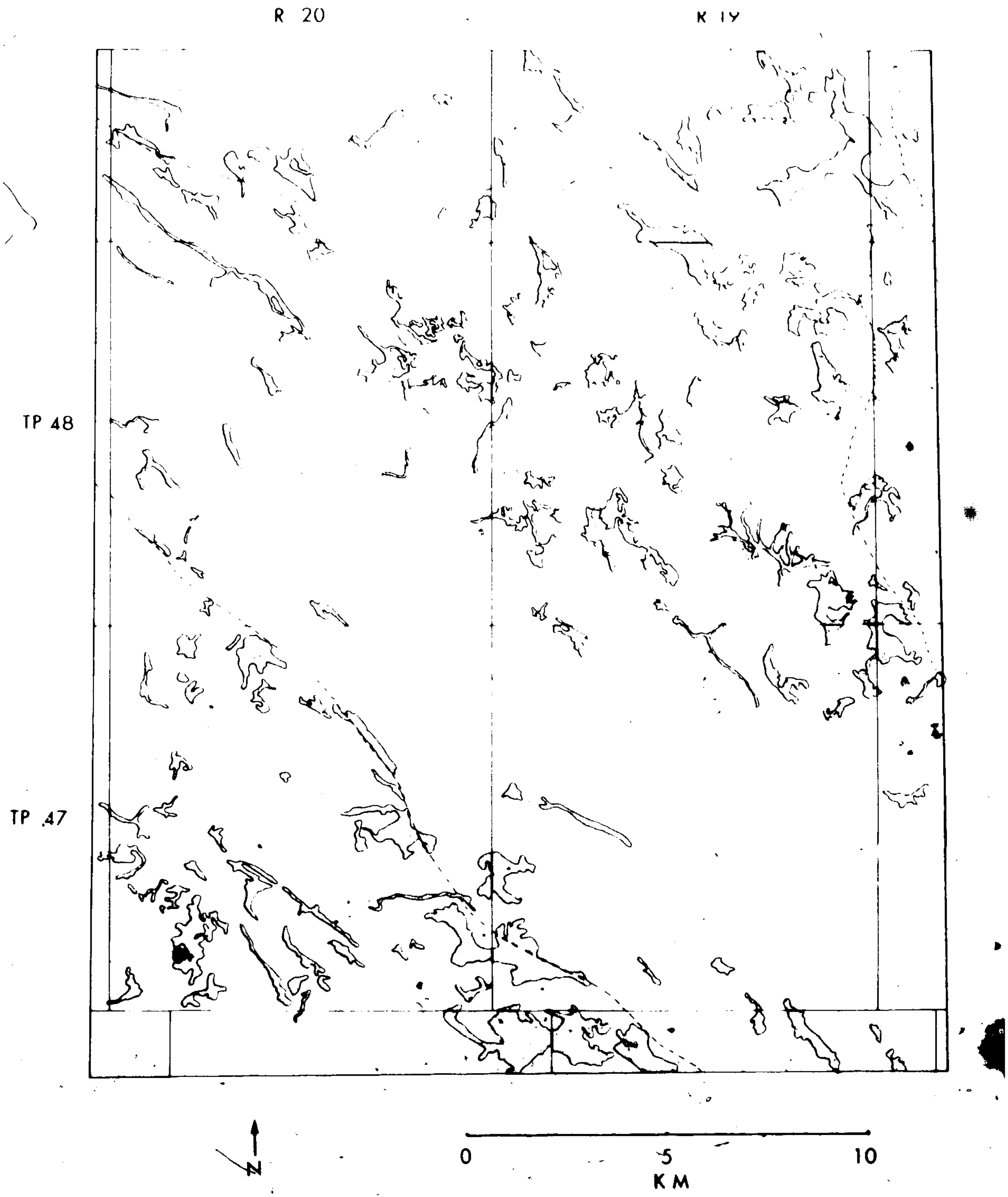


Figure 7.4. Single class map showing the distribution of organic soils in the study area. (source: Phase 3 Forest Inventory)

the variation in areas of high relief, but may be less so in describing some of the smaller landforms located in the low slope, undulating terrain. Landforms, and subsequently processes operating on them, will not be represented on the topographic map. These will, therefore, not be accounted for in the DEM and derivative products. The spectral values of the original 57 metre by 79 metre LANDSAT pixel will, however, reflect changes which occur at that scale. This problem may be accounted for by resampling all of the data sets to a coarser grid, or alternatively, using DTM sources which have a finer vertical resolution. The first solution may not always solve the problem as some of the landforms occurring in the areas of low relief are large in geographic area, but low in amplitude. The most useful solution would, therefore, be to use a DTM data source which has a contour interval which is less than 30 metres.

7.6 Summary

The results of the integration of the DEM - based drainage information, has significantly improved the classification accuracy of forest cover types in the study area. The classification has been broadened beyond the initial four classes obtained through the use of MSS data alone to include classes influenced by poor and moderate drainage. The initial hypothesis of this research, stated in Chapter 1, was that forest cover classifications based on MSS data could be improved through the addition of landscape drainage information, and that this drainage information could be derived through the use of digital elevation models. Although the results presented in this chapter have supported the validity of this hypothesis, there is a substantial room for improvement with respect to classification accuracy. There is also a need to investigate the potential contribution of higher resolution information for integration as ancillary data.

CHAPTER 8

Conclusions and Future Research

8.1 Conclusions

The objective in this research project was to investigate the integration of ancillary information with LANDSAT digital MSS data in an area of low relief, within the Boreal Mixedwood Forest ecozone. The aim in this integration was to improve the definition of forest cover composition through the use of digital classification techniques. It was recognized from a review of an extensive body of literature that the forest cover composition in these areas was not controlled by individual relief factors, as is the case in areas of high relief. Rather, the forest cover was influenced by a number of variables, including soil nutrients, landscape drainage and successional stage. Of the three, landscape drainage was the one variable most easily defined within a spatial context and was therefore easily integrated with the MSS data.

Two different sources of drainage information were available. The first source was an interpreted map product. This product was rejected based, on several factors. A first factor was the degree of cartographic generalization evident. A second factor was that the interpretation of the landscape units grouped several landform units into a single polygon so that the spatial representation of the process or landform was not preserved.

The rejection of this interpreted drainage map led to the investigation of a second potential source of information, the possibility of modelling landscape drainage through the use of DEM derivatives and landscape products. This modelling was carried out through the use of a multivariate approach, which was different from the published studies dealing with this type of drainage modelling. Recent studies have been attempts, with limited success, to describe landform drainage through the use of a single form variable or an index. It was found in this study that differing geometric properties had varying effects on the drainage depending on the position of the sample point within the landscape. The concept of flow paths was incorporated into the drainage modelling. In this concept it was recognized that there was an information transfer down slope. The moisture conditions at a site were affected by the slope shape, not only at the site, but also by the geometry of the terrain up slope from the site, along the path of steepest ascent from the sample point. This down slope information transfer has not been included in the previous landscape drainage modelling studies investigated. The result of this modelling (using linear discriminant analysis) was that a substantial amount of variation in landscape drainage could be explained through the use of terrain geometry. The classification error may have been due to a variety of factors including the scale of source documents and edaphic factors.

The forest cover was initially mapped through the use of MSS data alone. A combination of both unsupervised and supervised approaches to classification was used. Stands representing desired cover classes were identified and their locations digitized. It was recognized, however, that a continuum existed between pure forest stands, in terms of composition and other characteristics such as crown density. To account for this variability, training sites were clustered to isolate meaningful spectral clusters. The cluster analysis yielded six groups of spectral data which were assigned four cover class labels. An error check showed that the test area had a classification accuracy of 58%.

To improve on this accuracy, and to increase the classification resolution, ancillary data were introduced. The original six clusters were reclustered individually using the five DEM derivative and landscape variables identified in the drainage model as contributing substantially to the definition of the landscape drainage process. This reclustering process produced a total of 21 new clusters. The landscape drainage analyses yielded a description of the DEM variables within each drainage class, which were used to label the 21 clusters in the new classification. The test area was reclassified using discriminant analysis. A subsequent error check of this new classification indicated that a 73% accuracy was attained. In addition to the increased accuracy, the new classification procedure yielded a higher class resolution with an increase from an initial four classes in the MSS based classification to six.

Important steps have been taken in this research towards automating the mapping and inventory procedure of the Boreal forest resources. The first, and most notable, was the use of ancillary data in the digital classification of remotely sensed data. This approach was unique in that, where past studies have tended to use ancillary information to determine the position within the landscape relative to a small pixel neighbourhood (that is 3 X 3 or 5 X 5), this study viewed the DEM data as defining a landscape process which related the ancillary data to the entire study area.

A second notable contribution resulting from this research was in the area of process modelling based on the geometric properties of the landscape. Modelling of this type has been attempted in the past, with respect to soil erosion and landform drainage, but with minimal success. What has been recognized in this study and overlooked in many others is the multivariate nature of landscape forming processes. Also, the processes being modelled behaved differently at different points on the landform. Towards the summits of hillslopes the slope gradient was of greatest influence. At the base of the hill, where the slope gradients were gentler other variables such as slope length dominated.

A third contribution which was made in this dissertation was with respect to the nature of the mapped data, the mapping process, and the applicability of different data sources to integrate with remotely sensed data. It was noted that some mapping schemes, by grouping landforms and processes into single classes, tend to mask, or under-represent the spatial variability of the information which is being presented.

This final point leads to a comment regarding scale of the varying data sources in relation to the LANDSAT MSS data, and the implications for present and future high resolution sensors. The low classification accuracy achieved through the use of digital MSS data alone was not caused exclusively by the low spatial and spectral resolution of the sensing systems. Much of the error was caused by the use of spectral data alone in the classification of cover types. Several investigators in the past have sought to improve the classification accuracy through the increase in number of variables used in the classification process. With this research project the addition of ancillary data did improve the classification results, even higher accuracies could have been obtained given a finer contour interval of the topographic map source. If this was indeed the case, then the integration of ancillary information with higher resolution remotely sensed data may be of questionable use given current data sources.

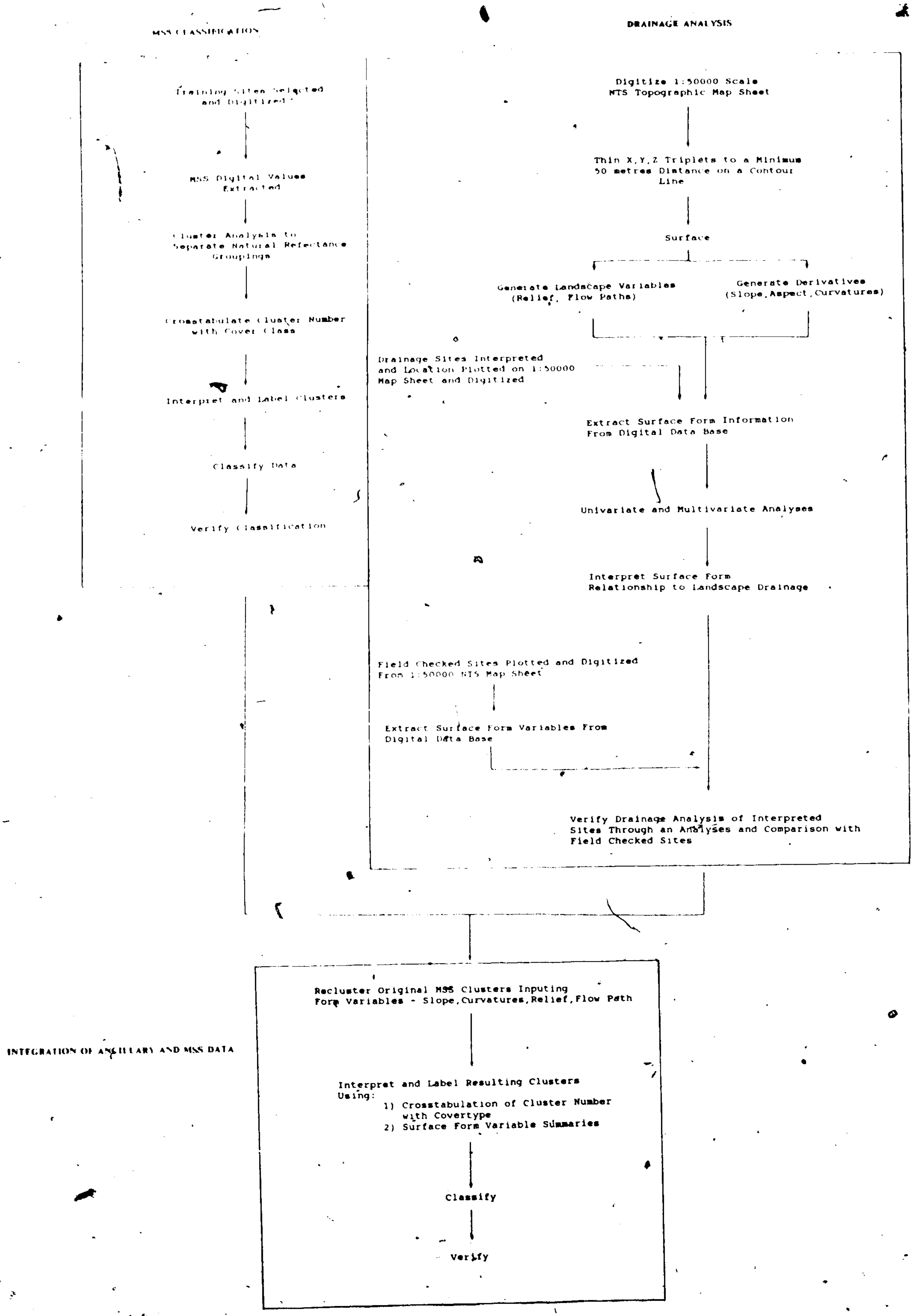
8.2 Future Research

There were a number of areas which this dissertation dealt with which can be identified as requiring further work. The first point is that of drainage modelling. A field oriented research program is necessary to more fully relate the drainage conditions at a point to the geometric properties of the landscape. Whereas the research reported in this dissertation used indicators, such as vegetation patterns and location within the landscape, to speculate on the drainage conditions, it will be important to relate numerically the drainage conditions at a site to the geometric properties. A second point regarding drainage modelling is the need to further develop DEM derivative products and landscape coverages which describe the landforms. One example is an extension to the flow path variable to relate the area of the pixel which contributes flow. For the actual area to be included, the slope of the surface must be considered as well. Additional research is necessary to investigate the effect higher resolution DTM data, especially with respect to elevation, would have on the classification accuracies. This problem of low height resolution was identified as being significant at several points in this dissertation.

The model results presented in this dissertation only reflect the characteristics on the study area chosen. It is necessary that the modelling proceedures which were used in this research be applied to other sites within the Boreal forest as well as other environments. The validity of these models and modelling techniques can be established by their application to other areas.

Further research is required in an area identified in Chapter 5 of this dissertation. Use of DEM derivative products, such as slope, aspect, and curvature, should be possible wi[illegible] more traditional methods of land evaluation. The interpretation methods which are used to produce these products may not be directly compatible with the digital DEM products, however. Investigations into the best methods of aggregating the data sets used and whether or not other information, such as frequency, may be of use in mapping exercises, should be undertaken as well.

Our understanding of the nature of the processes operating on the earth's surface and their effects on the vegetation cover are rudimentary. The research presented in this dissertation has only


MSS CLASSIFICATION
Training Sites Selected and Digitized
MSS Digital Values Extracted
Cluster Analysis to Separate Natural Refectance Groupings
Crosstabulate Cluster Number with Cover Class
Interpret and Label Clusters
Classify Data
Verify Classification
DRAINAGE ANALYSIS
Digitize 1:50000 Scale NTS Topographic Map Sheet
Thin X,Y,Z Triplets to a Minimum 50 metres Distance on a Contour Line
Surface
Generate Landscape Variables (Relief, Flow Paths)
Generate Derivatives (Slope,Aspect,Curvatures)
Drainage Sites Interpreted and Location Plotted on 1:50000 Map Sheet and Digitized
Extract Surface Form Information From Digital Data Base
Univariate and Multivariate Analyses
Interpret Surface Form Relationship to Landscape Drainage
Field Checked Sites Plotted and Digitized From 1:50000 NTS Map Sheet
Extract Surface Form Variables From Digital Data Base
Verify Drainage Analysis of Interpreted Sites Through an Analyses and Comparison with Field Checked Sites
INTEGRATION OF ANCILLARY AND MSS DATA
Recluster Original MSS Clusters Inputing Form Variables - Slope,Curvatures,Relief,Flow Path
Interpret and Label Resulting Clusters Using:
1) Crosstabulation of Cluster Number with Covertype
2) Surface Form Variable Summaries
Classify
Verify

scratched the surface of the problem. Rather than answering all of the questions, the findings of the study have acted as windows whereby efforts may be focused on more specific problems.

REFERENCES

Beaubien, J. (1979) Forest type mapping from LANDSAT digital data. Photogram. Eng. and Remote Sensing. V.45 1135-1144.

Bentz. J., D.Brierley, S. Nelson, S.Robertson, R. Wehrhahn (1985) Integrated Resource Inventory of the Coal Branch Study Area. Alberta Energy and Natural Resources Report #.T/83-No.1

Bernstein,R., D.G. Ferneybough (1975) Digital image processing. Photogramm. Eng. and Remote Sensing V.41. p. 1465- 1476.

Berry,J.K.,J.A. Smith, T.J. Heimes (1981) Effect of mixtures of primary scene components on spectral variability. Proceed. 47 Annual Meeting American Society Photogramm. p55-65.

Beven,K. (1978) The hydrological response of headwater and sideslope areas. Hydrological Sciences Bulletin. V23. p.419-437.

Boresjo, L. (1984) Classification of mires using multitemporal LANDSAT MSS and topographic map data. Proceed. 8th Canadian Symposium on Remote Sensing. p.581-590.

Bowersox,T.W.,W.W. Ward (1972) Prediction of oak site index in the ridge and valley region of Pennsylvania. Forest Science V.18 p.192-195.

Bryant, E., A.D.Dodge, S.D. Warren (1980) LANDSAT for practical forest type mapping: a test case. Photogramm. Eng. and Remote Sensing V46 p.1574-1584.

Burt,T.P., D.P. Butcher (1985) Topographic controls of soil moisture distributions. J. of Soil Science V.36 p.469-486.

Butlin,T.J., F.E. Guertin, S.S. Vishnubhalta (1978) The C.C.R.S. digital image correction system. Proceed. 5th Canadian Symp. on Re. Sen. p.271-283.

Calver., M.J. Kirkby, D.R. Weyman (1972) Modelling hillslope and channel flows. in Chorley, R.J. (ed.) Spatial Analysis in Geomorphology British Geomorphology Research Group p.197-218.

Canada Department of Energy, Mines and Resources (1982) National Topographic Service 1:50000 Map Sheet NTS 83F2.

Carson,M.A., M.J.Kirby (1972) Hillslope Form and Process. Cambridge U. Press.

Chrisman, N. (1982) Methods of Spatial Analysis Based on Error in Categorical Maps. unpub. Ph.D. Dissertation, U. of Bristol.

Coile T.S. (1952) Soil and the growth of forests. Advances in Agronomy V.4 p.330-398.

Colewell, J.E. (1974) Vegetation canopy reflectance. Rem. Sen. Env. V.3 p.175-183.

Corns, I.G.W. (1983) Forest community types of west-central Alberta in relation to selected environmental factors. Can. J. Forestry Research V.13 p.995-1010.

Dana, R.W. (1982) Background reflectance effects in LANDSAT data. Applied Optics V.21 p. 4106-4111.

Daubenmire, R. (1968) Soil moisture in relation to vegetation distribution in the mountains of northern Idaho. Ecology V.49 p.431-438.

Dryness, C.T. and D.F. Grigal (1979) Vegetation-soil relationships along a spruce forest transect in interior Alaska. Canadian J. Botany V.57 p.2644-2656.

Dumanski, J., T.M. Macyk, C.F. Veauvy, J.D. Lindsay (1972) Soil survey and land evaluation of the Edson Hinton area, Alberta. Alberta Institute of Pedology Report #S-72-31

Dunne, T., T.R. Moore, C.H. Taylor (1975) Recognition and prediction of runoff producing zones in humid regions. Hydrological Sciences Bull. V.20 p.305-327.

Evans, I.S. (1972) General geomorphometric derivations of altitude and descriptive statistics. in Chorley, R.J. (ed) Spatial Analysis in Geomorphology. Harper and Row, N.Y.

Eyton, J.R. (1974) An Analytical Surfacing Methodology for Application to the Geometric Measurement and Analysis of Landforms. unpub. Ph.D. Dissertation, U. of Illinois.

Eyton, J.R. (1987) Mapping the differential geometry of digital elevation models. (under review.)

Fleming, G. (1975) Computer Simulation Techniques in Hydrology. Elsevier, N.Y.

Franklin, S.E. (1985) The Significance of Geomorphometric Variables in LANDSAT MSS Analysis of a High Relief Environment. unpub. Ph.D. Dissertation, U. of Waterloo.

Franklin, S.E. (1987) Terrain analysis from digital patterns in geomorphometry and LANDSAT MSS spectral response. Photogramm. Eng. and Remote Sensing V. 53 p.59-66.

Franklin S.E., E.F. LeDrew (1984) High relief terrain classification using digital elevation model variables and LANDSAT MSS data in the Yukon, Territory, Canada. Proceed. 18th Int. Sympos. on Remote Sensing of the Environment. p.603-611.

Goodenough,D.G., J.Iisaka, K.Arai (1986) On the classification of Thematic Mapper data. (abstract only) Proceed. 10th Canadian Sympos. on Remote Sensing. p.993.

Graney,D.L., E.R.Ferguson (1971) Site quality relationships for shortleaf pine in the Boston Mountains of Arkansas. Forest Science 17 p.16-23.

Hame, T. (1984) LANDSAT aided forest site type mapping. Photogramm. Eng. and Remote Sensing V.50 p.1175-1184.

Hardy, R.L. (1971) Multiquadric equations of topography and other irregular surfaces. J. Geophysical Research V.76 p.1905-1915.

Harris, R. (1985) Contextual post classification processing of LANDSAT data using a probabilistic relaxation model. Int. J. of Remote Sensing. V.6 p.847-866.

Hartung, R.E., W.J. Loyd (1969) Influence of aspect on forests of the Clarksville soils in Dent County, Missouri. J. of Forestry. V. 26 p.178-182.

Heerdegen, R.G., M.A. Beran (1982) Quantifying source areas through land surface curvature and shape. J. of Hydrology. V.57 p.359-373.

Hegg,K.M, R.S. Driscoll (1982) Remote sensing applications to national forest inventories. in Johannsen,C.J. and J.L. Sanders (eds.) Remote Sensing for Resource Management. Soil Conservation Society of America. p.506-511

Hegyi,F., R.V. Quenet (1982) Updating the forest inventory data base in British Columbia. in Johannsen,C.J. and J.L. Sanders (eds.) Remote Sensing for Resource Management. Soil Conservation Society of America. p.512-518

Heringa,P.K., G.H.Cormack (1963) Relation of soils and ground cover vegetation in even-aged pine stands of central Alberta. Forestry Chronicles V.39 p.273-278

Hicks,R.R., P.S. Frank (1984) Relationship of aspect to soil nutrient, species importance and biomass in a forested watershed in West Virginia. Forest Ecology and Management V.8 p281-291.

Hutchinson,C. (1982) Techniques for combining LANDSAT and ancillary data for digital classification improvement. Photogramm. Eng. and Remote Sensing. V.48 p.123-130

Ivanov,K.E. (1981) Water Movement in Mirelands. Academic Press

Jeglum, J.K. (1974) Relative influence of moisture, aeration and nutrients on vegetation and black spruce growth in Northern Ontario. Can. J. of Forest Research. V.4 p.114-126

Kalensky, Z.D., W.C. Moore, G.A. Campbell, D.A. Wilson, A.J. Scott (1981) Summary forest resources data from LANDSAT images. Canadian Forestry Service Information Report P1-X-5.

Kleckla, W.R. (1980) Discriminant Analysis. Sage U. Paper Series on Quantitative Applications in the Social Sciences #07-019.

Krejcie, R.V., D.W. Morgan (1970) Determining sample size for research activities. Educa. and Psych. Measurements V. 30 p.607-620.

Larson, J.A. (1980) The Boreal Ecosystem. Academic Press

Loveland, T.R., G.E. Johnson (1983) The role of remotely sensed and other data for predictive modelling: the Umatilla, Oregon example. Photogramm. Eng. and Remote Sensing V.49 p.1183-1192.

Lowry, G.L. (1975) Black spruce site quality as related to soil and other site conditions. Soil Science Of Am. Proceed. V. 39 p.125-131.

Maxwell, D.A., R.D. Turpin (1968) Numeric ground image systems design. Texas Transportation Research Report 120-1

Monmonnier, M.S. (1982) Computer Assisted Cartography: Principles and Prospects. Prentice Hall Inc, N.J.

Moore P.D. and D.J. Bellamy (1974) Peatlands. Springer Verlag.

National Soil Survey Committee of Canada (1974) The system of soil classification for Canada. Canada Department of Agriculture Pub.1455.

Niemann, O., G. Langford, G. More (1983) Avalanche hazard mapping integrating LANDSAT digital data and digital topographic data. Proceed. 8th Canadian Sympos. on Remote Sensing. p.261-267

Niemann, O., G.Langford (1984) The use of LANDSAT imagery in wildlife habitat analysis. Internal report prepared for the Alberta Remote Sensing Center and the Alberta Environmental Centre.

Nix, L.E., W.A. Swain, K.O. Kelton (1984) Modelling forest biomass accessibility in S. Carolina with digital terrain data. Proceed. 1984 Symposium on Machine Processing of Remotely Sensed Data. p.389-394.

O'Callaghan, J.F., D.M. Mark (1984) The extraction of drainage networks from digital elevation data. Computer Vision, Graphics and Image Processing. V.23 p.323-344.

Orti, F. (1981) Optimal distribution of control points to minimize LANDSAT registration errors. Photogramm. Eng. and Remote Sensing V. 47 p.101-110.

Palylyk, C. (1985) LANDSAT M.S.S. Data for Peatland Inventory. unpub. M.Sc. Thesis, U. of Alberta.

Pendock, N., M. Sears, A. de Gasparis (1983) Reducing the spectral dimension of remotely sensed data and the effect on information content. Proceed. 17 Int. Symp. on Remote Sensing of the Env. p.1213-1222.

Pettinger, L.R. (1982) Digital classification of LANDSAT data for vegetation and landcover mapping in the Blackfoot River, Watershed, S.E. Idaho. U.S.G.S. Prof. Paper. 1219.

Place, J.L. (1985) Mapping of forested wetlands using Seasat radar images to complement conventional sources. Prof. Geographer V. 37 p.463-469.

Richards, N.A., E.L. Stone (1964) The application of soil survey to planting site selection: an example from the Alleghany Uplands of New York. J. Forestry V.62 p.475-480.

Rosenfield, G.H., K. FitzPatrick-Lins, H.S. Ling (1981) Sampling selection for thematic accuracy testing. Photogramm. Eng. and Remote Sensing V.48 p.131-138.

Rosenfield, G.H., K. Fitzpatrick-Lins (1983) Category analysis of the classification error matrix. Proceed. Machine Processing of Remotely Sensed Data. p.311-313.

Sampson, R.J. (1978) Surface Two Graphics System. Series on Spatial Analysis No.1, Kansas Geological Survey.

Sarle, W.S. (1983) Cubic clustering criterion. SAS Technical Report A-108.

SAS Institute (1985) SAS Users Guide. Cary, North Carolina.

Satterwhite, M., W. Rice, J. Shipman (1984) Using landform and vegetative factors to improve the interpretation of LANDSAT imagery. Photogram. Eng. and Remote Sensing. 50 p.83-91

Schowengerdt, R.A. (1983) Techniques for Image Processing and Classification in Remote Sensing. Academic Press, Toronto.

Shlien, S. (1979) Geometric correction, registration and resampling of LANDSAT imagery. Can. J. of Remote Sensing. V.5 p.74-89.

Smith, J.A., T.L. Lin, K.J. Ranson (1980) The Lambertian assumption and LANDSAT data. Photogramm. Eng. and Remote Sensing

V.46 p.1183-1189.

Spies, T.A., B.V. Barnes (1985) A multifactor ecological classification of the northern hardwood and conifer ecosystems of Sylvania Recreation Area, Upper Peninsula, Michigan. Can J. of For. Research. V. 15 p.949-960.

Spurr,S.H., B.V. Barnes (1980) Forest Ecology John Wiley, N.Y.

Strahler, A. (1980) The use of prior probabilities in maximum likelihood classification of remotely sensed data. Remote Sensing of Env. V.10 p.135-163.

Strahler, A.H., T.L. Logan, N.A. Bryant (1978) Improving forest cover classification accuracy from LANDSAT by incorporating topographic information. Proceed. 12th Int. Sympos. on Remote Sensing of the Env. p.927-942.

Strahler, A.H., J.E. Estes, P.F. Maynard, F.C. Mertz, D.A. Shaw (1980) Incorporating collateral data in LANDSAT classification and modelling procedures. Proceed. 14th Int. Sympos. on Remote Sensing, p.1009-1023.

Teillet, D.M., B. Guindon, D.G. Goodenough (1982) On the slope-aspect correction of multispectral scanner data. Can J. of Remote Sensing. V.8 p.84-106.

Tilton, J.C. (1983) Contextual classification of multispectral image data using compound decision theory. Proceed. 17th Int. Sympos. on Remote Sensing of the Env. p.863-871.

Todd, W.J., D.G. Gehring, J.F. Harman (1980) LANDSAT wildland mapping accuracy. Photogramm. Eng. and Remote Sensing V.46 p.509-520.

Tom, C.H., L.D. Miller (1980) Forest site mapping and modelling. Photogramm. Eng. and Remote Sensing V.46 p.1586-1596.

Toth, J.A. (1962) A theory of groundwater motion in small drainage basins in Central Alberta, Canada. J. of Geophysical Research. V.67 p.4375-4387.

Troeh, F.R. (1964) Landform parameters correlated to soil drainage. Proceed. Soil Science of America. V.28 p.808-812.

van Genderen, J.L., B.F. Lock (1978) Remote sensing: statistical testing of thematic map accuracy. Remote Sensing of the Environment. V.7 p.3-14.

Walsh,S.J. (1980) Coniferous tree species mapping using LANDSAT data. Remote Sensing of the Environment. V.9 p.11-29.

Weyman, D.R. (1973) Measurement of the down slope flow of water in a soil. J. of Hydrology. V.20 p.267-288.

Whipkey, R.Z., Kirkby, M.J. (1978) Flow with the soil. in Kirkby, M.J. (ed.) Hillslope Hydrology. J. Wiley, N.Y. p.121-144.

Wilde,S.A., G.K.Voigt, R.S. Pierce (1954) The relationship of soils and forest growth in the Algoma District of Ontario, Canada. J. Soil Science 5 p.22-38.

Yoeli, P. (1984) Cartographic contouring with computer and plotter. American Cartographer. V.11 p.139-155.

Young. A. (1972) Slopes Longman, N.Y.

Young, R.A., C.K. Mutchler (1969) Effect of slope shape on erosion and runoff. Trans. Am. Soc. Agric. Engineers. V.12 p.231-233.

Yuan,L.P., N.L. Vanderpool (1986) Drainage network simulation. Computers and Geosciences. V. 12 p.653-665.

Zobel, D.B., A. Mckee, G.Haw, C.T. Dyrness (1976) Relationship of environment to composition of forest communities of the central western Cascades of Oregon. Ecological Monographs. V.46 p.135-156.

APPENDIX 1

APPENDIX 2

Most of the work concerning the relationship of landform geometry and forest growth or integrating MSS and terrain data has used slope aspect as an independent variable influencing the nature of vegetation growth. Aspect is a circular distribution and while 1° and 359° are extremes of the distribution, they are in fact almost identical. Similarly, 0° and 180° are opposites on the compass rose, but not at extremes in the distribution. This point was recognized by Hartung and Loyd (1969) who noted that if the site index of trees, growing on a particular soils, were plotted against slope azimuth, a sine curve would be produced. Peaks and troughs in the curve were 180 degrees apart. Graney and Ferguson (1971) observed in their study that the lowest site index values were recorded on south facing aspects while the highest were on the north facing slopes. The sites measured on aspects from 0° to 180° were mirror images of 180° to 360°.

Hartung and Loyd (1969) suggested that the circularity of the distribution could be compensated for through a linearization. This linearization was accomplished through the relation:

$$La = (\sin a) + 1$$

where La was the linearized aspect and a the aspect. A value of 1 was added to the values to remove negative numbers.

It was noted that linearization must be carried out according to some rational reason. For example if the linearization was to be based on influence exerted by exposure to the sun or prevailing wind, then the nature of the transformation should reflect this influence.

For the purposes of this study, the aspects were linearized for the predominant wind direction. Bentz et al. (1985) reported that a majority of the precipitation which falls during the growing season in the west central portion of the foothills was the result of summer convectional storms. The direction of movement of these storms was determined by the predominant wind direction. As well, the lee slopes tended to retain the winter snows for a longer period of time, due to greater loading, than did the windward slopes and so were wetter earlier in the growing season. The data used for this analysis were based on recordings from Edson airport, to the north and east of the study area. As a result of the published climate data (Canada Department of the Environment, 1982) the aspects were linearized so that values occurring at 270 degrees in the original aspect distribution were assigned a value of 0, and 90 degrees a value of 180. Aspects of 0 and 180 degrees were, therefore, assigned values of 90 degrees.

The linearized aspect was analyzed together with the other DEM derivative and landscape products in the investigation of landscape drainage. The results obtained with the use of aspect, however, yielded a confusion matrix with a lower overall explanation than was the case when aspect was not included. The explanation for this may be that with the slopes being as low as they were in the study area, the direction of the wind was inconsequential. It was for this reason that aspect was omitted from the subsequent discussion in this dissertation.

R 20 R 19

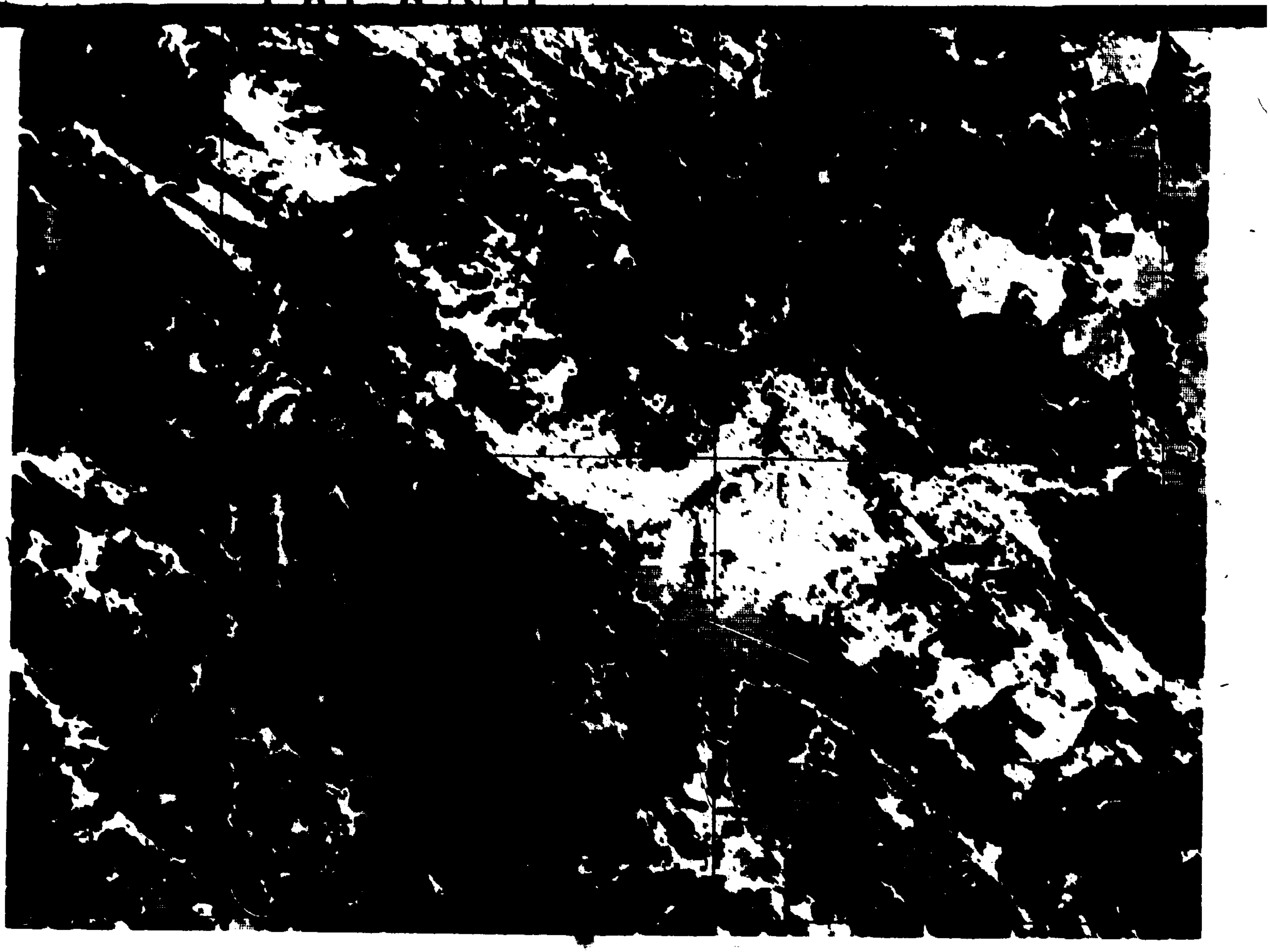
P48

R 20 R 19

TP 48

TP 47

LEGEND



LEGEND

OPEN SOIL

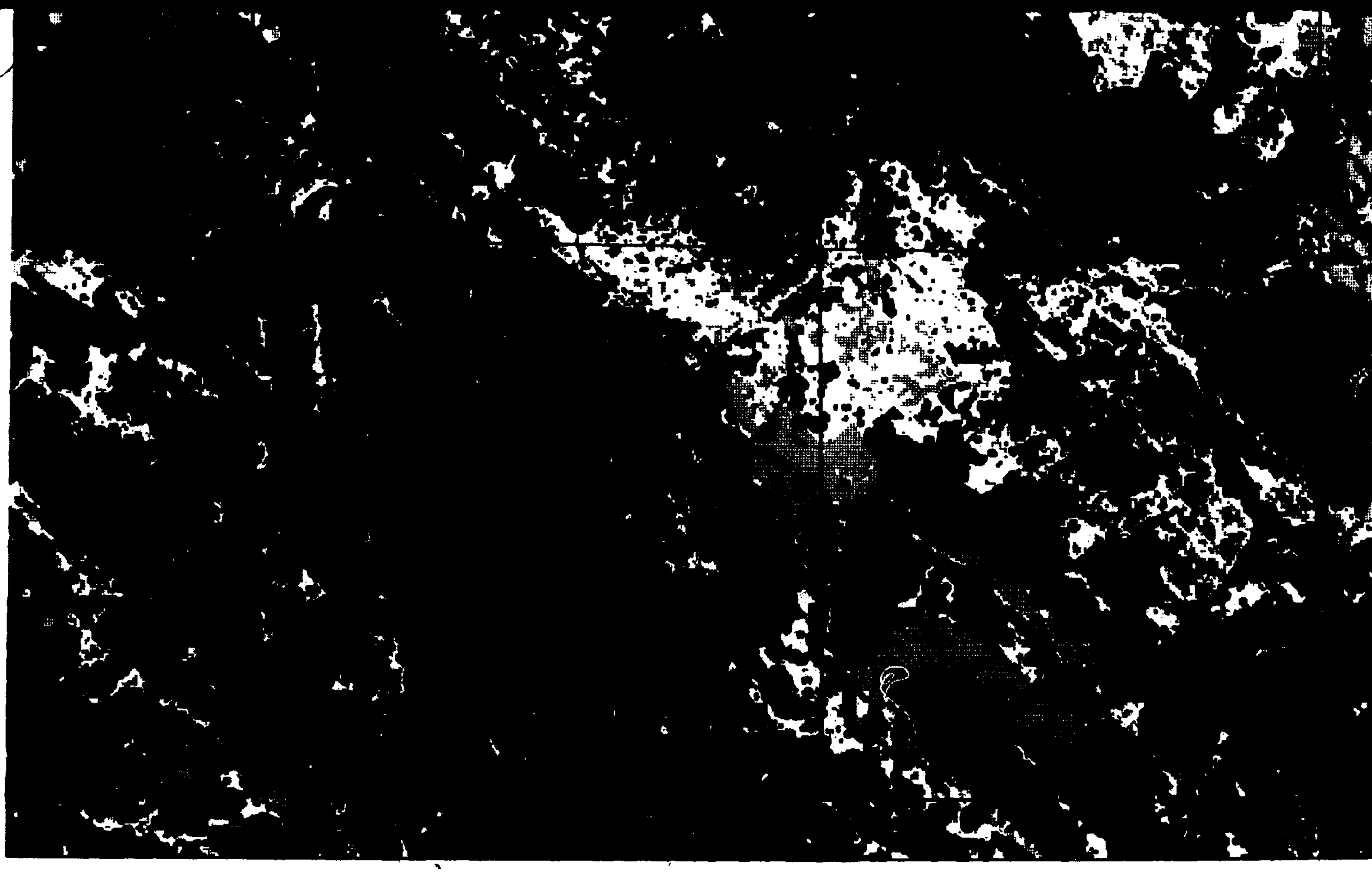
DECIDUOUS

■ CONIFER

■ MIXED CONIFER/DECID

LEGEND

P 47



LEGEND

OPEN SOIL

MUSKEG

DECIDUOUS

MIXED CONIFER

UPLAND CONIFER

MIXED CONIFER/DECID