PATHAK, PRASAD A., Ph.D. Geospatial Analysis of Lake and Landscape Interactions within the Toolik Lake Region, North Slope of Alaska. (2010)
Directed by Dr. Roy Stine. 190 pp.

The Arctic region of Alaska is experiencing severe impacts of climate change. The Arctic lakes ecosystems are bound to undergo alterations in its trophic structure and other chemical properties. However, landscape factors controlling the lake influxes were not studied till date. This research has examined the currently existing lake landscape interactions using Remote Sensing and GIS technology. The statistical modeling was carried out using Regression and CART methods.

Remote sensing data was applied to derive the required landscape indices. Remote sensing in the Arctic Alaska faces many challenges including persistent cloud cover, low sun angle and limited snow free period. Tundra vegetation types are interspersed and intricate to classify unlike managed forest stands. Therefore, historical studies have remained underachieved with respect thematic accuracies. However, looking at vegetation communities at watershed level and the implementation of expert classification system achieved the accuracies up to 90%.

The research has highlighted the probable role of interactions between vegetation root zones, nutrient availability within active zone, as well as importance of permafrost thawing. Multiple regression analyses and Classification Trees were developed to understand relationships between landscape factors with various chemical parameters as well as chlorophyll readings. Spatial properties of Shrubs and Riparian complexes such as complexity of individual patches at watershed level and within proximity of water

channels were influential on Chlorophyll production of lakes. Till-age had significant impact on Total Nitrogen contents. Moreover, relatively young tills exhibited significantly positive correlation with concentration of various ions and conductivity of lakes. Similarly, density of patches of Heath complexes was found to be important with respect to Total Phosphorus contents in lakes.

All the regression models developed in this study were significant at 95% confidence level. However, the classification trees could not achieve high predictabilities due to limited number of lakes sampled.

Keywords: Landscape factors, Lake primary productivity, Arctic, Climate change, Regression, CART

GEOSPATIAL ANALYSIS OF LAKE AND LANDSCAPE INTERACTIONS WITHIN THE TOOLIK LAKE REGION, NORTH SLOPE OF ALASKA

By

Prasad A Pathak

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2010

Approved by

Dr. Roy Stine
Committee Chair

Dedicated to my parents….

Mr. Avinansh Pathak

&

Mrs. Alka Pathak

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair: Dr. Roy Stine

Committee Members: Dr. Anne Hershey

Dr. Stephen Whalen

Dr. Zhi-Jun Liu

07-07-2010
Date of Acceptance by Committee

05-04-2010
Date of Final Oral Examination

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my committee chair, Dr. Roy Stine, for his continuous support. He not only supported me financially but showed complete faith in my abilities and helped me become more meaningful researcher. I am also thankful to Dr. Hershey, my committee member, who shared her knowledge and experience about the North Slope, Alaska, with me and helped me improve readability of this dissertation. The contribution by Dr. Stephen Whalen, who was my external committee member, cannot be summed up with words. He challenged limits in limnology and provided constant encouragement. Without his guidance this dissertation was not possible. I would like express my gratitude towards Dr. Zhi-Jun Liu, who also my committee member, for sharing a great deal of statistical expertise with me. His support made it possible to explore various statistical techniques, then new to me and his comments made my dissertation more meaningful.

Among my colleagues, I would like thank Dr. Debasish Chaudhuri who helped my smooth transition into UNCG and helped me in every possible aspect. I don't have any words for my really helpful and supportive friends, Mary Hall-Brown and Leanne Sulewski. They made me feel completely at home and constantly encouraged me to make my research better. I owe my deepest gratitude to Leanne Sulewski who took the painful job of editing this dissertation for me.

How could I complete my duties in the geography department without timely help and friendly support by Ms. Lois Carney and Ms. Ruth Canon? I also thank Head of the Department, Dr. Jeff Patton for his advice and constant support.

Additionally, I am thankful to my all my friends; Sarika Kshirsagar, Shirish Manchi, Sagar Mysorekar, and Archana Waran for constant emotional support. Last but not least, I would like to thank my sister Archana, niece Saisha, and brother-in-law Shripad for their support.

This research was funded by National Science Foundation research grant “A geomorphic-trophic hypothesis for arctic lake productivity” (0516043).

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

1.1 Overview

A variety of long term changes in climate have been reported at continental, regional, and local scales. The Intergovernmental Panel for Climate Change (IPCC) has published observations regarding such changes in its Assessment Report 4 (IPCC, AR4, 2007). It has been stated that over the last two decades, precipitation has increased in all major continents. In the last 50 years, nighttime temperatures are warmer than local averages, the frequency of tropical storms has increased, and drastic changes in typical patterns of various physical and biological systems have been reported. The same report has also pointed out that since 1970, the surface temperatures in the Arctic region have increased by 3.5 ^{0}C (IPCC, AR4, Synthesis report, pg 32). Thus, rapid warming of climate in the Arctic is well established. In fact, it has been predicted that in the area north of 60° N, temperatures will increase further by 3.7^{o} C (Arctic Climate Impact Assessment – (ACIA) designated five-model average, International Arctic Science Committee – (IASC), Section 4.4) in the near future. Other consequences of Arctic climate change are evident. Permafrost is melting rapidly in that region, leading to the deepening of the active soil layer and acceleration of thermokarst

activities (Serreze et al., 2000). The ACIA (2007) report also mentions that with the observed increases in precipitation, rivers and lakes in the Arctic are experiencing higher influxes of water. Many ecological processes are closely linked with these transformations in the Arctic landscapes.

With rising temperatures, it has been predicted that vegetation species currently existing in the southern region of the Arctic will encroach over tundra vegetation. It will also provide opportunities for more broad leaf species to grow in the Arctic. Already similar findings were obtained by Tape et al. (2006). The authors used repeat-photography in the North Slope of Alaska and discovered that the amount of shrubs has increased in low-lying areas.

Arctic lakes are integral part of the Arctic landscapes. Currently, they are known for their ultra-oligotrophic nature (Whalen and Alexander, 1986). With changes in the hydrological regimen and alterations in landscapes, these lakes may exhibit modifications in their water chemistry as well as trophic structure.

To predict the future trends of Arctic lakes, it is first necessary to understand current interactions between lakes and landscapes within the Arctic region. No such holistic attempts are documented in the available literature; however, research pertaining to various narrow aspects of nutrient exchanges between ecosystems is available. For example, Schimel et al. (1996) studied different vegetation communities in the Imnavait creek region for their abilities to uptake major nutrients from soils. A similar study was also carried out by Marion et al. (1989). However, Everett et al. (1989) focused on

impacts of snowmelt events on surface water chemistry. Oechel (1989) performed artificial nutrition addition to understand impacts of disturbances on different tundra vegetation types. On the other hand, Giblin et al. (1991) studied nitrogen and phosphorus concentrations across the toposequence of Sagavanirktok River in Arctic Alaska. Only recently has a more comprehensive study about geochemical weathering and its effects on soil and streams been carried out by Keller et al. (2007). It was clear from these studies that most of them were constrained to specific geographical area (e.g. Imnavait creek) and did not explore landscape factors along with water bodies; therefore, it is not yet clear how Arctic landscapes control lakes chemistry. Challenges in the field work and obtaining comprehensive information about landscapes might have restricted the scope of these studies.

One of the unavoidable challenges is that the Arctic region remains snow covered for most of the year; only the summer season is favorable for landscape study. Most of the area is inaccessible except by helicopter, making it expensive to carry out such studies (Hope and Stow, 1995). Use of satellite images is considered an alternative for the field work to encompass large regions in the Arctic. However, remote sensing technology faces several limitations, such as snow cover and persistent cloud cover. One of the major disadvantages relating to the challenges of remote sensing in the Arctic is that thematic accuracies of maps in the Arctic have remained underachieved compared to other regions (Noyel, 1999).

The current research focuses on the Arctic region of Alaska and has attempted to use moderate resolution satellite imagery to study landscape factors with respect lake chemistry. The lake chemistry data provided by Whalen et al. (2007, personal communication) was used to identify contemporary lake-landscape interactions. The primary goals of this research are described below.

1.2 Research Goals

1. **Accuracy of Classification** -Ecological studies use satellite data as a relatively inexpensive tool, which provides higher spatial and spectral resolution of a study area. At the same time, they demand an appropriate accuracy level of thematic classification obtained from these satellite data, because further analyses are dependent on accurate inputs. However previous remote sensing studies in the Arctic region do not exhibit high accuracies. Thus, improving accuracy of thematic maps at catchment level was one of the goals in this research.
2. **Identify the landscape factors** – Minimal human interference is observed in the Arctic Alaska region; therefore, it is crucial to identify relevant landscape factors, which are currently influencing lake water chemistry within the Arctic region.
3. **Statistical modeling** – After identification of landscape factors, it is necessary to employ robust and easy-to-interpret statistical methods to model relationships between lakes and the landscape factors.

1.3 Dissertation Structure

The dissertation is divided into 5 chapters. The literature review chapter provides necessary background about historical remote sensing studies and their limitations. It also discusses the landscape factors used in discovering lake-landscape interactions outside the Arctic and narrows down to important landscape factors utilized for this research. The next section introduces the study area. The Data and Methods section describes the satellite images used, lake chemistry data, and methods carried out to achieve the aforementioned research goals are described. The findings are divided into three subsections: the accuracy assessment results, significant statistical models for lake productivity, and major nutrients and ionic composition. Finally, the Conclusion chapter summarizes the major findings, illustrates limitations of the study, and highlights needs of future research work.

CHAPTER II

LITERATURE REVIEW

2.1 Problem Definition

2.1.1 Inaccessibility of the Arctic Region

Tundra ecosystems in the Arctic region of Alaska are markedly different than other widely studied ecosystems. "Tundra" is a word derived from the Finnish term "*tunturi,*" which describes an ecosystem with long and cold winters and very short summers, supporting only low lying herbs and shrubs (Wildlife Conservation, Alaska Division). With its vulnerability to climate change, these Arctic ecosystems are drawing the attention of scientists all over the world. A wide range of temperature changes are evident within this region. These changes are closely linked with the Carbon cycle and the recycling of other nutrients in these ecosystems. It has been predicted that large amounts of carbon stored in various ecosystems like bogs and peatlands in the Arctic, will be released to convert them from a carbon sink to a source. Permafrost melting will rapidly release stored nutrients, and the chemical weathering of parent material will release various ions (Keller et al., 2007). Terrestrial ecosystems in the Arctic, hence, will experience encroachment by broad leaf species as their growth will be supported by rising temperatures, expanding growing season, and frequent precipitation, along with an increased availability of nutrients. These alterations are bound to impact water bodies in the Arctic in terms of their productivity and trophic structure.

However, very limited knowledge of the landscape controls with respect to lake chemistry is currently available. The major reason could be the inadequate access to the region, along with extreme climate conditions (Hope and Stow, 1995). A gravelly Haul Road, also known as the Dalton Highway, is the only available ground transportation option in the Alaskan Arctic.

2.1.2 Satellite images as an alternative

Several of arctic ecosystem studies, especially the ones using remote sensing data, were reviewed by Stow et al., (2004). Research carried out by Walker (1977) using Television Scanning Densitometer was one of the early efforts to study geomorphology in the Arctic. Similar study was carried out by Stow et al. (1993) however; the authors used aerial photography and videography tools for the purpose. A plot level CO_2 flux study was carried out by McMichael et al. (1999). They used a handheld radiometer to study photosynthetic activity and its quantitative relationship to Normalized Difference Vegetation Index (NDVI) derived from radiometric measurements. NDVI has been widely used at regional levels, for example, Vourlities et al. (2000) used NDVI derived from Advanced Very High Resolution Radiometer (AVHRR) to develop a method of scaling up plot level CO_2 fluxes across the entire Arctic.

NDVI derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data was used by Markon et al. (2001) to study phenological changes in tundra vegetation. Jia et al. (2002) used bi-weekly AVHRR-NDVI data spanned across five years to study latitudinal trends within MAT and MNT vegetation types. Authors found

that NDVI values of MNT remained lower than those of MAT. In addition, a statistically significant relation was obtained between NDVI and elevation data. Hope et al. (2004) compared NDVI values of tundra at higher spatial resolution (aerial photography) to NDVI values at a regional level (obtained from AVHRR data). A significant agreement was found between NDVI values obtained at two different spatial resolutions.

These studies were confronted with various challenges such as (1) persistent cloud cover, (2) short growing season, (3) snow cover during winter, (4) water stagnated land covers, and (5) solar angle (Hope et al., 1995). Therefore, very few attempts were made to assess the accuracy of the thematic maps in this region (Fleming, 1988; Felix and Binney, 1989; Stow et al., 1989; Muller et al., 1998). It was also observed that the number of land cover categories created had significant impact on the accuracy along with method used for ground surveying.

Fleming (1988) identified a need for a rapid but accurate map production process for conservation and other management decisions. Hence, the author chose a computer-aided method to integrate Landsat MSS data, Digital Elevation data and Color Infra-red photographs for the same purpose. Based on aerial photographs and Landsat scenes, different regions which appeared homogenous spectrally were identified and stratified random samples for training purpose were selected. Size of the training site varied from 10 acres to 50 acres. Field observations made at these sample places were used to classify images and assess the accuracy of the method. Overall ten categories were derived, which were composed of certain forest types as well as types dwarf tundra vegetation. Overall accuracy was 78.2% after inclusion of DEM and Slope data.

In a similar study, Felix and Binney (1989) classified Landsat MSS data for Arctic National Wildlife Refuge. There were ten tundra vegetation categories and 3 water categories included in the classification scheme. Overall accuracy was only 37%. According to the authors, misclassification occurred due to close spectral resemblance of vegetation categories with each other. Moist/wet complex was not clearly identifiable with the Landsat data. Spectral overlap was found between Moist/wet complex and wet graminoid tundra as well as moist prostrate scrub classes. Highest accuracies were seen only for Clear Water land cover category.

SPOT images were used to map waterfowl habitats near Teshekpuk Lake area in Arctic coastal plains by Markon & Derksen (1994). Both unsupervised and supervised techniques were adopted to perform digital classification on the images dated summer of 1986. Field work was carried out in July of 1988, 1989 and 1991to convert spectral classes into land cover classes. . Total twelve land cover categories were used. However, map accuracy was not performed. Only intermittent cross checking and adjustments were done to the map. Generally, confusion occurred between flooded tundra areas and unvegetated tundra areas. These kind of differences were attributed to changed hydrological scenarios between data capture and ground observation periods.

Based on the classification scheme developed by Walker et al. (1994), a land cover map of the Kuparuk River basin was prepared using Landsat MSS data (Weller et al., 1995). Classes included in the scheme were Barren, MNT, MAT, Shrublands, Wet Tundra, Water, Clouds & Ice, and Shadows. Accuracy assessment of the map was carried

out by Muller et al. (1998). Considering the positional accuracy provided by the GPS receiver used for the field work, the plot size of sampling location was restricted to 3 by 3 pixels. Along pre-decided transects, sampling was done at an interval of every 250 meters. The field work was more concentrated on sampling MAT and MNT categories as they cover majority of the land within the watershed. Fuzzy sets logic was applied along with error matrix to assess overall accuracy as well as producer's and user's accuracies. Overall accuracy obtained for the classification was 87.1%. Barren, Shrublands, and Water exhibited highest User's and Producer's accuracies. MNT was confused with Wet Tundra whereas only occasionally MNT was confused with MAT. The authors explained that field work was difficult within the Arctic because of inaccessibility and expense. However, they claimed the fuzzy logic approach to be useful to improve accuracy of the map given the heterogeneity of tundra vegetation.

Stow et al. (2000) also studied the Kuparuk Watershed using single-date and seasonal time series AVHRR data. The major objective of the study was to determine the optimum spectral-temporal features for image classification. Unsupervised as well as Supervised algorithms were used to obtain three land cover categories i.e. MAT, MNT, and Wet Sedge Tundra. To assess the accuracy, reference map created by Muller et al. (1998) was used. The reference map was created using Landsat MSS at 50m spatial resolution. The overall accuracy of it was 86.7%. The accuracy of 86.1% obtained by Stow et al. (2000) was highest for supervised classification of single-date image integrated with NDVI. However multi-date image classification achieved accuracies up to 83.4%.

Outside of the Arctic region, there have been several attempts to improve the accuracies by incorporating ancillary data like elevation, slope, and aspect layers along with satellite images (Strahler et al., 1978; Gercek, 2002; Vatsai et al., 2005). Recently, though, Chaudhuri D. (2008) classified a SPOT image of the Toolik Lake region, Alaska using a hybrid approach. A knowledge base was created based on Normalized Difference Vegetation Index (NDVI), slope, and aspect layers. The author created rules using spectral values of the SPOT image as well as values of slope and aspect pertaining to each land cover class. Although this research achieved an overall accuracy of 75.57%, it highlighted that several land cover classes were spectrally overlapping and affected accuracies. The major confusion observed was between Moist Acidic Tundra complex and Moist Non-Acidic Tundra complex. These land covers showed significant mixing of signatures across north-south profile of the image probably because of low the sun angle.

Since the current research proposed use the same SPOT image, it was necessary to adopt a different method to overcome this limitation. It was decided to classify the image by dividing it into separate watersheds of each individual lake under consideration. Details about the watershed derivation as well as the classification scheme and the process are provided in the "Methods" section of this dissertation.

The classified images were further used to derive landscape factors useful for statistical modeling of lake landscape interactions in the study area. The following paragraphs of this section will illustrate the various landscape factors controlling the lake

chemistry. Further, it will narrow down the discussion to landscape factors relevant to the Arctic region.

2.2 Lake Landscape Interactions: A General Background

Terrestrial ecosystems were considered to be affecting streams and rivers, but not the lakes (Shindler and Scheuerell, 2002). However, recent limnological studies have shown that lakes share complex interactions with the surrounding landscape and cannot be studied as separate ecological units (Hasler, 1975, Oldfield, 1977). Willson et al. (1998) observed that these noticeably different habitats are coupled together by variety of forces such as gravity, water flow, and airflow. These forces are responsible for detritus matter and nutrients recycling between spatially separated habitats. Therefore, lakes are an integral part of many elemental cycles e.g. carbon, nitrogen, and phosphorus (Soranno et al., 1999). Cole (1994) illustrated that lakes receive organic and inorganic material from their watersheds. These allochthonous inputs are the driving force for primary production of the lakes. Multiple studies have established that, not only the primary production but other chemical properties and trophic levels of lakes are also governed by the nature of the surrounding landscape via these complex interactions (Cole, 2007; Pace et al., 1999; Auer et al., 2004). These natural landscape factors and human-dominated landscape factors affect the lakes differently.

2.2.1 Vegetation – Type and Age

Naturally, the vegetation exhibits control over nutrient inputs to adjacent water bodies due to its type and age. Zhu et al. (2008) explored this phenomenon for nitrogen export. The authors found that the amount of nitrogen contributed by different vegetation communities differ as per their type, leaf area, and nitrogen fixing capacity. They also mentioned that younger vegetation have lower nitrogen fixation rate and lower export to water bodies in the vicinity.

2.2.2 Landscape Position

Another well known factor related with lake water chemistry is the landscape position of a lake. Landscape position is defined by elevation of a lake and its connectivity with other nearby lakes. While studying effects of landscape position on the lakes, Kratz et al. (1997) observed that the low lying and connected lakes exhibited high concentration of silica and other nutrients; however, the lakes at higher elevations had lower nutrient concentration. The observed chemical properties of the lakes were linked to the respective source of water to them. According to the authors, rain water was crucial for the lakes at higher elevations, where as lower lying lakes were more influenced by ground water; therefore, they showed higher silica contents.

2.2.3 Physical properties of watersheds and lakes

The area of watershed is also an important factor controlling the amount of nutrient flowing into lakes. The larger the watershed, the greater the amounts of flow

would be produced reaching into the lakes (Schindler, 1971). Similarly, the lake area determines the amount of nutrients received by lakes through dry or wet fall (McColl and Grigal, 1977). It was observed that water retention time also affected the nutrient levels in lakes. Although it is not practically feasible each time, ratio of watershed area to lake area has been successfully used as a surrogate index for water retention time (Soranno et al. 1999).

2.2.4 Topography

Another critical landscape factor is topography of the watershed. Topography determines general hydrological properties such as soil moisture and overland flow. These properties control the nutrient transport from the watershed to the lake. Hence, indirectly topography influences the lake chemistry (Veith et al., 2003). To portray spatial properties of soil moisture within a watershed various topographical indices are commonly used (Burt and Butcher, 1986). The topographical wetness index, depicted by the ratio of upslope area to the slope at any given location within the watershed, is commonly used to quantify topographic control on hydrological processes (Sørensen et al., 2005).

2.2.5 Human Influence

Certain human activities are responsible for changes in the natural properties of landscapes. Deforestation and agricultural practices are among the widely studied human activities altering innate lake landscape interactions.

2.2.5.1 Deforestation

Each year approximately 13 million hectares of forest are lost from the Earth surface (Kourous, G. 2005). The major impact of deforestation is soil erosion and changes in nutrient recycling within terrestrial systems (Southgate and Whitaker, 1992). Increased soil erosion contributes suspended matter into water bodies and affects ionic composition of them.

It was also observed that forested lands regulate the temperature of lakes. For this reason, when watersheds experience deforestation, the water temperature was found to be increased, which enhances biological processes within lakes (Lombardzzi, undefined).

2.2.5.2 Agriculture

According to the United States Environmental Protection Agency (USEPA, 2000), run off from agricultural land use is the major non-point source of pollution, besides deforestation and urban land use. Catchments with agriculture impart large amount of particulate matter and dissolved phosphorus to their lakes. Therefore, those lakes show higher rates of productivity (Vanni et al*.,* 2005). Various studies have found that lake variables like Acid Neutralizing Capacity (ANC), Total Phosphorus (TP), and Total Organic Carbon (TOC) were positively correlated with land use in respective catchments (Stendera and Johson, 2006; Crosbie and Chow-Fraser, 1999; Arbuckle and Downing, 2001; Soranno et al., 1996).

2.2.6 Landscape metrics

Human activities like agriculture and deforestation affect the structural properties of natural landscapes. It was observed that the altered structural properties play a crucial role in limnological studies (Forman and Godron, 1986; Turner 1989; Gustafson 1998). Landscape metrics are useful in converting these structural properties into the numerical format. Even though a variety of metrics are available, they can be broadly categorized into the following categories: area (percentage/proportion), edge, shape, core area, diversity, contagion, and interspersion (Haines-Young & Chopping, 1996). According to the authors, the area, the core area, the contagion, and the interspersion metrics are relatively easy to interpret, but the shape metrics are intricate to link with their functional roles. As such, the authors have suggested using shape metrics very carefully.

Most of these metrics have also been instrumental in understanding impacts of human development on wildlife distribution (Andreassen et al., 1996; Matter 1996). On the other hand, limnological studies carried out in human dominated areas have also adopted these metrics, acknowledging the importance of the spatial arrangement of land use/land cover (King et al., 2005; Allan & Johnson, 1997; Griffith et al., 2002). Gergel et al. (2002) found that, percentage of impervious areas and distribution of riparian habitat in terms of continuity and width were useful indicators of water chemistry. Jones et al. (2001) also observed that the proportion land cover was a major landscape metric explaining variations in the nitrogen and phosphorus loadings in selected water reservoirs. Stewart et al. (2001) employed several indices, such as percentage and fragmentation, for riparian land use within the agriculture dominated watersheds located

in the eastern Wisconsin. The authors claimed that landscape metrics were very useful in predicting trophic structure of the water bodies. Bott et al. (2006) used the percentage of land use/land cover to predict productivity and chlorophyll contents of 8 reservoirs in New York City. Johnson et al. (1997) and Richards et al. (1996) found significant correlation between patch density of certain land uses and water quality in Michigan, in general, but other more generic landscape factors like geology had more of an impact.

2.3 Significant Factors for Arctic lakes

2.3.1 Human activities

Unlike described above, negligible human activities are evident within the Arctic region of Alaska. Examples of such human activity include the Dalton Highway, the Oil pipeline, and other mining and oil drilling activities (Oechel, 1989). Initial research about recycling of nutrients was carried out to understand the impacts of these disturbances on the nearby water bodies. According to Chapin et al. (1988), these human activities disrupted the natural water flow patterns and altered the nutrient transportation directed to surrounding water bodies. They also stated that changed water flow patterns enhanced the down slope movement of several nutrients. However, these impacts were observed explicitly near the disturbances but not in other regions. Hence, the authors suggested that natural factors were more significant in controlling nutrient exchanges.

2.3.2 Natural Factors

2.3.2.1 Precipitation and Snowmelt:

Everette et al. (1989) explored hydrology and geochemistry of Imnavait creek. They found that precipitation is responsible for significant inputs of certain cations and anions, including Ca, Mg, SO_4, Cl and others. Also all the cations and anions achieved their peak concentration immediately after snow melt and decrease drastically later. Again, at the end of growing season, some of them exhibited higher concentrations. The authors attributed the initial peaks of ions to organic material in upper soil layer being leached by snow melt. However, the authors claimed that during growing season ions experienced uptake by vegetation communities and could not reach the water bodies.

2.3.2.2 Natural vegetation communities and interactions with soil:

Apart from the precipitation and snowmelt phenomena, natural land cover composed of different vegetation alliances were studied as drivers of nutrient circulation. The relationship between nutrients levels observed in soils and plant communities were studied by Marion et al. (1989). The authors grouped the vegetation communities broadly into dry, moist, and wet categories. Generally a higher concentration of nutrients was found in the wet type of vegetation. They also found that deciduous shrubs have higher proportions of nitrogen (N), phosphorus (P), potassium (K), and magnesium (Mg). Similarly, soil was divided into three horizons i.e., upper two organic horizons and the third, the “A” horizon. Within soil horizons, the second organic horizon exhibited higher concentrations of nutrients, cations, and anions indicating their origin from the soil

horizons below it. Only K ions were higher in the top horizon, indicating their strong association with biological processes within root zone.

Chapin et al. (1988) emphasized the role of soil water in nutrient recycling. While studying the productivity of the Arctic tundra, they concentrated on *Eriophorum vaginatum* sp. Within this community it was observed that recycling of N and P occurred more rapidly along water tracks, as they provide more active soil depth and higher temperatures. *Eriphorum vaginatum* also facilitated nitrogen mineralization. The study also highlighted the difference between the root structures of *Eriophorum* to that of other tundra vegetation types. The deep root structure of *Eriophorum* was able to exploit more nutrients from soil water. The observed flowing soil water was also responsible for the higher productivity of riparian shrub communities. Root structures were studied in detail by Schimel et al. (1996) (Figure 1). The authors stated that like microbial mineralization and immbolization, root structure also played important role in nutrient utilization within terrestrial system. According to the authors, *Eriophorum vaginatum* have a thick non-branched and deep root structure. This enables them to uptake nutrients from freshly thawed soils. However, the roots of species like *Ledum* do not grow deep and can be found only in upper 5 cm of soils, which indicates a higher availability of nutrients and organic matter in upper soil horizons.

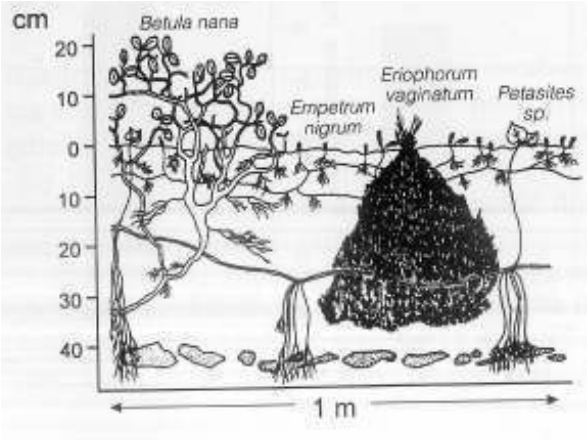


Figure 1. Root structures of different tundra vegetation types. (Source: Schimel et al., 1996, pp. 212)

More comprehensive knowledge about vegetation communities, soil, and permafrost thaw was presented by Giblin et al. (1991) with respect major nutrients i.e. nitrogen and phosphorus.

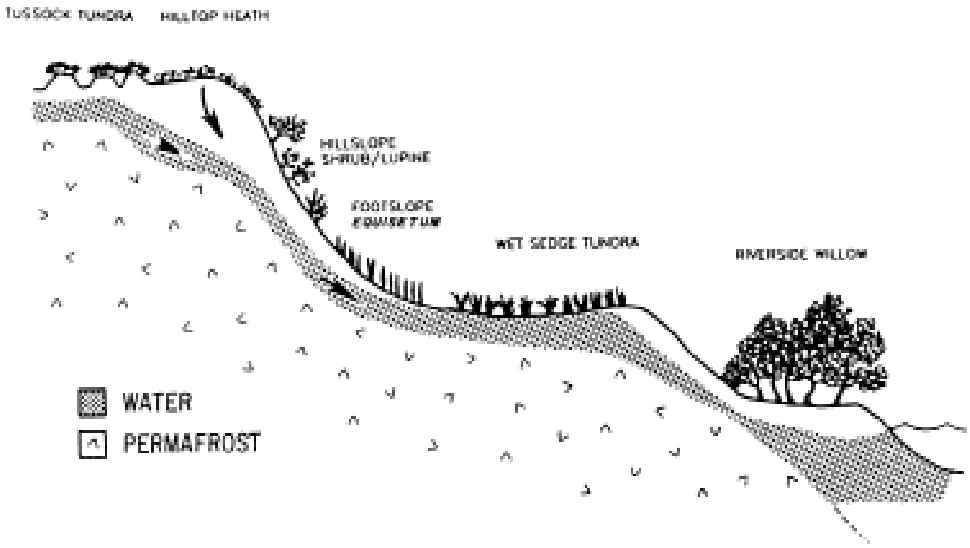


Figure 2. Toposequence (Souce: Giblin et al., 1991, pp. 107)

While studying a toposequence along a particular slope in the Sagavanirktok River, the authors explained that occurrence of different vegetation communities was controlled by active soil depth, which in turn controlled the nature of subsurface flow. Moreover, vegetation communities controlled the nutrients levels in soil with differential rate of uptake (figure 2). The authors explained that wet sedge showed higher concentrations of ammonium where as nitrate contents were relatively higher in tussock tundra and heath. Riverside willows showed the highest amount of extractable N whereas Hilltop Heath had the lowest. The same trend was exhibited for areal measurements of inorganic N along the toposequence. The extractable P pool was very high in Hilltop Heath, and it relatively decreased down the slope. The higher P in Heath was due to thaw depth reaching the mineral layer of soil. The authors attributed the higher dissolved inorganic nitrogen (DIN) concentrations to its release from upslope tussock tundra as they found higher nitrification rate within that vegetation community.

Another crucial source of nutrients was studied by Keller et al. (2007), i.e. permafrost. The authors mentioned that permafrost thawing leads to the release of stored nutrients as well as enhancing the chemical weathering of parent material. Accordingly, it may affect the nutrient release. Their analysis showed that the extractable fraction of calcium (Ca), potassium (K), and phosphorus (P) concentrations were significantly higher in the permafrost layer than active soil. The authors used ratio of 87Sr to 86Sr as an indicator of weathering. The results showed that there was a lower rate of weathering within younger tills like Itkilik II initial (Itk2) but higher rates within older tills like Sagavanirktok Main phase (Sag1). They also found that Ca concentrations were higher

within young tills. For stream chemistry, higher concentration of carbonate and Ca were observed where active soil layer was deep. With these findings authors predicted that P and K will be released in more quantities with rising temperatures and thawing of permafrost.

A summary of the interactions within different factors is represented by Figure 3:

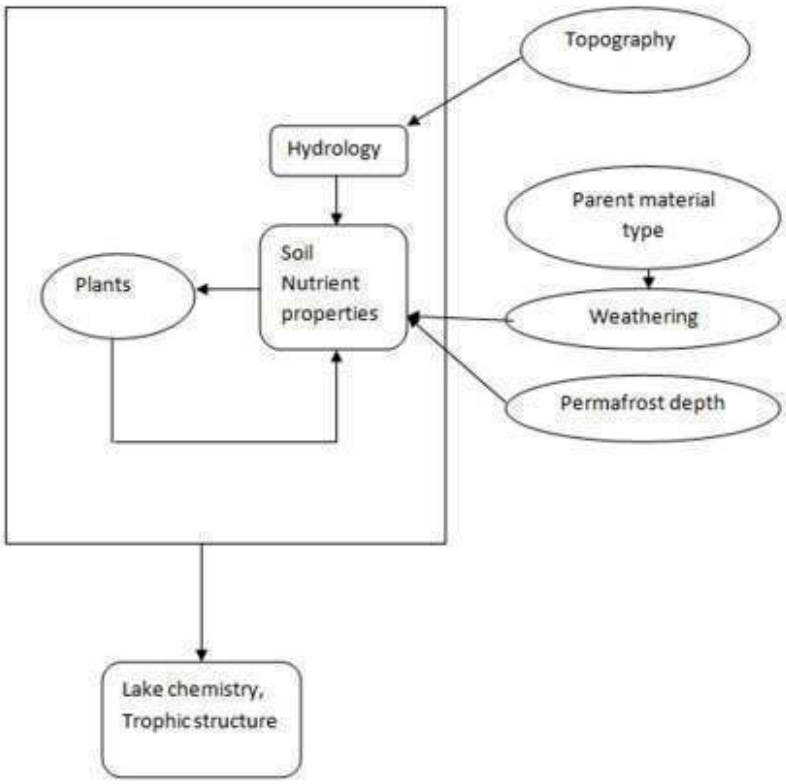


Figure 3. Diagrammatic representation of interactions between external factors affecting lakes

2.4 Proposed use of landscape metrics in the current research

It is obvious from the aforementioned studies that spatial aspects of landscape factors were not considered with respect to nutrient recycling. The unavailability of such datasets may be the reason behind it. Nevertheless, structural/spatial properties of land covers were crucial for this research as they would indirectly depict conditions of active soil depth and soil moisture. Being that the current study uses satellite data, it is limited to above-ground landscape coverage. Landscape metrics derived from these land covers were used to overcome this constraint.

2.4.1 Initial steps to understand landscape influence on lakes:

Investigation of relationships between landscape factors and lake trophic structure was initiated by Hershey et al. (1999). The authors studied six different fish communities in certain Arctic lakes. The authors observed that geomorphology had a great control on the fish access to lakes via the gradient of outflow and connectivity between lakes. In further research, Hershey et al. (2006) found more landscape factors affecting fish colonization and extinction. Factors like the lake size, the depth, the outflow gradient, the distance between lakes, and the order of lake exhibited influence on either colonization or extinction of variety of fish species.

Eight Arctic lakes were studied by Whalen et al. (2006) with respect to benthic productivity. It was observed that benthic production was largely controlled by lake morphometric properties. The authors also stated that landscape settings such as glacial geology were responsible for lake morphometry, thus, having indirect control over benthic production. In their research it was observed that behavior of lakes situated on

solifluction deposits and meltwater deposits was found to be different with respect to primary production. Thus, terrestrial landscapes play a major role in lake productivity.

In recent studies of arctic lakes, it was also noted that allochthonous material was largely the source of dissolved organic carbon (DOC) affecting primary productivity. DOC was acting as an important source of carbon for pelagic zooplanktons as well as benthic trophic structures of the lakes (Kritzberg et al., 2004; Pace et al., 2004; Grey et al., 2004). It was converted into depleted dissolved inorganic carbon (DIC), which algae used for photosynthesis (Lennon et al., 2006). Neff and Hooper (2002) stated that lakes in the Arctic Foothills experienced high lake-catchment interaction rate and received more DOC. The amount of DOC was proposed to be correlated with amount of shrub cover in the catchments (Sturm et al., 2001). It was also observed that the arctic lakes are nitrogen limited. The nitrogen availability was found to vary with landscape geology and water retention time of the lakes (Whalen et al., 2006).

Using the existing knowledge about the landscape factors outside the Arctic region as well as extracting information from the limited literature available for the Arctic region, the following landscape factors were included in the current research:

1. Proportion of different vegetation communities (expressed in terms of percentage)
2. Shape Indices – Patch Shape Index, Landscape Shape Index, Fractal dimension index
3. Fragmentation and distribution of vegetation communities – Patch density, Edge density

4. Buffer level indices – The proportion, Shape indices and Fragmentation indices calculated at watershed level were also calculated just for 20 meter buffer zones of probable ephemeral streams

Detailed description about derivation these indices is provided in the “Methods” section of the dissertation.

CHAPTER III

METHODOLOGY

3.1 Study Area description

Figure 4. Location map of the study area

The area under consideration was the Toolik Lake region (68° 38’ N/ 149° 36’ W), situated in foothills of the Brooks Range in the northern slope, Alaska (Figure 4). The average local relief of the area is 750 meters. Hillocks, exposed barren areas, and

moraines characterize the entire landscape. Water tracks, streams, and rivers, along with different types of lakes, dissect the area. The Sagavanirktok, Toolik, Itlkilik, and Kuparuk rivers comprise the major rivers in the study area.

This area experienced multiple glaciations in mid-Pleistocene and late Pleistocene era. The Sagavanirktok river area is situated on mid-Pleistocene glacial drift, whereas the Toolik Lake area is on a younger drift of late Pleistocene (Hamilton, 1986). Age of these glacial drifts has played an important role with respect to soil pH, with older drifts showing more acidic soils (Walker et al., 1994).

Looking at the average from the past twenty years, the mean annual temperature of the region is 10.6^{o} F whereas the average temperature for June is up to 39.2^{o} F. Temperature generally raises to 46.0^{o} F in July and starts cooling down after August. The mean temperature in August was recorded as 42.7^{o} F and the mean annual precipitation for past twenty years is 5.97 inches, out of which 33% was received in the form of snow. Hydrologic activity can be observed only during summer (Alaska Climatology, 2008).

3.2 Lake characteristics

The Arctic Alaska is a land of lakes. They are prevalent in, both, coastal and inland areas. Arctic lakes vary in the nature of their origin i.e. kettle, moraine, ice-scour lakes, thermokarst lakes are common types of lakes (Woo and Xia, 1995; Hartman and Carlson, 1973; Woo et al., 1981). These lakes are very closely linked with climatic conditions (Doran et al., 1996; Welch et al., 1987). Long duration of ice cover, lake-ice

melting, snowfall and other climatic conditions control the dynamics of lake ecosystems in the Arctic region and consequently their primary productivity (Magnuson et al. 2006; Smol et al., 2005; Schindler and Smol, 2006). Arctic lakes are the most oligotrophic lakes in the world (Whalen and Alexander, 1986). The sensitive arctic lakes are warming up more rapidly because of global warming (Prowse et al., 2006).

3.3 Data

A SPOT (Satellite Pour L'Observation de la Terre; 5 HRVIR) image acquired in July 25th, 2005 was used for this research. This image was PAN sharpened to a 5 meter spatial resolution and it had 3 spectral bands: Band1 (0.50 – 0.59 μm), Band2 (0.61 – 0.68 μm) and Band3 (0.79 – 0.89 μm). The image encompassed approximately a 60 x 60 square kilometer area surrounding the Toolik region. A UTM projection, zone 6, was applied to geo-rectify the image (Figure 1). 5 meter x 5 meter resolution Digital Elevation Model (DEM) was used in this study to delineate watersheds.

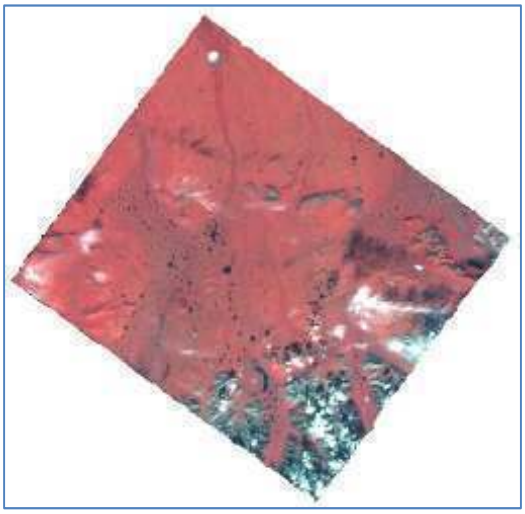


Figure 5. Snapshot of the SPOT image of the study area

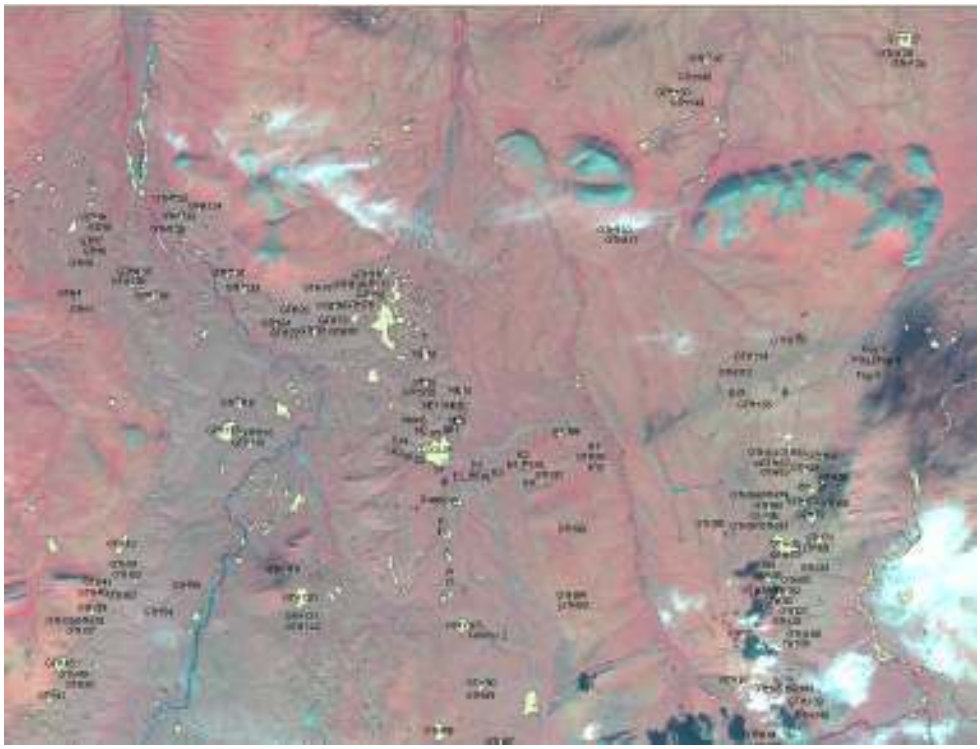


Figure 6. Locations of lakes included in the research

3.3.1 Lakes sampled

Sampling dates for all 41 study lakes	
Lake #	Sample Date
115, 116, 117, 118	29 June 2001
119, 120, 121, 122	7 July 2001
127, 128, 129, 130	8 July 2002
131, 132, 133, 134	10 July 2002
135, 136, 137, 138	12 July 2002
139, 140, 141, 142	28 June 2003
143, 144, 145, 146	3 July 2003
147, 148, 149, 150	14 July 2003
151, 152, 153, 154	24 July 2003
E4	28 July 2003
123, 124, 125, 126	4 August 2003

Table 1. Lakes and Sampling season

3.3.2 Lake parameters

Morphometric characteristics	Maximum depth in meters, surface Area in hectares, perimeter in meters, and shoreline development factors
Landscape position	Lake order
Physical and biological characteristics	Fish richness, light attenuation coefficient in m^{-1} euphotic zone depth in meters
Chlorophyll a content ($\mu g\ L^{-1}$)	
Total Nitrogen $\mu m\ L^{-1}$	
Total Phosphorus $\mu m\ L^{-1}$	
Chemical properties	Surface temperature in °C, dissolved oxygen (DO) in $mg\ L^{-1}$, pH, alkalinity in $meq\ L^{-1}$, conductivity in $\mu S\ cm^{-1}$
Cations and Anions $mg\ L^{-1}$	

Table 2. Lake Properties

Lakes (Table 1) were sampled over three different seasons i.e. summer season of 2001, 2002, and 2003.

Following section illustrates various landscape parameters derived using the SPOT image, NDVI, and Digital Elevation Model of the study area.

3.4 Land cover derivation

As there were no man-made structures present in the given watersheds, only land cover categories composed of different vegetation communities were crucial for this study. Land cover categories were further used to estimate parameters such as landscape

metrics using Fragstat 3.3. The adopted land cover scheme was a combination of classification systems used by Walker et al. (1994) and field modifications done by Roy Stine and Peter Ray (Table 3).

Class	Description
Barren	barren surfaces, sparsely vegetated, rocks covered with lichens
Moist Acidic tussock tundra complex	*Eriophorum vaginatum, Carex bigelowii, Betula nana, Salix pulchra, Sphagnum spp.*
Moist Non-acidic tussock tundra complex	*Salix reticulate,Saxifraga oppositifoli, Cares bigelowii, Carex membranacea, Dryas integrifolia, Ledum decumbens, Equisetum arvens*
Shrub complex	*Betula spp., Salix spp., Sphagnum spp.*
Riparian complex	*Eriophorum anguistifolium, Salix pulchra, Salix alaxensis , Salix richardisonii,*
Fen complex	*Carex rariflora, Carex rotundata, and mosses like Sphagnum spp, Carex chordorrhiza, Carex aquatilis and mosses like Tomentypnum nitens*
Heath complex	*Festuca altaica, Empetrum hermaphroditum, Loiseleuria procumbens, Dryas octopetala, along with Cassiope tetragona, Salix phlebophylla*
Snowbed complex	*Cassiope tetragona, Salix rotundifolia, Arnica frigida*
Mountain Meadow complex	*Carex biglowii, Cassiope tetragona, Salix Chamissonis*
Aquatic vegetation complex	Similar to fen but on lake-fringes
Water	Lakes, streams and rivers
Clouds	
Shadows	

Table 3. Vegetation community categories and species composition

These categories were derived using the SPOT data and secondary data, including a geology map, the Normalized Difference Vegetation Index (NDVI), and others.

Initially, watersheds were developed using DEM and then the SPOT image was cut according to them.

3.4.1 Ground truth data collection

The current research involved forty watersheds, which were sampled for lake water chemistry (GTH 115 to GTH 154 and E4). To visit every watershed was not feasible economically and logistically. Therefore, to obtain representative samples, the following watersheds were selected on the basis of pre-existing knowledge of vegetation communities and till surface variability.

GTH 135 - Itkillik II initial

GTH 133 - Itkillik II initial

GTH 153 – Sag Main Phase

GTH 149 – Sag Main Phase

GTH 120 – Itkillik II initial

GTH 144 – Itkillik I

While creating the database, till age categories were assigned ranks from 1 to 5. 5 being the oldest and 1 being the youngest among these categories.

Stratified random sampling method was adopted to generate X and Y coordinates of sampling locations within each watershed. For that a tentative unsupervised classification was performed on each image using 10 classes and minimum of 6 points

per class. During actual field work the maximum possible sampling points were visited. There were no sampling points generated for Snowbed complex as it was not classified at this stage.

At each sampling point, photos of major vegetation complexes were taken along with some landscape view photographs of watersheds. Field notes about the vegetation communities observed and photos were used later to determine the category of vegetation community at each sampling point. Total 201 points were collected during summer 2008 field work.

3.4.2 Derivation of watersheds from DEM

Hydrology tools and Conditional tools in the Spatial Analyst extension of ArcGIS 9.3 were used to derive watersheds. For each lake an appropriate portion of the entire DEM was cut. It saved the processing time for further processes. Using the portion of DEM as input, Flow direction and Flow Accumulation surfaces were derived. Flow direction creates a raster of flow direction from each cell to its steepest downslope neighbor. Flow accumulation creates a raster of accumulated flow to each cell based on the flow direction raster. As the higher values depict likely places of higher flow accumulation, they were filtered using a threshold value i.e. Con Flow process. The threshold value opted for was 500. By overlaying the conditional layer on the SPOT image, pour point or exit location for outflow of each lake was determined separately. Finally using the pour point and the flow direction raster, watersheds were derived for each lake.

The watershed for GTH 146 was not derived as it did not generate the required flow accumulation network properly and pour point for the lake could not be established. The reason behind the inability to derive the watershed was the poor quality of the DEM. For GTH 143 and 145, DEM at 30m X 30m was used as they were out of the coverage of finer 5m x 5m resolution DEM. Figure 7 outlines the steps used to create the watersheds:

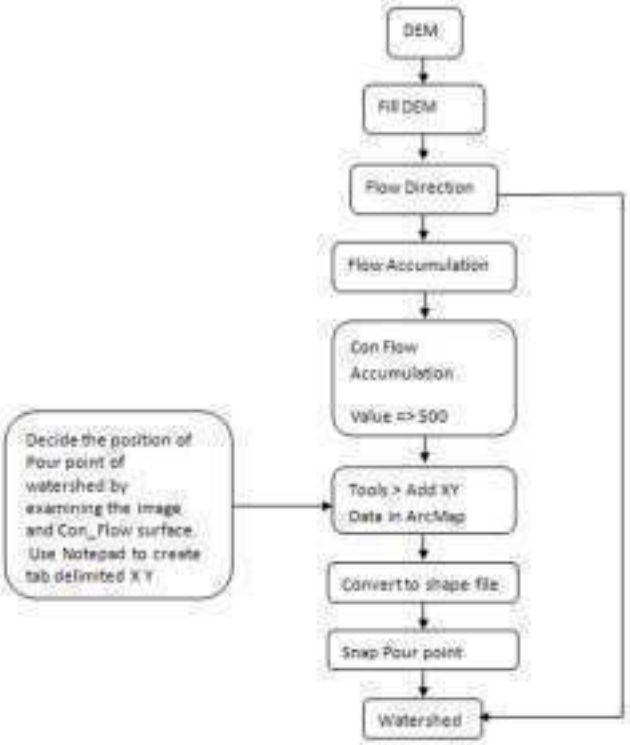


Figure 7. Steps to derive Watersheds

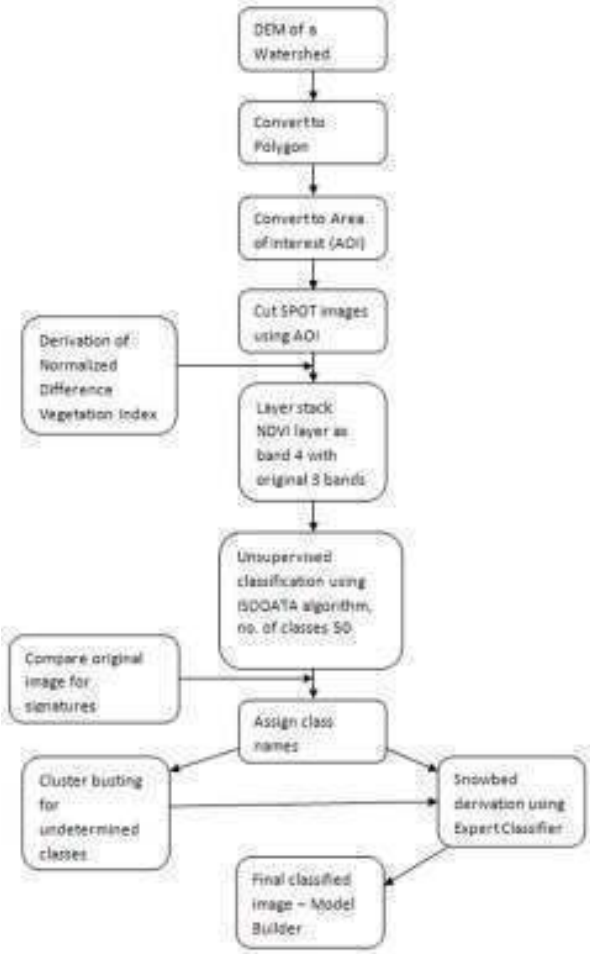


Figure 8. Steps for Land cover classification

3.4.3 Classification of land covers

Using the watersheds derived from the DEMs for each lake, the SPOT image was cut. Normalized Difference Vegetation Index (NDVI) layer for each watershed image was calculated and stacked together with its three spectral bands. Several studies have shown that inclusion of vegetation indices like NDVI layer has improved the quality of vegetation classification (Wolter et al., 1995; Friedl et al., 1999; Hansen et al., 2000). NDVI ratio reduces multiplicative noise such as Sun illumination, topographic variation,

and differential Sun illumination (Jensen, 2005). It is also directly related to physical properties of vegetation such as vigor and above ground biomass (Gamon et al., 1995; Dong et al., 2003).

NDVI for the SPOT image was calculated for each image.

$$\text{NDVI} = \frac{\text{Band 3 - Band 2}}{\text{Band 3 + Band 2}} \quad \text{or} \quad \frac{\text{NIR - Red}}{\text{NIR + Red}} \qquad \text{(Jensen, 2005)}$$

The values of NDVI ratio range between -1 to +1; negative values indicating no vegetation and positive values indicating vegetation. Positive values, as mentioned earlier, are positively correlated with vegetation physical properties like biomass.

Based on the previously carried out field work (data collected during summer 2007 and summer 2008) field knowledge, spectral signature of each class was determined (Figure 9).

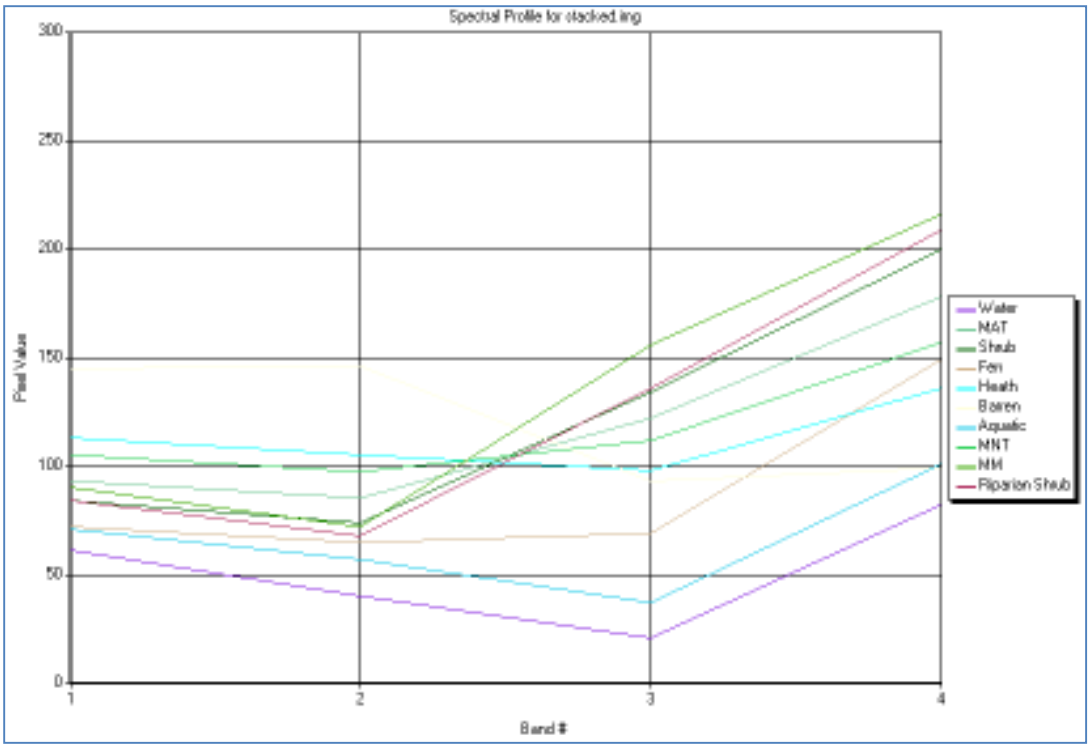


Figure 9. Spectral signatures of land covers

Unsupervised classification of remote sensing data has been useful in deriving land use/land cover (Loveland et al., 1999). Jensen (2005) has described this method as dividing the remote sensing data using its inherent spectral grouping. A *posteriori* assignment of clusters to real world land cover is carried out by analyst. ISODATA clustering is the most commonly used and effective algorithm. It was adopted in the current study. This approach requires minimal inputs from user but interpretation of the results is the major task for him/her (ERDAS Field Guide). The book also explains the ISODATA as "Iterative Self-Organized Data Analysis Technique". Spectral distance between different classes is the basis to define clusters. In the first iteration, ISODATA compares the Euclidean distance of each pixel from preliminary mean vectors and assigns

them to respective cluster. From second iteration onwards, mean vector is recalculated which is used to rearrange the pixels for forthcoming iteration. Jensen (2005) has cautioned that enough number of iterations should be allowed to obtain good classification.

On the basis of field knowledge and spectral signatures, clusters were assigned to a respective class. But some cluster were overlapping in certain watersheds e.g. confusion between MAT and MNT. In that case only clusters with confusion were separated and classified again in appropriate number of classes and were assigned to most logical land cover class. The classified images were then used to run Expert classifier and Snowbed complex class was derived (Figure 10).

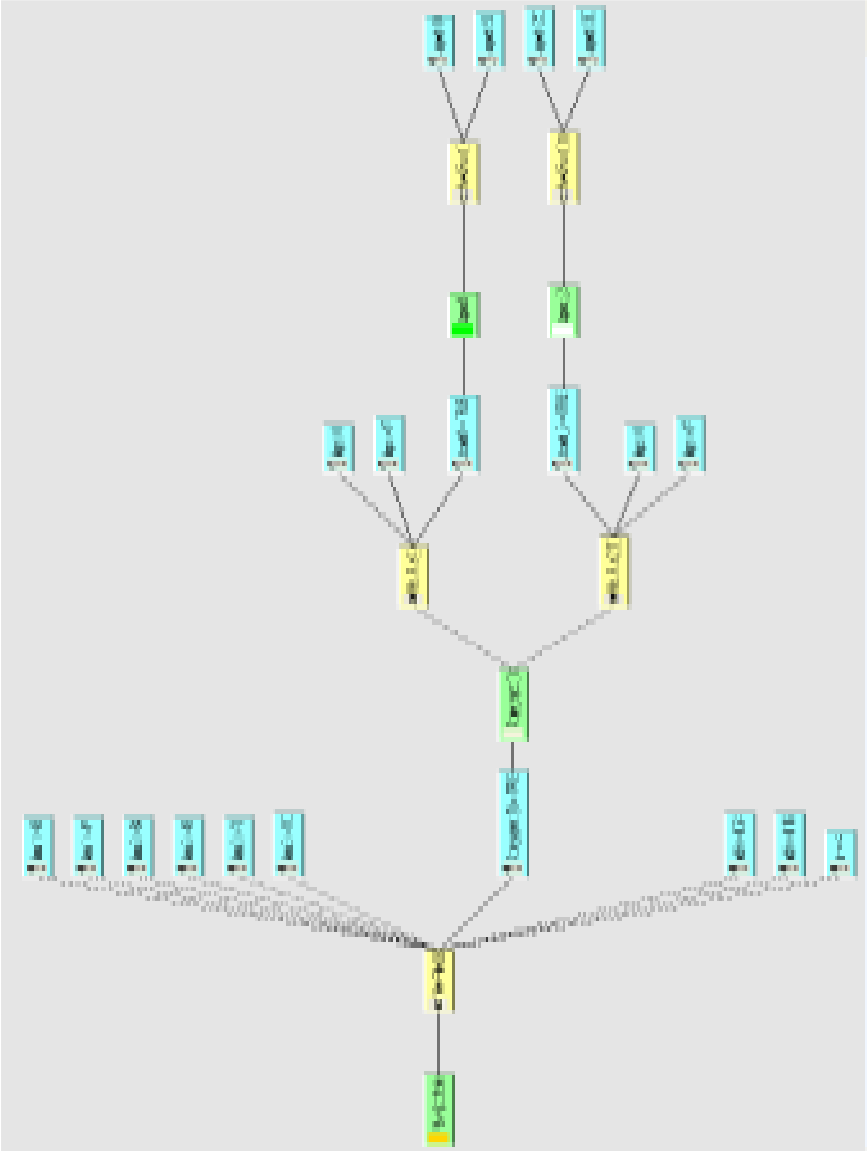


Figure 10. Expert Classification Rule used for Snowbed Derivation

It was decided to use expert classifier in ERDAS Imagine 9.3 to obtain snowbed land cover class. Walker et al. (1994) have described the ground conditions for occurrence of snowbed. Usually snowbeds are north facing and gentle slopes, which allows the portion of the ground to be away from Sun and retain snow. Areas of snowbed

have *Cassiope tetragona* along with *Ledum decumbens* and other prostrate shrubs. Settings of snowbed complex on ground make it difficult to identify them on satellite image especially in oblique sun angels experienced at Arctic. Hence, as mentioned by Stine et al. (2010), spectral characteristics of snowbed complex are strongly overlapping that with Fen complex and sometimes shadows and water.

However, the knowledge about snowbed was used to negate the problem by implementing building knowledge base and using it for classification purpose. The North facing characteristic was adopted through Aspect layer derived from DEM of the study area. Aspect values indicated by following range values depicted North in the expert rule for snowbed: (0 – 90), (270 – 360). Gentle slope conditions were incorporated by including slope layer. Slope values below 17 degree were used for the purpose. NDVI values and Spectral value ranges for 3 bands in SPOT image were adaptation from Stine et al. (2010). They were as follows:

Band 1 - >= 68 and <= 97

Band 2 - >= 54 and <= 88

Band 3 - > 70 and <=110

NDVI - > = 0.064 and <=0.136

Class = 4 (Fen complex)

As mentioned earlier, Fen complex was overlapping with snowbed complex. Hence, condition to extract snowbed from fen complex was added.

3.4.4 Landscape metrics derivation

Basics concepts of landscape metrics were obtained from Fragstats documentation (Fragstats, Conceptual Background).

1. Percentage:

Once every watershed was completely classified an Area column was calculated. The same column was used to calculate percentage of each land cover within watersheds.

The following metrics depicting Structural properties of the land covers were derived using Fragstats 3.3. To recognize .IMG format of the watershed images, Fragstats required Spatial Analyst extension of ArcGIS.

2. Patch Density -

Patch density is expressed in terms of number of patches of a vegetation type per 100 hectares. It is calculated at class level. Applicable mostly in wildlife studies, it is used as an indicator of fragmentation of a particular habitat which may be important for a specific animal (Gilpin & Hanski, 1991). To analyze if overall distribution properties of any particular land cover class would be controlling nutrient inputs to lakes, patch density was incorporated in this research (Figure 11).

$$PD = \frac{n_i}{A}(10{,}000)(100)$$

(n_i represents number of patches and A is total area of landscape. Later it is multiplied by 10,000 and 100 to convert that into hectares.)

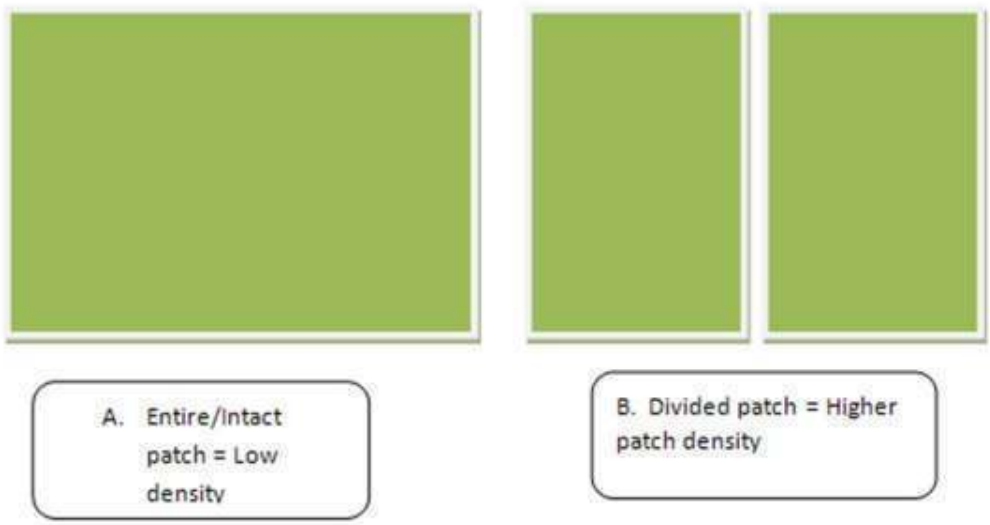


Figure 11. Patch Density: Hypothetical illustration

3. Edge Density –

Edge density is the number of edges of the given class per unit area (meters per hectare). More the edge density, more fragmented the class under consideration is. Similar to Patch Density, Edge density index indicates how fragmented any land cover class is, within the given watershed.

$$ED = \frac{\sum_{k=1}^{m} e_{ik}}{A}(10{,}000)$$

(e_{ik} represents sum of total edge in meters. A is area of landscape in square meteres and it is multiplied by 10,000 to convert into hectares.)

4. Landscape Shape Index –

Landscape Shape Index depicts the total edge of a particular class (if this index is calculated at class level), divided by the minimum possible edge length for that class. Overall it is more direct measure of disaggregation of the land cover class than Patch density or Edge density index.

$$LSI = \frac{E}{\min E}$$

5. Shape Index (Mean) –

Mean shape index (MSI) measures the average patch shape, or the average perimeter-to-area ratio, for a particular patch type (class) or for all patches in the landscape. SHAPE = 1 when the patch is circular (vector) or square (raster) and increases without limit as patch shape becomes more irregular.

$$SHAPE = \frac{0.25\, p_{ij}}{\sqrt{a_{ij}}}$$

(For raster format of inputs p_{ij} represents sum of the patch perimeter, a_{ij} represents patch area, 0.25 is a constant to adjust for a square standard.)

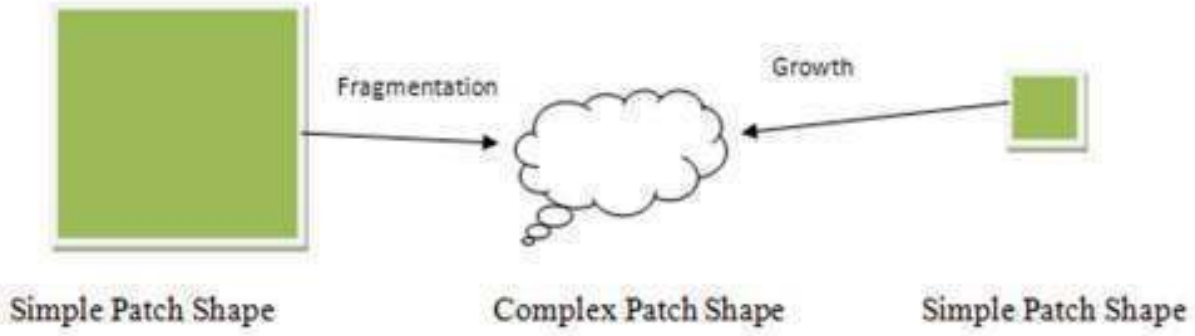


Figure 12. Shape Index: Hypothetical illustration

6. Fractal Dimension Index (Mean) –

Units: None. Range: 1 FRACT 2.

A fractal dimension greater than 1 for a 2-dimensional patch indicates a departure from Euclidean geometry (that is, an increase in shape complexity). FRACT approaches 1 for shapes with very simple perimeters such as circles or squares, and approaches 2 for shapes with highly convoluted, plane-filling perimeters.

Shape indices were included in the current research to obtain indirect clues about active soil depth and probable moisture content of soil. For example, riparian complex zones would occur only near major streams and have greater soil depth compared to other vegetation classes (Giblin et al., 1995). Structural properties e.g. very small patches (Mean Patch Size index) of riparian complex would indicate that only small areas near streams have deeper active soil zone and may either be source or sink of nutrients.

$$FRACT = \frac{2 \ln (0.25\, p_{ij})}{\ln a_{ij}}$$

(p_{ij} represents patch perimeter, a_{ij} represents patch area).

3.4.5 Derivation of water channels and buffer zones:

Buffer zones, streams as well as ephemeral water channels were attributed as hydrologically active zones within watershed by Hunsaker and Levine (1995). The authors stated that landscape metrics such contagion and edge density calculated for land use within buffer zones were positively correlated with pollutant levels. According to the authors, the direct transportation of pollutants via streams as well as interactions within of vegetation root zone in the buffers control the inputs to larger water bodies.

To verify the role of hydrologically active zones within watersheds in the current study, all of the aforementioned landscape indices were also derived for images of stream buffers. Stream network was derived from the DEM. Conditional flow accumulation raster surface for each watershed at value of 500 was matched with field observation. It was found that it follows streams observable from satellite image as well as small streams observed in the field. The smaller streams which were detected in the field may represent ephemeral streams/ streams formed after precipitation (Figure 13).

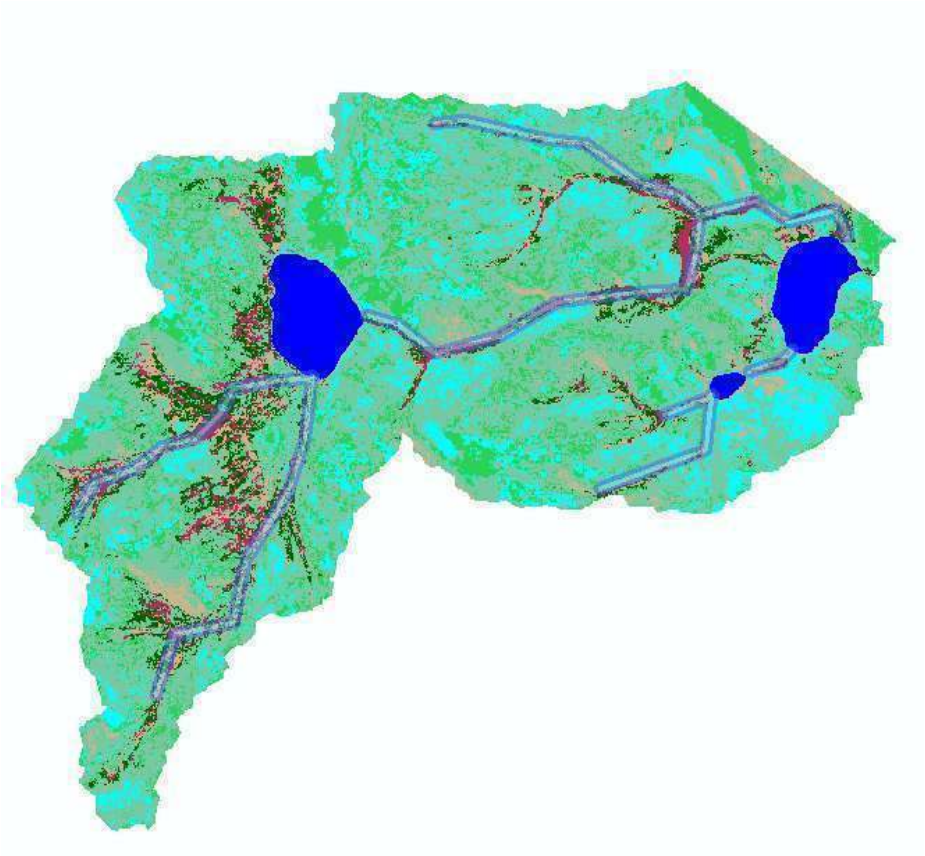


Figure 13. Area of GTH 144 depicting buffer zones derived using DEM

Horizontal distances of 10, 20, and 50 meters from water channel network were derived for buffer zones. Buffer size of 50 meters was going out of the extent of certain watersheds whereas the landscape measurements in 10 meter buffer zones were too small to be useful in modeling. Hence, 20 meter buffer size was considered as an optimal for the use.

3.4.6 Lake order

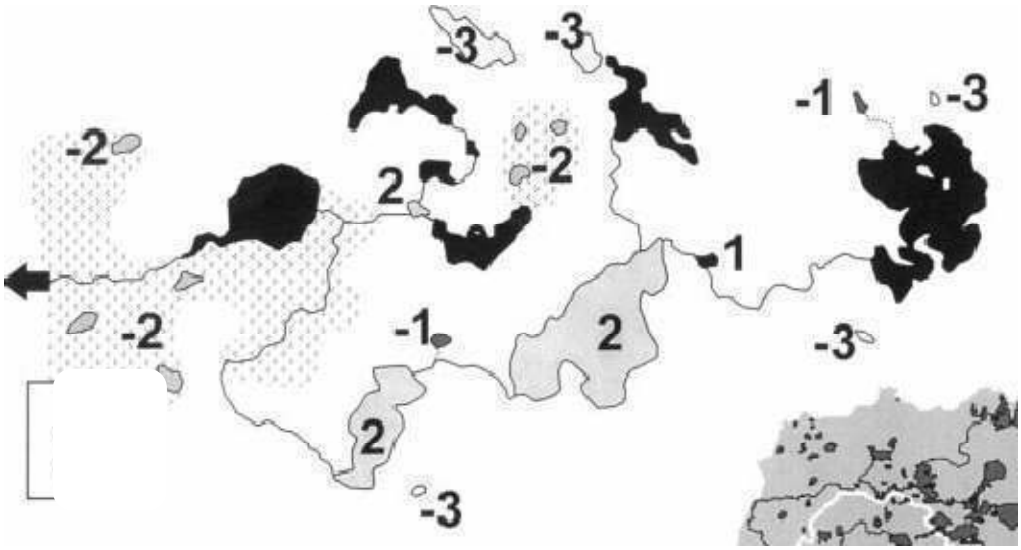


Figure 14. Lake Order

As explained by Riera et al. (2000) the streams which are clearly visible on the image as well as topomap were ordered using Strahler's scheme. Then the lake whose outlet was of stream order 3, received an order 3. Similarly lakes with outlet of order 2 were assigned lake order 2. Lakes, whose outlet was order 1, were assigned to lake order 1. Following the illustration provided by the authors, headwater lakes whose catchment is small compared to lake order 1, were classified as of order 0. Lakes which drained by temporary surface drainage were assigned -1 order whereas lakes which were drained into other streams via wetland were of order -2. Lake order -3 was assigned to lakes disconnected with surrounding hydrological units.

Using topomaps of the area, lake orders were decided (Hershey et al., 2006). Thirteen lakes were of order 0, 14 lakes were of order 1 and 7 received order 2. Only 2 lakes were of order 3. GTH 154 was of order -3 where as GTH 121 and 122 got order -1.

3.4.7 Topographic Wetness Index

Based on Wolock, (1993) a python program was developed to calculate the wetness index. The script is provided in the Appendix D.

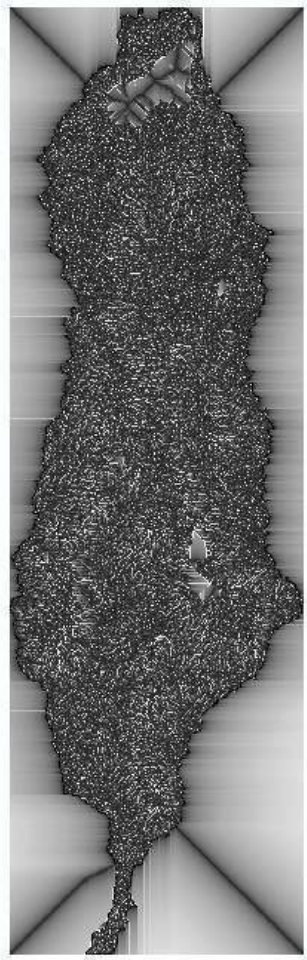


Figure 15. GTH 138 Topographic Wetness Index

Higher brightness values represented higher TWI values indicating probability of more soil moisture in those pixels. Overall the topographic wetness index followed the pattern of probable stream network in watersheds. But it had very severe salt-paper effect. The value of TWI ranged -2.05 to 23.41. But there was not much variation in TWI from watershed to watershed and did not show significant correlation with any lake chemistry parameter. Hence, it was decided to omit the TWI from further analysis. Better performance could be achieved using higher quality DEM e.g. LiDAR data.

3.4.8 Lake Surface Area (LA), Watershed Area (WA) and Watershed Area to Lake Area ratio

From the Area column of each image lake surface area was calculated. If there were more than 1 lake within watershed, then measurement polygon tool was used to derive the surface area. To derive watershed area total of the Area column was used. Using these readings, ratio of WA and LA was calculated, which was used as surrogate index for water retention time.

3.5 Methods

3.5.1 Accuracy assessment

Verification of the accuracy of a thematic map before its application in any ecological study is essential (Jensen, 2005). Jensen maintains that when a thematic classification is used for scientific studies and policy-making purposes, a statistical figure explaining the reliability of the data are required; if the data will not be cited for such purposes, visual inspection of the data's reliability is enough.

Accuracy assessments were performed in ERDAS 9.3, resulting in an error matrix. An error matrix represents a systematic comparison between the class depicted by a pixel and the ground reference for same location. It also provides information about the type of error (errors of commission and omission) and helps refine classification. If there are K classes, then error matrix is represented by a K X K matrix. Columns of the matrix indicate the ground reference, whereas the rows represent remote sensing image derived class information. The diagonal values indicate the correctly classified number of samples and the other values indicate misclassified pixels. Error matrix results in three types of accuracies: overall accuracy, producer's accuracy, and user's accuracy. *Overall accuracy* is the ratio of correctly classified sample pixels to the total number of samples used for accuracy. When the number of accurately classified pixel of a class is divided by total number of pixel in that column, it is called the *producer's accuracy*. This indicates to the analyst how accurately an area could be classified using this particular classification. When the correct number of pixels in a row is divided by row total for a

category, it represents the *user's accuracy*. It corresponds to the probability that thematic classification matches the ground reference data.

The error matrix is accompanied by Kappa statistics in ERDAS. Kappa statistics represent the agreement between the reference data and classification results (Congalton, 1981). K values between 0.40 and 0.80 represent moderate agreement, whereas values below 0.40 represent poor agreement (Landis and Koch, 1977).

Before the classifications were utilized for the landscape metrics, an accuracy assessment was carried for all of the six watersheds utilized in this research. Details about the accuracies are provided in the "Results and Discussion" chapter.

3.5.2 Statistical analysis methods

Johnson and Gauge (1997) reviewed statistical methods and different landscape approaches to study linkages between landscape factors and stream, river, or lake ecology. The authors mentioned that factors affecting water bodies occur at multiple levels. For example, climate change is generally viewed as global scale phenomenon, or the conversion of certain forest patch to agriculture is considered a local level phenomenon. It is very difficult to analyze such a heterogeneous data in order to understand complex processes. Initially, Hynes (1975) mentioned a strong influence of valleys on streams. Afterwards, with the emergence of remote sensing and GIS technology, it became possible to capture spatial data at various scales with relative ease (Johnson, 1990). Quantitative assessment of landscape factors via understanding their lateral, longitudinal, and vertical properties has become possible because of these

technologies. Combination of them with multivariate statistical analysis packages has provided strong tools for ecological studies (Petts et al., 1995; Puckett, 1995).

As mentioned by Carpenter et al., (1989), a major objective of such ecological studies is to establish empirical relationship between landscape factors and observed limnological phenomenon. But at the same time multicollinearity within landscape factors should not be overlooked. The use of statistical methods, such as ANOVA or PCA, begins with the elimination of highly correlated factors to obtain a set of factors which explain most variance in the data but are independent of each other. Liu et al. (1997) exemplified the use of advanced methods such as Path Analysis in a similar study. However, Johnson and Guage (1997) indicated that regression analysis is the most widely used of the analytical methods, because it is easy to interpret and replicate.

3.5.2.1 Regression analysis

Limnological studies often used regression models to establish relationship between external factors and chemical parameters of water bodies. A regression model was found useful by Webster et al., (1996) to understand impacts of drought on Ca and Mg ion concentration within certain lakes situated in different landscape positions in Wisconsin. Dillon and Molot (1997) used a regression model to successfully establish a relationship between land covers and long term average of dissolved organic carbon, Total Phosphorus, and iron contents in lakes in Ontario, Canada. Linear regression was also implemented by Hiscock et al. (2003) to understand influence of landscape factors on phosphorus loadings into Lake Okeechobee, Florida.

Basics of Regression: According to Rogerson (2006: 170), "Regression analysis is used to specify and test a functional relationship between variables". The linear regression analysis assumes existence of a relationship between the dependent variable and independent variables. Then it proceeds to find a best fitting line through a set of given points to define that relationship. Regressions are useful for predicting changes in dependent variable with respect to changes in independent variables. The difference between the predicted value and actual value of a dependent variable is called a *residual*. The *constant* value in any regression equation represents the minimum possible value of the dependent variable when all independent variables have a value of zero. How much variance is explained by the regression model is illustrated by the r^2 value. The *F test* is used to identify if the regression model is significant or not, at an expected level.

3.5.2.2 Classification and Regression Tree Analysis

Another method used in the study was Classification and Regression tree (CART). CART is also trade name for the software provided by Salford Systems, claiming that it is the real Classification and Regression Tree software. CART method builds a decision tree to suggest a way to classify data or predict values of dependent variable based on the pattern observed. It is also known as binary recursive partitioning method as it always divides parent node into exactly two child nodes using a condition called splitting criteria. A classification tree can be developed using categorical dependent variable whereas regression trees require continuous or discrete numerical dependent variable.

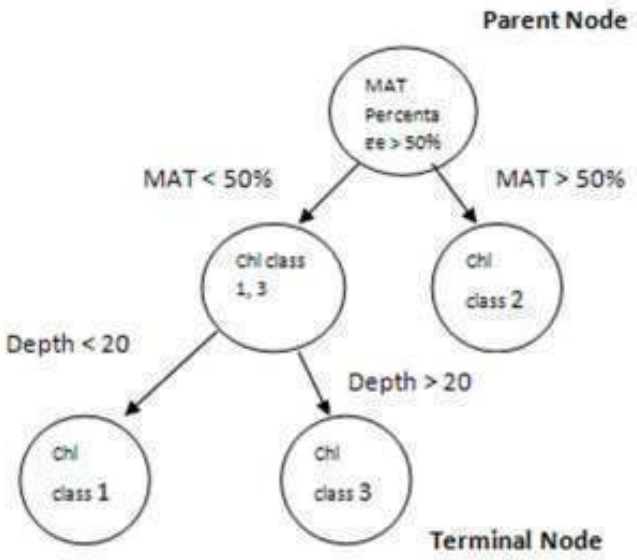


Figure 16. Hypothetical example of CART structure and related terms

CART allows the user to specify target and predicting variables. To interpret the developed trees following terms are used:

Parent node: Any node which could be split further to diversity the tree is Parent node.

Child node: Except the initial parent node, each node which is formed because of division is a child node.

Pure node: A terminal node which represents a condition or cases pertaining to only one particular class or criterion.

Splitting criterion: Value used to divide parent node into child nodes is called splitting criterion. It may be either a True/False condition or a numerical value.

V-Fold cross-validation: According to CART manual, 3000 or more cases are considered to be enough to separate data into learning and testing samples. But when the records are not enough, cross validation method is used. This allows for the use of the entire dataset to build a tree and to use for testing. Recommended value for V-fold cross validation is 10 i.e. 10 different trees are built and tested on 10% of randomly selected data. Results of the validation are summarized in a table.

Tree development models: Different models are available in CART software which allow user to specify splitting rule. The splitting rule is a strategy used to grow a tree. CART manual has a brief explanation for different rules. Gini is default rule for classification tree and has been suggested as the best method for it.

Class probability: It is based on Gini rule but deals with probability of tree structure formation. Major differences between classification tree and regression tree are that the classification tree uses a value of predicting the variable to split a node. Unlike the classification tree, the regression tree calculates statistics for the parent node and uses mean values to form child nodes. When the classification tree is generated, the navigator shows relative cost of tree formation. However when the regression tree is developed, the relative error is shown.

Relative cost/Relative error: Tree navigator displays cost or error value curve for each tree formed. The value ranges from 0 to 1. 0 being no error or perfect fit and 1 being totally random event of tree formation, give an idea about all the trees formed.

ROC (Receiver Operating Characteristic Curve): ROC curves are used to summarize the performance of a model. ROC curves are generated when cross validation method is used for testing. ROC can range between 0 and 1, and unlike the error value, when ROC is higher, the model performance is better.

Gain: Indicates how many percent cases are identified at any particular node. It usually focuses on one class at a time when showed in a Gain chart.

Profit: Similar to gain and it is used to explain profit of information at each node. Thus, each record that is sorted out for a particular node adds to the profit of that node. It is generated when regression tree is calculated.

Prediction success: It is a confusion matrix representing the actual class and predicted class by a tree. Prediction success table for Learn and Test sample is generated if the classification tree is developed. The confusion matrix helps identify the misclassified cases.

An optimal tree is decided based on multiple criteria like Relative cost, ROC value, as well as prediction success.

There are many advantages of classification and regression tree over traditional regression. CART does not assume normal distribution of data. Tree methods are known as non-parametric and non-linear; therefore, they provide more flexibility to user. It can also handle missing values, and it can incorporate categorical variables as well.

Additionally, it provides a hierarchical structure of the data, which makes the results easier to interpret.

There are many examples showing usefulness of CART in the medical field (Poon et al., 2001; Camp and Slattery, 2001; Jazbec et al., 2007). However, CART is useful in ecological studies; for example, Pesch et al., (Article in Press) employed CART to classify terrestrial and marine environments using data within GIS domain. Spruill et al. (2002) found classification tree very useful in indentifying source of nitrogen pollution in ground water from multiple sources such as contributions from agriculture, poultry, septic system, and other similar sources. Using certain indices like the ratio of sodium to potassium, just N concentration, nitrate to ammonia ratio and zinc as predicting variables, authors achieved more than 80% success in identifying the sources of nitrogen. Impacts of environmental factors within Everglades Agricultural Area (EAA) on phosphorus loading into oligotrophic lakes in Florida were studied by Grunwald et al. (2009) using the tree method. Using 10 different predictors like water management, nature of agriculture, and others, a tree based analysis was carried out. It suggested that hydrology/water management is the key factor controlling P loadings into surrounding lakes. The authors claimed that not only successful prediction, but identification of influencing factors, was a major advantage of the tree-based method.

Hershey et al. (2006) used CART to predict presence or absence of certain fish species in Arctic lakes. The authors explored various landscape factors like outflow

gradient, lake connectivity, age of glacial till, lake depth, and lake size with respect to survival and distribution of certain fish species.

In the current research, to explore the possibility of simultaneously affecting landscape factors at different priority levels, CART was included. Both, regression and CART analyses methods were employed to identify important landscape factors. In the "Results" chapter, more details could be found about implementation and interpretation of both methods.

CHAPTER IV

RESULTS AND DISCUSSION

As described in the Methods chapter, six watersheds were visited to collect ground truth data of land cover types. Prior to calculation of landscape metrics and statistical analyses, accuracy assessments of classified images of those six watersheds were carried out. Results of the accuracy assessment are provided in the initial portions of this chapter. In the later sections of this chapter, regression models and classification trees obtained for each lake chemistry variable are described.

4.1 Results of Accuracy Assessment

For the six watersheds, overall accuracies ranged from 82.29% to 95%, while overall Kappa values ranged from 0.78 to 0.92. In general misclassifications were observed between MAT complex and MNT complex. Assessment outcome for individual watershed has been illustrated in the following paragraphs.

a. GTH 120

As depicted by Figure 16, in total there were eight land cover categories besides Water in GTH120. The error matrix obtained for the watershed has been represented by Table 4 while Table 5 represented the accuracy details.

Figure 17. Thematic map of GTH 120

An overall accuracy for GTH 120 was 95% with a Kappa value of 0.93. However, certain land cover types experienced misclassification. Shrub complex was classified as Fen, whereas Snowbed complex was assigned to Fen complex (Table 4).

Classified Data	Water	AV	Shrub	Fen	MAT	MNT	Heath	Therm	Snowbed	Row total
Water	**3**	0	0	0	0	0	0	0	0	3
AV	0	**1**	0	0	0	0	0	0	0	1
Shrub	0	0	**4**	0	0	0	0	0	0	4
Fen	0	0	1	**9**	0	0	0	0	1	11
MAT	0	0	0	0	**14**	0	0	0	0	14
MNT	0	0	0	0	0	**4**	0	0	0	4
Heath	0	0	0	0	0	0	**2**	0	0	2
Therm	0	0	0	0	0	0	0	**2**	0	2
Snowbed	0	0	0	0	0	0	0	0	**1**	1
Column Total	3	1	5	9	14	4	2	2	2	

Table 4. Error Matrix for GTH 120

Class Name	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy	User's Accuracy
Water	3	3	3	100.00%	100.00%
AV	1	1	1	100.00%	100.00%
Shrub	5	4	4	80.00%	100.00%
Fen	9	11	9	100.00%	81.82%
MAT	14	14	14	100.00%	100.00%
MNT	4	4	4	100.00%	100.00%
Heath	2	2	2	100.00%	100.00%
Therm	2	2	2	100.00%	100.00%
Snowbed	2	1	1	50.00%	100.00%

Overall Classification Accuracy = 95%

Table 5. Accuracy Table for GTH 120

Overall Kappa Statistics = 0.9282

Class Name	Kappa
Water	1.0000
AV	1.0000
Shrub	1.0000
Fen	0.7597
MAT	1.0000
MNT	1.0000
Therm	1.0000
Snowbed	1.0000

Table 6. Kappa statistics for GTH 120

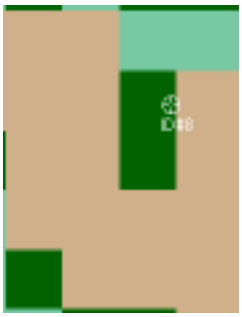


Figure 18. Misclassification GTH 120 Shrub (Reference) to Fen (Assigned)

Figure 19. Misclassification GTH 120 Snowbed (Reference) to Fen (Assigned)

As illustrated in Figure 17, the ground truth point was located in the Shrub complex patch occupying merely two pixels. It was surrounded by the Fen complex and hence, ERDAS include the ground truth point into Fen complex due to majority rule. Similar condition was noticed for the misclassification of Snowbed complex (Figure 18).

b. GTH 133

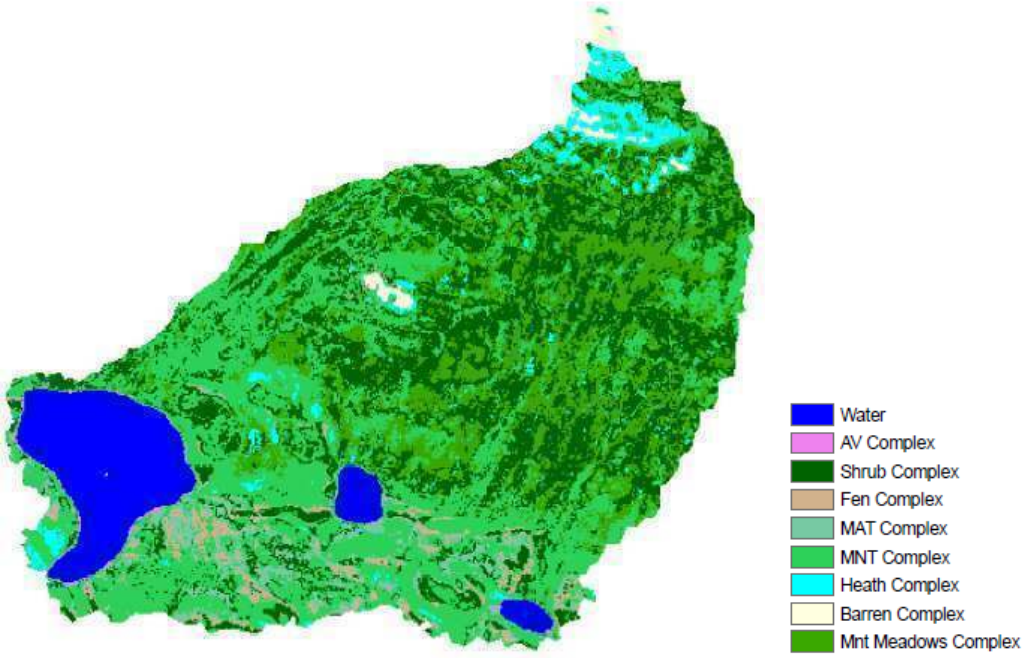


Figure 20. Thematic map of GTH 133

Classified Data	Water	Shrub	Fen	MAT	MNT	Heath	Barren	MM	Row Total
Water	3	0	0	0	0	0	0	0	3
Shrub	0	10	0	0	0	1	0	1	12
Fen	0	0	1	0	0	0	0	0	1
MAT	0	0	0	6	0	0	0	0	6
MNT	0	0	2	0	7	0	0	0	9
Heath	0	0	0	0	0	6	0	0	6
Barren	0	0	0	0	0	1	2	0	3
MM	0	0	0	0	0	1	0	3	4
Column Total	3	10	3	6	7	9	2	4	

Table 7. Error Matrix for GTH 133

Class Name	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy	User's Accuracy
Water	3	3	3	100.00%	100.00%
Shrub	10	12	10	100.00%	83.33%
Fen	3	1	1	33.33%	100.00%
MAT	6	6	6	100.00%	100.00%
MNT	7	9	7	100.00%	77.78%
Heath	9	6	6	66.67%	100.00%
Barren	2	3	2	100.00%	66.67%
MM	4	4	3	75%	75%

Overall Classification Accuracy = 86.36%

Table 8. Accuracy Table for GTH 133

Overall Kappa Statistics = 0.8221

Class Name	Kappa
Water	1.0000
Shrub	0.7796
Fen	1.0000
MAT	1.0000
MNT	0.7320
Heath	1.0000
Barren	0.6496
MM	0.7230

Table 9. Kappa Statistics for GTH 133

Thematic map of GTH 133 exhibited an overall accuracy 86.36% (Figure 19, Table 8), and the Kappa value was 0.8221 (Table 9). Shrub complex was confused with Mountain Meadow complex in one instance. This may primarily be due to the fact that both of the classes reflected high in Infra red band and have similar NDVI range. Similarly, spectral overlap was responsible for misclassifying certain Heath pixels as

Barren. Fen complex was classified as MNT complex in two instances. These misclassifications between Fen and MNT complex were due to wrong assignment of "Class Value" of pixels by the software.

c. GTH 135

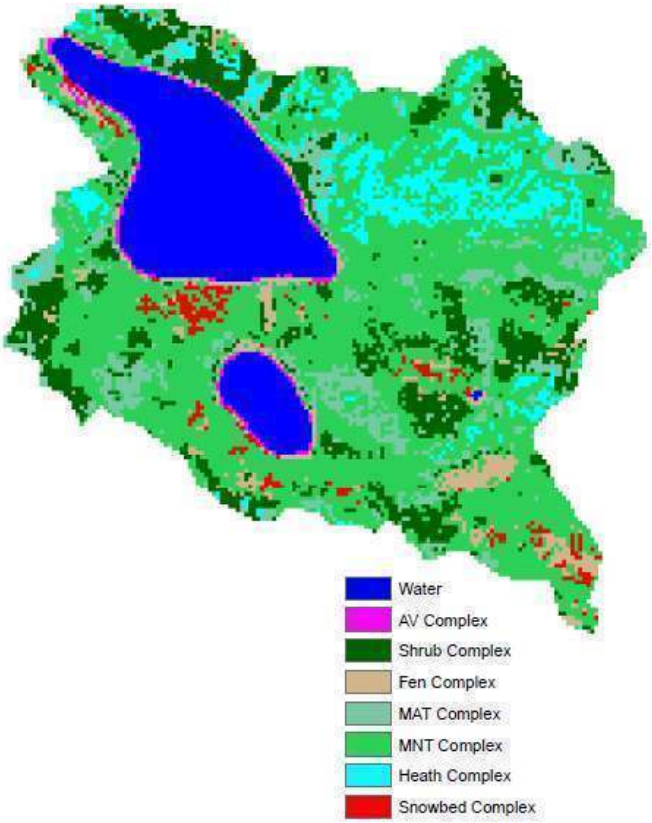


Figure 21.Thematic map of GTH 135

Classified Data	Water	Shrub	Fen	MAT	MNT	Heath	Snowbed	Row Total
Water	2	0	0	0	0	0	0	2
Shrub	0	5	0	0	0	0	0	5
Fen	0	0	4	0	0	0	0	4
MAT	0	0	0	3	0	0	0	3
MNT	0	0	1	1	6	0	1	9
Heath	0	0	0	1	1	2	0	4
Snowbed	0	0	0	0	0	0	2	2
Column Total	2	5	5	5	7	2	3	

Table 10. Error Matrix for GTH 135

Class Name	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy	User's Accuracy
Water	2	2	2	100.00%	100.00%
Shrub	5	5	5	100.00%	100.00%
Fen	5	4	4	80.00%	100.00%
MAT	5	3	3	60.00%	100.00%
MNT	7	9	6	85.71%	66.67%
Heath	2	4	2	100.00%	50.00%
Snowbed	3	2	2	66.67%	100.00%

Overall Classification Accuracy = 82.75%

Table 11. Accuracy Table for GTH 135

Overall Kappa Statistics = 0.7833

Class Name	Kappa
Water	1.0000
Shrub	1.0000
Fen	1.0000
MAT	1.0000
MNT	0.5556
Heath	0.4615
Snowbed	1.0000

Table 12. Kappa Statistics for GTH 135

For GTH 135, the overall accuracy obtained was 82.75% (Figure 20, Table 10) with a kappa value of 0.78 (Table 12). The major misclassification was observed in MAT complex being confused with MNT complex as well as Heath complex.

Being situated on the Itkillik II initial till surface, MNT complex was the dominant land cover type with intermittent patches of MAT complex, in GTH 135. As mentioned by Chaudhuri (2008), the lower sun angle could be responsible for overlapping spectral properties of both land cover types. Similar to GTH 133, there was a misclassification between Fen complex and MNT complex. Along with that, Snowbed complex and MNT complex also exhibited misclassification.

d. GTH 144

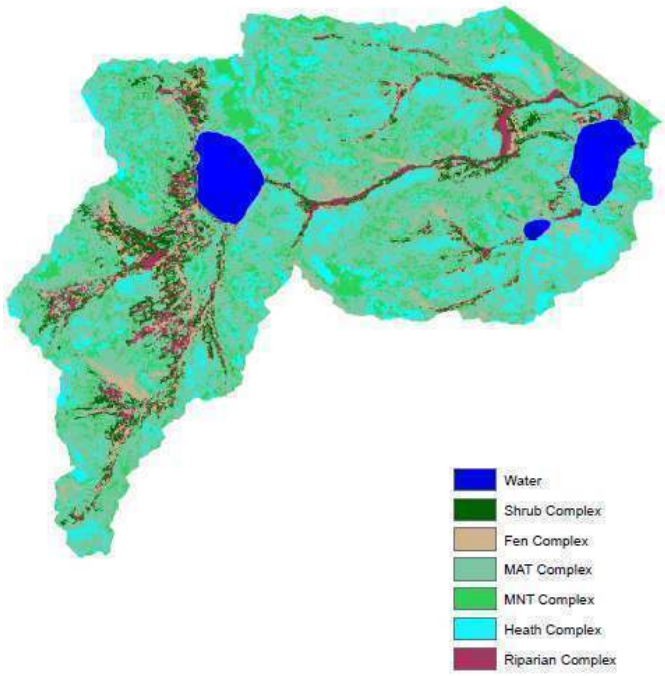


Figure 22. Thematic map of GTH 144

Classified Data	Water	Shrub	Fen	MAT	MNT	Heath	Riparian	Row Total
Water	4	0	0	0	0	0	0	4
Shrub	0	3	0	0	0	0	0	3
Fen	0	1	2	0	0	0	0	3
MAT	0	0	1	14	0	2	0	17
MNT	0	0	0	1	3	0	1	4
Heath	0	0	0	0	0	2	0	2
Riparian	0	0	0	0	0	0	1	1
Column Total	4	4	3	15	3	4	1	

Table 13. Error Matrix for GTH 144

Class Name	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy	User's Accuracy
Water	4	4	4	100.00%	100.00%
Shrub	4	3	3	75.00%	100.00%
Fen	3	3	2	66.67%	66.67%
MAT	15	17	14	93.33%	82.35%
MNT	3	4	3	100.00%	75.00%
Heath	4	2	2	50.00%	100.00%
Riparian	1	1	1	100.00%	100.00%

Overall Classification Accuracy = 85.29%

Table 14. Accuracy Table for GTH 144

Overall Kappa Statistics = 0.7512

Class Name	Kappa
Water	1.0000
Shrub	1.0000
Fen	0.6296
MAT	0.6471
MNT	0.7222
Heath	1.0000
Riparian	1.0000

Table 15. Kappa Statistics for GTH 144

For GTH 144, an overall accuracy of 85.29% (Figure 21, Table 14) was obtained, with a Kappa value of 0.75 (Table 15). The ISODATA classifier could not completely distinguish Fen complex from MAT complex (Table 13). Moreover, two of the Heath complex ground truth points were misclassified as MAT complex. Moreover, in one instance there was confusion between Shrub complex and Fen complex. The probable reason behind these incorrect classifications could be attributed to highly heterogeneous distribution of land cover types within low lying areas of the watershed, which affected the class assignment during assessment process.

e. GTH 149

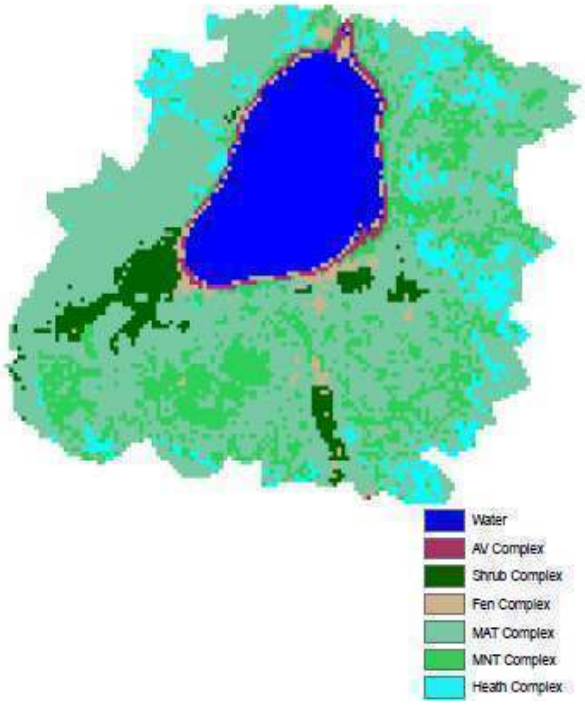


Figure 23. Thematic map of GTH 149

Classified Data	Water	AV	Shrub	Fen	MAT	MNT	Heath	Row Total
Water	2	0	0	0	0	0	0	2
AV	0	1	0	0	0	0	0	1
Shrub	0	0	1	0	0	0	0	1
Fen	0	0	0	3	0	0	0	3
MAT	0	0	0	0	16	2	0	18
MNT	0	0	0	0	0	3	0	3
Heath	0	0	0	0	0	0	4	4
Column Total	2	1	1	3	16	5	4	

Table 16. Error Matrix for GTH 149

Class Name	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy	User's Accuracy
Water	2	2	2	100.00%	100.00%
AV	1	1	1	100.00%	100.00%
Shrub	1	1	1	100.00%	100.00%
Fen	3	3	3	100.00%	100.00%
MAT	16	18	16	100.00%	88.89%
MNT	5	3	3	60.00%	100.00%
Heath	4	4	4	100.00%	100.00%

Overall Classification Accuracy = 93.75%

Table 17. Accuracy Table for GTH 149

Overall Kappa Statistics = 0.8947

Class Name	Kappa
Water	1.0000
AV	1.0000
Shrub	1.0000
Fen	1.0000
MAT	0.7619
MNT	1.0000
Heath	1.0000

Table 18. Kappa Statistics for GTH 149

The GTH 149 classified image obtained an overall accuracy of 93.75% (Figure 22, Table 17) with a Kappa value of 0.89 (Table 18). Misclassification was mainly observed between MAT complex and MNT complex (Table 16). The inherent limitation of the image, i.e. spectral overlap between the two land covers could be the reason behind observed inaccuracy.

f. GTH 153

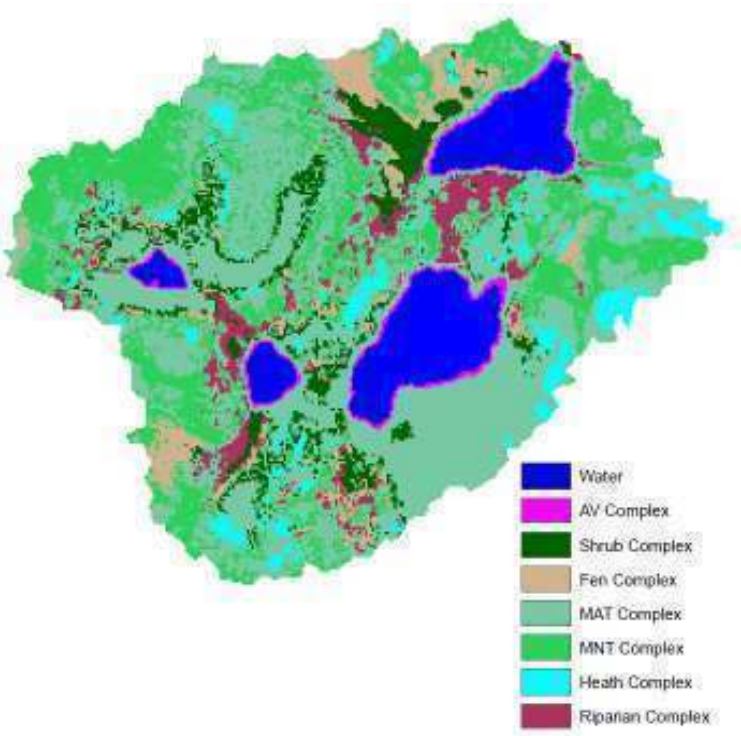


Figure 24. Thematic map of GTH 153

Classified Data	Water	Shrub	Fen	MAT	MNT	Heath	Riparian	Row Total
Water	3	0	0	0	0	0	1	4
Shrub	0	4	0	0	0	0	0	4
Fen	0	0	4	0	0	0	0	4
MAT	0	0	0	7	1	1	0	9
MNT	0	0	0	0	6	0	0	6
Heath	0	0	0	0	0	4	0	4
Riparian	0	0	0	0	0	1	4	5
Column Total	3	4	4	7	7	6	5	

Table 19. Error Matrix for GTH 153

Class Name	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy	User's Accuracy
Water	3	4	3	100.00%	75.00%
Shrub	4	4	4	100.00%	100.00%
Fen	4	4	4	100.00%	100.00%
MAT	7	9	7	100.00%	77.78%
MNT	7	6	6	85.71%	100.00%
Heath	6	4	4	66.67%	100.00%
Riparian	5	5	4	80.00%	80.00%

Overall Classification Accuracy = 88.88%

Table 20. Accuracy Table for GTH 153

Overall Kappa Statistics = 0.8538

Class Name	Kappa
Water	0.7123
Shrub	1.0000
Fen	1.0000
MAT	0.7179
MNT	1.0000
Heath	1.0000
Riparian	0.7643

Table 21. Kappa Statistics for GTH 153

An overall accuracy for this watershed was 88.88% (Figure 23, Table 20), with a Kappa value of 0.85 (Table 21). Even though an acceptable accuracy was achieved, it was highly affected by the shadows of low level clouds. This resulted into misclassification of Heath complex into MAT complex and Riparian complex (Table 19).

4.2 Checking consistency of the Classification based on Classification Tree predictability

It was necessary to check the applicability of these results over other non-sampled watersheds. Hence, consistency of the classification was checked using Classification Tree method. Values of Green (Band1), Red (Band2), and IR (Band3) as well as NDVI at sample points were used as inputs for building Classification Tree.

A classification tree was build for Shrub complex, Fen complex, MAT complex, MNT complex, and Heath complex. As there were fewer points for other classes, they might have affected tree formation process and hence were eliminated for the tree building process.

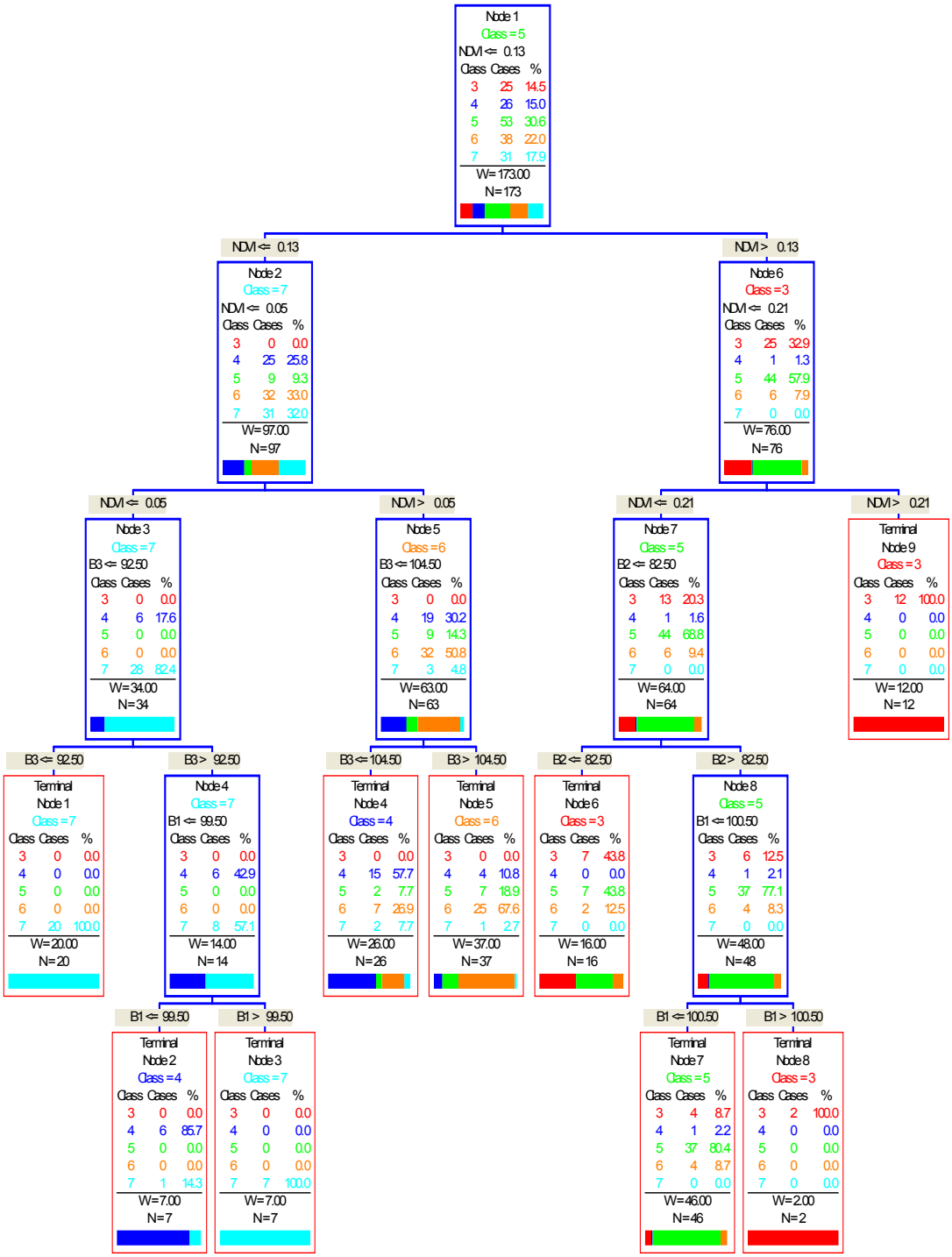


Figure 25. Classification tree for accuracy consistency inclusive of MAT and MNT complex

At the relative cost of 0.435 the classification tree was formed with 9 terminal nodes. Terminal nodes 1, 3, 8 and 9 were pure nodes. A 10 fold cross-validation method was used to verify prediction of success of the classification tree. Prediction success on test data was overall 64.16%. Looking at the misclassification values for the learning dataset it can be concluded that class 5 and Class 6 (MAT and MNT, respectively) were classified at relatively higher cost of 0.3 and 0.34. But for the test data, prediction cost of all the classes remained above 0.3 except for Class 7 (Fen). Relative cost for fen class was only 0.16.

Class	NCases	N Mis-Classed	Pct Error	Cost
3	25	4	16.00	0.16
4	26	5	19.23	0.19
5	53	16	30.19	0.30
6	38	13	34.21	0.34
7	31	4	12.90	0.13

Table 22. Misclassification for Learn Data (Accuracy Assessment extension)

Class	N Cases	N Mis-Classed	Pct Error	Cost
3	25	10	40.00	0.40
4	26	10	38.46	0.38
5	53	24	45.28	0.45
6	38	13	34.21	0.34
7	31	5	16.13	0.16

Table 23. Misclassification for Test Data (Accuracy Assessment extension)

It can be concluded that confusion between MAT and MNT complex was contributing towards higher learning and predicting cost. Thus it follows the observations made from accuracy assessment tables of watersheds.

To check the possibility of improving learning and predictability of the classification tree method, two different files were created. In one file MNT was retained with Shrub, Fen, and Heath (Figure #26), while in other MAT was retained with Shrub, Fen, and Heath points (Figure #27).

Separate classification trees were built using these files. As expected, performance of tree was improved for these segregated files. The tree with MNT was formed at very low relative cost of 0.293. Prediction success of test sample was overall 77.5%.

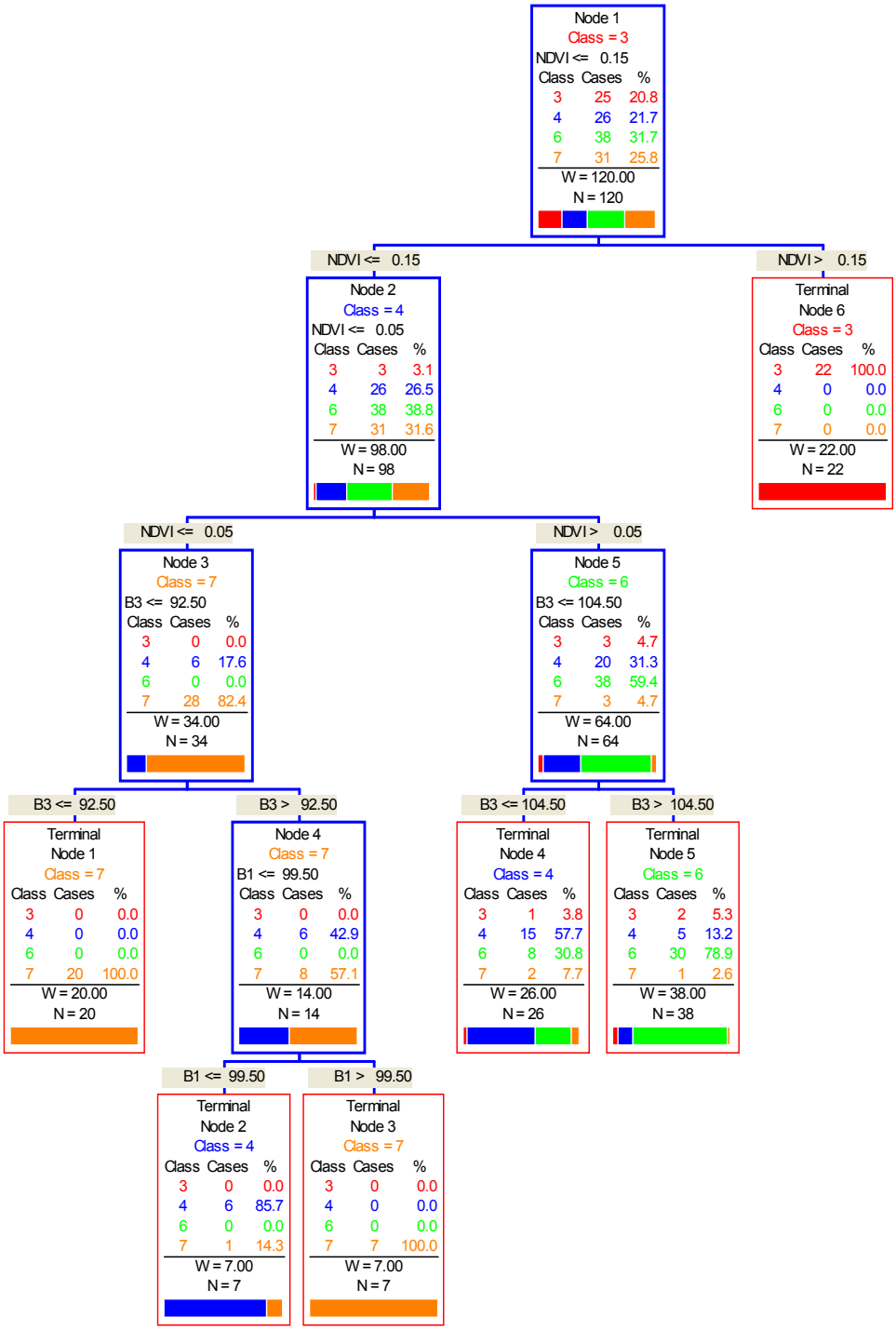


Figure 26. Classification tree (without MAT complex)

The classification tree had 5 terminal nodes. As expected NDVI was the first splitting criterion. An NDVI value above 0.15 indicated presence of Shrub complex. This was logical, as Shrub complex displayed large leaf area and higher biomass compared to any other complex. The pure terminal node 6 represented 22 cases of shrubs out of 25.

NDVI values lower than 0.15 were related to other classes. As per terminal node 1, 20 cases of Heath complex were separated when NDVI values even below 0.05 and Band 3 values were equal to or below 92.5. Terminal node 3 was also pure node for Heath complex. For terminal node 3, NDVI values were lower than 0.05, but Band 3 values were higher than 92.5 and Band 1 values were greater than 99.5. Terminal node 5 represented approximately 79% of its cases as MNT complex. They were obtained when NDVI values were greater than 0.05 and Band 3 values were greater than 104.5.

Fen complex was distributed between terminal nodes 2, 4, and 6. Terminal node 4 represented 57% of its cases as Fen complex. Thus 15 out of 26 were classified in that node. The conditions to identify the Fen complex were NDVI value higher than 0.05, but Band 3 values lower than 104.5.

The Classification tree with MAT was developed at the relative cost of 0.364. ROC training was 0.93 where as ROC test was 0.86 indicating very good performance of the classification tree. Prediction success for test sample was overall 70.37%.

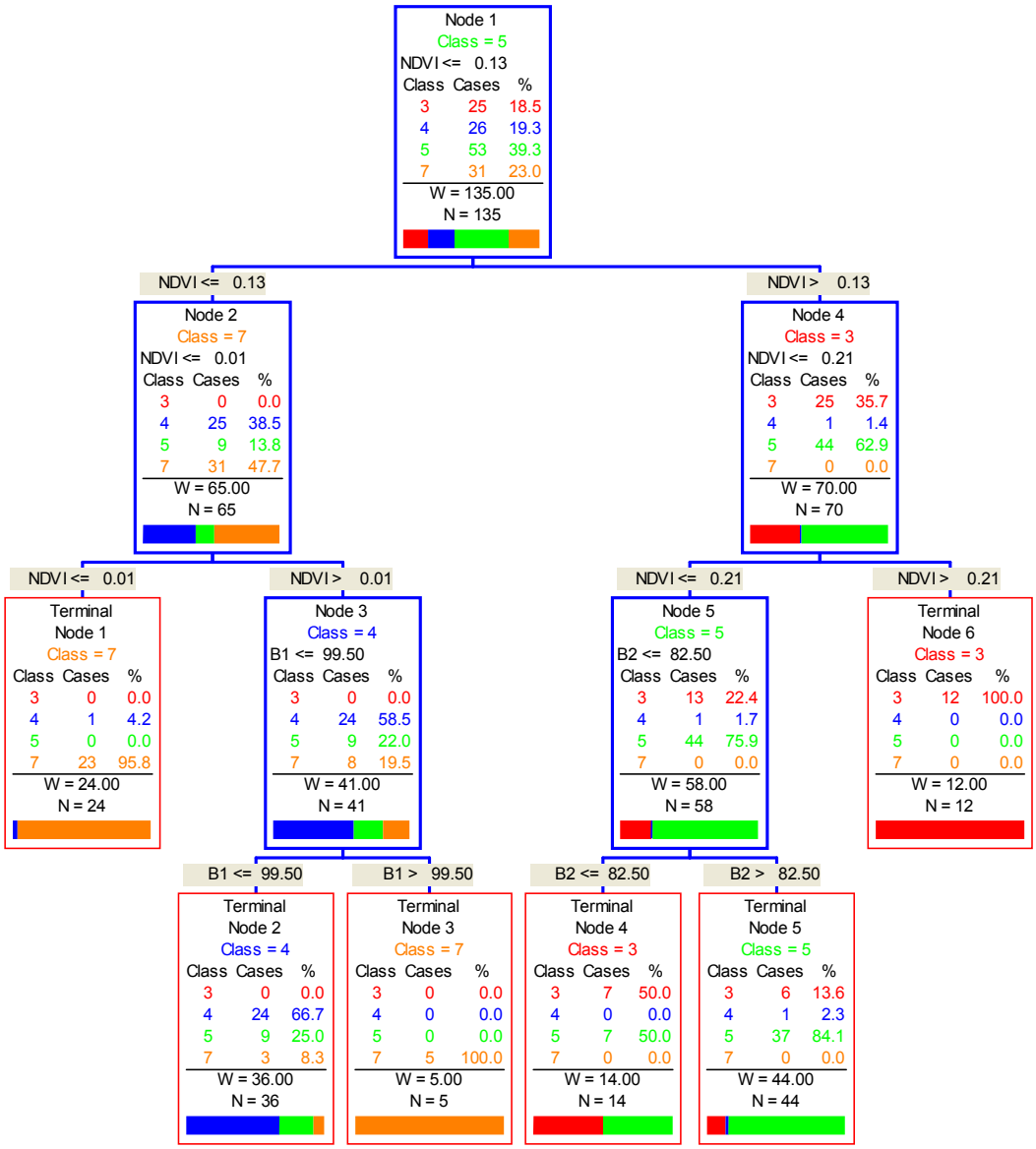


Figure 27. Classification tree (without MNT complex)

This tree was also developed in the similar fashion. Values of NDVI were the most important factor to differentiate between Shrub complex and other classes. Moreover, NDVI values were also used in the second hierarchical layer to differentiate between MAT complex and Shrub complex. The threshold value of NDVI to separate Shrub complex and MAT complex was 0.13 at the first level. When values of NDVI were higher than 0.21 then 12 cases of Shrub complex, out of 25 were identified by terminal node 6.

Other Shrub complex cases were mixed with MAT. Terminal node 5 represented 37 cases of MAT complex when the values of NDVI were more than 0.13 but below 0.21 and Band 2 values were greater than 82.5. Terminal node 1 represented 23 out of 31 cases of Heath complex. These cases were identified when the NDVI was equal to or below 0.01. However, when NDVI value was between 0.13 and 0.01 and Band 1 value was below 99.5, 24 cases out of 26 for Fen complex were identified.

The prediction success of these trees indicated that the spectral signatures derived from the ground truth data was reliable and should produce thematic maps for other watersheds at acceptable accuracies.

After accuracy assessment, the remaining watersheds were classified. Landscape metrics were derived for each watershed from these classified images. Multiple regression analysis and Classification tree analysis were carried out.

4.3 Statistical Analysis Results and Discussion

In general, estimation of primary productivity is affected by solar insolation at the time of measurement. The Chlorophyll a estimates are related to biomass produced over a certain time period; therefore, they are considered as more logical indicators of productivity rate. Due to the same reason, field observations of the primary productivity were not included in the analysis. Instead, the Chlorophyll estimates were used.

The chemical variables i.e. Chlorophyll a (Volumetric and Areal estimates), Total Nitrogen, Total Phosphorus, Conductivity, Calcium, Magnesium, Potassium, Chlorides, Sulfates, and Dissolved Inorganic Carbon (DIC) were treated as dependent variables for the regression analysis and classification tree analysis. On the other hand, all the landscape factors derived from the satellite image, lake area, watershed area, ratio of the watershed area to lake area, lake order, maximum depth, euphotic zone depth, and shoreline development factor were treated as predicting variables.

It has been stated that landscape metrics are often highly correlated with each other (Frohn and Hao, 2006). Hence, correlation analysis was carried out on all predicting variables. It was found that Edge Density metrics were highly correlated with Patch Density metrics (0.9, Pearson's correlation coefficient). Hence, it was eliminated from further analyses.

Separate regression models were developed for each chemical variable. Initially, all the landscape factors were used by the Enter method of regression provided in SPSS 17.0. Although, a large amount of variance was explained by models at this stage, they

were statistically insignificant. Therefore, based on the multi-collinearity exhibited by the variables, their significance level, and partial regression plots; variables were shortlisted and used to refine the models. The resultant models were then cross-checked with the outcomes of Stepwise method, where each variable was entered in the process based on its statistical significance. Both of the methods yielded exactly the same models for each chemical variable.

Similarly, a classification tree for each chemical variable was developed. As explained in the Methods chapter, originally regression trees were going to be developed. However, no significant regression model was identified by the CART software. The probable reason behind that could be the limited number of samples. Therefore, classification tree method was adopted.

Classification tree method required dependent variables in categorical form. Therefore, estimates of each chemical variable were divided into categories based on natural breaks in their value range.

The most significant regression model (significant at 95% confidence interval) and classification tree developed for each lake chemical variable has been illustrated in the following section. For better representation of the regression equations, acronyms for landscape variables have been adopted. Please refer to Appendix A for the details about the acronyms.

4.3.1.a Chlorophyll a Areal (Chla_A) Regression Model:

The estimates of Chla_A depicted concentration of Chlorophyll a pigments per unit area of the lake (mg m^{-2}). A regression model developed for Chla_A with predicting variables: EZD, PD_Heath, and MS_B_Shrub was significant at the level 0.001. The regression model was as follows:

Chla = 1.62 + 1.498 (EZD) + 0.021 (PD_ Heath) – 0.232 (MS_B_Shrub)

It explained approximately 72% of variability in the Chla_A concentration within the lakes. Out of the physical parameters of lakes and watersheds, only euphotic zone depth was positively correlated with Chla_A content. According to the regression model, increase in EZD by 1 meter would increase Chlorophyll Areal estimate by 1.498 mg m^{-2}. This indicated that the greater the sunlit zone, the greater photosynthetic activity present, resulting in more Chla_A concentration. It was an obvious relationship.

Euphotic zone depth in the current study varied between 1.3 meters and 12 meters. GTH 135 had the euphotic zone depth of 12 meters and the highest Chla_A concentration with 27.6 mgm^{-2}. On the other hand, GTH 116 showed the Chla_A concentration of 0.9 mgm^{-2} with euphotic zone depth of merely 1.3 meters.

Patch density of Heath and Chla_A showed a weak positive correlation. According to the model, when the number of Heath patches per 100 hectares increases by 1, then Chla_A concentration would increase by 0.021 mg m^{-2}. The patch density indicated how fragmented and well-spread the heath patches were. Greater patch density would be the equivalent of greater fragmentation.

As shown in the Figure (2) below, PD_Heath in GTH126 was 34.85 whereas 166.74 in GTH135. It indicated that Heath complex in GTH126 was composed of limited number of contiguous patches whereas in GTH135 Heath complex exhibited more number of distributed patches.

Heath complex soils exhibited relatively higher amounts of loosely bound phosphorus during the survey carried out by Giblin et al. (1991). In that perspective, Heath complex within GTH126 might not act as a source of phosphorus but fragmented patches of it in GTH135 might release relatively more phosphorus contents to the lake enhancing the productivity and Chlorophyll a concentration.

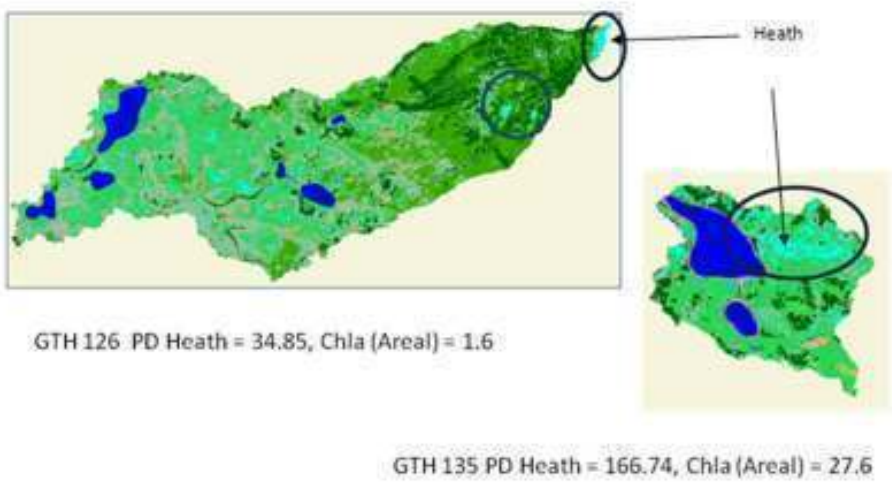


Figure 28. Example of Patch density of Heath complex and Chla_A relationship

Shape index of Shrub complex patches within buffer distance showed negative correlation with Chla_A. According to the regression model, increase in Patch shape complexity of Shrub complex by 1 unit would result in decrease of Chlorophyll Areal estimate by 0.232 mg m^{-2}.

As shown in the example (Fig 3), when Shrub patches were occupying large portions of the buffer area, the shape was considered as simple (GTH124) where as the smaller patches with a greater number of edges indicated relatively complex shapes (GTH 149). To understand the probable relationship between shape complexity of Shrub patches and Chlorophyll a pigments, one has to take into consideration that the broad leaf Shrub complex has displayed higher nutrient recycling and productivity rates (Giblin et al., 1991). Thus it is highly likely that when the Shrub complex was occupying the entire buffer zone, the nutrient exudates from root zone would easily reach the nearby lake and increase the production of Chlorophyll a pigments. However, when smaller patches of Shrub complex were present near water channels, the amount of nutrients received from their root zones might be lower resulting in lower productivity and Chlorophyll a concentration.

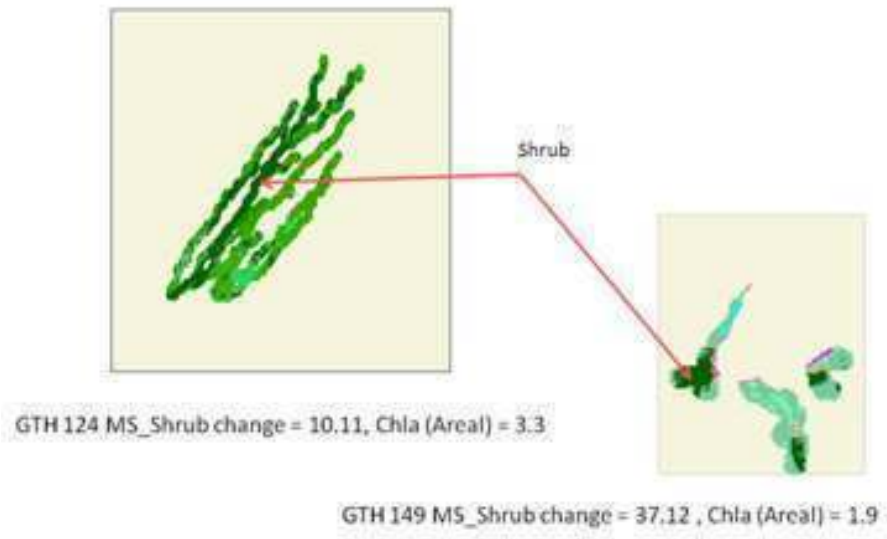


Figure 29. Example of Shrub complexity at buffer level and Chla_A relationship

Relative importance of the predicting variables:

The Standardized Coefficients (Beta) obtained for EZD, PD_Heath and MS_B_Shrub are 0.767, 0.336, and -0.278, respectively. These coefficients indicated that among these variables, EZD exhibited greater influence on Chla compared to other two.

4.3.1.b Classification Tree for Chla_A estimates:

The Chla values were divided into equal intervals. Each interval was assigned a numerical category and was treated as a Chla_A concentration class. Classification Tree was developed based on these categorized Chla_A concentrations.

The class details are as follows:

Class 1 - 0.9 mg m^{-2} to 7.4 mg m^{-2}

Class 2 - 7.5 mg m^{-2} to 14.1mg m^{-2}

Class 3 - 14. 1 mg m^{-2} and above

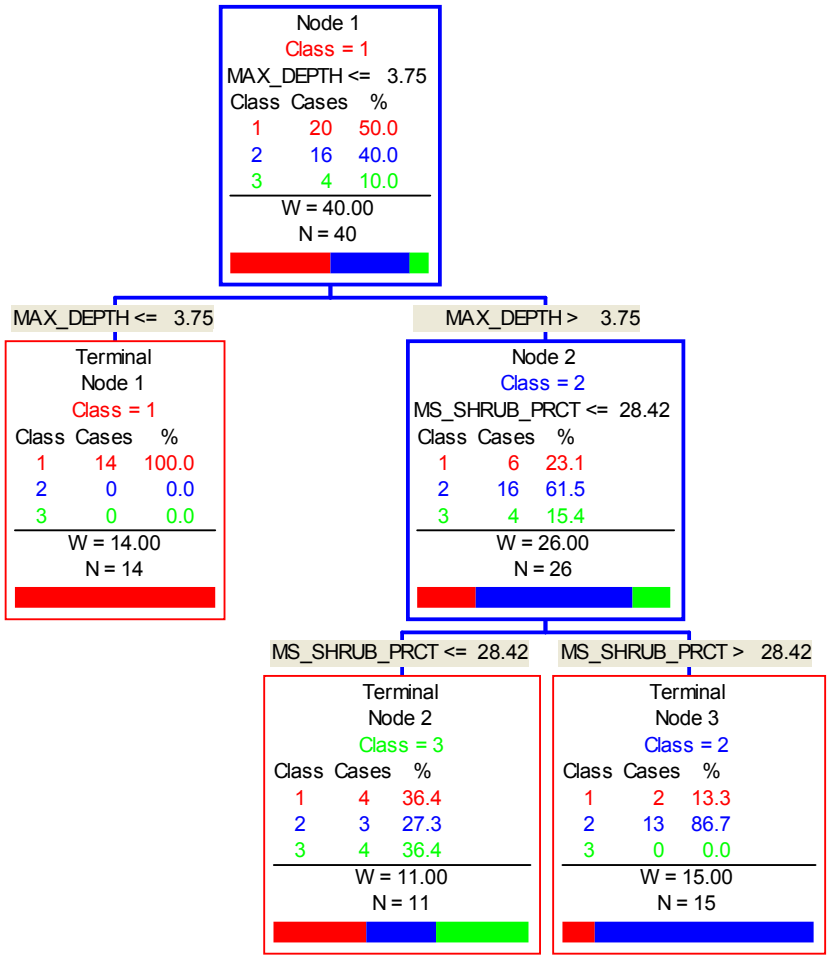


Figure 30. Classification tree for Chla_A

A significant tree for Chla was formed at relative cost of 0.806. The relatively higher cost indicated that the tree has found a pattern but with a higher misclassification rate. The tree had three terminal nodes out of which only terminal node 1 was a pure node and segregated cases of Class 1.

Prediction success rate based on the learning sample and test was 55%. Highest success was obtained for class1 of Chla_A, where out of 20 cases, 14 were successfully identified (70%). Only 43.75 % cases (7 out of 16) were correctly identified for class 2. Out of 4 cases, only 1 case was defined under the tree structure to assign 25% predictability for class 3.

The confusion matrix of misclassification indicated that highest cost of 0.75 was due to class 3. Cost of misclassification for class 2 was 0.56 where as the lowest cost of misclassification was 0.30 for class 1.

For class 1, 14 cases were correctly identified based on maximum depth of the lake. Lakes having a maximum depth of less than 3.75 meters were identified at terminal node 1. As euphotic zone depth extended up to maximum depth of lake in most of the cases, the shallow lakes might be linked with limited euphotic zone availability for chlorophyll a formation.

For class 2, 13 cases out of 16 were identified using criteria of maximum depth and percent change in mean shape index of Shrub complex. When maximum depth was greater than 3.75 meters and percent change in mean shape index of Shrub was greater than 28.42, the majority of the Class 2 cases occurred.

Class 2 indicated higher Chlorophyll a values than Class 1. Hence, separation of Class 2 cases based on higher range of maximum depth is logical. It could be explained with greater maximum depth and consequently higher euphotic zones. On the other hand segregation of Class 2 cases based on shape complexity of Shrub (greater than 28% change) could be related to their higher nutrient contribution with increasing complexity.

4.3.2.a Chlorophyll a Volumetric(Chla_V) Regression Model:

Chlorophyll volumetric (**Chla_V**) estimates represented concentration of Chlorophyll a pigments per unit volume of lake water. It was measured in μg L^{-1}. The regression model developed was as follows:

Chla = 1.130 + 0.043 (LSI_Rip) + 0.003 (PD_B_Heath)

Approximately 52 % of variance within **Chla_V** values was explained by the model. Landscape Shape Index of riparian complex was positively related to **Chla_V**. Increase in LSI_Rip by 1 unit would increase the volumetric concentration of Chlorophyll a pigments by 0.043 μgL^{-1}. Riparian complexes were observed in low lying areas. They exhibited deep and active root zones with higher water fluxes. Giblin et al. (1991) observed that soils of Riparian complexes received a large amount of nitrogen generated by upslope MAT complex and Shrub complex. Therefore, it can be inferred from the aforementioned relationship that when Riparian complex had a relatively simple landscape shape, indicated by a few simple patches; it might have utilized the received amount of nitrogen for its own growth. Otherwise, with relatively complex Riparian patches, retention and utilization of nitrogen contents could have reduced, allowing it to reach the nearby lake enhancing its productivity and Chla_V concentration.

It could be further explained with Fig1 where GTH 153 depicts only 16 % change in LSI of riparian complex from 1. Hence, the Chla_V concentration is 2.1 μg L^{-1}. In case of GTH 144, the change in LSI is 43% and concentration of Chla_V is 2.7.

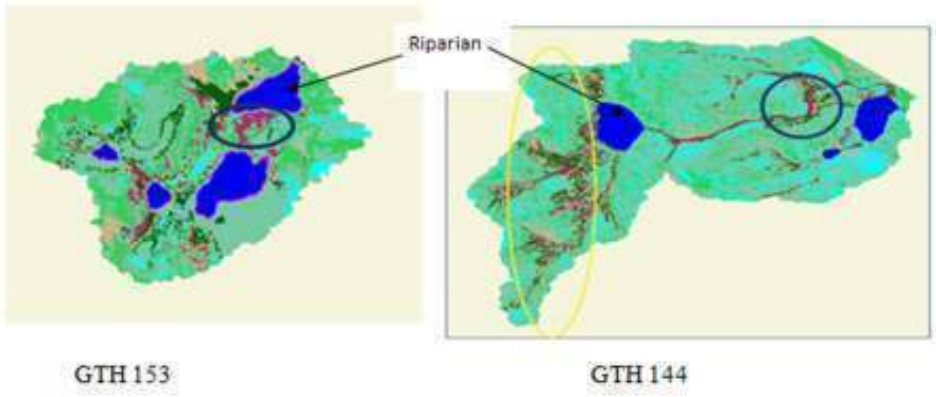
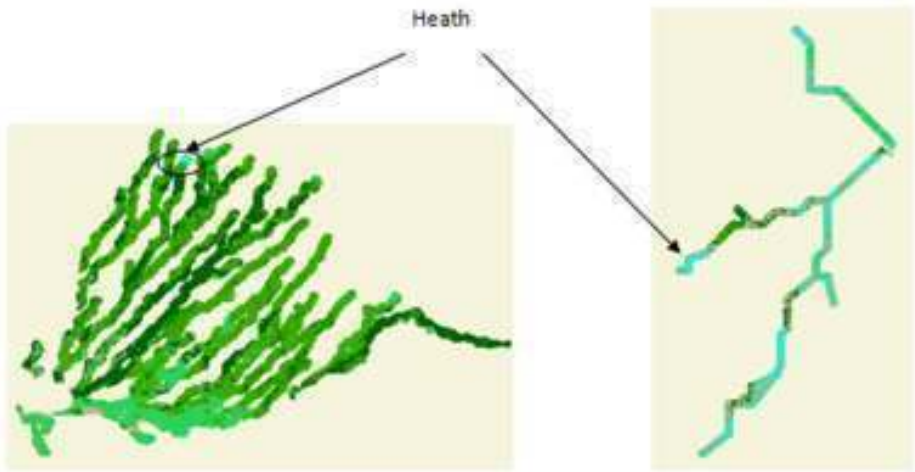


Figure 31. Example of LSI of Riparian complex and Chla_V relationship

Within a distance of 20 meters along water tracks, Patch density of Heath complex had positive relationship with Chla_V estimates. According to the regression model, increase in PD_B_Heath by 1 patch per 100 hectares would enhance volumetric Chla estimates by 0.003 μgL^{-1}. In the soil study carried out by Giblin et al. (1991), a relatively high amount of loosely bound phosphorus was found within heath complex. Therefore, fragmented Heath patches present near water tracks could act as a source of phosphorus and contribute to phosphorus contents of lakes. Phosphorus reaching the corresponding lake would experience relatively higher productivity rates indicating higher chlorophyll estimates.



GTH 124 PD_B_Heath = 6.7 Chla = 1, GTH 143 PD_B_Heath = 100.2 Chla = 3.6
Figure 32. Example of Patch density of Heath complex at buffer and Chla_V relationship

Relative importance of predicting variables:

The standardized coefficients for LSI_Rip and PD_B_Heath were 0.649 and 0.305, respectively. Based on these coefficients it can be said that LIS_Rip was more influential.

4.3.2.b Classification Tree for Chla_V estimates:

The class details are as follows:

Class 1 < 0.9 μg L^{-1}

Class 2 1.0 μg L^{-1} to 1.5 μg L^{-1}

Class 3 1.6 μg L^{-1} to 2.1 μg L^{-1}

Class 4 > 2.2 μg L^{-1}

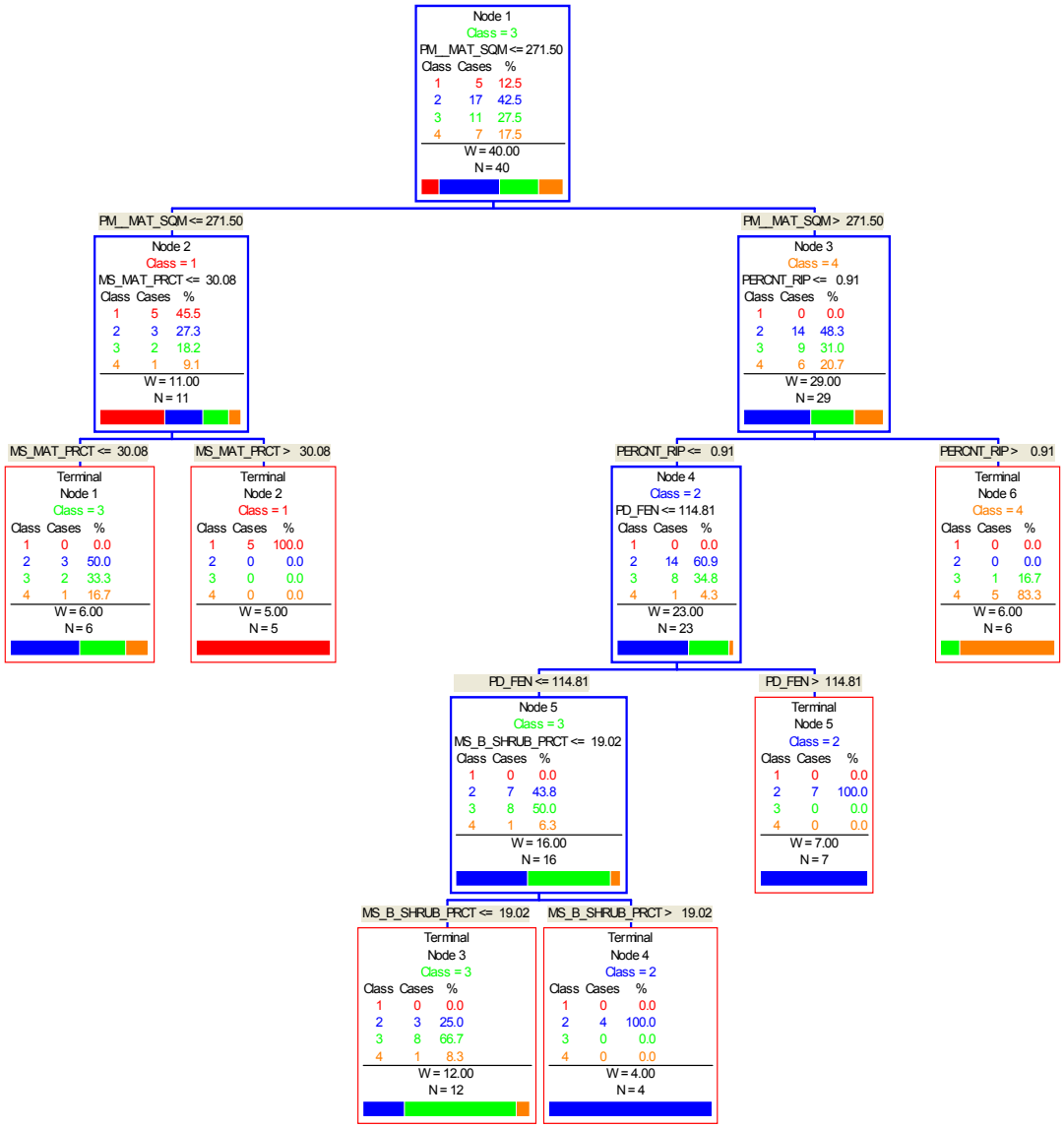


Figure 33. Classification tree for Chla_V

The Classification Tree for Chla_V was developed at the relative cost of 0.71. It showed six terminal nodes, where only 2nd, 4th, and 5th terminal nodes were pure.

Prediction success rate on the learning sample was 77.5%, whereas it was only 42.5% on the test sample. This indicated that, even though CART was able identify a pattern, it was limited in prediction given the limited sample size.

While looking at the misclassification rate for test data, it could be concluded the highest cost was experienced by class 3, as 8 out 11 cases for this class were incorrectly classified. Also, for Class 2, 10 out of 17 cases were misclassified causing a cost of 0.59. Class 1 and 3 experienced a cost around 0.4.

As depicted by terminal node 2, there was a clear pattern to identify cases of Class 1 of Chla_V. It suggested that when the mean patch size of MAT was smaller than 271.5 square meters and when the percent change in mean shape of MAT patches is was lower than 30.08, Class 1 category of lakes was identified. In simple terms, it was indicating highly fragmented MAT patches. According to Giblin et al. (1991), MAT soils support higher rates of nitrification. However, if they were fragmented and comprised of smaller patches, then nitrate production might be relatively lower, leading to lower amounts delivered to the lakes and hence the lower Chla_V concentrations.

At the terminal node 5, 7 out of 17 cases of Class 2 were segregated given that mean size of MAT patches were higher than 271.50 m^2 and percentage of Riparian complex was less than 0.91 and the patch density of fen complex was higher than 114.81,. This indicated that the lower the amount of riparian shrubs, the greater the chances are of nutrients reaching lakes, as well as fragmented patches of fen would be the source of nutrients, mainly nitrogen.

Also increasing complexity of mean shape of Shrubs patches (terminal node 4) indicated that Shrub complex would not act as a sink of nutrients allowing nutrients to be delivered to the lakes and increasing chlorophyll estimates. Overall, these conditions were pointing towards the watersheds with more nitrogen sources than sinks. However, no clear rule was obtained for other classes.

4.3.3.a Total Nitrogen Regression Model:

According to Wetzel et al., (2001) total nitrogen represents concentration of nitrogen in all the forms i.e. organic and Ammonia, Nitrite, and Nitrate. In this study, Total Nitrogen (TN) estimates were regresses against all the landscape factors to obtain following model:

$$TN = 14.65 + 0.58\ (Prct\_B\_Rip) + 0.011\ (PD\_B\_Fen)$$

The model explained approximately 48% of variance in the remaining TN estimates. The model indicated that the amount of Riparian complex had a positive correlation with TN estimates. It indicated that an increment in the complexity of Riparian patches within the buffer zone of channels, by 1 percent would enhance the TN content in lakes by 0.58 mol L^{-1}. As observed by Giblin et al. (1991), Riparian complexes were zone of accumulation for nitrogenous compounds and they experienced higher water fluxes being near to streams. Therefore, with higher amount of flowing water along with subsurface flow, the nitrogen in all form was likely to contribute to total nitrogen found in lakes. This finding was supported by observations within riparian willows in other regions of Alaska (Van Cleve et al., 1993). This research indicated that

the willow community, even though it was claimed to be a nitrogen removing agent, may experience higher microbial mineralization activity with respect to higher water fluxes. It may also lead to release high quantities of organic nitrogen compounds.

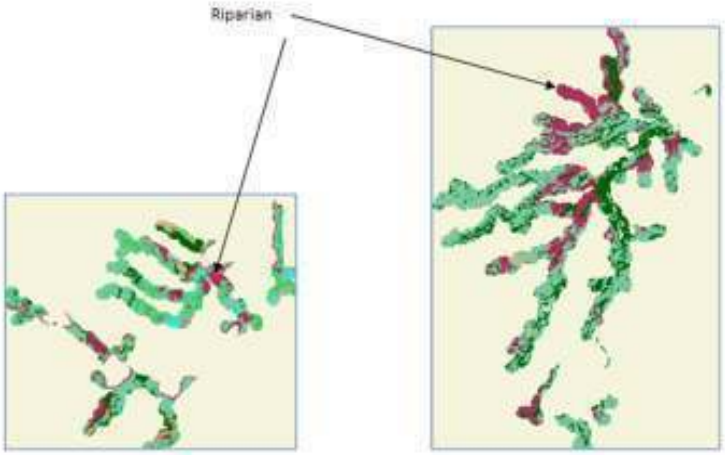


Figure 34. Example of Amount of Riparian complex in buffers and TN relationship

A weak positive relationship was observed between patch density of fen complex within buffer zone and TN. It could be said that an increase in patch density of Fen complex within a 20-meter buffer of channels by 1 patch per 100 hectares would increase the TN content by 0.011 mol L^{-1}. Giblin et al. (1991) have found that Fen areas were saturated with nitrogen and displayed higher N/P ratios. The denitrification activity was higher in areas of fen cover than any other vegetation communities (Giblin et al., 1991). It might add inorganic forms of nitrogen to surrounding lakes if it is adjacent to streams and more fragmented.

Relative importance of predicting variables:

The standardized coefficients of these variables depicted that Prct_B_Rip contributed the most in this regression model. The coefficients are as follows: Prct_B_Rip (0.552) and PD_BFen (0.196).

4.3.3.b Classification Tree for Total Nitrogen estimates:

TN estimates were divided into four classes. The range values for those classes were as follows:

Class 1 – 10.9 mol L^{-1} to 14.6 mol L^{-1}

Class 2 – 14.7 mol L^{-1} to 18.4 mol L^{-1}

Class 3 – 18.5 mol L^{-1} and above

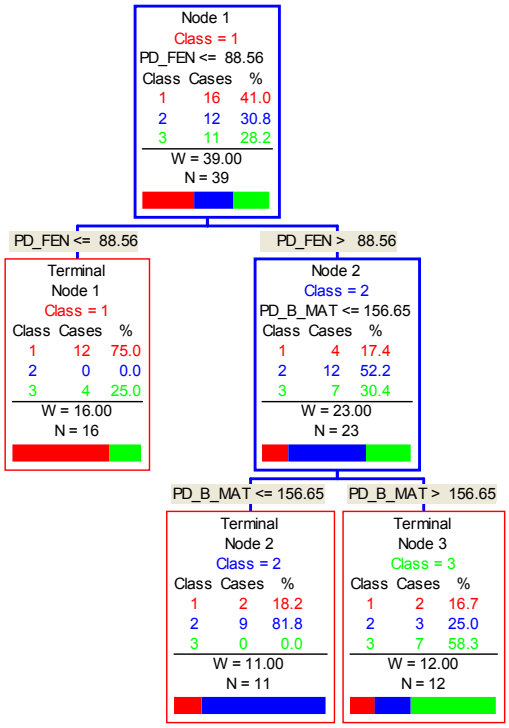


Figure 35. Classification tree for Total Nitrogen

The classification tree was formed at 0.846. The tree for TN had 3 terminal nodes, containing no pure nodes. Prediction success on the test data was merely 43.69%. Misclassification cost was higher than 0.55 for all classes. The criterion to differentiate majority of Class 1 cases from other classes was patch densities of less than 88.56 for Fen complex (terminal node 1). Out of 16 cases, 12 cases were identified based on this landscape factor. This highlighted again relationship between higher N/P ratio within Fen complex and nitrogen content. When the patch density was less than 88.56, the nitrogen release might be lower, but if patch density was higher than 88.56, then it might enhanced nitrogen release.

Patch density of MAT complex within the buffer was the next splitting criterion. As indicated by terminal node 2, when patch density of MAT complex was less than 156.65, then the tree identified cases of Class 2, but when it was greater than 156.65, then most of class 3 cases were identified. As mentioned by Giblin et al. (1991), MAT complex was found responsible for release of nitrogen in soil solution in the upslope areas of watersheds. Using same logic, it can be said that if more fragmented MAT was present within buffer zone, it may lead to nitrogen release as well as cause effective transport of nitrogen to lakes.

4.3.4.a Total Phosphorus Regression Model:

Total phosphorus (TP) represents dissolved as well as particulate phosphorus content in water. In the current research, following landscape factors were found to be correlated with total phosphorus estimates of the selected Arctic lakes:

$$TP = 0.094 - 0.001\ (PD\_AV) + 0.001\ (PD\_Shrub) + 0.004\ (MS\_MAT)$$

The model was able to explain approximately 52% of variance in the TP content found within lakes. A very weak negative relation was exhibited by Aquatic Vegetation (AV). The model depicted that increase in patch density of AV by 1 i.e. 1 patch per 100 hectares, would decrease the TP contents of lakes by 0.001 mol L^{-1}. The near shore zone experiences more sedimentation from allochthonous phosphorus loadings Giblin et al., 1991). In this study, Aquatic vegetation complex represent the near shore zone, which experiences high water fluxes as well as probable high organic soil zone acting as an interface between lake water and external landscape. When the patch density of AV

increases, it would experience more fragmentation leading small size of patches. Hence, the contribution of AV complex in phosphorus recycling would lessen with its decreasing size.

However, a positive relationship was observed between Patch density of Shrub complex and TP concentration in lakes. The regression model indicated that if PD_Shrub increases by 1 patch per 100 hectares then TP contents would increase by 0.001 mol L^{-1}. Chapin III et al. and Shaver (1988) observed that per unit area, the uptake rate of phosphorus by deciduous as well as evergreen shrubs was higher in nontrack areas of watersheds. Marion et al. (1989) observed that higher nutrient concentrations occur within wet zones such as Shrub complex, as more thaw depth and nutrient recycling occur in their soils. They also found higher P content in leaves of *Salix* and *Betula* species. Thus Shrub complex was established as a sink of phosphorus but if it was more fragmented or composed of complex patches, then it would show lower phosphorus uptake rate and contribute to TP contents of lakes.

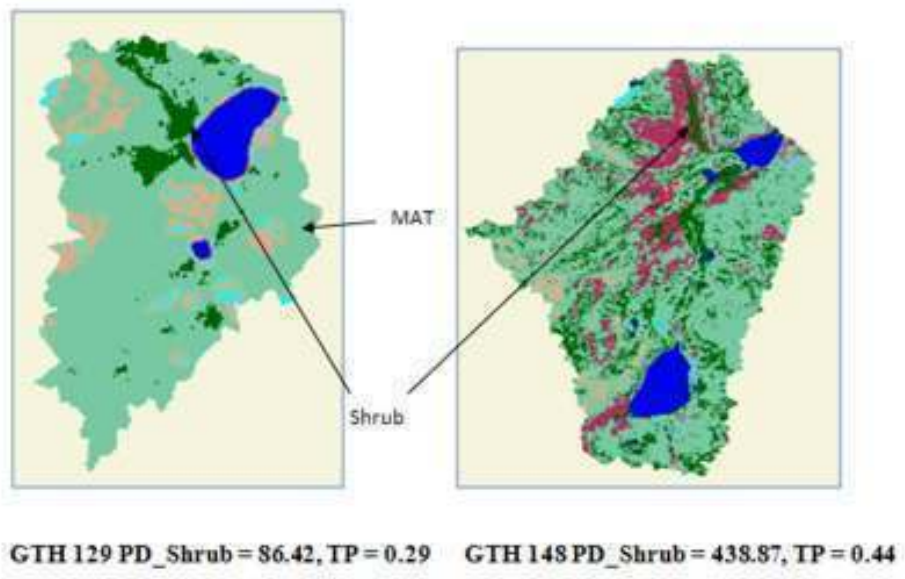


Figure 36. Example of Patch density of Shrub complex and TP relationship

It was also observed that when Shape complexity of MAT increases by 1 unit then the TP contents of lakes increase by 0.004 0.001 mol L^{-1}. Chapin III et al. (1988) studied role of *Eriophorums* (Major vascular tundra species, and defining species for MAT) with respect to track areas and nontrack areas of watersheds. According to these authors, the deep root structure of Eriophorums enabled them to interact with subsurface flow and absorb nutrients. Overall nutrient recycling rates were higher within MAT sites. With reference to Figure 20, it could be said that increasing complexity of the shape MAT patches means more edges or more discontinuity. It may enhance to chances of nutrient especially Phosphorus to escape from being absorbed by MAT.

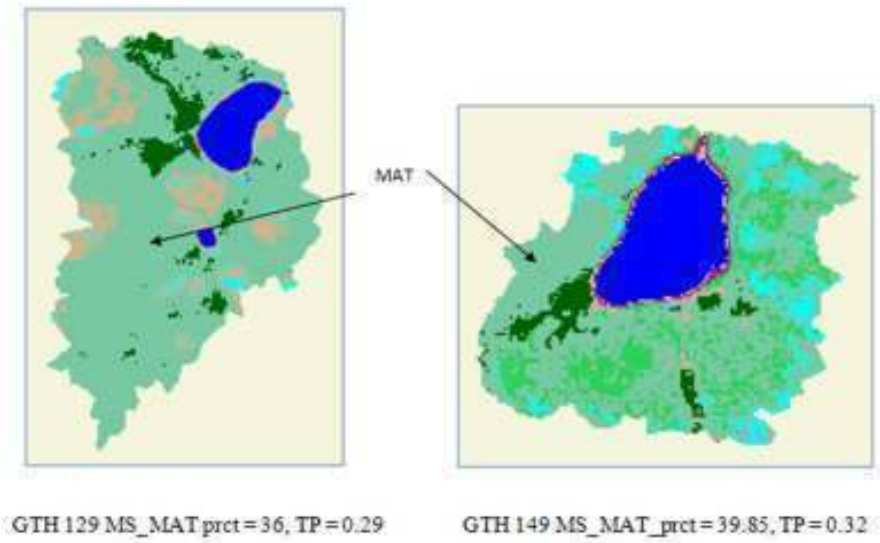


Figure 37. Example of Shape complexity of MAT complex and TP relationship

Relative importance of predicting variables:

Standardized coefficients for them were: PD_AV(-0.334), PD_Shrub(0.52), and MS_MAT_Prct (0.459). Thus, PD_Shrub had relatively more contribution in the model.

4.3.4.b Classification Tree for Total Phosphorus estimates:

TP estimates varied within very short range. Hence, only they were divided into only two classes, which were as follows:

Class 1 < 0.25 mol L^{-1}

Class 2 > 0.25 mol L^{-1}

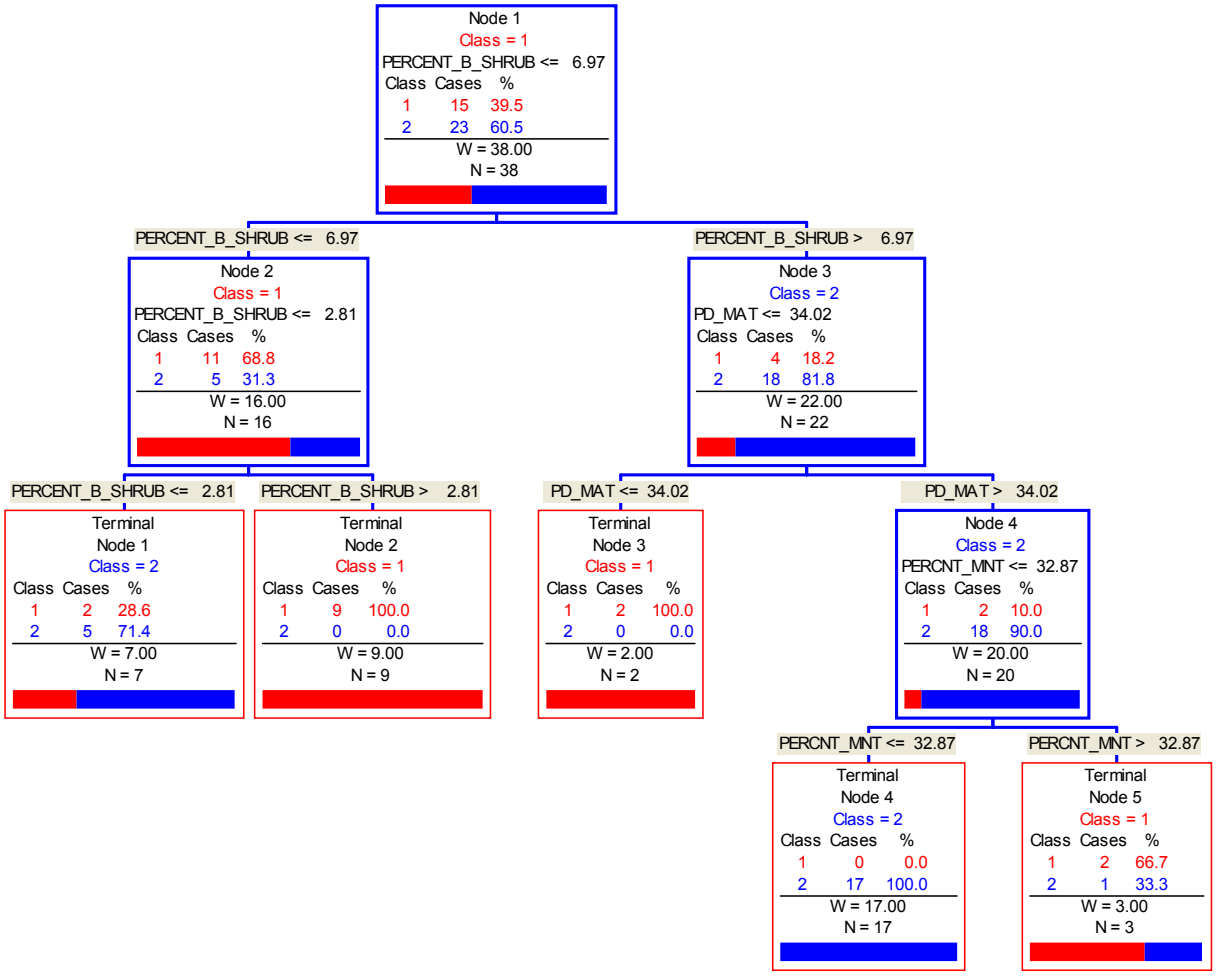


Figure 38. Classification tree for Total Phosphorus

The cost of tree formation for TP concentrations was 0.525. There were five terminal nodes obtained, out of which node 2 and 3 were pure nodes for Class 1 and node 4 was pure node for Class 2.

The success rate of prediction on the test data was approximately 65%. Looking at the misclassification rate for the test data, it can be concluded that Class 2 experienced

highest rate of misclassification with cost of 0.48 as 11 out 23 cases were wrongly identified. On the contrary, Class 1 showed only 0.13 as a cost of misclassification.

As per terminal node 2, 9 cases for Class 1 of TP were identified when amount of Shrub complex within watershed was between 2.91% and 6.97 %. Therefore, 2.91% of Shrub complex within buffer may indicate a lower threshold value to have influence on Phosphorus transportation process. However, when the amount of Shrub within buffer was between 2.91% and 6.97%, it could act as a sink for phosphorus and reduce the amount flowing towards lakes.

When percentage of Shrub complex within buffer was higher than 6.97, then most of the Class 2 cases were segregated. As explained by Marion et al. (1989), the presence and amount of vegetation is closely linked with availability of nutrients and water. Based on this study, it could be proposed that Shrub complex more than 6.97% was indicator of nutrient rich wet soil along with higher contents of phosphorus. In that case, lakes would receive more phosphorus from such soils.

The above mentioned condition was further refined at terminal node 4. It indicated that along higher amounts of Shrub complex near water channels, if the patch density of MAT complex was higher than 34.02 and the amount of MNT complex within watershed was less than 32.97 percent, then 17 out 23 cases of the Class2 were identified.

This might indicate that at the watershed level, when the MAT complex was more fragmented, a greater release of phosphorus could occur. At the same time, the splitting criterion for the amount of MNT complex might be a pointer towards watersheds

dominated by MAT complex instead of MNT to experience higher amounts of phosphorus contents in the lakes.

4.3.5.a Conductivity Regression Model:

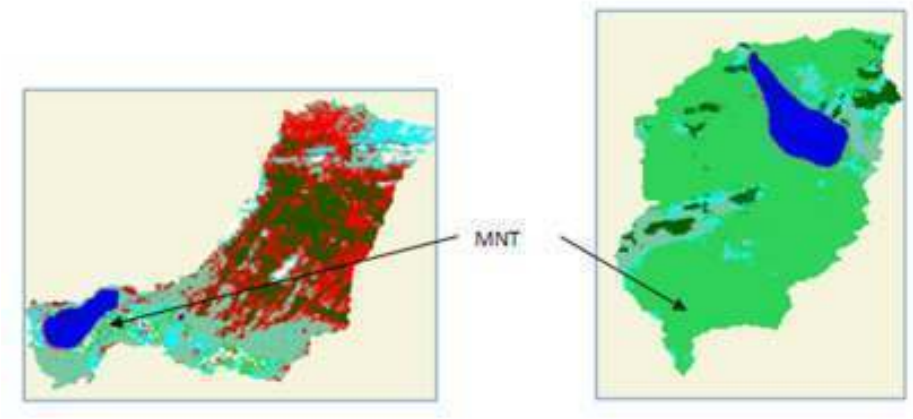
Conductivity of water is defined as a measure of water to conduct electricity. It is a property of water attributed cations and anions, hence, a cumulative indicator of their concentration in water.

Conductivity = -29.426 + 2.246 (Prct_MNT) + 7.17 (EZD) + 0.195 (PD_B_Shrub)

The regression model obtained was able to explain more than 80% of variance using Prct_MNT, EZD, and PD_B_Shrub. The positive correlation between amount of MNT complex and Conductivity indicated that if the amount of MNT increases by 1 percent then Conductivity increases by 2.2 $\mu S\ cm^{-1}$. With an increase in EZD of 1 meter, the Conductivity would increase by 7.17 $\mu S\ cm^{-1}$. On the other hand Conductivity would increase by 0.195 $\mu S\ cm^{-1}$ when PD_B_Shrub by 1 patch/100 hectares.

The only explainable relationship of conductivity was with MNT complex. Walker et al. (1994) mentioned that relatively young tills in the Arctic were dominated by MNT complex. Moreover, younger tills exhibit the dominance of physical weathering process, which may release different ions to corresponding lakes. Hence, the influence of till surface age might have been reflected through the relationship between the Moist Non-Acidic Tundra complex and Conductivity.

Watersheds without MNT displayed conductivity of average 10.88 µS cm^{-1} whereas watersheds with MNT had an average of 98.7 µS cm^{-1}.



GTH 123 Percent of MNT 4.33, Conductivity 46.9 µS cm^{-1}. GTH 119 Percent of MNT 74.28, conductivity 149.1 µS cm^{-1}

Figure 39. Example of Percentage of MNT complex and Conductivity relationship

Relative importance of predicting variables:

The standardized coefficients obtained were as follows: Prct_MNT (0.772), EZD (0.364), and PD_B_Shrub (0.329). It can be concluded that Prct_MNT has more contribution towards the Conductivity whereas EZD and PD_B_Shrub had approximately equal effect.

4.3.5.b Classification Tree for Conductivity estimates:

Estimates of the Conductivity were divided as follows:

Class 1 3.7 $\mu S\ cm^{-1}$ to 58.7 $\mu S\ cm^{-1}$

Class 2 58.8 $\mu S\ cm^{-1}$ to 113.7 $\mu S\ cm^{-1}$

Class 3 113.7 $\mu S\ cm^{-1}$ to 168. 7 $\mu S\ cm^{-1}$

Class 4 168.8 $\mu S\ cm^{-1}$ and above

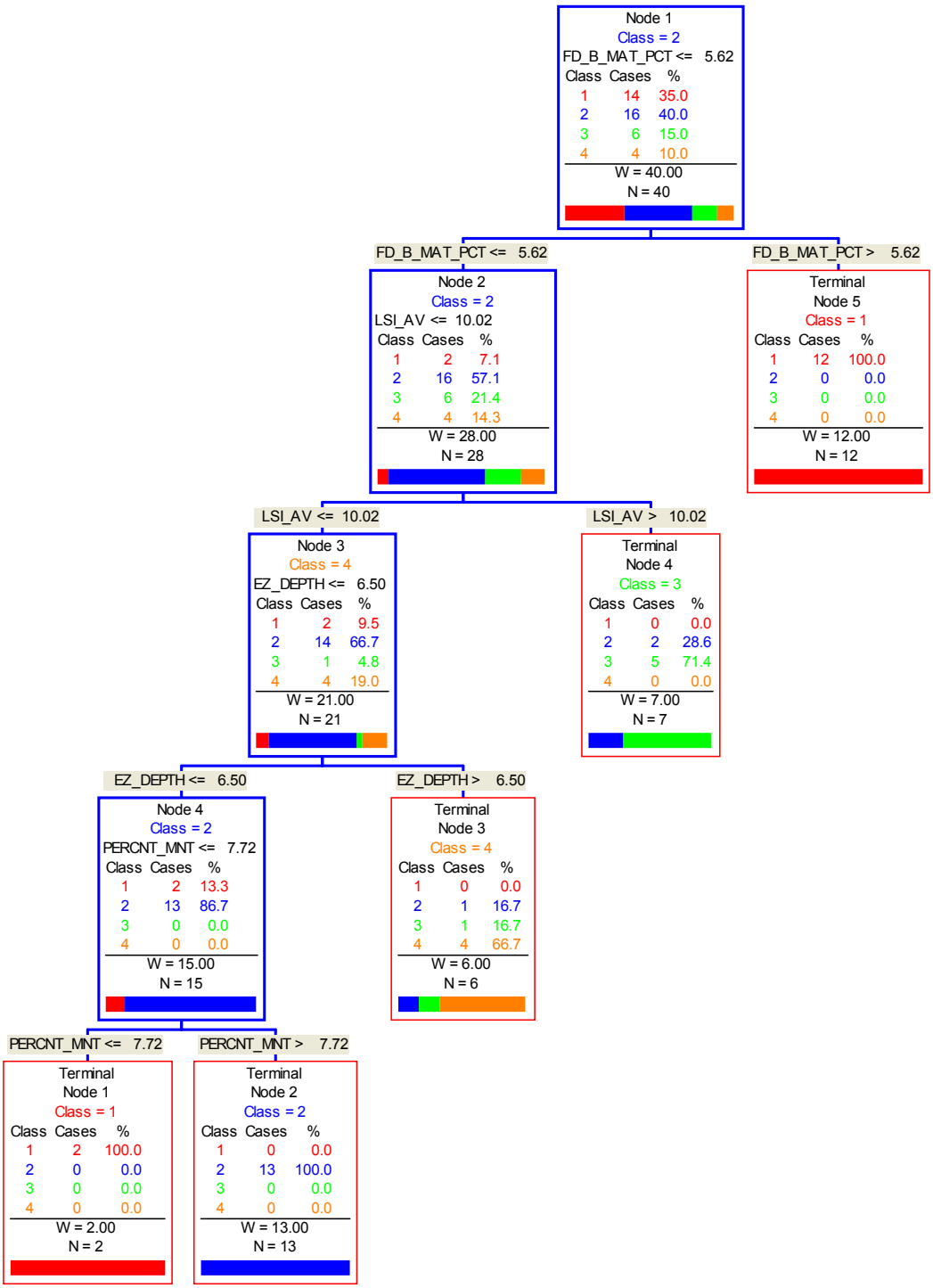


Figure 40. Classification tree for Conductivity

A Classification tree with 5 terminal nodes was formed at the relative cost of 0.624. Out of 5, only 3 terminal nodes were pure.

Class 1 had the lowest cost of misclassification (0.14, with only 2 cases of 14 were misidentified). For Class 2, the misclassification cost was 0.31, as 5 out of 16 cases were misclassified. Class 3 had the misclassification cost of 0.67. Only 2 out of 6 cases for Class 3 were correctly classified. The highest misclassification cost of 0.75 was assigned to class 4 as only 1 out of 4 cases was identified rightly.

At terminal node 5, 12 out of 14 cases of Class 1 were segregated when the percent change in fractal dimension (FD) of MAT in buffer zone was greater than 5.62. Higher FD of the MAT within buffer zones indicated that the MAT patches were growing and dominating within the buffer zone. Conversely, when FD of the MAT was lower, it was the MNT complex dominating buffer zones. As mentioned earlier, MNT dominance was observed in relatively younger tills with relatively higher inputs of ions to nearby lakes (Walker, 1994). Hence, the watersheds with the MAT in their buffer zones showed less conductivity with lower ions inputs and therefore Class 1 was segregated on this basis.

According to the terminal node 2, 13 cases out 16 for Class 2 were identified when percent change in fractal dimension of the MAT complex within the buffer zone was less than 5.62 and LSI of Aquatic vegetation complex was less than 10.02 and euphotic zone depth was less than 6.5 meters and amount of MNT was more than 7.7 percent.

The amount of MNT complex might be an indicator of till surface age.

4.3.6.a Calcium Regression Model:

The study carried out by Keller et al. (2007) has established that large amounts of Calcium were stored in the permafrost zone. Furthermore, Everett et al. (1989) mentioned that overland flow and interflow coming out of organic rich soil horizons contribute to calcium concentrations of the arctic streams. The authors also found that the precipitation (dry fall) was significantly responsible for calcium ion inputs into lakes.

The regression model for TN exhibited following relationship with landscape factors:

Calcium = -2.202 + 0.368 (Prct_MNT) + 2.13 (EZD) + 0.317 (LSI_MM)

This regression model obtained an R^2 values of 0.75 thus illustrated 75% variance within Calcium estimates. The model indicated that if the amount of MNT increases by 1 percent then Calcium contents in lakes would increase by 0.368 mgL^{-1}. Similarly, Calcium contents would be increased by 2.13 $\mu S\ cm^{-1}$when EZD increases by 1 meter. Additionally, if complexity of MM at the landscape level as indicated by LSI increases by 1 unit then Calcium concentration would be enhanced by 0.317 $\mu S\ cm^{-1}$.

According to the model, Calcium content would increase with increase in the amount of Moist Non-Acidic Tundra (MNT). This finding is opposite to the observations made by Marion et al (1989) or Everett et al (1989) that Calcium is contributed by a deep active zone exhibiting higher organic and water contents i.e. Shrub complex. However,

similar to the observed relationship between conductivity and MNT complex; Calcium contents might exhibit influence of till surface age in the form of MNT complex. This inference is supported by the findings of Keller et al (2007). The authors found higher acid digestible Calcium concentrations within the Itk 2, the relatively younger till. . Hence, not the direct influence of MNT complex but chemical weathering processes within young tills may be the reason behind the positive relationship. The lakes encompassing the GTH 119, 120, 121, and 122 watersheds are an excellent example of lakes situated on Itk 2 initial surface. The amount of MNT complex ranged from 13 to 74% in those watersheds. The average value of calcium observed was 27 mg L^{-1} where as lakes outside Itk 2 till exhibited 7 mg L^{-1} of average Calcium concentration.

Additionally, the Calcium estimates were positively related with LSI of Mountain Meadow (MM) complex however it was present only in nine watersheds. The average value of calcium concentration within those watersheds was 21 mg L^{-1}, whereas average of watersheds without MM was 12 mg L^{-1}. Within the nine watersheds, two watersheds exemplifying the effect of LSI of MM are exhibited in Figure 23.

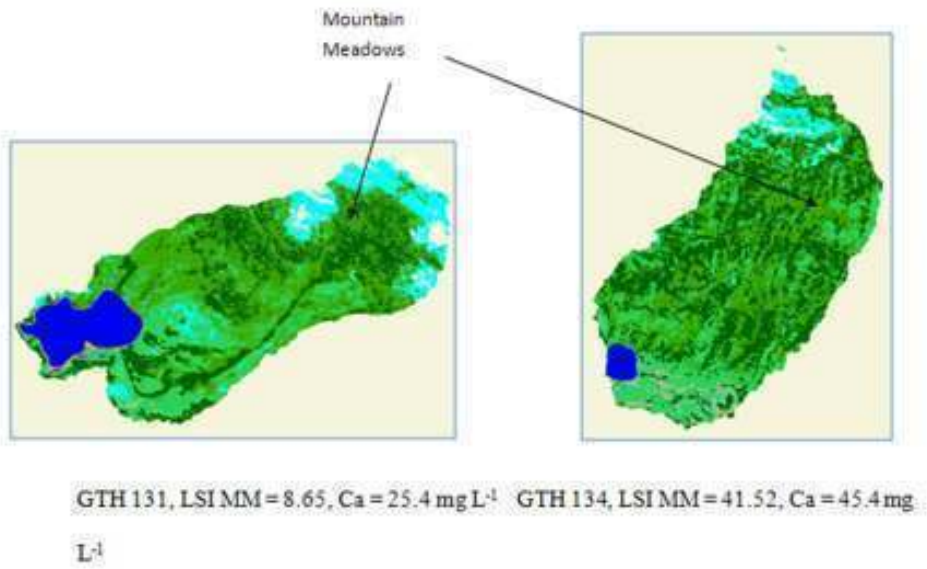


Figure 41. Example of LSI of Mountain Meadow complex and Ca relationship

The MM complex in GTH 131 is more continuous and less intercepted by the Shrub complex where as MM within GTH 134 watershed has been highly dissected by the Shrub complex. Both complexes add large amount of organic matter to the soils. According to Marion et al (1989) the upper layers of soils have cation like Ca and K attached to organic matter and can easily loose them with subsurface water movement or interflow. Therefore, when MM complex is intermixed with Shrub complex it may enhance the Ca concentration.

Relative importance of predicting variables:

For these predicting variables the following standardized coefficients were obtained: Prct_MNT (0.571), EZD (0.496), and LSI_MM (0.28). Thus, Prct_MNT has been the major contributor towards Calcium estimates with EZD being slightly less

responsible. However, LSI_MM showed relatively lower impact on the Calcium concentration.

4.3.6.b Classification Tree for Calcium estimates:

These were the class categories formed for analysis purpose:

Class 1 – 0.9 mg L^{-1} to 11.9 mg L^{-1}

Class 2 – 12 mg L^{-1} to 22.9 mg L^{-1}

Class 3 – 23 mg L^{-1} to 33.9 mg L^{-1}

Class 4 – 34 mg L^{-1} and above

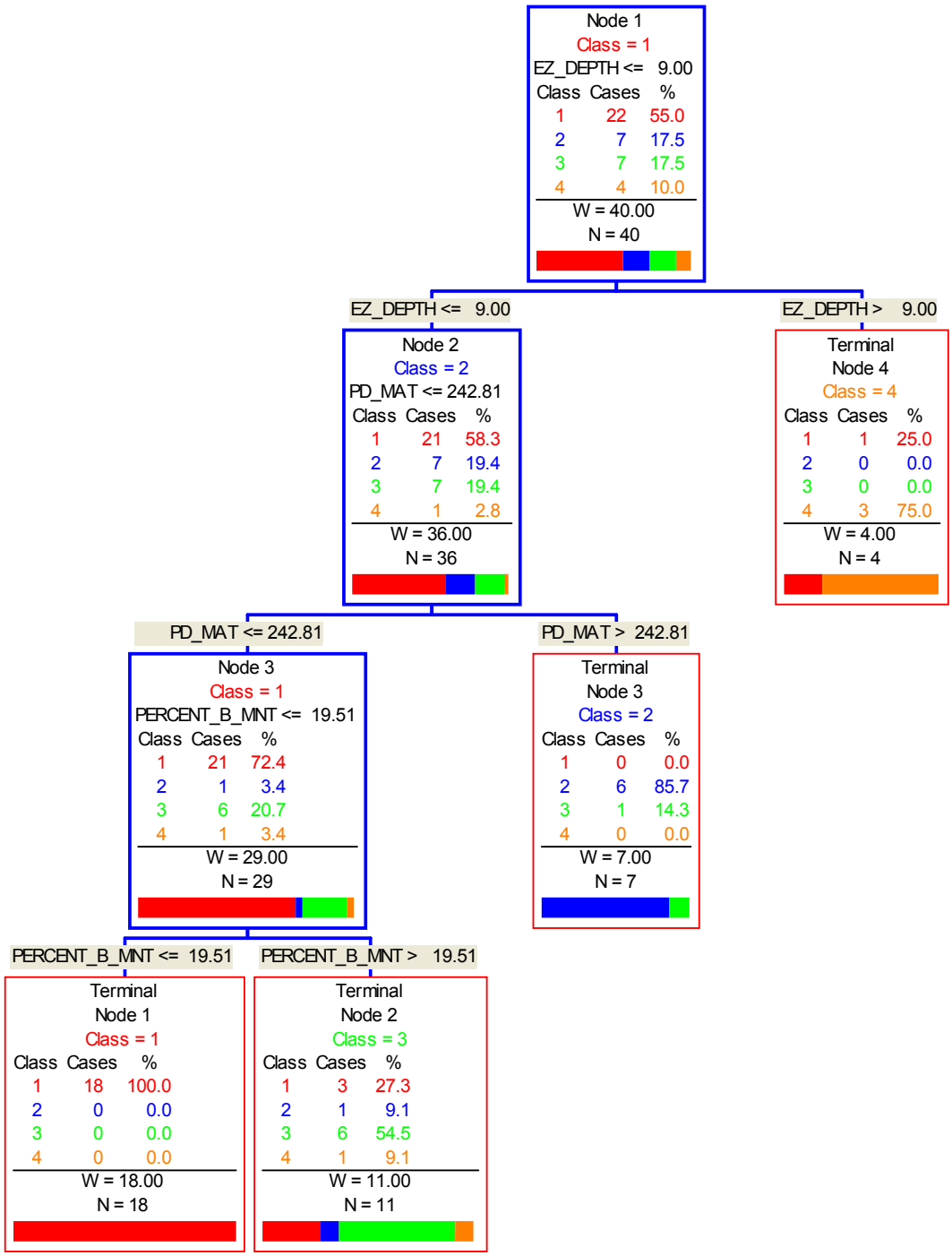


Figure 42. Classification tree for Calcium

An optimal tree was obtained at the relative cost of 0.732 with 4 terminal nodes. Compared to the learning success of the model, the prediction success was low. Overall, 50% predictability was obtained. Class 3 showed higher misclassification cost of 0.75, whereas other classes showed misclassification cost of almost 0.50.

EZD was used as a major splitting criterion; however, the reasoning behind it could not be established. The splitting criterion used at the second level was the Patch Density of Moist Acidic Tundra (PD MAT) complex. Values of PD MAT greater than 242.81 were related to class 2. When PD MAT was lower than 242.81 and amount of MNT within buffer was lower than 19.51%, most of the class 1 cases were identified in terminal node 1. These conditions cumulatively suggested that watersheds situated on the older tills i.e. dominated with MAT would experience lower Ca contents compared to lakes on the younger tills, similar to findings of Keller et al. (2007).

Hence, lower values of Ca ions could be observed in watersheds with less but continuous MAT patches with lower amounts of MNT in the buffer.

Class 3 did not emerge in any of the pure nodes hence; no specific rule was derived for class 3.

4.3.7.a Magnesium Regression Model:

The regression equation developed for the magnesium estimates was as follows:

Mg = 2.415 – 0.037 (Prct_MAT) + 0.983 (Till Age 2) + 0.117 (EZD)

This model was able to explain approximately 74% of the variance in the magnesium concentration using the amount of the MAT in watersheds, till-age, and euphotic zone depth.

According to the regression model developed for magnesium, an increase in the amount of MAT by 1 percent would cause a decrease in magnesium concentration in lakes by 0.037 mg L^{-1}. It also indicated that lakes situated in till age surface category 2 had 0.983 times more magnesium than other lakes. Along with that, increase in EZD by 1 meter would enhance the magnesium concentration in lakes by 0.117 mg L^{-1}.

The negative correlation between MAT and Mg could be an indicator of till age. According to Walker et al. (1995), MAT is more dominant on older till surfaces. As shown in Keller et al. (2007), older tills had lower amount of cations compared to relatively younger tills.

Keller et al. (2007) observed comparatively greater concentration of magnesium within Itk2 till surface. It supports the relationship observed in the current study suggesting that lakes situated on Itk2 would experience higher magnesium concentration compared to the lakes located on other tills.

Relative importance of predicting variables:

The standardized coefficients obtained during the regression analysis indicated that Prct_MAT (-0.546) contributed relatively higher than other variables. The standardized coefficient for till age 2 category was 0.312 whereas for EZD it was 0.227.

4.3.7.b Classification Tree for Magnesium estimates:

Class 1 – 0.1 mg L^{-1} to 1.5 mg L^{-1}

Class 2 – 1.6 mg L^{-1} to 2.9 mg L^{-1}

Class 3 – 3.0 mg L^{-1} to 4.3 mg L^{-1}

Class 4 – 4.4 mg L^{-1} and above

A classification tree was developed at the relative cost of 0.714.

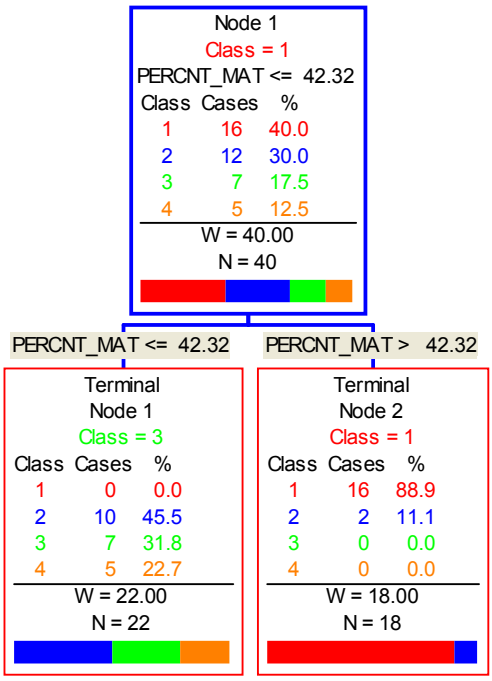


Figure 43. Classification tree for Magnesium

Prediction success of the tree was only 55%. Class 1 had no misclassification cost, Class 3 had only a 0.14 misclassification cost. But Class 2 and Class 4 were completely misclassified with test data.

The only splitting criterion used was percentage of MAT within watersheds. The watersheds with MAT complex greater than 42.32% displayed magnesium values falling within Class 1. The terminal node 2 was composed of mostly all the cases of Class 1 and only 2 cases of Class 2.

Only Marion et al. (1989) has studied magnesium concentration in the Arctic soils. The authors stated that, distribution of magnesium within watershed was moisture

dependent and hence soils with Riparian complex had more magnesium concentration than MAT complex soils. However, for the current research, the relationship between MAT complex and magnesium could be linked with age of till surface i.e. older tills would exhibit lower concentrations of magnesium.

Terminal node 1 was found to be impure and was composed of all the remaining cases of magnesium classes.

4.3.8.a Sodium Regression Model:

The regression model developed was able to explain only 50% of variance in sodium estimates using a single landscape factor: amount of Moist Acidic Tundra.

$$Na = 0.309 - 0.006\ (Prct\_MAT)$$

As per the regression model obtained for Sodium, its concentration would deplete by 0.006 mg L^{-1} if the amount of MAT increased by 1 percent. Again, this relationship might be pointing towards influence of till age.

None of the studies carried out earlier were able to establish relationship between landscape factors and Sodium concentrations within arctic watersheds. Only Everett et al (1989) mentioned that Sodium ions were relatively greater in concentration immediately after melting, but the concentration was below the level of detection throughout the growing season. Similar to other cations, this regression model might be emphasizing the role of till age, being related to chemical weathering processes, which are dominant in younger tills.

4.3.8.b Classification Tree for Sodium estimates:

Class 1 – 0.1 mg L^{-1} to 1.0 mg L^{-1}

Class 2 – 1.1 mg L^{-1} and above

The value ranges of sodium estimates were very small. The classification tree was developed at the relative cost of 0.762. The only splitting criteria used, was the amount of Moist Acidic Tundra expressed in terms of percentage.

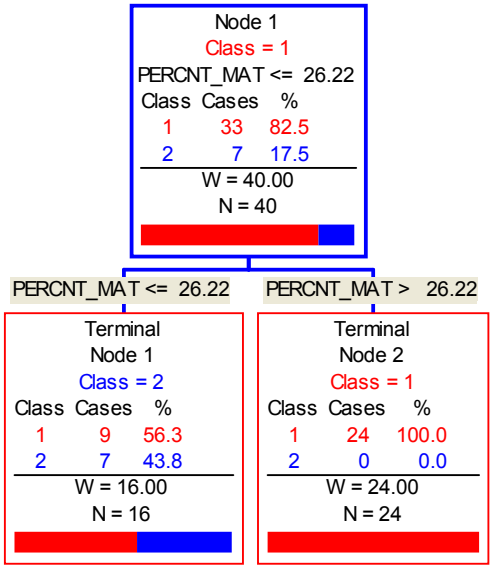


Figure 44. Classification tree for Sodium

Prediction success on the test data was 65%. The misclassification cost for Class 1 was 0.33, as 11 cases out of 33 were incorrectly classified. The misclassification cost for Class 2 was 0.43 as 3 out of 7 cases were not identified correctly.

As per the tree developed, Class1 (less than 1 mg L^{-1}) was identified when the amount of MAT was above 26.22 percent. However, if the amount of MAT was less than 26.22 percent then sodium concentration would be higher than 1 mg L^{-1}, as indicated by terminal node 2. Thus, the classification tree also highlights the role of till surface age.

4.3.9.a Potassium Regression Model:

The overall range of potassium estimates was very short. Potassium was not detected in five out of forty lakes, whereas other lakes showed concentrations ranging from 0.1 to 0.5 mg L^{-1}. The following relationships were obtained through the regression model:

$$K = 0.212 - 0.003\ (Prct\_MAT) + 0.019\ (EZD) + 0.114\ (Till\ age\ 1)$$

Approximately 63% of variance was explained by the currently established regression model. It indicated that 1 percent increase in the amount of MAT would result in reduction of Potassium contents by 0.003 mg L^{-1}. However, with increase in EZD by 1 meter would result in an increase of the potassium contents in lakes by 0.019 mg L^{-1}. The model also showed that lakes situated on age category 1 of the till surface then they should exhibit potassium contents higher by 0.114 mg L^{-1}.

A very weak negative relationship was observed between amount of MAT and potassium concentration. Also, lakes on Itk2 re-advance till surfaces (the youngest) i.e. Till age 1, were found to positively relate to K ion concentration. The probable reason behind both relationships might be that youngest till is experiencing more chemical weathering and may have parental material rich in K ions e.g. Sandstone.

Relative importance of predicting variables:

Among these predicting variables Prct_MAT has relatively higher influence than other two, based on following standardized coefficients:

Prct_MNT (-0.511), EZD (0.393), and Till age 1 (0.331).

4.3.9.b Classification Tree for Potassium estimates:

K ion estimates were divided into two classes. They were as follows:

Class 1 – 0 mg $L^{-1}$to 0.25 mg L^{-1}

Class 2 – 0.26 mg $L^{-1}$and above

Classification tree was obtained at the relative cost of 0.56 with two terminal nodes. The only splitting criterion emerged was Mean Patch size of MAT complex within buffer.

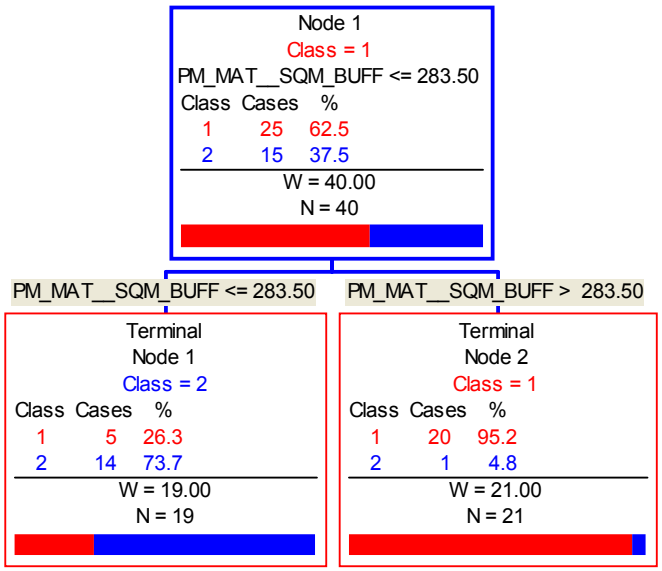


Figure 45. Classification tree for Potassium

Very good prediction success was obtained i.e. 70%. Looking at misclassification cost on the test data, Class 1 had 0.36 and Class 2 had 0.2 cost.

Terminal node 2 was more on pure side, identifying 20 out of 25 class 1 cases. It showed that lower values of K ions (Class 1) occurred when Mean Patch size of the MAT complex within a buffer was greater than 283.5 sq meters. At the same time, terminal node 1 had 14 out of 15 cases of Class2. Marion et al. (1989) observed that MAT vegetation community had higher rate of K uptake. Moreover, as illustrated by Giblin et al . (1991), Eriophorums in the MAT community had relatively deep root structure allowing more access to K rich soil horizons. Based on these attributes of MAT

complex, it is logical that watersheds with larger MAT complex patches would experience low amounts of K being released into the lakes.

4.3.10.a Sulfate Regression Model

The regression model developed for Sulfate concentrations was as follows:

SO4 = 4.351 + 0.059 (PD_MM) – 0.407 (LSI_B_Fen) + 0.215 (LSI_ Shrub)

The model indicates that if patch density of Mountain Meadows increases by 1 patch/100 hectares then Sulfate concentration in lakes would increase by 0.059 mg L^{-1}. However, if the Landscape Shape index of Fen within the buffer zone increases by 1 then it would reduce the sulfate contents by 0.407 mg L^{-1}. In contrast, if the LSI of Shrub at watershed increases by 1 then sulfates would be increased by 0.215 mg L^{-1}.

This model explained only 37% of variance of SO4 estimates using landscape factors. But this model could be claimed a poor one as most of the watersheds have sulfate values less than 4 mg L^{-1}, whereas only 5 watersheds have values greater than 16 mg L^{-1}.

Relative importance of predicting variables:

Out of these predicting variables LSI_B_Fen has the relatively higher influence on sulfates as it has the highest standardized coefficient of (-0.512). The standardized coefficients for PD_MM and LSI_Shrub were (0.401) and (0.398), respectively.

4.3.10.b Classification Tree for Sulfate estimates:

Class 1 1.3 mg L^{-1}– 2.9 mg L^{-1}

Class 2 3.0 mg L^{-1} and above

A classification tree was developed at the relative cost of mere 0.286. The tree exhibited only 2 terminal nodes, one of which was a pure node classifying most of the class 1 cases.

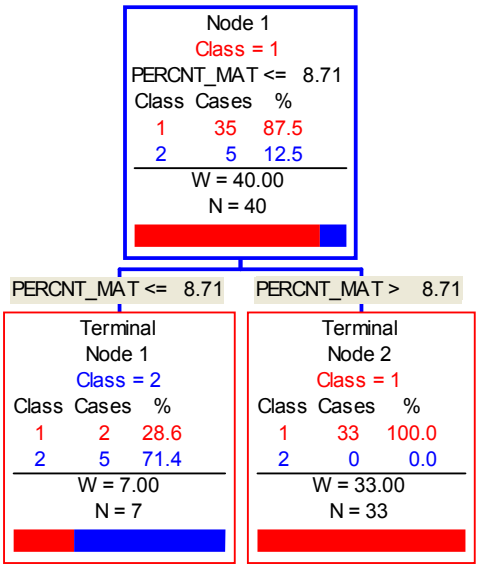


Figure 46. Classification tree for Sulfate

Prediction success of the tree on the test sample was 90%. Misclassification cost for Class 1 was only 0.09 where as the misclassification cost for Class 2 was 0.2.

The only criterion used to achieve the split was the percentage of Moist Acidic Tundra complex. If the percentage of MAT complex was greater than 8.7, then Class 1 of

SO_4 was identified. Whereas, all the higher values of SO_4 (Class 2) were related to lower MAT percentage than 8.7. Similar to other cations and anions, the influence of till age might have been expressed via the amount of MAT suggesting lower chemical weathering and relatively lower amounts of SO_4 present in older till soils.

4.3.11.a Chlorides Regression Model:

With use of landscape factors, the regression model developed was able to explain approximately 60% of variance of chloride estimates. The model was as follows:

Chloride = 0.186 + 0.042 (Prct_Heath) – 0.304 (Till age 2) + 0.001 (PD_ Fen) + 0.023 (MS_B_Fen) + 0.023 (MS_B_MM_Prct)

As per the regression model, if the amount of Heath complex increases in watershed then chloride contents of lakes would raise by 0.042 mg L^{-1}. Contrary to that, if a lake situated on till surface of age category 2, then it would experience deficit of chloride contents by 0.304 mg L^{-1}. If PD_Fen increases by 1 patch/100 hectares, the concentration of chlorides would increase by 0.001 mg L^{-1}. Likewise, if the complexity of Fen patches within the buffer zone expressed as MS_B_Fen if increase by 1 then it would result in increased Chlorides by 0.023 mg L^{-1}. Also, if the complexity of MM patches within buffer zone increases by 1 percent then it would cause the Chloride contents to increase by 0.023 mg L^{-1}.

Lakes present on Itk 2 (initial) have lower concentration of chloride ions, hence the negative relationship was observed. Lakes present on this till are expected to have

0.304 times less Cl ion concentration than lakes present on other glacial tills. Cl ion estimates for lakes present on Itk 2 (initial) was 0.1 mg L^{-1}, except GTH 126 with 0.8 units. Otherwise the average for other lakes was 0.6 mg L^{-1}.

Relative importance of predicting variables:

The standardized coefficients of these predicting variables were as follows: Prct_Heath (0.568), Till_age2(-0.322), PD_Fen(0.212), and MS_B_MM_Prct(0.364). Thus it can be concluded that Prct_Heath was more influential amongst the predicting variables.

4.3.11.b Classification Tree for Chlorides estimates:

Cl ion concentrations were divided into following classes:

Class 1 – 0.1 mg L^{-1} to 1.0 mg L^{-1}

Class 2 – 1.1 mg L^{-1} and above

The cost at which the classification tree developed was 0.821.

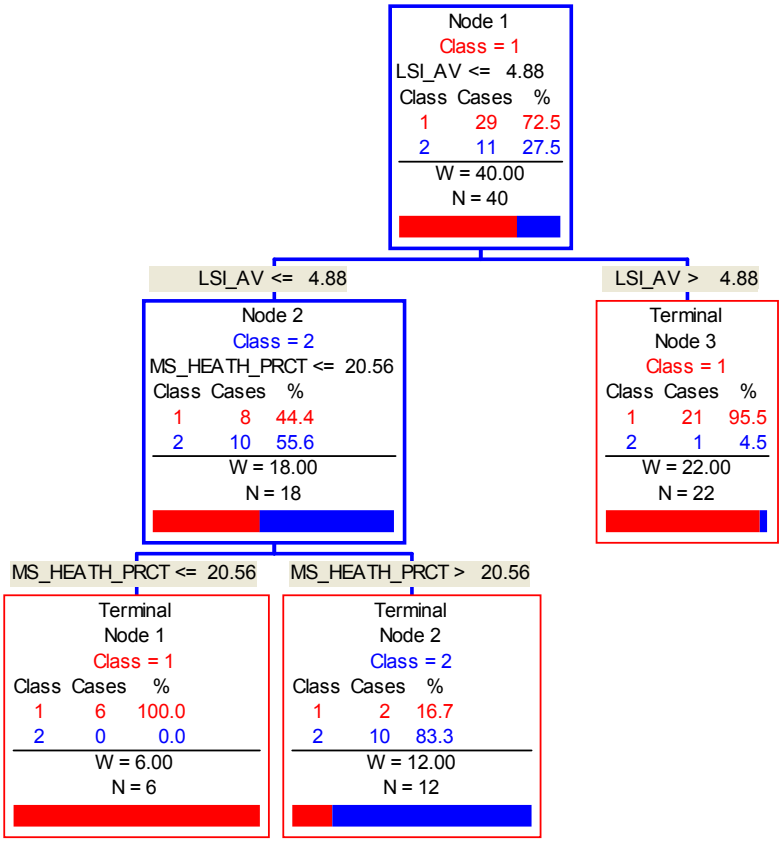


Figure 47. Classification tree for Chlorides

Prediction success rate on the test data was 65%. Class 1 had very low misclassification cost during test i.e. 0.28 as only 8 out of 29 cases were misclassified. But for class 2, 6 out of 11 were wrongly classified and cost of it was 0.56.

It displayed 3 terminal nodes and had two splitting criteria. First criterion was Landscape Shape Index of Aquatic Vegetation. According to the tree, when LSI of AV complex was higher than 4.33, most of the class 1(relatively lower Cl concentration cases) were classified. Terminal node 3 represents the situation. But when LSI of AV

complex is less than 4.33 and percent change in Mean Shape of Heath complex is higher than 20.56 then most of the cases of Class 2 (higher Cl concentration) were classified.

4.3.12.a Dissolved Inorganic Carbon Regression Model:

DIC = 22.966 + 0.893 (Prct_MNT) + 27.86 (Till Age 2) – 5.169 (FD_B_Shrub) + 3.146 (EZD) + 0.596 (LSI_MM)

According to the developed regression model for Dissolved Inorganic Carbon (DIC), if the amount of MNT complex within the watershed increases by 1 percent then the DIC concentration of lake would be raised by 0.893 mg L^{-1}. Similarly, if the lake is present on till surface age category 2 then it would experience a high concentration of DIC i.e. by 27.89 mg L^{-1}. On the contrary if the Fractal Dimension of Shrub patches within the buffer zone increases by 1 percent that would reduce the DIC concentration by 5.169 mg L^{-1}. However, an increase in EZD by 1meter would result in an increase in DIC contents of lakes by 3.146 mg L^{-1}. Also, an increase in LSI_MM by 1 would cause an increase in DIC by 0.596 mg L^{-1}.

Lakes situated on the Till_age2 (Itk 2 Initial) have shown higher contents of DIC. The average value of DIC for these lakes was 88.65 mg L^{-1}, whereas the average for lakes present on other tills was 30 mg L^{-1}. This relationship supports the findings of Keller et al. (2007), who observed higher cold acid digestible mineral ion concentrations especially carbonates in the It2 till surface and streams flowing on it.

The relationship with MNT also should refer to the till age, as MNT is dominant on relatively younger till surfaces such as Itk2.

Relative importance of predicting variables:

Prct_MNT was found to be the relatively more influential landscape factor than others. The conclusion was drawn on the basis of following standardized coefficients of these predicting variables: Prct_MNT(0.493), Till age 2(0.372), FD_B_Shrub_Prct(-0241), EZD(256), and LSI_MM(0.224).

4.3.12.b Classification Tree for DIC estimates:

Class ranges for DIC used for CART are as follows:

Class 1 - < 3 mg L^{-1}

Class 2 - 3 mg L^{-1} to 37 mg L^{-1}

Class 3 - 38 mg L^{-1} to 71 mg L^{-1}

Class 4 - 72 mg L^{-1} to 105 mg L^{-1}

The classification tree was developed at 0.663 value.

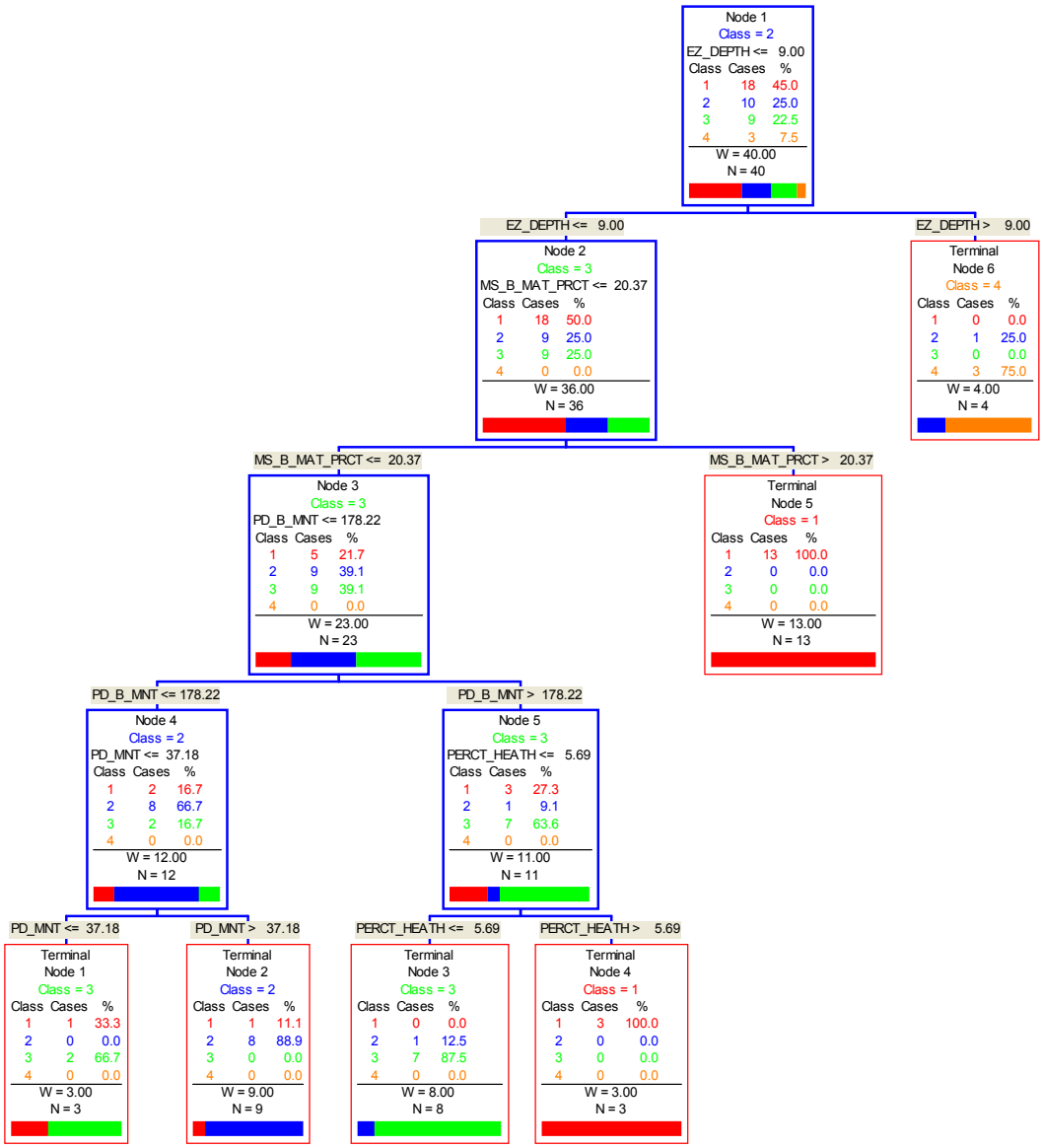


Figure 48. Classification tree for DIC

Prediction success on the test sample was 57.5%. The lowest misclassification cost of 0.28 was attributed class 1, where out of 18 records 13 were identified correctly. Only 3 cases out of 10 for class 2 were correctly classified costing 0.7. The case was

similar with Class 4, where out of 3, only 1 was correctly classified. Misclassification cost for Class 4 was relatively lower i.e. 0.33 because out of 9, 6 cases were identified accurately.

The classification tree has also highlighted the importance of the euphotic zone depth i.e. when euphotic zone depth was greater than 9 meters, highest value class of DIC i.e. class 4 was segregated from other classes. But when euphotic zone depth was less than 9 meters and percent change in mean shape index of MAT within buffer was more than 20.37, the class 1 cases were identified.

If the percent change in the Mean Shape Index of MAT within buffer was less than 20.37 and the patch density of MNT within the buffer was less than 178.22 and the patch density of MNT at the class level was greater than 37.18 then majority of cases for Class 2 were classified. On the contrary, if the patch density of MNT within buffer was higher than 178.22, and amount of heath was less than 5.69%, then the majority of Class 3 cases were identified.

According to Keller et al (2007), younger tills in the Arctic exhibit relatively higher storage of carbonates and the capacity to deliver them in streams and other waterbodies as the permafrost melts. The shape complexity and patch density conditions of MAT and MNT may indicate that watersheds with MAT as a major land cover may not contribute to DIC. For example, terminal node 5 indicated that when MAT patches were complex within buffers water channels i.e. if they were in MAT dominant watersheds then DIC was in the lowest category or Class1.

4.4 Discussion

4.4.1 Comparison with earlier attempt of classification using the entire SPOT image:

Chaudhuri (2008) classified the Toolik region using the same SPOT image with a hybrid classification technique. The output derived initially using a knowledge based classifier was further refined with Artificial Neural Network. The Overall accuracy achieved by combining various classifiers was 75.57% with overall Kappa value of 0.684.

On the other hand, the current study adopted a watershed level classification approach. Moreover, keeping in mind the probable role of different vegetation communities as nutrient sink or source, the land cover scheme was modified. For example, Aquatic complex of Chaudhuri (2008) was divided into Water and Aquatic vegetation complex categories. Heath complex was either included in Barren or MNT complex by Chaudhuri (2008), but the Heath complex was found to be a source of phosphorus. Hence, in the current research Heath was treated as a different vegetation community. At the watershed level, ISODATA clustering method was able to differentiate between Heath and Barren as well Heath and MNT complexes. This resulted in higher accuracies for individual classes. For the Barren complex, the hybrid method by Chaudhuri (2008) was approximately 86% correct with Kappa value of 0.83 whereas at watershed level, Heath complex showed overall accuracy of more than 90%.

Fen complex, which was called as Wet Sedge Tundra by Chaudhuri (2007), experienced higher accuracies at watershed level. The confusion between shallow waterbodies and the Snowbed complex was avoided by the watershed level approach.

Even though watershed level approach was not completely successful in differentiating MAT complex and MNT complex, it was relatively more successful than classifying them using the entire image. The hybrid classifier had a greater Producer's accuracy for MAT complex (86%) compared to User's accuracy (68%), while MNT complex showed lower Producer's accuracy of 66% and 81% of User's accuracy. Watershed level classification showed more than 90% of Producer's accuracy and more than 77% User's accuracy for MAT complex. Also, MNT complex obtained more than 66% of User's accuracy and more than 85% of Producer's accuracy.

This research was also successful in differentiating between Shrubs and Mountain Meadows. Similar to that, Shrub complexes occurring near large prominent streams were classified as Riparian Shrub complexes. This allowed for the exploration of their role in nutrient recycling process and probable impacts on lake productivity.

4.4.2 Comparison between Regression models and Classification trees

Direct comparison between two the analysis methods (Regression analysis and Classification Tree analysis) was not possible. The usefulness of a regression model was depicted by its R^2 value, indicating how much variance within the dependent variable has been explained by the model. The importance of independent variables was judged by the significance level, multi-collinearity, and change in R^2 with addition of each variable.

On the other hand, CART was not able to establish a Regression tree for any of the variables. Hence, Classification trees were developed. The success of a classification tree was evaluated based on the relative cost at which the tree was formed and the prediction success rate of the testing data. As this was a hierarchical approach of data classification, the final output was inclusive of relative importance of each variable used for splitting.

When regression and classification tree models were compared for each chemical parameter, there were some similarities and differences observed. For example, a relationship between Areal estimates of Chlorophyll a and MSI of Shrub were important at the buffer level but MSI Shrub at the watershed level was used by the Classification tree method. The regression model for Chlorophyll a Areal estimates used Euphotic Zone but Maximum depth of lake was used by the classification tree method. Patch density of Fen was a common variable used in both methods for Total Nitrogen. Patch density of Shrub complex was used in the Regression model of Total Phosphorus, whereas percentage of Shrub complex within the buffer zone was important in the Classification tree. Patch density of MAT complex was involved in both methods for Magnesium and Sodium estimates. Euphotic zone depth was used by both methods for Calcium and DIC. Thus, it can be concluded that both the methods recognized influence of the same or similar landscape factors in most of the cases, but in certain instances they were considered at different levels i.e. either watershed level or buffer level.

Similarities and differences could be attributed to different approaches these analytical methods use. However, CART would be more useful when a large number of cases are available to find patterns and verity the tree, but regression will be useful even with limited cases like this study. Thus each method has its own usefulness and limitations. To predict productivity of arctic lakes more precisely, more research will be needed with more number of cases.

4.4.3 Limitations

1. Accessibility – Out of forty watersheds in this study only one watershed was accessible by road. Other watersheds did not have a direct access from the Toolik Field Station. Therefore, field work was confined to the helicopter use. As the helicopter was shared with other research groups, limited flying time was allotted for this study.
2. Time limitation – Time period of two weeks in the field was sufficient to visit only six watersheds. Weather conditions such as frequent rain and cloud cover affected the working conditions outdoor as well as hindered the GPS reception.
3. Cost limitation – Flying to the study area, stay at the Toolik Field Station, and using the helicopter to visit the watersheds were very expensive. That was the reason why field work was limited to only two weeks.
4. Extent of Ancillary data – DEM available at 5 m X 5 m spatial resolution did not cover the watersheds for GTH 143, GTH 145 and GTH 146 lakes. The coarse resolution DEM (30 m X 30 m) was helpful in deriving watersheds for GTH 143 and GTH 145, although it was unable to generate watershed for GTH 146 because of its poor quality.

The scale of ancillary datasets such as Geology and Geomorphology was 1:25000. When these layers were incorporated in the classification, they created very sharp boundaries for vegetation types and provided classified images a blocky appearance. Hence, they were not useful in improving classification accuracy.

5. Image availability and Quality – The SPOT image had only 3 spectral bands i.e. Green, Red, and Near Infrared. This image was pan-sharpened, although the method of that process was not provided by the SPOT Corporation. Therefore, the reasons behind spectral overlaps could not be addressed. Certain portions of the image were also affected by clouds and shadows. Even though, no clouds were present in the watersheds utilized by this research, shadows influenced classification of the watersheds near GTH 153.

6. Limited numbers of points were collected for classes other than MAT and MNT, which may be due to the fact that the other classes occupied less area in watersheds. This affected the accuracy of the classes dramatically. For example, GTH 133 had only three points collected for Fen complex and out of which only 1 point was correctly identified. The result was only a 33% Producer's accuracy for that watershed. The study was carried out with very limited number of lakes. Therefore, Regression trees could not be generated. Moreover, while categorizing the data for Classification tree method, a balanced representation of each category within dependent variables was not possible. Therefore, predicting the success of Classification tree was difficult. It was also not possible to develop separate dataset for tree develo*p*ment and testing processes.

7. As mentioned earlier, the season of data collection and satellite image acquisition were different (chemistry data collection was carried out during summer of 2001, 2002,

and 2003 yet, the satellite image used was from summer 2005). As no drastic changes in landscape factors or lake chemistry were expected in such a short period of time, the relationship between them might not have been affected. However, one cannot ignore the fact that lake chemistry is linked with contemporary weather conditions and associated phenology of vegetation communities. Due to the difference in the season of data collection and satellite image acquisition, such impacts were not addressed in this research. Therefore, results of this study could only be perceived as more generalized.

8. Limited literature availability – Earlier studies (Chapin et al., 1989; Oechel, 1989) were carried out to study effect of disturbances. Some of the studies only concentrated on major nutrients (Chapin et al., 1988; Schimel, 1996) or considered nutrient levels in soils and plants (Marion et al., 1989). A Toposequence study carried out by Giblin et al. (1991) formed a broad basis for this research, but the study was generalized and did not capture local variations at the watershed level.

CHAPTER V

CONCLUSION

This dissertation has examined probable functional roles of certain landscape factors in controlling primary productivity of Arctic lakes. Spatial characteristics of landscapes were obtained using thematic maps of watersheds derived from a SPOT image. Hence, developing a method to achieve acceptable thematic accuracies was one of the objectives of this research. Using statistical analyses methods i.e. Multiple Regression Analysis and Classification Tree method, relationships between certain chemical properties of the lakes and selected landscape factors were explored. This chapter summarizes the major findings and illustrates the success of it in the view of limitations and scope of future research work.

5.1 Conclusions

Objective 1: Classification Accuracy:

Overall accuracies ranged from 82.75% to 95%. Water class experienced the accuracy of 100%. However, the other major land cover classes, namely, Moist Acidic Tundra complex and Moist Non-Acidic Tundra complex displayed considerable spectral overlap even at watershed level.

Hence, Moist Acidic Tundra complex exhibited average Producers' Accuracy of 92% and User's accuracy of 91%. Similarly, Moist Non-Acidic Tundra complex displayed an average Producer's Accuracy and User's Accuracy of 88%. The average Producer's accuracy for Shrub complex was 92% along with User's Accuracy of 97%. The Aquatic vegetation complex achieved Producer's and User's Accuracy of 100%. The Fen complex showed relatively lower Producer's Accuracy i.e. 80% but its User's Accuracy was 91%. The Heath complex which experienced lower accuracies being spectrally similar with barren areas displayed lower Producer's Accuracy of 74%.while its User's Accuracy average was 92%. Thermokarst areas and Barren Complexes experienced Producer's Accuracy of 100% and User's Accuracy was 83%. Riparian complexes exhibited average of 90% for both, Producer's and User's Accuracy. The Snowbed complex suffered from limited number of ground truth points and had the lowest Producer's Accuracy of 58%. However, its average User's Accuracy was 100%. The Mountain Meadows complex was present only in one watershed out of six assessed ones, displayed 75% of Producer's and User's Accuracy.

In general, the watersheds were well above the recommendation value of 85% accuracy with kappa value of 0.75 (Congalton et al., 1995), with the only exception of GTH 135, which showed an overall accuracy of 82.75%. Thus, the research was successful in achieving its objective of producing higher thematic accuracies compared to historic studies while adopting a comparatively detailed scheme of land covers.

Objective 2: Identification of a set of landscape factors suitable for the research and exploring relationships between the landscape factors and lake chemistry variables

After reviewing the available literature, nine vegetation communities/land cover types were included in this study. Five types of landscape metrics were derived for each land cover type at the watershed level as well as buffer level (20 meter from water channels).

Among the land cover types studied, Shrub, MAT, MNT, and Heath complexes emerged as significant in most of the models. Patch Shape complexity of Shrub complex at buffer level exhibited strong influence on Chlorophyll a (Areal) estimates.

The percentage and patch size of Moist Acidic Tundra complex were inversely correlated with Cations such as Sodium and Potassium. On the other hand, percentage of Moist Non-Acidic Tundra complex was positively correlated with DIC and Conductivity estimates. Both, Moist Acidic Tundra complex and Moist Non-Acidic Tundra complexes, indicated role of till surface age as a potential source of ions.

The Heath complex was recognized as one of the indicators for Chlorophyll a estimates, both, Volumetric and Areal. For the areal estimates, fragmentation (Patch density) of Heath complex at watershed level was influential. However, Patch density of Heath complex within buffer was more significant for Volumetric estimates. Even though Heath complex was considered as a source of phosphorus, no direct relationship could be established with between Heath complex and Total Phosphorus estimates.

Patch density of Fen complex within the buffer distance was critical for total Nitrogen contents of lakes. Mountain Meadows were present in only nine watersheds but they were strongly related to Ca ion concentrations and Chlorides of the lakes related to those particular watersheds.

Among the lake physical properties, euphotic zone depth was found to be correlated with Cations like Ca, K, and Mg, as well as with Chlorophyll a Areal estimates. However, no literature is available to illustrate the observed relationships.

In general, Regression Analysis was more useful in deriving relationships between landscape factors and lake chemistry parameters. The models developed were significant at 95% confidence level. Classification tree method needs more number of watersheds in order to develop significant trees. Complete potential of this method was not exploited in this research because of limited sample size. However, both methods pointed towards same land cover types even though each method might have differed for the scale to determine the influence on lake chemistry variable.

5. 2 Significance of the findings in the view of Climate Change

It has been predicted that with rising temperatures in the Arctic, more Shrub complex could be observed in low lying areas of basins (Tape et al., 2006). The current research findings suggest that the probable increase in amount of shrub and changes in soil-water regimen may enhance the nutrient influx to arctic lakes. Therefore, in future, it is likely that the lakes will exhibit higher productivity.

The amount of MAT complex increases with age of till surfaces (Walker et al., 1994), and warming of the soil will support growth MAT. Thus the MNT complex will gradually lose its dominance over the younger tills as they age. This may lead to higher release of nitrogen compounds and cations from upslope MAT regions to water bodies.

With expected rise in soil temperatures and greater permafrost thawing, younger tills such as Itk 2 may release more Cations and Anions in near future affecting lake properties e.g. conductivity.

5.3 Future Research

1. Moderate spatial resolution is enough for obtaining thematic map accuracies; however, use of hyperspectral dataset may result in precise signature establishment and classification of relevant vegetation communities within Arctic watersheds.
2. Image acquisition should be carried in tandem with lake chemistry data collection. Images captured just prior to lake data collection will provide a logical landscape scenario of influential factors on lake water chemistry.
3. Visiting more watersheds and more frequently (Pre and post classification) would help creating more precise maps of the area.
4. Availability of precise Digital Elevation Model will be useful in detailed topographic analysis.
5. A study should be carried out using multi-season data and images to confirm the relationships between landscape factors and lake chemistry.
6. In general, nutrients, cations, and anions are more released during the initial snow melt or before leafing of the vegetation communities and at the end of the season when

they undergo senescence. Therefore, early season and late season lake chemistry data should be collected and analyzed along with contemporary satellite data.

To carry out more holistic research, surface water and sub surface water coming out of different vegetation patches should be studied for its chemical or nutrient contents throughout growing season.

REFERENCES

Alaska Department of Fish and Game. Alaska's Arctic Tundra. Retrieved Jan 6, 2010, from http://www.wc.adfg.state.ak.us/index.cfm?adfg=ecosystems.tundra

Alaska Climate Research Center. (2008). Alaska Climatogoly: Climate Summaries. Retrieved Jan 20, 2010, from http://climate.gi.alaska.edu/Climate/

Allan, J.D., & Johnson, L.B. (1997). Catchment-scale analysis of aquatic ecosystems. *Freshwater Biology, **37***, 107–112.

Andreassen, H.P., Ims R.A. & Stenseth N. C. (1996). Discontinuous habitat corridors: effects on male root vole movements. *Journal of Applied Ecology*, *33*, 555–560.

Arbuckle, K.E., & Downing, J.A. (2001). The influence of watershed land use on lake N:P in a predominantly agricultural landscape. *Limnology and Oceanography, 46,* 970-975.

Auer B., Elzer, U., & Ardnt, H. (2004). Comparison of pelagic food webs in lakes along a trophic gradient and with seasonal aspects: influence of resource and predation. *Journal of Plankton Research. 26,* 697-709.

Bott, T. L., Montgomery, D. S., Arscott, D. B., & Dow, C. L. (2006). Primary productivity in receiving reservoirs: links to influent streams. *Journal of the North American Benthological Society, 25,* 1045–1061.

Burt, T. & Butcher, D. (1986). Stimulation from simulation – a teaching model of hillslope hydrology for use on microcomputers. *Journal of Geography in Higher Education, 10*, 23–39.

Camp, N.J., & Slattery, M.L. (2002). Classification tree analysis: a statistical tool to investigate risk factor interactions with example for colon cancer. *Cancer Causes Control, 13*(9),813-823.

Carpenter, S. R., Frost, T. M., Heisey, D., & Kratz, T. (1989). Randomized intervention analysis and the interpretation of whole ecosystem experiments. *Ecology, 70,* 1142-1152.

CART documentation: Desktop Help

Chapin, F.S. III, Fetcher, N., Kielland, K., Everett, K.R., & Linkins, A.E. (1988). Productivity and nutrient cycling of Alaskan tundra: enhancement by flowing soil water. *Ecology, 69*, 693-702.

Chapin, F.S. III, & Shaver, G.R. (1988). Differences in carbon and nutrient fractions among arctic growth forms. *Oecologia, 77*, 506-514.

Chaudhuri D. (2008). *Hybrid Image Classification Technique for Land-Cover Mapping in the Arctic Tundra, North Slope, Alaska*. Unpublished doctoral dissertation, University of North Carolina, NC. (Personal Communication).

Cole J.J. (1994). Carbon dioxide supersaturation in the surface waters of lakes. *Science, 265,* 1568-1570.

Cole J.J., Prairie Y.T., Caraco N.F., Mcdowell W.H., Tranvik L.J., Striegl R.G., Duarte C.M., Kortelainen P., Downing J.A., Middelburg J.J., & Melack J. (2007). Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems, 10* (1), 172-185.

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

Crosbie, B., & Chow-Fraser, P. (1999). Percentage land use in the watershed determines the water and sediment quality of 22 marshes in the Great Lakes basin. *Canadian Journal of Fisheries and Aquatic Sciences*, 56,1781–1791.

Dillon, P. J., & Molot, L. A. (1997). Effect of Landscape form on Export of Dissolved Organic Carbon, Iron, and Phosphorus from Forested Stream Catchments. *Water Resources Research*, *33*(11), 2591–2600.

Dong, J. , Kaufmann, R.K. , Myneni, R.B. , Tucker, C.J. , Kauppi P. & Liski, J., *et al.* (2003). Remote sensing estimates of boreal and temperate forest woody biomass: Carbon pools, sources, and sinks, *Remote Sensing of Environment,* ***84*****,** 393–410.

Doran, P.T., McKay, C.P., Adams, W.P., English, M.C., Wharton, R.A., & Meyer, M.A. (1996). Climate forcing and thermal feedback of residual lake-ice covers in the high Arctic. *Limonogical Oceangraphy, 41*, 839-848.

ERDAS Imagine Field Guide. (2005).

Everette, K.R., Marion, G.M., & Kane, D.L. (1989). Seasonal Geochemistry of an Arctic Tundra Drainage Basin. *Holarctic Ecology, 12*(3), 279 – 289.

Felix, N.A., & Binney, D.L. (1989). Accuracy assessment of a Landsat-assisted vegetation map of the coastal plain of the Arctic National Wildlife Refuge. *Photogrammetric Engineering & Remote Sensing, 55,* 475 - 478.

Fleming, M.D. (1988).An integrated approach for automated cover-type mapping of large inaccessible areas in Alaska. *Photogrammetric Engineering & Remote Sensing, 54, 357*- 362.

Forman, R. T. T., & Godron, M. (1986). *Landscape Ecology*. New York, NY: John Wiley.

Fragstats Conceptual Background: *Landscape Pattern Metrics*. Retrieved Dec 22, 2007, from http://www.umass.edu/landeco/research/fragstats/documents/Conceptual%20Background/Background%20TOC.htm

Friedl, M. A., Brodley, C. E., & Strahler, A. H. (1999). Maximizing land cover classification accuracies produced by decision trees at continental to global scales. *IEEE Transactions on Geoscience and Remote Sensing, 37*(2), 969–977.

Frohn, R.C., & Hao, Y. (2006). Landscape metric performance in analyzing two decades of deforestation in the Amazon Basin of Rondonia, Brazil. *Remote Sensing of Environment, 100(2)*, 237-251.

Gamon, J. A., Field, C. B., Goulden, M. L., Griffin, K. L., Hartley, A. E., Joel, G., Peñuelas, J., & Valentini, R. (1995). Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecological Applications, 5*, 28–41.

Gerçek, D. (2002). *Improvement of Image Classification with the integration of Topographical data*. Master of Science Thesis, The Graduate School of Natural and Applied Sciences, Middle East Technical University.

Gergel, S.E., Turner, M.G., Miller, J.R., Melack, J.M., Stanley, E.H. (2002). Landscape indicators of human impacts to riverine systems. *Aquatic Science, 64*, 118–28.

Giblin, A. E., Nadelhoffer, K. J., Shaver, G. R. , Laundre, J. A., & McKerrow, A. J. (1991). Biogeochemical diversity along a riverside toposequence in arctic Alaska. *Ecological Mono-graphs, 61,* 415-436.

Griffith, J.A., Martinko, E.A., Whistler, J.L., Price, K.P. (2002). Interrelationships among landscapes, NDVI, and stream water quality in the US central plains. *Ecological Applications, 12*, 1702–18.

Grey, J., Kelly, A., & Jones, R.I. (2004). High intraspecific variability in carbon and nitrogen stable isotope ratios of lakes chironomid larvae. *Limnology and Oceanography, 49*, 681-689.

Grunwald, S., Daroub, S.H., Lang, T.A. & Diaz, O.A. (2009). Tree-based modeling of complex interactions of phosphorus loadings and environmental factors. *Science of The Total Environment, 407(12)*, 3772-3778.

Gustafson, E. J. (1998). Quantifying landscape spatial pattern: what is the state of the art? *Ecosystems,1*, 143-156.

Haines-Young, R. & Chopping, M. (1996). Quantifying landscape structure: A review of Landscape Indices and their application to forested landscapes. *Progress in Physical Geography*, 20(4), 418-445.

Hamilton, T,D. (1986). Late Cenozoic glaciation of the Central Brooks Range. In Hamilton, T. D., Reed, K. M., & Thorson, R. M.(Ed.). *Glaciation in Alaska: the Geologic Record. Alaska Geological Society* (pp. 9-49).

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., & Sohlberg, R. (2000). Global land cover classi cation at 1km spatial resolution using a classi cation tree approach. *International Journal of Remote Sensing, 21*, 1331–1364.

Hartman, C.W., & Carlson, R.F. (1973). Water Balance of a Small Lake in a Permafrost Region. University of Alaska, Fairbanks, Institute of Water Resources.

Hasler, Arthur D. & International Association for Ecology. & International Limnological Congress, 1975 Coupling of land and water systems / edited by Arthur D. Hasler Springer-Verlag, New York.

Hershey A. E., Gettel, G.M., McDonald, M.E., Miller, M.C., Mooers, H., O'Brien, W.J., Pastor, J., Richards, C., & Schuldt, J.A. (1999). GTH model for landscape control of arctic lakes food web. *Bioscience, 49*, 887-889.

Hershey, A.E., Beaty, S., Forino, K., Keyse, M., Mou, P.P., O'Brien, W.J.O., Ulseth, A.J., Gettel, G.A., Lienesch, P.W., Luecke, C., McDonald, M.E., Mayer, C.H., Miller,

M.C., Richards, C., Schuldt, J.A., & Whalen, S.C. (2006). Effects of Landscape factors on fish distribution in arctic Alaskan lakes. *Freshwater biology, 51*, 39-55.

Hiscock, J.G., Thourot, C. S., & Zhang, J. (2003). Phosphorus budget—land use relationships for the northern Lake Okeechobee watershed, Florida. *Ecological Engineering, 21*(1), 63-74.

Hope, A., & Stow, D. (1995). Shortwave reflectance properties of Arctic tundra. In Reynolds, J., & Tenhuenen, J. (Eds.), Landscape function and disturbance in Arctic tundra. Ecological Studies, 120,155– 164. Heidelberg: Springer.

Hope, A. S., Pence, K. R., & Stow, D. A. (2004). NDVI from low altitude aircraft and composited NOAA AVHRR data for scaling Arctic ecosystem fluxes. *International Journal of Remote Sensing, 25(20)*, 4237 – 4250

Hunsaker, C. T., & Levine, D. A. (1995). Hierarchical approaches to the study of water quality in rivers. *BioScience, 45*(3),193-203.

Hynes, H.B.N. (1975). The stream and its valley. *Internationale Vereinigung für theoretische und angewandte Limnologie Verhandlungen, 19*, 1-15.

International Panel on Climate Change, 2007: Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland, 104

International Arctic Science Committee (Content Partner); Mark McGinley (Topic Editor). 2007. "Arctic climate change scenarios for the 21st century projected by the ACIA-designated models." In: Encyclopedia of Earth. Eds. Cutler J. Cleveland (Washington, D.C.: Environmental Information Coalition, National Council for Science and the Environment). [First published in the Encyclopedia of Earth October 23, 2007; Last revised October 24, 2007; Retrieved April 25, 2008]. <http://www.eoearth.org/article/Arctic_climate_change_scenarios_for_the_21st_century_projected_by_the_ACIA-designated_models>

Jazbec, J., Ečimović, P., Jereb, B.(2004). Second neoplasms after treatment of childhood cancer in Slovenia**.** *Pediatric Blood and Cancer, **42***(7),574-581.

Jensen, J. R. (2005). *Introductory Digital Image Processing: A Remote Sensing Perspective*. NJ: Prentice Hall.

Jia, G.J., Epstein, H.E., & Walker, D. A. (2002). Spatial characteristics of AVHRR-NDVI along latitudinal transects in northern Alaska. *Journal of Vegetation Science, 13,* 315-326.

Jones, K.B., Neale, A.C., Nash, M.S., Van Remortel, R.D., Wickham, J.D., et al. (2001). Predicting nutrient and sediment loadings to streams from landscape metrics: a multiple watershed study from the United States Mid-Atlantic Region. *Landscape Ecology, 16*, 301–312.

Johnson, L. B. (1990). Analyzing spatial and temporal phenomena using geographic information systems: a review of ecological applications. *Landscape Ecology, 4*, 31-43.

Johnson, L. B., & Gauge, S. H. (1997). Landscape approaches to the analysis of aquatic ecosytems. *Freshwater Biology, 37*, 113-132.

Johnson, L. B., Richards, C., Host, G. E., & Arthur, J. W. (1997). Landscape influence on water chemistry in Midwestern stream ecosystems. *Freshwater Biology, 37*(1), 193–218.

Keller, K., J. Blum, & Kling, G. W. (2007). Geochemistry of soils and streams on surfaces of varying ages in arctic Alaska. *Arctic, Antarctic, & Alpine Research, 39,*84-98.

King R.S., Baker, M.E., Whigham, D.F., Weller, D.E., Jordan, T.E., Kazyak, P.F., & Hurd, M.K. (2005). Spatial considerations for linking watershed land cover to ecological indicators in streams. *Ecological Applications, 15*,137–153.

Kourous, G. (2005). Deforestation continues at an alarming rate Retrieved on Nov 4, 2009, from http://www.fao.org/newsroom/en/news/2005/1000127/index.html

Kratz, T.K., Webster, K.E., Bowser, C.J., Magnuson, J. J., & Benson, B.J. (1997). The influence of landscape position on lakes in northern Wisconsin. *Freshwater Biology*, *37*, 209 – 217.

Kritzberg, E.S., Cole, J.J., Pace, M.L., Graneli, W., & Bade, D.L. (2004). Autochthonous versus allochthonous carbon sources of bacteria: Results from whole-lake 13C addition experiments. *Limonological Oceanography, 49*, 588-596.

Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Lennon, J.T., Faiia, A.M., Feng, X.H. & Cottingham, K.L. (2006) Relative importance of CO_2 recycling and CH_4 pathways in lake food webs along a dissolved organic carbon gradient. *Limnology and Oceanography*, *51*,1602–1613.

Liu, Z., Hallberg, G.R., & Malanson, G.P. (1997). Structural equation modeling of dynamics of nitrate contamination in ground water. *Journal of The American Water Resources Association, 33*(6), 1219-125.

Lombardzzi, D. (n.d.) The effects of Deforestation on Nutrient Concentrations in Tributaries of Lake Tanganyika. Retrieved on June 16, 2008, http://www.geo.arizona.edu/nyanza/pdf/Lombardozzi.pdf

Loveland, T.R., Sohl, T., Sayler, K., Gallant, A., Dwyer, J., Vogelmann, J., Zylstra, G., Wade, T., Edmonds, C., Chaloud, D. and Jones, K.B. (1999). *U.S. land use and land cover trends: rates, causes, and consequence of late-twentieth century change, EPA Research Plan*. Retrieved on June 10, 2009, from http://www.epa.gov/nerlesd1/land-sci/pdf/trend.pdf.

Markon, C. J., & Derksen, D.V. (1994). Identification of tundra land cover near Teshekpuk Lake, Alaska using SPOT satellite data. *Arctic, 47*(3), 222–231.

Markon, C.J., Fleming, M.J., & Binnian, E.F. (1995). Characteristics of vegetation phenology over the Alaskan landscape using AVHRR time-series data. *Polar Record, **31***(177), 179–190.

Marion, G.M., Hastings, S.J., Oberbauer, S.F., Oechel, W.C. (1989). Soil-plant element relationships in a tundra ecosystem. *Holarctic Ecology*, *12*, 296–303.

Matter, S.F. (1996). Interpatch movement of the red milkweed beetle, *Tetraopes tetraophthalmus*: individual responses to patch size and isolation. *Oecologia, 105*, 447–453.

McColl, J.G., & Grigal, D.F. (1977). Nutrient changes following a forest wildfire in Minnesota: effects in watersheds with differing soils. *Oikos, 28*, 105–112.

McMichael, C. E., A. S. Hope, D. A. Stow, J. Fleming, G. Vourlitis, & Oechel, W. (1999). Estimating CO2 exchange at two sites in arctic tundra ecosystems during the growing season using a spectral vegetation index. *International Journal of Remote Sensing,20*,683–698.

Muller, S.V., Walker, D.A., & Nelson, F.E., et al. (1998). Accuracy assessment of a land-cover map of the Kuparuk River basin, Alaska: considerations for remote regions. *Photogrammetric Engineering and Remote Sensing, 64,* 619-628.

Neff, J.C., & Hooper, D.U. (2002). Vegetation and climate controls on potential CO_2, DOC and DON production in northern latitude soils, *Global Change Biology,* ***9,*** 872–884.

Noyle, B. M. (1999). *The Vegetation and a Landsat Assisted Vegetation Map of the Barrow Peninsula, Northern Alaska*. Master of Science Thesis, Michigan State University, 183.

Oechel, W.C. (1989). Nutrient and Water Flux in a small Arctic Watershed: An Overview. *Holarctic Ecology, 12*(3), 229 – 237.

Oldfield, F. (1977). Lakes and their drainage basins as units of sediment-based ecological study, *Progress in Physical Geography,1*, 460-540.

Pace M. L., Cole, J. J., Carpenter, S. R., & Kitchell, J. F. (1999). Trophic cascades revealed in diverse ecosystems. *TREE, 14,* 483-488.

Pace, M.L., Cole, J.J., Carpenter, S.R., Kitchell, J.F., Hodgson, J.R., Van de Bogert, M.C., Bade, D.L., Kritzberg, E.S., & Bastviken, D. (2004). Whole-lake carbon-13 additions reveal terrestrial support of aquatic food webs. *Nature, 427*, 240-243.

Pesch, R., Schmidt, G., Schroeder, W., & Weustermann, I. (Article in Press). Application of CART in ecological landscape mapping: Two case studies, *Ecological Indicators*.

Petts, G. , Maddock, J. , Bickerton M. & Ferguson , A.J.D. (1995). Linking hydrology and ecology: the scientific basis for river management. In Harper, D.M., & Ferguson, A.J.D. (Ed.),*The Ecological Basis for River Management* (pp. 1-16) , Chichester:Wiley.

Poon, T.C., Chan, A.T., Zee, B., Ho, S.K., Mok, T.S., Leung, T.W., Johnson, P.J.(2001). Application of classification tree and neural network algorithms to the identification of serological liver marker profiles for the diagnosis of hepatocellular carcinoma. *Oncology, 61*(4),275-283.

Prowse, T.D., Wrona, F.J., Reist, J.D., Hobbie, J.E., Levesque, L., & Vincent, W.F. (2006). General Features of the Arctic Relevant to Climate Change in Freshwater Ecosystems. *Ambio, 35*(7), 330 – 338.

Puckett, L.J. (1995). Identifying the major sources of nutrient water pollution. *Environmental Science and Technology, 29*(9), 408-414.

Riera, J. L., Magnuson, J. J., Kratz, T. K., & Webster, K. E. (2000). A geomorphic template for the analysis of lake districts applied to the Northern Highland Lake District, Wisconsin, USA. Freshwater Biology, 43, 301–318.

Richards, D., Johnson, L. B., & Host, G. (1996). Landscape-scale influenceson stream habitats and biota. *Canadian Journal of Fisheries and Aquatic Science*, *53*(Suppl. 1), 295–311.

Rogerson, P. (2006). *Statistical methods for geography: a student guide*. Second Edition (pp. Thousand Oaks, CA: Sage Publications Inc. pp. 170.

Schimel, J.P., & Clein, J.S. (1996). Microbial response to freeze-thaw cycles in tundra and taiga soils. *Soil Biology & Biochemistry, 28*, 1061-1066.

Schindler, D. W. (1971). A hypothesis to explain differences and similaritiesa mongl akes in the Experimental Lakes Area, northwestern Ontario. Proc. Fish. Res. Bd Canada 28: 295-301.

Schindler, D.E., & Scheuerell, M.D. (2002). Habitat coupling in lake ecosystems. *Oikos, 98*, 177-189.

Schindler, D. W., & Smol. J. P. (2006). Cumulative effects of climate warming and other human activities on freshwaters of Arctic and subarctic North America. *Ambio, 35*,160–168.

Serreze, M.C., Walsh, J. E., Chapin, F.S., Osterkamp, T., Dyurgerov, M., Romanvosky, V., Oechel, W. C., Morison, J., Zhang, T., & Barry, R.G. (2000). Observation evidence of rencent change in the northern high-latitude environment. *Climatic Change, 46,* 159 – 207.

Smol, J. P., Wolfe, A. P., Birks, H. J. B., Douglas, M. S. V., Jones, V. J., Korhola, A., Pienitz, R., Rühland, K. , Sorvari, S., Antoniades, A. , Brooks, S. J., Fallu, M. A. , Hughes, M., Keatley, B., Laing, T. , Michelutti, N., Nazarova, L., Nyman, M., Paterson, A. M., Perren, B. B., Quinlan, R., Rautio, M., Saulnier-Talbot, E., Siitonen, S., Solovieva, N., & Weckström, J. (2005). Climate-driven regime shifts in the biological communities of arctic lakes. *Proceedings of the National Academy of Sciences,* 102, 4397–4402.

Soranno, P.A., Hubler, S.L., Carpenter, S. R. , & Lathrop, R. C. (1996). Phosphorus Loads to Surface Waters: A Simple Model to Account for Spatial Pattern of Land Use. *Ecological Applications*, *6* (3), 865-878.

Soranno, P. A., Webster, K.E., Riera, J.L., Kratz, T.K., Baron, J.S., Bukavechas, P.A., Kling, G.W., White, D.S., Caine, N., Lathrop, R.C, & Leavitt, P.R. (1999). Spatial variation among lakes within landscapes: Ecological organization along lake chains. *Ecosystems, 2*, 395-410.

Sørensen, R., Zinko, U., & Seibert, J. (2005). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrology and Earth System Sciences Discussions, 2, underopen review of Hydrology and Earth System Sciences, 1807–1834.

Southgate, D., & Whitaker, M. (1992). Promoting resource degradation in Latin America: tropical deforestation, soil erosion, and coastal ecosystem disturbance in Ecuador. *Economical Development and Cultural Change, 40*, 787–807.

Spruill, T.B., Showers, W.J., & Howe S.S. (2002) Application of Classification Tree Method to identify Nitrate Sources in Ground Water. *Journal of Environmental Quality,* 31, 1538-1549.

SPSS Statistics Base 17.0 User's Guide

Stendera, S. & Johnson, R.K. (2006). Multiscale Drivers of Water Chemistry of Boreal Lakes and Streams. *Environmental Management, 38*, 760-770

Stewart, J.S., Wang, L.Z., Lyons, J., Horwatich, J.A., Bannerman, R. (2001). Influences of watershed, riparian-corridor, and reach-scale characteristics on aquatic biota in agricultural watersheds. *Journal of American Water Resources Association, 37*,1475–1487.

Stine R., Chaudhuri, D., Ray, P., Pathak, P., Hall-Brown, M. (2010) Comparing Image Classification Techniques for Arctic Tundra Land Cover, Toolik Lake, Alaska. GIScience and Remote Sensing.

Stow, D., Burns, B., & Hope, A. (1989). Mapping arctic tundra vegetation types using digital SPOT/HRV-XS data: a preliminary assessment. *International Journal of Remote Sensing, 10,* 1451-1457.

Stow, D. A., Burns, B. H., & Hope, A. S. (1993). Spectral, spatial and temporal characteristics of arctic tundra reflectance. *International Journal of Remote Sensing, 14,* 2445–2462.

Stow, D., Hope, A., Boynton, W., & Daeschner, S. (2000). Arctic tundra functional types by classification of single-date and AVHRR bi-weekly NDVI composite datasets. *International Journal of Remote Sensing, 21,* 1773–1779.

Stow, D., Hope, A., McGuire D., Gamon, J., Huemmrich, F., Houston, S., Racine, C., Sturm, M., Tape, K., Hinzman, L., Yoshikawa, K., Tweeide, C., Noyle, B., Silapaswan, C., Douglas, D., Griffith, B., Jia, G., Epstein, H., Walker, D., Daeschner, S., Petersen, A., Zhou L., & Myneni, R. (2004). Remote Sensing of vegetation and Land-Cover Change in Arctic Tundra Ecosystems. *Remote Sensing of Environment, 89*, 281-308.

Strahler, A. H., Logan, T. L., & Bryant, N. A. (1978). Improving forest cover classification accuracy from Landsat by incorporating topographic information. *Proceedings of the Twelfth International Symposium on Remote Sensing of Environment*, 927- 942. Environmental Research Institute of Michigan.

Sturm M., McFadden, J.P., Liston, G.E., Chapin, F.S., Racine, C.H., Holmgren, J. (2001) Snow-shrub interactions in Arctic tundra: a hypothesis with climatic implications. *Journal of Climatology, 14*, 336–344.

Tape, K., Sturm, M., & Racine, C. (2006). The evidence for shrub expansion in Northern Alaska and the Pan-Arctic. *Global Change Biology, 12*, 686–702.

Turner, M.G. (1989). Landscape Ecology: The Effect of Pattern on Process. *Annual Review of Ecological Systems, 20*, 171–197.

U.S. Department of Agriculture, Natural Resources Conservation Service. (2000). *1997 National Resources Inventory: Highlights.*
Retrived on Jan 13, 2008, from
http://www.nhq.nrcs.usda.gov/land/pubs/97highlights.html

US Environmental Protection Agency. (2000). Water quality conditions in the United States: a profile from the 1998 National water quality inventory report to congress. EPA 841/F-00/006

VanCleve, K., Yarie, J., Erickson, R,. Dyrness, C.T. (1993). Nitrogen mineralization and nitrification in successional ecosystems on the Tanana River floodplain, Interior Alaska. *Canadian Journal of Research, 23*, 970–978.

Vanni, (2005). Linking landscapes and food webs: effects of omnivorous fish and watersheds on reservoir ecocyctems. *BioScience, 55,* 155-167.

Vatsavai, R.R., & Shekhar, S. (2002). An Efficient Hybrid Classification System for Mining Multi-spectral Remote Sensing Imagery Guided by Spatial Database. etreived on January 13, 2007, from http://www-users.cs.umn.edu/~vatsavai/papers/prrs02f-vatsavai.ps.

Veith T.L., Srinivasan, M.S., & Gburek, W.J. (2003). Process representation in watershed-scale hydrologic models: an evaluation in an experimental watershed. In Paper Read at First Interagency Conference on Research in the Watersheds, October 27–30 2003, Benson, AZ.

Vourlitis, G.L., Oechel, W.C., Hope, A., Stow, D., Boynton, B., Verfaillie, J., Zulueta, R. & Hastings, S.J. (2000). Physiological models for scaling plot measurements of CO2 flux across an arctic tundra landscape. *Ecological Applications, 10,* 60–72.

Walker, D.A. (1977). Analysis of the effectiveness of a television scanning densitometer for indicating geobotanical features in an ice-wedge polygon complex at Barrow, Alaska, 129 pp. M.A. Directed by: P.J. Webber.

Walker, M. D., Walker, D. A., & Auerbach, N. A. (1994). Plant communities of a tussock tundra landscape in the Brooks Range Foothills, Alaska . *Journal of Vegetation Science, 5*, 843–866.

Webster, K.E., Kratz, T.K., Bowser, C.J., Magnuson, J.J., & Rose, W.J. (1996). The Influence of Landscape Position on Lake Chemical Responses to Drought in Northern Wisconsin. *Limnology and Oceanography*, *41*(5), 977 – 984.

Welch, H.E., Legault, J.A., & Mergmann, M.A. (1987). Effects of snow and ice on the annual cycles of heat and light in Saqvaqjuac lakes. *Canadian Journal of Fisheries and Aquatic Sciences. 44*, 1451-1461.

Weller,G ., Chapin, F. S. HII, Everett, K. R., Hobbie, J. E., Kane, D., Oechel, W. C., Ping, C.-L., Reeburgh, W. S., Walker, D., & Walsh, J. (1995). The Arctic Flux Study: a regional view of trace gas release. *Journal of Biogeography, 22*, 365-374.

Wetzel, R.G. (2001). Limnology: Lake and River Ecosystems. Third Edition. Academic Press.

Willson, M. F., Gende, S. M. & Marson, B. H. (1998). Fishes and the forest. *Bioscience, 48*, 455–462.

Whalen, S.C., & Alexander, V. (1986). Chemical influences on 15N and 14C primary production in an arctic lake. *Polar Biology, 5,* 211-219.

Whalen S.C., & Alexander, V. (1986). Seasonal inorganic carbon and nitrogen transport in an arctic lake. *Canadian Journal of Fisheries and Aquatic Sciences*, *43*, 1177 - 1186.

Whalen, S.C., Chalfant, B.A., Fischer, E.N., Fortino, K.A., & Hershey, A.E. (2006). Comparative influence of resuspended glacial sediment on physicochemical characteristics and primary production in two arctic lakes. *Aquatic Science, 68*(1), 65-77.

Wolock, D.M., (1993). Simulating the variable-source-area concept of streamflow generation with the watershed model TOPMODEL. U.S.G.S. Water-resource investigation report 93-4124, Lawrence, KS.

Wolter, P.T., Mladenoff, D.J., Host, G.E., &Crow, T.R., 1995, Improved forest classification in the northern lake states using multi-temporal Landsat imagery. *PhotogrammetricEngineering & Remote Sensing, 61*, 1129-1143.

Woo, M.K., Heron, R., & Steet, P. 1981. Catchment Hydrology of a High Arctic Lake. *Cold Region Science and Technology, 5*, 29-41.

Woo, M. K., & Xia, Z. J. (1995). Suprapermafrost groundwater seepage in gravely terrain, Resolute, N.W.T., Canada. *Permafrost and Periglacial Processes, 6,*57–72.

Zhu, J. Z. & Mazumder A. (2008). Estimating nitrogen exports in response to forest vegetation, age and soil types in two coastal-forested watersheds in British Columbia. *Forest ecology and management, 255*, 1945 – 1959.

APPENDIX A: ACRONYMS FOR LANDSCAPE FACTORS

Landscape Metrics	Vegetation Community	Acronym
	Aquatic Vegetation Complex	Prct_AV
Percentage	Moist Acidic Tundra Complex	Prct_MAT
	Moist Non-Acidic Tundra Complex	Prct_MNT
	Shrub Complex	Prct_Shrub
	Fen complex	Prct_Fen
	Heath Complex	Prct_Heath
	Mountain Meadow complex	Prct_MM
	Riparian Complex	Prct_Rip
	Snowbed Complex	Prct_SB
	Barren/Thermokarst	Prct_BT
Patch Density	Aquatic Vegetation Complex	PD_AV
	Moist Acidic Tundra Complex	PD_MAT
	Moist Non-Acidic Tundra Complex	PD_MNT
	Shrub Complex	PD_Shrub
	Fen complex	PD_Fen
	Heath Complex	PD_Heath
	Mountain Meadow complex	PD_MM
	Riparian Complex	PD_Rip
	Snowbed Complex	PD_SB
	Barren/Thermokarst	PD_BT
	Aquatic Vegetation Complex	MS_AV
Mean Shape Index (Percent change)	Moist Acidic Tundra Complex	MS_MAT
	Moist Non-Acidic	MS_MNT

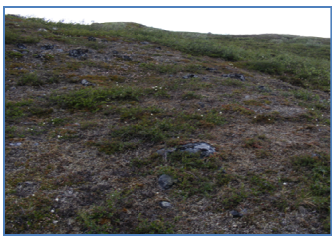
	Tundra Complex	
	Shrub Complex	MS_Shrub
	Fen complex	MS_Fen
	Heath Complex	MS_Heath
	Mountain Meadow complex	MS_MM
	Riparian Complex	MS_Rip
	Snowbed Complex	MS_SB
	Barren/Thermokarst	MS_BT
Landscape Shape Index	Aquatic Vegetation Complex	LSI_AV
	Moist Acidic Tundra Complex	LSI_MAT
	Moist Non-Acidic Tundra Complex	LSI_MNT
	Shrub Complex	LSI_Shrub
	Fen complex	LSI_Fen
	Heath Complex	LSI_Heath
	Mountain Meadow complex	LSI_MM
	Riparian Complex	LSI_Rip
	Snowbed Complex	LSI_SB
	Barren/Thermokarst	LSI_BT
Fractal Dimension (Converted into Percent Change)	Aquatic Vegetation Complex	FD_AV
	Moist Acidic Tundra Complex	FD_MAT
	Moist Non-Acidic Tundra Complex	FD_MNT
	Shrub Complex	FD_Shrub
	Fen complex	FD_Fen
	Heath Complex	FD_Heath
	Mountain Meadow complex	FD_MM
	Riparian Complex	FD_Rip
	Snowbed Complex	FD_SB
	Barren/Thermokarst	FD_BT
	Aquatic Vegetation Complex	MP_AV
	Moist Acidic Tundra Complex	MP_MAT

Mean Patch Size	Moist Non-Acidic Tundra Complex	MP_MNT
	Shrub Complex	MP_Shrub
	Fen complex	MP_Fen
	Heath Complex	MP_Heath
	Mountain Meadow complex	MP_MM
	Riparian Complex	MP_Rip
	Snowbed Complex	MP_SB
	Barren/Thermokarst	MP_BT
	Euphotic Zone Depth	EZD
	Lake Order	Lake Order
	Maximum Depth	Max_Depth
	Shoreline Development Factor	SDF

Note: When these indices were calculated at buffer level, a suffix of "_B_" was added in their acronyms.

APPENDIX B: FIELD PHOTOS OF THE VEGETATION COMMUNITIES

Barren Complex



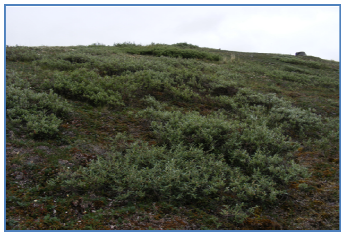
Heath Complex

Aquatic Vegetation Complex

MAT Complex

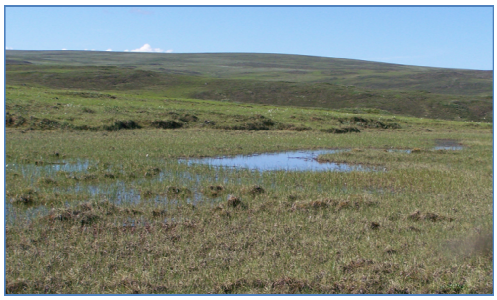
MNT Complex

Riparian Complex



Shrub Complex

Snowbed Complex

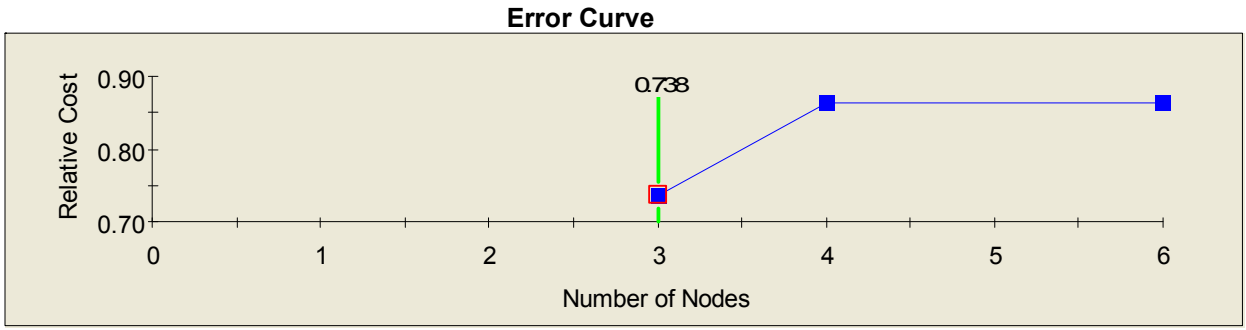


Fen Complex

Mountain Meadows

APPENDIX C: CLASSIFICATION TREE DETAILS

Chlorophyll a Areal Estimates



Navigator 1 (3): Tree Summary Reports: Gains Data for 1

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
1	14	100.00	70.00	70.00	35.00	35.00	14	2.00	2.00
2	4	36.36	20.00	90.00	62.50	27.50	11	1.44	0.73
3	2	13.33	10.00	100.00	100.00	37.50	15	1.00	0.27

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=14	2 N=15	3 N=11
1	20	70.00	14	2	4
2	16	81.25	0	13	3
3	4	100.00	0	0	4
Total:	40.00				
Average:		83.75			
Overall % Correct:		77.50			

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=16	2 N=10	3 N=14
1	20	65.00	13	2	5
2	16	37.50	3	6	7
3	4	50.00	0	2	2
Total:	40.00				
Average:		50.83			
Overall % Correct:		52.50			

Misclassification for Learn Data

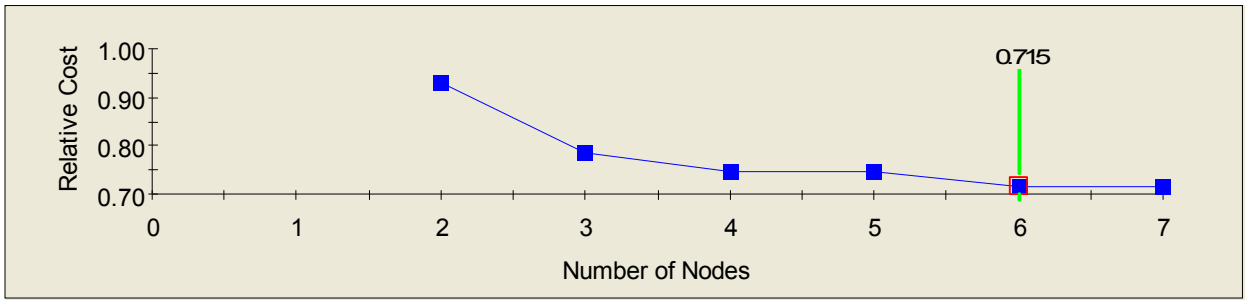
Class	N Cases	N Mis-Classed	Pct Error	Cost
1	20	6	30.00	0.30
2	16	3	18.75	0.19
3	4	0	0.00	0.00

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	20	7	35.00	0.35
2	16	10	62.50	0.63
3	4	2	50.00	0.50

Chlorophyll a Volumetric estimates

Error Curve



Navigator 2 (6): Tree Summary Reports: Gains Data for 1

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
2	5	100.00	100.00	100.00	12.50	12.50	5	8.00	8.00
3	0	0.00	0.00	100.00	42.50	30.00	12	2.35	0.00
5	0	0.00	0.00	100.00	60.00	17.50	7	1.67	0.00
1	0	0.00	0.00	100.00	75.00	15.00	6	1.33	0.00
6	0	0.00	0.00	100.00	90.00	15.00	6	1.11	0.00
4	0	0.00	0.00	100.00	100.00	10.00	4	1.00	0.00

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=5	2 N=11	3 N=18	4 N=6
1	5	100.00	5	0	0	0
2	17	64.71	0	11	6	0
3	11	90.91	0	0	10	1
4	7	71.43	0	0	2	5
Total:	40.00					
Average:		81.76				
Overall % Correct:		77.50				

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=10	2 N=9	3 N=13	4 N=8
1	5	60.00	3	1	1	0
2	17	41.18	2	7	6	2
3	11	27.27	5	1	3	2
4	7	57.14	0	0	3	4
Total:	40.00					
Average:		46.40				
Overall % Correct:		42.50				

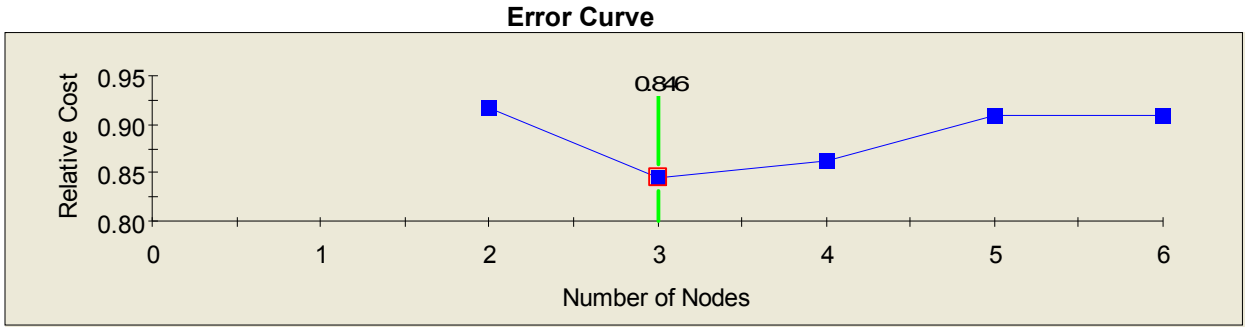
Misclassification for Learn Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	5	0	0.00	0.00
2	17	6	35.29	0.35
3	11	1	9.09	0.09
4	7	2	28.57	0.29

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	5	2	40.00	0.40
2	17	10	58.82	0.59
3	11	8	72.73	0.73
4	7	3	42.86	0.43

Total Nitrogen estimates



Navigator 1 (3): Tree Summary Reports: Gains Data for 1

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
1	12	75.00	75.00	75.00	41.03	41.03	16	1.83	1.83
2	2	18.18	12.50	87.50	69.23	28.21	11	1.26	0.44
3	2	16.67	12.50	100.00	100.00	30.77	12	1.00	0.41

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=16	2 N=11	3 N=12
1	16	75.00	12	2	2
2	12	75.00	0	9	3
3	11	63.64	4	0	7
Total:	39.00				
Average:		71.21			
Overall % Correct:		71.79			

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=14	2 N=11	3 N=14
1	16	43.75	7	5	4
2	12	41.67	2	5	5
3	11	45.45	5	1	5
Total:	39.00				
Average:		43.62			
Overall % Correct:		43.59			

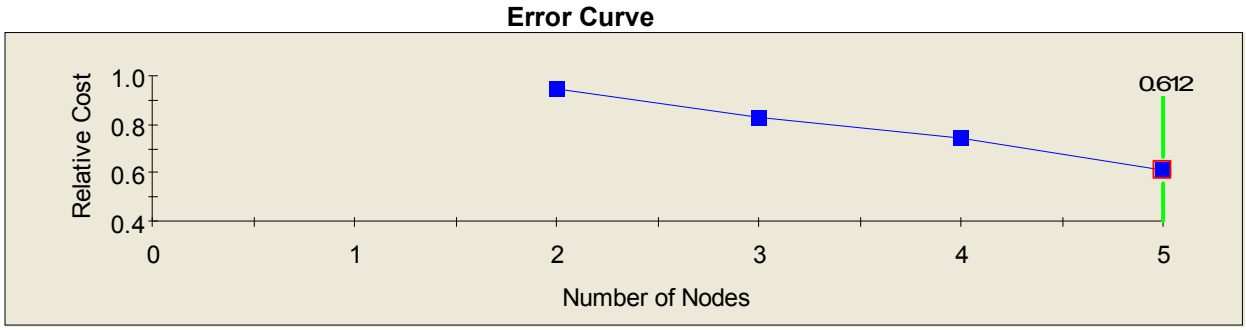
Misclassification for Learn Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	16	4	25.00	0.25
2	12	3	25.00	0.25
3	11	4	36.36	0.36

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	16	9	56.25	0.56
2	12	7	58.33	0.58
3	11	6	54.55	0.55

Total Phosphorus estimates



Navigator 7 (5): Tree Summary Reports: Gains Data for 2

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
4	17	100.00	73.91	73.91	44.74	44.74	17	1.65	1.65
1	5	71.43	21.74	95.65	63.16	18.42	7	1.51	1.18
5	1	33.33	4.35	100.00	71.05	7.89	3	1.41	0.55
2	0	0.00	0.00	100.00	94.74	23.68	9	1.06	0.00
3	0	0.00	0.00	100.00	100.00	5.26	2	1.00	0.00

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=14	2 N=24
1	15	86.67	13	2
2	23	95.65	1	22
Total:	38.00			
Average:		91.16		
Overall % Correct:		92.11		

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=24	2 N=14
1	15	86.67	13	2
2	23	52.17	11	12
Total:	38.00			
Average:		69.42		
Overall % Correct:		65.79		

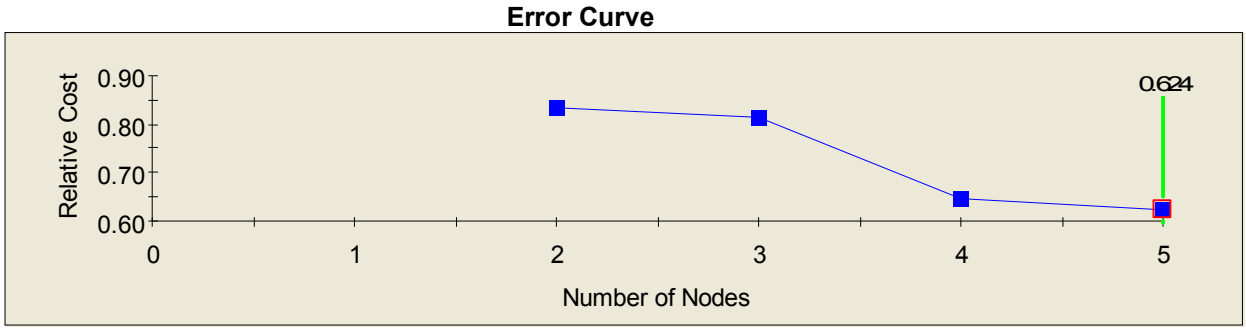
Misclassification for Learn Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	15	2	13.33	0.13
2	23	1	4.35	0.04

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	15	2	13.33	0.13
2	23	11	47.83	0.48

Conductivity estimates



Navigator 5 (5): Tree Summary Reports: Gains Data for 1

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
5	12	100.00	85.71	85.71	30.00	30.00	12	2.86	2.86
1	2	100.00	14.29	100.00	35.00	5.00	2	2.86	2.86
2	0	0.00	0.00	100.00	67.50	32.50	13	1.48	0.00
4	0	0.00	0.00	100.00	85.00	17.50	7	1.18	0.00
3	0	0.00	0.00	100.00	100.00	15.00	6	1.00	0.00

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=14	2 N=13	3 N=7	4 N=6
1	14	100.00	14	0	0	0
2	16	81.25	0	13	2	1
3	6	83.33	0	0	5	1
4	4	100.00	0	0	0	4
Total:	40.00					
Average:		91.15				
Overall % Correct:		90.00				

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=14	2 N=15	3 N=6	4 N=5
1	14	85.71	12	1	1	0
2	16	68.75	2	11	1	2
3	6	33.33	0	2	2	2
4	4	25.00	0	1	2	1
Total:	40.00					
Average:		53.20				
Overall % Correct:		65.00				

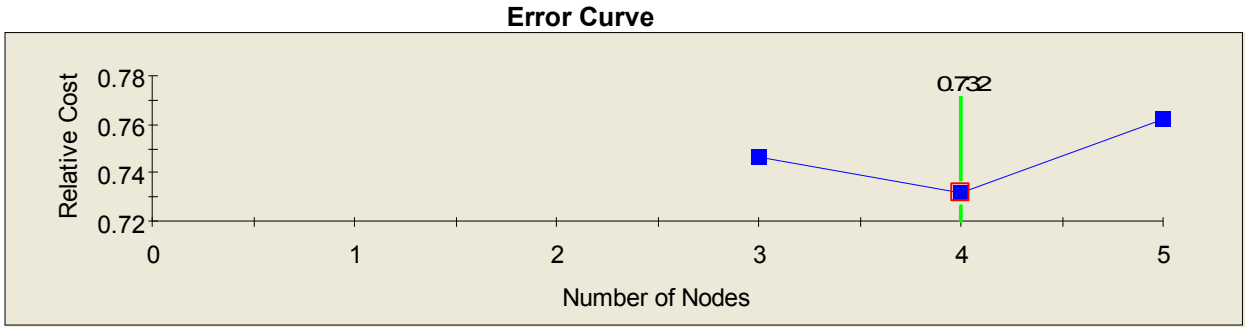
Misclassification for Learn Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	14	0	0.00	0.00
2	16	3	18.75	0.19
3	6	1	16.67	0.17
4	4	0	0.00	0.00

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	14	2	14.29	0.14
2	16	5	31.25	0.31
3	6	4	66.67	0.67
4	4	3	75.00	0.75

Calcium estimates



Navigator 3 (4): Tree Summary Reports: Gains Data for 1

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
1	18	100.00	81.82	81.82	45.00	45.00	18	1.82	1.82
2	3	27.27	13.64	95.45	72.50	27.50	11	1.32	0.50
4	1	25.00	4.55	100.00	82.50	10.00	4	1.21	0.45
3	0	0.00	0.00	100.00	100.00	17.50	7	1.00	0.00

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=18	2 N=7	3 N=11	4 N=4
1	22	81.82	18	0	3	1
2	7	85.71	0	6	1	0
3	7	85.71	0	1	6	0
4	4	75.00	0	0	1	3
Total:	40.00					
Average:		82.06				
Overall % Correct:		82.50				

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=16	2 N=6	3 N=8	4 N=10
1	22	59.09	13	2	2	5
2	7	42.86	1	3	3	0
3	7	28.57	2	0	2	3
4	4	50.00	0	1	1	2
Total:	40.00					
Average:		45.13				
Overall % Correct:		50.00				

Misclassification for Learn Data

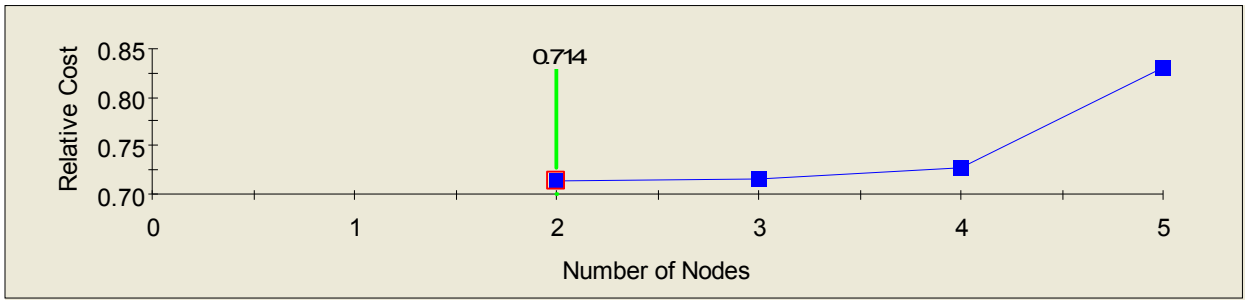
Class	N Cases	N Mis-Classed	Pct Error	Cost
1	22	4	18.18	0.18
2	7	1	14.29	0.14
3	7	1	14.29	0.14
4	4	1	25.00	0.25

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	22	9	40.91	0.41
2	7	4	57.14	0.57
3	7	5	71.43	0.71
4	4	2	50.00	0.50

Magnesium estimates

Error Curve



Navigator 8 (2): Tree Summary Reports: Gains Data for 1

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
2	16	88.89	100.00	100.00	45.00	45.00	18	2.22	2.22
1	0	0.00	0.00	100.00	100.00	55.00	22	1.00	0.00

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=18	2 N=0	3 N=22	4 N=0
1	16	100.00	16	0	0	0
2	12	0.00	2	0	10	0
3	7	100.00	0	0	7	0
4	5	0.00	0	0	5	0
Total:	40.00					
Average:		50.00				
Overall % Correct:		57.50				

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=19	2 N=0	3 N=20	4 N=1
1	16	100.00	16	0	0	0
2	12	0.00	3	0	9	0
3	7	85.71	0	0	6	1
4	5	0.00	0	0	5	0
Total:	40.00					
Average:		46.43				
Overall % Correct:		55.00				

Misclassification for Learn Data

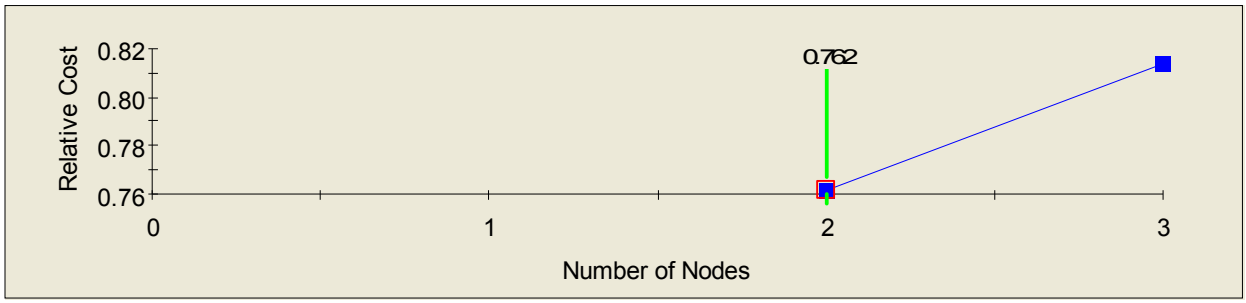
Class	N Cases	N Mis-Classed	Pct Error	Cost
1	16	0	0.00	0.00
2	12	12	100.00	1.00
3	7	0	0.00	0.00
4	5	5	100.00	1.00

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	16	0	0.00	0.00
2	12	12	100.00	1.00
3	7	1	14.29	0.14
4	5	5	100.00	1.00

Sodium estimates

Error Curve



Navigator 3 (2): Tree Summary Reports: Gains Data for 2

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
1	7	43.75	100.00	100.00	40.00	40.00	16	2.50	2.50
2	0	0.00	0.00	100.00	100.00	60.00	24	1.00	0.00

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=24	2 N=16
1	33	72.73	24	9
2	7	100.00	0	7
Total:	40.00			
Average:		86.36		
Overall % Correct:		77.50		

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=25	2 N=15
1	33	66.67	22	11
2	7	57.14	3	4
Total:	40.00			
Average:		61.90		
Overall % Correct:		65.00		

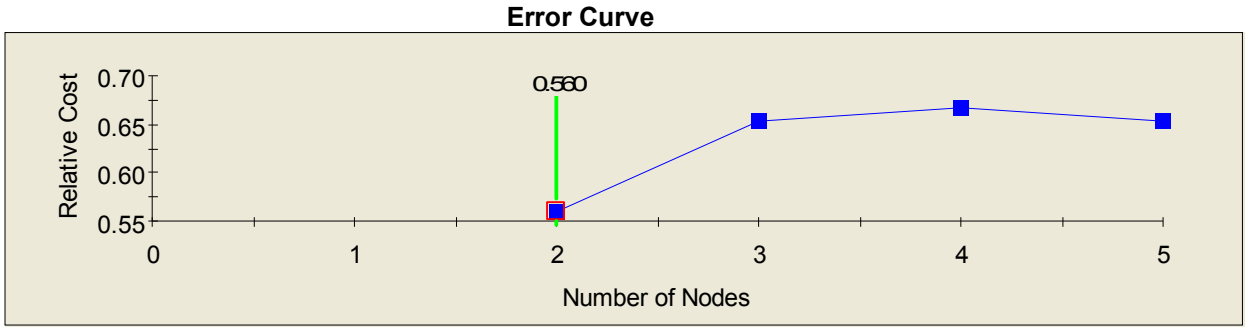
Misclassification for Learn Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	33	9	27.27	0.27
2	7	0	0.00	0.00

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	33	11	33.33	0.33
2	7	3	42.86	0.43

Potassium estimates



Navigator 6 (2): Tree Summary Reports: Gains Data for 2

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
1	14	73.68	93.33	93.33	47.50	47.50	19	1.96	1.96
2	1	4.76	6.67	100.00	100.00	52.50	21	1.00	0.13

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=21	2 N=19
1	25	80.00	20	5
2	15	93.33	1	14
Total:	40.00			
Average:		86.67		
Overall % Correct:		85.00		

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=19	2 N=21
1	25	64.00	16	9
2	15	80.00	3	12
Total:	40.00			
Average:		72.00		
Overall % Correct:		70.00		

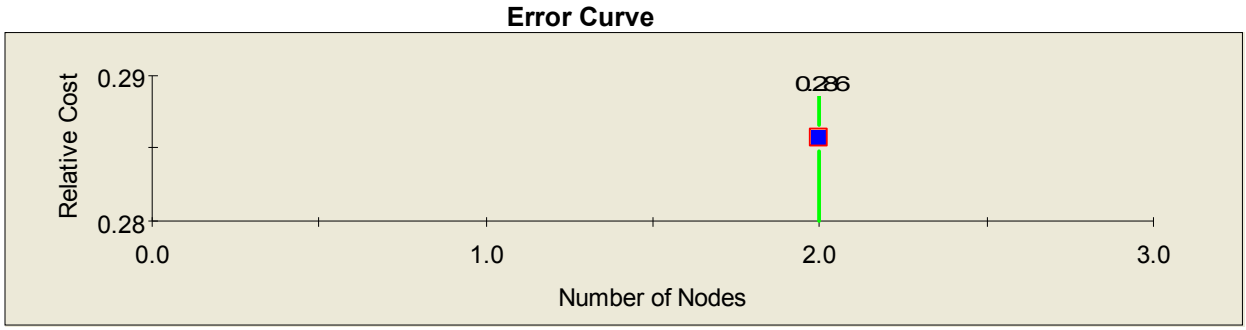
Misclassification for Learn Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	25	5	20.00	0.20
2	15	1	6.67	0.07

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	25	9	36.00	0.36
2	15	3	20.00	0.20

Sulfate estimates



Navigator 10 (2): Tree Summary Reports: Gains Data for 2

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
1	5	71.43	100.00	100.00	17.50	17.50	7	5.71	5.71
2	0	0.00	0.00	100.00	100.00	82.50	33	1.00	0.00

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=33	2 N=7
1	35	94.29	33	2
2	5	100.00	0	5
Total:	40.00			
Average:		97.14		
Overall % Correct:		95.00		

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=33	2 N=7
1	35	91.43	32	3
2	5	80.00	1	4
Total:	40.00			
Average:		85.71		
Overall % Correct:		90.00		

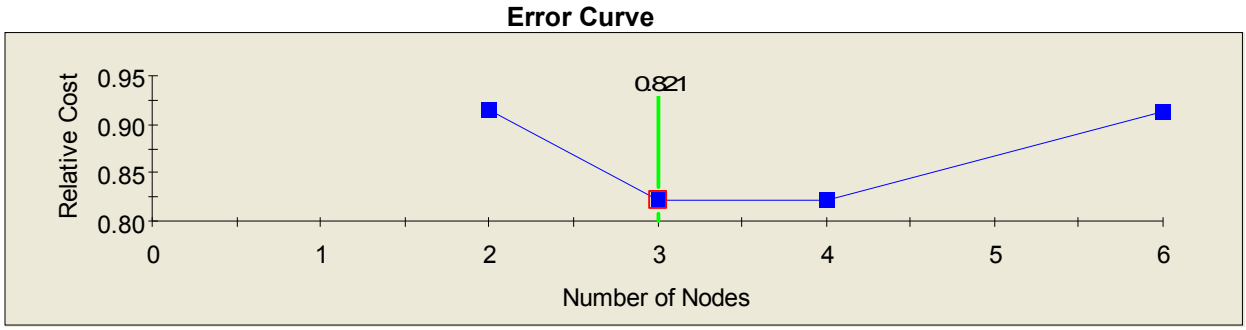
Misclassification for Learn Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	35	2	5.71	0.06
2	5	0	0.00	0.00

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	35	3	8.57	0.09
2	5	1	20.00	0.20

Chlorides estimates



Navigator 5 (3): Tree Summary Reports: Gains Data for 2

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
2	10	83.33	90.91	90.91	30.00	30.00	12	3.03	3.03
3	1	4.55	9.09	100.00	85.00	55.00	22	1.18	0.17
1	0	0.00	0.00	100.00	100.00	15.00	6	1.00	0.00

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=28	2 N=12
1	29	93.10	27	2
2	11	90.91	1	10
Total:	40.00			
Average:		92.01		
Overall % Correct:		92.50		

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=27	2 N=13
1	29	72.41	21	8
2	11	45.45	6	5
Total:	40.00			
Average:		58.93		
Overall % Correct:		65.00		

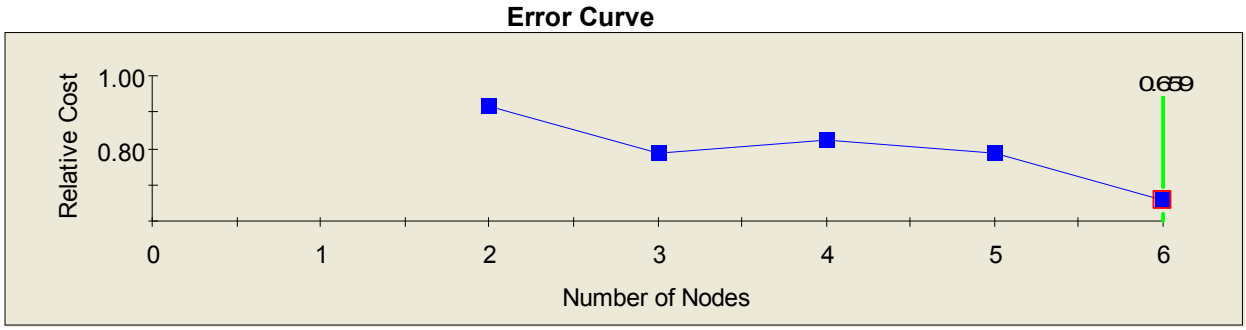
Misclassification for Learn Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	29	2	6.90	0.07
2	11	1	9.09	0.09

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	29	8	27.59	0.28
2	11	6	54.55	0.55

DIC estimates



Navigator 6 (6): Tree Summary Reports: Gains Data for 1

Node	Cases Tgt. Class	% of Node Tgt. Class	% Tgt. Class	Cum % Tgt. Class	Cum % Pop	% Pop	Cases in Node	Cum lift	Lift Pop
5	13	100.00	72.22	72.22	32.50	32.50	13	2.22	2.22
4	3	100.00	16.67	88.89	40.00	7.50	3	2.22	2.22
1	1	33.33	5.56	94.44	47.50	7.50	3	1.99	0.74
2	1	11.11	5.56	100.00	70.00	22.50	9	1.43	0.25
3	0	0.00	0.00	100.00	90.00	20.00	8	1.11	0.00
6	0	0.00	0.00	100.00	100.00	10.00	4	1.00	0.00

Prediction Success--Learn--Count

Actual Class	Total Cases	Percent Correct	1 N=16	2 N=9	3 N=11	4 N=4
1	18	88.89	16	1	1	0
2	10	80.00	0	8	1	1
3	9	100.00	0	0	9	0
4	3	100.00	0	0	0	3
Total:	40.00					
Average:		92.22				
Overall % Correct:		90.00				

Prediction Success--Test--Count

Actual Class	Total Cases	Percent Correct	1 N=16	2 N=7	3 N=15	4 N=2
1	18	72.22	13	1	4	0
2	10	30.00	1	3	5	1
3	9	66.67	1	2	6	0
4	3	33.33	1	1	0	1
Total:	40.00					
Average:		50.56				
Overall % Correct:		57.50				

Misclassification for Learn Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	18	2	11.11	0.11
2	10	2	20.00	0.20
3	9	0	0.00	0.00
4	3	0	0.00	0.00

Misclassification for Test Data

Class	N Cases	N Mis-Classed	Pct Error	Cost
1	18	5	27.78	0.28
2	10	7	70.00	0.70
3	9	3	33.33	0.33
4	3	2	66.67	0.67

APPENDIX D: PYTHON SCRIPT FOR TOPOGRAPHIC WETNESS INDEX

```
#   Created By: Prasad A Pathak
# Purpose: Calculate Topographic Wetness Index: natural logarithm of area
divided by slope
# It indicates the probable water saturation level of the ground
# By default, ArcGIS calculates slope by considering just 3 X 3 neighborhood
# For this index a slope is calculated from the topmost pixel to any particular
pixel is needed
# This is acheived by calculating the minimum elevation raster
# Import system modules
import sys, string, os, arcgisscripting, win32com.client
# Geoprocessor object
gp = arcgisscripting.create(9.3)
gp.overwriteoutput = 1
# Check out license for Spatial Analyst
gp.CheckOutExtension("spatial")
# Load toolboxe
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Spatial
Analyst Tools.tbx")
# Get the name of folder
gp.workspace = gp.GetParameterAsText (0)
#Input DEM
```

```
InDEM = gp.GetParameterAsText(1)

#Output TWI raster

OutTWI = gp.GetParameterAsText(2)

# Variables used

Output_surface_raster = gp.workspace + "/eldodem_fill"

Output_flow_direction_raster = gp.workspace + "/eldodem_flw"

Output_drop_raster = gp.workspace + "/eldodem_drp"

Output_accumulation_raster = gp.workspace + "/eldodem_acc"

Output_Plus_raster = gp.workspace + "/eldodempls"

Output_Times_raster = gp.workspace + "/eldodemtim"

Output_Mean_elev_raster = gp.workspace + "/mean_elev"

Output_edrop = gp.workspace + "/edrop"

Out_change = gp.workspace + "/change_elev"

Out_distance = gp.workspace + "/distance"

Out_slope = gp.workspace + "/slope"

Out_preatan = gp.workspace + "/preatan"

# Fill DEM to make it depressionless

gp.Fill_sa(InDEM, Output_surface_raster, "")

print "DEM filled"

gp.addmessage  ("DEM filled")

# Flow Direction
```

```
gp.FlowDirection_sa(Output_surface_raster, Output_flow_direction_raster,
"NORMAL", Output_drop_raster)
print "Flow direction raster created successfully"
gp.addmessage ("Flow direction raster created successfully")
# Flow Accumulation
gp.FlowAccumulation_sa(Output_flow_direction_raster,
Output_accumulation_raster, "", "FLOAT")
print "Flow accumulation raster created successfully"
gp.addmessage ("Flow accumulation raster created successfully")
# One is added to each pixel to get an count of how many pixel including the
current are contributing the flow
gp.Plus_sa( Output_accumulation_raster, "1", Output_Plus_raster)
print "addition raster created successfully"
gp.addmessage ("addition raster created successfully")
# Contributing Area using the number of pixels
gp.times_sa(Output_Plus_raster, "25", Output_Times_raster)
print "multiplication raster created successfully"
gp.addmessage ("multiplication raster created successfully")
# Process: Block Statistics...
gp.BlockStatistics_sa(Output_surface_raster, Output_Mean_elev_raster,
"Rectangle 3 3 CELL", "MINIMUM", "DATA")
print "MIN elevation raster created successfully"
```

```
gp.addmessage ("MIN elevation raster created successfully")
gp.Minus_sa (Output_surface_raster, Output_Mean_elev_raster, Output_edrop)
print "EDROP raster created successfully"
gp.addmessage ("EDROP raster created successfully")
Input_false_raster_or_constant_value = "0.005"
gp.Con_sa(Output_edrop, Output_edrop, Out_change,
Input_false_raster_or_constant_value, "\"VALUE\" >= 0.005")
print "Change elevation raster created successfully"
gp.addmessage ("Change elevation raster created successfully")
Input_true_raster_or_constant_value = "5"
false_raster_or_constant_value = "7.07"
gp.Con_sa(Output_flow_direction_raster, Input_true_raster_or_constant_value,
Out_distance, false_raster_or_constant_value, "\"VALUE\" = 1 OR \"VALUE\"
= 4 OR \"VALUE\" = 16 OR \"VALUE\" = 64")
print "Distance raster created successfully"
gp.addmessage ("Distance raster created successfully")
gp.Divide_sa (Out_change, Out_distance, Out_slope)
print "Slope raster created successfully"
gp.addmessage ("Slope raster created successfully")

gp.Divide_sa (Output_Times_raster, Out_slope, Out_preatan)
print "Pre-atan raster created successfully"
```

```
gp.addmessage ("Pre-atan raster created successfully")

gp.Ln_sa (Out_preatan, OutTWI)

print "TWI raster created successfully"

gp.addmessage ("TWI raster created successfully")
```