

CHAUDHURI, DEBASISH, Ph.D. Hybrid Image Classification Technique for Land-Cover Mapping in the Arctic Tundra, North Slope, Alaska. (2008)
Directed by Dr.Roy Stine. 182 pp.

Remotely sensed image classification techniques are very useful to understand vegetation patterns and species combination in the vast and mostly inaccessible arctic region. Previous researches that were done for mapping of land-cover and vegetation in the remote areas of northern Alaska have considerably low accuracies compared to other biomes. The unique arctic tundra environment with short growing season length, cloud cover, low sun angles, snow and ice cover hinders the effectiveness of remote sensing studies. The majority of image classification research done in this area as reported in the literature used traditional unsupervised clustering technique with Landsat MSS data. It was also emphasized by previous researchers that SPOT/HRV-XS data lacked the spectral resolution to identify the small arctic tundra vegetation parcels. Thus, there is a motivation and research need to apply a new classification technique to develop an updated, detailed and accurate vegetation map at a higher spatial resolution i.e. SPOT-5 data.

Traditional classification techniques in remotely sensed image interpretation are based on spectral reflectance values with an assumption of the training data being normally distributed. Hence it is difficult to add ancillary data in classification procedures to improve accuracy. The purpose of this dissertation was to develop a hybrid image classification approach that

effectively integrates ancillary information into the classification process and combines ISODATA clustering, rule-based classifier and the Multilayer Perceptron (MLP) classifier which uses artificial neural network (ANN). The main goal was to find out the best possible combination or sequence of classifiers for typically classifying tundra type vegetation that yields higher accuracy than the existing classified vegetation map from SPOT data.

Unsupervised ISODATA clustering and rule-based classification techniques were combined to produce an intermediate classified map which was used as an input to a Multilayer Perceptron (MLP) classifier. The result from the MLP classifier was compared to the previous classified map and for the pixels where there was a disagreement for the class allocations, the class having a higher kappa value was assigned to the pixel in the final classified map. The results were compared to standard classification techniques: simple unsupervised clustering technique and supervised classification with Feature Analyst. The results indicated higher classification accuracy (75.6%, with kappa value of .6840) for the proposed hybrid classification method than the standard classification techniques: unsupervised clustering technique (68.3%, with kappa value of 0.5904) and supervised classification with Feature Analyst (62.44%, with kappa value of 0.5418). The results were statistically significant at 95% confidence level.

Keywords: Arctic tundra, hybrid classification, artificial neural network, kappa analysis.

HYBRID IMAGE CLASSIFICATION TECHNIQUE FOR LAND-COVER
MAPPING IN THE ARCTIC TUNDRA,
NORTH SLOPE, ALASKA

By

Debasish Chaudhuri

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

2008

Approved by

DR. ROY STINE

Committee Chair

© 2008 DEBASISH CHAUDHURI

Dedicated to my parents....

Mr. Mihirbaran Chaudhuri

&

Mrs. Madhuri Chaudhuri

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. HAMID NIMATI

DR. RICK L. BUNCH

DR. Z. J. LIU

Date of Acceptance by Committee

Date of Final Oral Examination

ACKNOWLEDGEMENTS

I would like to express my heartiest gratitude to Dr. Roy Stine, chair of my dissertation committee for all his valuable guidance, comments, help and suggestions to complete this dissertation research successfully. I tender my gratefulness and thanks to my professors, Dr. Zhi-Jun Liu and Dr. Rick Bunch for their advice and comments on the research paper. I would like to thank Dr. Hamid Nemati for being a part of my research committee from the ISOM department and his immense help and suggestions in implementing the neural network classification. I am also grateful to Dr. Peter Ray for helping me to identify the vegetation species during the field work, in Alaska. I would also like to take this opportunity to thank the Department of Geography, University of North Carolina, Greensboro for providing financial support for the PhD program and considering me a part of their family through out these years.

I am especially grateful to Ms. Mary Hall Brown and Mr. Prasad Pathak for their help in data collection, encouragement and care. It was a wonderful experience to be with you all and share the difficulties and anxiousness of graduate student life. I cannot forget to thank Mr. Jason Ortegren for encouragement and help in improving the readability of this dissertation.

Special thanks to Mr. Somesh Jana and Dr. Shamim S. Mondal for their help during my application for admission in the PhD program. I am thankful to Baba, Ma, Raja, my in-laws, Rajada, my childhood friends, and many others for their love, moral support, and blessings that helped me to complete this research. Last, but not the least, my beloved wife Koel deserves my heartiest thanks to be with me through out these days, for her care, love, support and providing me with enthusiasm and motivation that helped to complete this dissertation successfully.

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES	ix
 CHAPTER	
I. INTRODUCTION	1
1.1. Overview	1
1.2. Remotely Sensed Image Classification and the Arctic Tundra...	2
1.3. Traditional (Spectral Based) Classification Techniques	3
1.4. Non-Traditional Image Classification Techniques	4
1.5. Problem Definition	5
1.6. Research Objectives	5
1.7. Research Questions.....	6
1.8. Dissertation Structure.....	8
II. LITERATURE REVIEW	10
2.1. Traditional Classification Methods.....	10
2.2. Remote Sensing with Satellite Imagery and Traditional Classification Techniques Applied in the Arctic Tundra.....	11
2.3. Problems in Remote Sensing and Traditional Image Classification Techniques with SPOT Satellite Image in Alaska.....	15
2.4. Incorporation of Spatial Data in Classification	18
2.5. Knowledge (rule) Based Classification.....	20
2.6. Data Mining Techniques in Remotely Sensed Image Classification.....	21
2.6.1 Classical Data Mining and Spatial Data Mining	21
2.6.2 Spatial Data Mining and Knowledge Discovery	22
2.6.3 Artificial Neural Networks	23
2.6.4. Decision Trees	25
2.6.5. Hybrid Classification Techniques	26
III. METHODOLOGY.....	29
3.1. Data and Study Area.....	29
3.1.1. Ancillary Data	34

3.1.2. Normalized Difference Vegetation Index (NDVI)	34
3.1.3. Slope and Aspect	36
3.1.4. Texture	39
3.2. Software Used.....	44
3.3. Geometric Correction	45
3.3.1. Collection of Ground Control Points	46
3.3.2. Post-processed Differential Correction	47
3.3.3. Geometric Correction model.....	48
3.4. Cloud and Shadow pixels removal	51
3.5. Classification Scheme	54
3.6. Collection of sample data for training and accuracy (testing) assessment	58
3.7. Unsupervised Classification (ISODATA clustering).....	60
3.8. Classification (supervised) with Feature Analyst	64
3.9. Hybrid Classification.....	71
3.9.1. Hybrid Methodology	71
3.9.2. Knowledge Based (Expert) Classification	73
3.9.2.1. Expert Classifier in ERDAS Imagine	73
3.9.2.2. Advantages and Disadvantages of Expert Classification	74
3.9.2.3. Preliminary Classification Rules.....	75
3.9.3. ISODATA clustering of the remaining pixels.....	82
3.9.4. Final Classification rules	84
3.9.5. MLP Classifier	90
3.9.5.1. Artificial Neural Network and its advantages	90
3.9.5.2. The Multilayer Perceptron.....	92
3.9.5.3. Neural Network Architecture and Parameters	95
3.9.6. The Tundra Index	101
3.9.7. Integration of the two classifiers: rule-based and MLP	108
3.10. Post-processing of the classified images for the area under haze	109
IV. RESULTS AND DISCUSSIONS	113
4.1. Results	113
4.1.1. Accuracy Assessment	113
4.1.1.1. Error Matrix	114
4.1.1.2. Overall Accuracy.....	116
4.1.1.3. User's and Producer's Accuracy.....	116
4.1.1.4. The Kappa Statistic.....	119

4.1.2. Statistical Significance of the Accuracy Assessment.....	121
4.2. Discussions.....	125
4.2.1. Comparison of classification performance for each class: error matrices, producer's and user's accuracies, and kappa values.....	125
4.2.2. Overall Accuracy and Statistical Significance.....	129
4.2.3. Confusion between vegetation classes	130
4.2.4. Solution to the research questions	135
4.2.5. Research Concerns and Limitations.....	139
V. CONCLUSIONS	147
5.1. Conclusions.....	148
5.2. Future Research Recommendations.....	151
REFERENCES	154
APPENDIX A. FIELD PHOTOS COLLECTED ILLUSTRATING THE DIFFERENT LAND-COVER CLASSES.....	166
APPENDIX B. EXAMPLE OF FIELD FORMS AND FIELD DATA COLLECTION METHODOLOGY.....	173
APPENDIX C. FIGURE SHOWING THE DIFFERENT RULES OF THE RULE-BASED CLASSIFIER.....	174
APPENDIX D. TABLE SHOWING THE TRIAL RUNS FOR THE MLP CLASSIFIER WITH DIFFERENT INPUT LAYERS AND NUMBER OF NODES IN THE HIDDEN LAYERS VALUES TO FINALIZE THE NETWORK ARCHITECTURE.....	181

LIST OF TABLES

	Page
Table 1. Description of the image data properties	32
Table 2. Statistics for residual and RMS errors	49
Table 3. Iterations showing the improvements in training accuracies and kappa values with the inclusion of the tundra index in the MLP classifier	104
Table 4. Error Matrix for unsupervised ISODATA clustering	115
Table 5. Error Matrix for supervised classification with Feature Analyst.....	115
Table 6. Error Matrix for hybrid classification.....	115
Table 7. Comparison of overall accuracies for the three different classifiers ...	116
Table 8. User's and Producer's accuracies for unsupervised ISODATA clustering.....	117
Table 9. User's and Producer's accuracies for supervised classification with Feature Analyst	118
Table 10. User's and Producer's accuracies for proposed hybrid classifier	118
Table 11. Comparison of the kappa values for the three classifiers	120
Table 12. Statistical significance between hybrid classifier and unsupervised classification with ISODATA clustering.....	124
Table 13. Statistical significance between hybrid classifier and supervised classification with Feature Analyst.....	124

LIST OF FIGURES

	Page
Figure 1. General location of the research area around Lake Toolik, in Alaska	31
Figure 2. The subset of SPOT image (cloud and shadow pixels removed) used and the subset on the entire image in the inset (RGB 3, 2, 1). ..	33
Figure 3. NDVI Image used (cloud and shadow pixels removed)	35
Figure 4. Slope map for the image area (cloud and shadow pixels removed) ...	37
Figure 5. Aspect map for the image area (cloud and shadow pixels removed) ..	38
Figure 6. Band 1 texture (cloud and shadow pixels removed)	40
Figure 7. Band 2 texture (cloud and shadow pixels removed)	41
Figure 8. Band 3 texture (cloud and shadow pixels removed)	42
Figure 9. NDVI texture (cloud and shadow pixels removed)	43
Figure 10. Collection of ground control points in the study area	46
Figure 11. Example of geometric distortion of the SPOT-5 image	50
Figure 12. Spectral profile for barren (brown) and cloud (white) pixels showing the spectral similarity between the two	51
Figure 13. Spectral profile for water (blue) and shadow (black) pixels showing the spectral similarity between the two	52
Figure 14. Figure showing the shadow cast (black) by the terrain due to the sun angle estimated by ATCOR in ERDAS Imagine 9.1	54
Figure 15. Training sites from field knowledge and sample sites collected in the form of lines and polygons	59
Figure 16. Overall distribution of training and testing (accuracy) sample points in the study area	60
Figure 17. Classified Image for unsupervised ISOADATA clustering	63

Figure 18. Showing the general work flow of the Feature Analyst.....	65
Figure 19. Comparing extraction of the features in the Toolik Lake with the different input representations	67
Figure 20. Figure showing the accepted, incorrect, correct and current signatures for snowbed complex	68
Figure 21. Flow chart showing the work flow for supervised classified image with Feature Analyst	69
Figure 22. Classified Image (Supervised) with Feature Analyst	70
Figure 23. Flowchart for proposed hybrid classification methodology	72
Figure 24. Snapshot of the Expert Classifier of the Expert Classifier in ERDAS Imagine.....	74
Figure 25. Pixels unclassified from the Expert classifier.....	80
Figure 26. Partially classified image by Expert classifier	81
Figure 27. Unclassified pixels from Expert Classifier, classified and recoded.....	83
Figure 28. Classified Image from Expert Classifier.....	89
Figure 29. An Artificial Neural Network with single hidden layer.....	93
Figure 30. Script in ERDAS Imagine used for scaling the input raster.....	97
Figure 31. The MLP classifier interface in IDRISI Andes.....	100
Figure 32. Script in ERDAS Imagine for modeling the tundra index	102
Figure 33. Spectral profile for MAT and MNT in the SPOT-5 image.....	105
Figure 34. The tundra index layer (with cloud and shadow pixels removed). ...	106
Figure 35. The classified image from the MLP classifier	107
Figure 36. Script in ERDAS Imagine for modeling the tundra index	109

Figure 37. The area under haze (top) and the clipped area reclassified and recoded (bottom)	111
Figure 38. Script in ERDAS Imagine for model merging of the recoded haze area classification into the output of the MLP classifier.....	112
Figure 39. Showing the comparison of the accuracy values for the three classifiers.....	121
Figure 40. SAS Program to calculate kappa variance from error matrix for hybrid classifier.....	122
Figure 41. Comparison of the three classified images.....	126
Figure 42. Spectral similarity between shrubs (riparian) on the left and wet graminoid tundra complex (rich fen) on the right.....	134
Figure 43. Comparison of the three classifiers in terms of individual kappa values of the classes.....	137
Figure 44. Spectral mixing in the transition zones: the effect of heterogeneity in the area	142
Figure 45. Sample points discarded which were collected within 300m (buffer) from the Dalton road	145

CHAPTER I

INTRODUCTION

1.1. Overview

Arctic ecosystems are considered to be particularly sensitive to disturbances in the form of a change either in vegetation or the underlying substrate caused by some external factors which range from localized events, such as energy exploration or lightning induced fires, to global climate change (Walker, 1996). Walker et al. (1991) pointed out that although most anthropogenic disturbances are microscale ($10^{-7} - 10 \text{ km}^2$) phenomenon, but cumulatively they can cause mesoscale ($10 - 10^4 \text{ km}^2$) disturbances which in turn can affect the tundra ecosystems in the macroscale ($10^4 - 10^6 \text{ km}^2$) level. The arctic provides a test bed to provide a better understanding and evaluation the effects of threshold changes in regional system dynamics (Chapin et al. 2005). Thus it is very important to understand the consequences of global climate change on the mesoscale patterns of the arctic ecosystem. Shifts in arctic tundra ecosystem functioning due to global climate change are likely to be expressed through changing vegetation phenology and species combinations, since vegetation will respond most rapidly to climatic change (Epstein et al. 2004; Calef et al. 2005, Vierling et al. 1997).

1.2. Remotely Sensed Image Classification and the Arctic Tundra

Remote sensing is the best tool for looking at vast areas of the Earth's surface to analyze, map, and monitor ecosystem patterns and processes (Gould, 2000). Several researchers have used remote sensing image classification techniques to understand the vegetation pattern in the arctic region (Walker, 1999; Stow et al. 1993; Shippert et al. 1995). The use of optical remote sensing systems in arctic regions faces a number of challenges, including frequent cloud cover (Stow et al. 1998, 2004; Hope et al. 1995).

Noyle (1999) pointed out several studies showing considerably low classification accuracies for mapping of land-cover and vegetation in remote areas of northern Alaska (Fleming, 1998; Stow et al. 1989; Pacific Meridian Resources, 1995; Felix et al. 1989). The only exception to this was Muller et al. (1998) who performed an accuracy assessment on a land-cover map (Auerbach et al. 1997b) derived from a Landsat MSS data (50 meter spatial resolution, resampled) of the Kuparuk river basin and achieved considerably high accuracy results. Geographical remoteness of the area and cold climate hinders ground truth data collection for post-classification accuracy assessments of vegetation classification studies in Alaska.

1.3. Traditional (Spectral Based) Classification Techniques

Traditional image classification techniques in remote sensing involve the acquisition and interpretation of spectral based remote measurements to obtain information about the Earth's surface. The classification process assigns each pixel in a number of spectral bands of an image to a particular class of interest, such as water, barren, vegetated, urban etc. The resulting image is referred to as a classified or thematic map. Many different approaches have been proposed for performing the classification task.

There are many existing standard classification techniques and algorithms. Classes may be specified a priori by an analyst (supervised classification) or automatically clustered (unsupervised classification) into number of information classes, where the number of desired classes is specified by the analyst. In all the existing traditional image classification procedures, spectral brightness values of the different spectral bands are used as the numerical basis for classification. All these classification algorithms are based on the assumption that objects on the Earth's surface will have unique spectral values and hence belong to one of the several distinct and exclusive spectral classes concerned. But in reality, the spectral responses of surface features in an image are dependant on many other factors including terrain, slope, aspect, soil type and moisture content, and atmospheric conditions. Thus, multi-spectral image information by itself has sometimes proven insufficient for differentiating

land-cover classes in a satellite image (Carpenter et al. 1997). As a result spectral based classifications will not be able to capture the complexities of the Earth's surface. For example, Stow et al. (1989) pointed out maps generated by classifying spectral data of SPOT/HRV-XS data alone are unsuitably inaccurate for mapping arctic tundra vegetation types.

1.4. Non-Traditional Image Classification Techniques

To help differentiate land-cover classes that are not easily separated using spectral brightness values, ancillary data have often been used. Ancillary data such as elevation, slope, aspect, soil, and hydrology have been incorporated directly into modern classification algorithms such as expert systems (knowledge based and rule based) and neural networks. Parametric methods such as unsupervised cluster busting and maximum likelihood classifier (MLC), nonparametric methods such as nearest-neighbor classifiers, fuzzy classifier and neural network and non-metric methods such as rule-based decision tree classifiers are widely being used by recent remote sensing researchers (Duda et al. 2001, Liu et al. 2002, Skidmore et al. 1997; Jensen et al. 2001, Stow et al. 2003). Availability of spatial databases and incorporation of data mining techniques have opened several new opportunities to improve traditional classifiers and develop new classification systems that can incorporate these spatial databases into the decision process.

1.5. Problem Definition

The unique arctic tundra environment with short growing season, cloud cover, snow and ice cover hinders effectiveness of remote sensing studies (Hope et al. 1995). As pointed out by Noyle, (1999) the classification accuracy of arctic biomes falls below the accuracy of other biomes of the world. Although, a vegetation map made from a Landsat MSS Image (resampled to 50m pixel) classification map by, gave an accuracy measure of 87.1% (Auerbach et al. 1997b, Muller et al. 1998). On the other hand Stow et al. (1989) emphasized the fact that SPOT/HRV-XS data were not fine enough to identify the small arctic tundra vegetation parcels while the panchromatic band of SPOT/HRV could resolve spatially most of the vegetation parcels but lacked enough spectral resolution to discriminate the vegetation types. This research is motivated and guided by the research need to apply a new classification technique to develop an accurate vegetation map at a higher spatial resolution i.e. SPOT-5, which will be very useful for scientific researchers in the area.

1.6. Research Objectives

The purpose of this dissertation was to develop a hybrid image classification approach that effectively integrates ancillary information into the classification process. The proposed classification approach combines ISODATA clustering (unsupervised), expert classifier (rule-based) and the Multilayer Perceptron (MLP) classifier that uses artificial neural network (ANN). The main

goal was to examine the best possible combination or sequence of classifiers for typically classifying tundra type vegetation in the SPOT-5 satellite image. The proposed classification approach aims to produce higher accuracy than the existing classified vegetation map of the arctic tundra made from SPOT-5 data using traditional classification techniques.

The specific objectives include: (1) application of data mining techniques with geo-spatial and spectral knowledge in SPOT-5 satellite image data to develop a new hybrid classification technique for unique tundra vegetation in Alaska, and (2) compare and contrast the image classification performance between the proposed classification and the standard (spectral) classification techniques: unsupervised ISODATA clustering and supervised classification with Feature Analyst for the arctic tundra vegetation environment with accuracy measures.

1.7. Research Questions

- How to extract spatial and spectral knowledge for the unique arctic tundra vegetation type that can be utilized for expert classification?
- How can a hybrid classifier be used to classify SPOT-5 data (resampled to 5 meter pixel) to achieve higher classification accuracy than traditional classification techniques used using actual ground truth data?

- What are the statistical significances of the classification accuracy obtained from the proposed method as compared to traditional spectral classifiers used in that area?

This research used a hybrid classification technique that combined three classifiers, namely, unsupervised (ISODATA) clustering, rule-based classifier, and a Multilayer Perceptron (MLP) classifier. Unsupervised (ISODATA) clustering and rule-based classifiers were used to produce an intermediate classified map which was used as an input to the Multilayer Perceptron (MLP) classifier. The result from the MLP classifier was compared to the classified map obtained in the previous step (combination of the rule-based and ISODATA clustering) and for the pixels where there was a disagreement for the class allocations, the class having a higher kappa value was assigned to the pixel in the final classified map. The results were compared to standard classification techniques: unsupervised clustering technique and supervised classification with Feature Analyst for accuracy measures with ground truth data. The results indicated higher classification accuracy (75.6%, with kappa value of .6840) for the proposed hybrid classification method than the standard classification techniques: unsupervised clustering technique (68.3%, with kappa value of 0.5904) and supervised classification with Feature Analyst (62.44%, with kappa value of 0.5418). The results were statistically significant at 95% confidence level.

1.8. Dissertation Structure

This dissertation is divided into 5 chapters explaining literature review and the background, the data preparation steps, the methods and concepts used, the methods used and the results of the research, related discussions, conclusions and future research probabilities. The contents of the rest of the chapters are outlined below.

Chapter 2 describes the background literature review and concepts of image classification, remote sensing in Alaska and applied traditional classification methods and problems, incorporation of spatial data in image classification, knowledge (rule) based classifiers, data mining techniques in classification including ANN, Decision Tree and Hybrid Classifiers.

Chapter 3 explains the methods applied in this research with the information about the image data and the ancillary data used. The chapter elaborates the geometric correction of the image, cloud and shadow pixel removal, classification scheme, data collection procedures, unsupervised ISODATA clustering, classification with Feature Analyst, the proposed hybrid classifier with rule-based classification and MLP classifier and finally post-processing of the area under the haze.

Chapter 4 discusses the results of the research. The classification accuracies of the classified images from the three different classifiers are

presented and comparison between them is explained with statistical significance. This chapter also points out the research limitations and concerns.

Conclusions of the results of the dissertation, and future research recommendations are presented in chapter 5.

CHAPTER II

LITERATURE REVIEW

2.1. Traditional Classification Methods

The use of spectral or pixel-based classifiers with multi-spectral data began in the 1970s (Anderson et al. 1976). Both traditional unsupervised and supervised classification techniques often applied to remotely sensed data, are spectral based approaches, to match the spectral classes in the data to the information classes of interest. Rarely a simple one-is-to-one match is found in the real world between pixel groups and concerned information classes and traditional techniques neglect the spatial arrangement of the pixels. In reality, either unique spectral classes not corresponding to any information class of interest, or, one broad information class (e.g. cultivated field) containing a number of sub-classes with unique spectral signatures can be found. For example in a cultivated field, spectral sub-classes may be formed due to variations in age, species, and water content, shadowing or variations in scene illumination due to different sun-angle (Canada Centre for Remote Sensing, n.d.).

Supervised image classification is a method of classification in which the analyst defines small homogenous areas, known as training sites which

represents each land-cover category of interest. This delineation of training areas representative of an information class is most effective when an analyst has sufficient knowledge of the geography of the region and experience with the spectral properties of the cover classes concerned (Skidmore, 1989). The analyst then trains the software used to recognize spectral signatures associated with the training sites. The software then uses those defined signatures for each land-cover category for to classify the remaining pixels (ERDAS field guide, 2005).

Unsupervised image classification is a method in which the analyst uses the software to separate the image into X number of classes (or clusters). No prior information is needed from the analyst regarding the information classes. Once this process is completed, the analyst identifies and relabels the land-cover type for each class (cluster) based on image interpretation, ground truth data, previous maps and field reports to combine the spectral clusters into information classes. (ERDAS field guide, 2005)

2.2. Remote Sensing with Satellite Imagery and Traditional Classification Techniques Applied in the Arctic Tundra

The history of remote sensing in the arctic with digital image goes back in 1973 when Anderson et al. used Landsat 1 (ERTS -1) for land conservation and mapping in Alaska. Morrissey et al. (1981) used 10 Landsat scenes to map over 23 million acres of vegetation within the National Petroleum Reserve in Alaska. In the mid eighties, several other private and government agencies used remote sensing to map vegetation maps in vast expanses of land in Alaska e.g. Alaska

Bureau of Land Management (BLM), NASA, USGS (Shasby et al. (1986), Walker et al. (1985), Markon (1992; 1995).

Besides vegetation mapping, remote sensing is potentially used to identify ecosystem changes, changes in land-cover, structure, phenological growth characteristics, and ecotones (boundaries) (Stow et al. 2004). A normalized difference vegetation index (NDVI) in biophysical remote sensing studies was used for arctic tundra regions at different geographic scales in order to measure phytomass in bioclimate subzones and vegetation units (Hope et al. 1993, McMichael et al. 1999, Shippert et al. 1995). Not only Landsat data but data from other sensors were equally useful in other biological interests in the region. Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS) sensor data were used to characterize changes in the phenological growth characteristics of Arctic vegetation (Zhou et al. 2001; Markon et al. 1995). Remote sensing was also used in the arctic region to identify land-cover change and vegetation characteristics using time-series NDVI data, to study intra-seasonal dynamics on arctic vegetation, inter-annual growth dynamics with NDVI, biotic controls over spectral reflectance of arctic tundra vegetation, and primary productivity, spatial variation in carbon dioxide flux (Stow et al. 2003; Jia et al. 2004; Hope et al. 2005; Riedel et al. 2005; Williams et al. 2001; Vourlitis et al. 2000).

Satellite image classification and mapping are necessary in order to study the large expanses of arctic tundra lands which are difficult to access and are seldom covered by suitable aerial photos (Walker et al. 1995). The first vegetation analysis and description of plant communities according to the Braun-Blanquet approach in the tundra landscape of the Toolik region in the northern slope of Brooks Range in Alaska, was done by Walker et al. (1994). Several researchers worked with Landsat MSS and SPOT multi-spectral (XS) data in the arctic region and used traditional image classification techniques (Walker et al. 1987; Shasby et al. 1986; Stow, 1989)

Markon et al. (1994) used a mosaic of three SPOT MSS satellite scenes to apply clustering techniques to develop statistical parameters by which the SPOT data were spectrally classified to map Tundra vegetation in the Teshekpuk Lake area of the Alaskan Arctic Coastal Plain. A maximum likelihood algorithm that correlated spectral classes with land-cover types was applied to the SPOT data. Field data were used to assist in spectral class labeling and vegetation descriptions. In the next year, Walker et al. (1995) used twelve land-cover classes which were spectrally identified and mapped using both supervised and unsupervised clustering techniques. In their research Walker et al. (1995) developed standardized systems of arctic vegetation classification and used classification techniques which exploited moderate resolution satellite data from Landsat Multi-Spectral scanner (MSS).

Muller et al. (1998) classified a Landsat MSS mosaic using a K-means unsupervised algorithm. Forty three cluster classes were initially generated which were later merged into eight land-cover classes. Using first-hand experience with the area and other local areas maps from the North Slope each cluster was interpreted and grouped into eight land-cover categories (Walker et al. 1987; Walker et al. 1991; 1996). The overall map accuracy was 71%, from error matrix analysis done in 2000.

Stow et al. (2000) determined the optimal spectral radiometric and temporal features derived from single-date and seasonal time series AVHRR imagery for classifying three arctic tundra functional types: acidic tundra, moist non-acid tundra and wet sedge tundra. Both supervised classification and unsupervised classification techniques were compared. An ISODATA clustering routine was used with 30 cluster classes, with a maximum likelihood decision rule and interactive cluster labeling to identify the spectral classes. A single-date, three-band (VIS, NIR and NDVI) input yielded a map with the highest agreement (86.1% for supervised and 87.8% for unsupervised approaches) compared to the reference map made by Auerbach et al. (1997b). However, the wet sedge tundra class for the different AVHRR inputs varied the most and were least similar to the reference data, in other words, had the least accuracy.

Vegetation map of the Hood River region of the Central Canadian Arctic was prepared by Gould (2000) which was derived from Landsat Thematic

Mapper (TM) bands 1–5 and 7, using a maximum likelihood algorithm for supervised classification. Training sites for the supervised classification were chosen from homogeneous areas for which detailed vegetation descriptions were available (Gould et al. 1999). However the accuracy assessment was yet to be accomplished for the produced vegetation map.

2.3. Problems in Remote Sensing and Traditional Image Classification Techniques with SPOT Satellite Image in Alaska

Among all the satellite image classification studies done so far in this region, only few researchers have tested the accuracy of the classified map with quantitative accuracy assessments from actual ground truth data (Fleming, 1988; Felix and Binney 1989; Stow et al. 1989; Kempka et al. 1995; Muller et al. 1998). This may be due to the factors that are mostly unique to Arctic tundra environments which limit the effectiveness of remote sensing studies in general, specifically multi-temporal optical sensing: (1) short growing season (2) persistent cloud cover, (3) solar geometry, (4) standing water and shallow lakes, and (5) snow and ice cover (Hope et al. 1995). Besides, the effects of roadside disturbance on the substrate and vegetation properties might be a cause of lower classification accuracy rate (Auerbach et. al., 1997; Walker et. al. 1987; Forbes et. al. 1999).

Fleming (1988) used a unsupervised clustering technique with Landsat MSS data (50 meter resampled) and ancillary data DEM, slope, and aspect to map broad land-cover (1:125,000 scale) in large inaccessible areas in Alaska.

The classification had 13 preliminary classes: open spruce forest, closed mixed forest, open birch forest, closed birch forest, open tall shrub, closed tall shrub, dwarf shrub/graminoid tussock, prostrate dwarf shrub tundra, closed dwarf shrub tundra, aquatic forb marsh, water, cloud shadow, and cloud/snow. The overall accuracy achieved was 78.2%, with a kappa value of 57.2.

Felix et al. (1989) did an accuracy assessment study with 126 sites for ground truth data collection, for a vegetation map based on Landsat MSS data for coastal plain and foothills of north east Alaska. The 13 classes in consideration were clear water, offshore water, shallow water, very wet graminoid tundra, wet graminoid tundra, moist/wet tundra, moist prostrate dwarf scrub, moist graminoid tussock, mesic erect dwarf scrub, alluvial deciduous scrub, dry prostrate dwarf scrub, scarcely veg. floodplain, and barren floodplain. The overall accuracy was 37% and classes with least accuracies were very wet graminoid tundra, wet graminoid tundra, and moist/wet tundra.

Stow et al. (1989) pointed out that the unique characteristics of the landscape in the Foothills of Alaska affects the interpretation of the SPOT/ High Resolution Visible (HRV) multispectral image (XS) images. Firstly, the micro-relief of the dominant vegetation, tussock tundra, causes shadowing and bidirectional reflectance properties. Secondly, images acquired at this location are mostly associated with solar elevations less than 45 degrees, which is generally considered insufficient quantitative image analysis. Thirdly, the short

season of acceptable illumination and snow free conditions (June-August) and the infrequent clear sky conditions make satellite-based remote sensing studies difficult. The research also emphasized on the fact that SPOT/HRV-XS with its 20m spatial resolution was not fine enough to identify the small arctic tundra vegetation parcels while, the 10 m spatial resolution panchromatic band SPOT/HRV could resolve spatially most of the vegetation parcels but lacked enough spectral resolution to discriminating the vegetation types. The overall accuracy of the classification for this study by Stow et al. (1989) was about 56%.

The study done by Kempka et al. (1995), in the national petroleum reserve area, in North Slope of Alaska involved the use of two Landsat TM images for classification. The 7 major classes mapped were: water, aquatic, flooded tundra, wet tundra, moist tundra, shrub, and barren ground. The overall accuracy achieved was 50.72%.

The only exception was, Muller et al. (1998) who studied the accuracy assessment of a vegetation map made from a mosaiced Landsat MSS image (resampled to 50m pixel) by Auerbach et al. (1997b). Unsupervised ISODATA clustering approach was applied for the classification and the initial 40 spectral clusters were identified into eight land-cover classes using first hand field knowledge, geobotanical and other Landsat derived maps of the area. The classes were: barren, moist acidic tundra, moist non-acidic tundra, shrublands, wet tundra, water, clouds and ice, shadows. A post-classification sorting was

applied with ancillary data to get the final classified map. For the accuracy assessment, 3 X 3 homogenous pixel blocks were used for sampling in 178 sites and an overall accuracy measure of 87.1% was achieved.

2.4. Incorporation of Spatial Data in Classification

The spectral response of thematic classes is dependent on many factors including terrain, slope, aspect, soil type, and atmospheric conditions present during the image acquisition. Spectral data alone cannot be used to classify a satellite image to get accurate image classification. Strahler et al. (1978) showed that accuracies of computer classification of species-specific forest cover types from Landsat imagery can be improved by 27% or more through the incorporation of topographic information from digital terrain tapes registered to multirate Landsat imagery. Thus, it is possible to exploit the knowledge derived from ancillary spatial data to improve the classification accuracies (Vatsavai et al. 2005).

In a study by Gerçek (2002), an approach for integrating topographic data including elevation, slope and aspect in land-cover classification was implemented. Training sets were used to perform standard maximum likelihood classification of spectral data together with topographical raster data. The results conveyed that procedure provided an improvement of 10% in overall accuracy for the classification with the integration of topographical data over spectral data only.

In the light of traditional spectral based classification techniques, Visual Learning Systems, Inc. (VLS), of Missoula, Montana, has developed a commercial software application called Feature Analyst which utilizes multiple spatial attributes (size, shape, texture, pattern, and spatial association) with spectral information and incorporates advanced machine learning techniques to supply higher levels of accuracy in feature extraction (O' Brien, 2003). The benefit of this machine learning approach over standard supervised image classification techniques, such as the maximum likelihood method, lies in the ability to improve feature classification using inductive learning techniques (Kader et al. 2002). It is quite evident that with the consequent improvement in existing technologies, researchers are trying to improve remote sensing classification techniques.

A similar data mining environment for interactive exploration and analysis of remotely sensed data was suggested by Koperski et al. (2001) who described the usage of DEM data and DEM derived information such as aspect and slope with the system for data mining and statistical analysis of remotely sensed imagery. Aksoy et al. (2005) suggested the used of statistical summaries of spectral, textural and shape properties of pixels to model clusters and assigned memberships to those clusters in multiple resolution levels are used to classify the corresponding pixels into land-cover/land-use categories using decision tree classifiers. This research using region based spatial information was proved to

be effective over traditional spectral based techniques in terms of overall accuracy.

2.5. Knowledge (rule) Based Classification

Skidmore (1989) defined an expert system as a computer system that attempts to solve complex real-world problems by reasoning. The expert's knowledge about the slope, aspect, geomorphology, geology etc. of the area under consideration can be used in the classification procedures along with the spectral knowledge. The knowledge can be aggregated into hierarchical rules (IF – THEN) to classify image data. The knowledge base is represented as a tree diagram consisting of final and intermediate class definitions (hypotheses), rules (conditional statements concerning variables), and variables (raster, vector, or scalar). Such classification is known as knowledge (or rule) based classification. The gathered knowledge can then be repeated by someone who may not be an expert consistently producing reliable and repeatable analysis results (ERDAS Field Guide, 2005). Avci et al. (2004) reassured the importance of using additional spectral and spatial knowledge in order to improve the classification accuracy and used a knowledge based hierarchical approach to classify and detect forest types in the Ömerli Dam Lake Region. Hazarika et al. (n.d.) used the rule-based classification technique to identify rhino habitats in India where several GIS data and remote sensing data were integrated to develop the knowledge base.

However, the disadvantages of using knowledge base classifier is the difficulty faced in the creation of the knowledge base , availability of reliable training data , and knowledge acquisition bottleneck (Gonzalez et al., 1993; Avci et al. 2004; Huang et al. 1997). Several other researchers have successfully used rule-based classification techniques, integrating GIS and remote sensing data (Jensen, 1978; Hansen et al. 2000; Stow et al. 2003).

2.6. Data Mining Techniques in Remotely Sensed Image Classification

Classification and clustering are two of the most common operations associated with classical data mining. Classification refers to a learning function that maps data into one or more predefined classes of interest. Data clustering (unsupervised learning) arrange data into clusters based on some attributes which minimizes the interclass similarity and maximizes the intraclass similarity. Traditional land-use/land-cover classification techniques which are applied to multi-spectral remotely sensed data for extraction of information classes are based on statistical pattern recognition techniques (Narumalani et al. 2002). But there is an essential difference between classical statistical methods and data mining.

2.6.1. Classical Data Mining and Spatial Data Mining

Hand (1998) pointed out that data mining techniques can handle large sets of data, contaminated or “dirty” data (i.e. anomalies in the data), selection

bias, dependent observations, find interesting patterns (patterns having high conditional probability as well as reasonably large marginal probabilities for the conditioning variables), and non-numeric data, which are not possible by traditional statistical techniques. Thus data mining can be defined as a technique that draws on techniques from machine-learning, database management, and statistics to rapidly search for patterns, and which allows researchers to discover potentially interesting, useful and unexpected patterns of information embedded in a large database (Shekhar et al. 2003). There is, however, a major difference between classical data mining and spatial data mining. Shekhar et al. (2003) pointed out that classical data mining fundamentally assumes that the data are independent but spatial data shows high degrees of spatial autocorrelation. Therefore many classical data mining algorithms often perform inadequately when applied to spatial data sets.

2.6.2. Spatial Data Mining and Knowledge Discovery

Several recent research studies have focused on incorporating spatial data mining techniques in remote sensing image classification with the help of ancillary data, e.g. DEM, slope, soil, and hydrology (Soh et al. 1998; Carpenter et al. 1997; Quinlan, 2000). Spatial data mining has been acknowledged as a useful technique in analyzing large volumes of geo-spatial data and remotely sensed imagery to identify patterns and their respective attributes in an image (Soh et al. 1998). Spatial data mining and knowledge discovery (SDMKD) can be defined as

the extraction of implicit, interesting spatial or non-spatial patterns and general characteristics. Spatial data mining is used in: 1) intelligent analysis of GIS data, and 2) knowledge driven interpretation and analysis of imagery. SDMKD thus provides a new method of knowledge acquisition for remotely sensed image classification. Li et al. (2004) pointed out that most existing remote sensing image retrieval systems use only simple queries based on sensor, location, and date of image capture. In this paper, Li et al. (2004) introduced an integrated approach to retrieving spectral and spatial patterns from remotely sensed imagery using state-of-the-art data mining and advanced database technologies allowing spatial queries that permit efficient retrieval of useful hidden information from large image databases.

2.6.3. Artificial Neural Networks

Neural networks simulate the thinking process of human beings, with interconnected neurons processing incoming information (Jensen et al. 1999; Hengl, 2002). Neural networks can improve classification accuracy by 10-30% compared to traditional classification techniques (Carpenter et al. 1997). Several researchers have applied neural network techniques in remote sensing image classification (Benidicktsson et al. 1993; Foody et al. 1995; Skidmore, 1988; Roli et al. 1996). These studies show that neural network classifiers make no a priori assumptions on the data probability distribution and are able to learn from nonlinear and discontinuous data samples. Moreover, neural networks can

readily accommodate ancillary data and are proven to be more accurate than traditional classifiers having a flexible architecture which can adapt to improve classification performance in particular situations (Carpenter et al. 1997). Being introduced in the 1970s, Adaptive Resonance Theory (ART) based on human cognitive information processing, paved the way for application of neural network models for unsupervised and supervised category learning and pattern recognition (Grossberg, 1976). ARTMAP systems, a supervised network architecture, self-organize arbitrary mappings from input vectors representing features such as pixel brightness values and ancillary data, to output vectors representing predicted information classes of interests.

Carpenter et al. (1997) developed new methodology for automatic mapping from Landsat TM and terrain data, based on the fuzzy ARTMAP neural network. Results were compared to those of maximum likelihood classifiers, as well as back propagation neural networks and k-nearest-neighbor (kNN) algorithms. ARTMAP dynamics are fast, stable, and scalable, overcoming common limitations of back propagation. Best results were obtained using a hybrid system based on a convex combination of fuzzy ARTMAP and maximum likelihood predictions. The research showed how the network automatically constructs a minimal number of recognition categories to meet accuracy criteria.

2.6.4. Decision Trees

In machine learning, the process of inductive learning can be viewed as a heuristic search through a space of symbolic descriptions for plausible general descriptions, or concepts, that explain the input training data and are useful for predicting new data (Jensen, 2005). There are a number of inductive learning algorithms among which C5.0 (a system that extracts informative patterns from data) and its predecessor C4.5 are mostly commonly used in image classification techniques (Quinlan, 1993; 2000). C5.0 is flexible (has no dependence on the probability distribution of the attributes) and is based on a decision-tree algorithm that is one of the most efficient forms of inductive learning. The three basic steps of applying this inductive algorithm to build a knowledge base system for image analysis with the incorporation of GIS data are: 1) training, 2) generating the decision-tree, and 3) creating production rules. Several researchers employed this inductive learning technique of decision-tree classifier to incorporate ancillary GIS data for multi-spectral image classification (Eklund et al. 1998; Huang, et al. 1997; Zhang et al. 2005).

Di et al. (2000) studied data mining techniques to discover knowledge from GIS database and remote sensing image data in order to improve land-use classification. The approach was to combine inductive learning with conventional image classification methods (Bayes classification) in the Beijing area using SPOT multi-spectral image and GIS data. A C5.0 inductive learning algorithm

was used to discover rules about spatial distribution patterns and shape features. Comparing with the result produced only by Bayes classification, the overall accuracy increased to 11 percent, thus indicating that inductive learning can resolve the problem of spectral confusion to a great extent. Combining Bayes method with inductive learning also extended the classification by subdivision of some classes with the discovered knowledge.

2.6.5. Hybrid Classification Techniques

Kanellopoulos et al. (1993) combined a MLC with two neural network classifier to get an enhanced performance where the second neural net was used to train those pixels which were mismatch between classes produced by the MLC and the first neural net. Brown et al. (1998) suggested the highest sum of the class membership values for each class derived from two different classification methods could be assigned the class to the pixel. Liu et al. (2002) pointed out that it is also possible that an expert system could be used in combination with a neural network and cited (Caudill, 1990; Wilkinson et al. 1992) to bolster the fact that the concept of integration of neural network and expert system already existed outside and within the remote sensing field. In their research, Liu et al. (2002) also used a consensus builder to adjust classification output in the case of disagreement in classification between maximum likelihood classifier, expert system classifier and neural network classifier.

Vatsavai et al. (2001) presented a new classification approach which combines knowledge based (KB) systems and maximum likelihood classifier (MLC) utilizing knowledge derived from ancillary spatial databases. This approach claimed to minimize the limitation of KB by simplifying the rule-base. In this simplified approach, the rule-base is used to stratify the image into homogeneous regions rather than classifying individual pixels. The stratified regions minimized the overlap among the classes and thus provided a robust environment for MLC. A semi-automated learning process was used to acquire training samples in each of the stratified regions, and classification was performed using standard MLC. This classification fusion approach yielded an overall accuracy of 85% for classes like water, high density urban, hardwood confers, and crop land but accuracy was low for lowland conifer, wetland, and low density urban because of high spectral overlap among these classes.

In line with this research, Vatsavai et al. (2002) proposed an efficient hybrid classification technique, based on statistical and knowledge based classifiers for mining remote sensing images. A traditional unsupervised technique based on the C-means clustering algorithm was applied to extract spectral clusters which were later classified into information classes using a decision-tree classifier made from ancillary geo-spatial data. The initial results of this research showed more efficient and accurate results than traditional MLC or decision-tree classifier.

All of these studies with implementation of spatial data mining techniques are in the early stage of experimentation and needs improvements. Also these studies address the research need of implement and test these techniques in different geographic environments (Vatsavai et al. 2001; 2002). This research focused on implementing the hybrid approach of image classification as suggested by Vatasavai et al. (2001; 2002) and Liu et al. (2002), using spectral and spatial knowledge, implementing data mining techniques to classifying the arctic tundra land-cover in a SPOT-5 imagery. The methodology combined together the two different approaches stated by Liu et al. (2002): 1) using a classified map obtained from one classifier (rule-based) as an input to a neural network classifier, and 2) the use of a consensus builder to improve classification accuracy. Specifically, spectral and spatial knowledge was gathered and implemented in a knowledge base of a rule based classifier to classify the SPOT-5 image. Pixels that were not classified by the rule based classifier were classified by unsupervised clustering (ISODATA) technique. The resulting classified image along with the available spectral and spatial data layers were fed into a MLP (neural network) classifier. Finally a consensus builder was used to choose the pixels having higher kappa values between the two classified images obtained from the MLP classifier and the rule-based classifier.

CHAPTER III

METHODOLOGY

3.1. Data and Study Area

A SPOT-5 image acquired in July, 2005, was used for the research. The image covers an area of about 650 square km near Lake Toolik. The research was conducted in the Toolik Lake region (68.63 °N/ 149.6 °W), foothills of the Brooks Range in the northern or arctic slope, Alaska. The area is a younger landscape glaciated during the late Pleistocene era and is a heterogenous area with small glacial lakes, kames and moraines including large areas from the Itkillik I (60000 years) and Itkillik II (10000 years) glacial drifts (Walker et al. 1994). The area has a rolling topography with an elevation range of 400 meters to 1300 meters, covered with mostly, acidic loamy soil with poorly drained surface layer (Walker et al. 1989). Typically, being in the north slope of the Brook's Range, only the surface (active layer, about 0.6 to 4 meters deep) of the tundra thaws each summer, while most of the soil remains permanently frozen i.e. the permafrost. The average yearly precipitation (1989 - 1999) is approximately 318 mm, while the average temperatures in July and December are around 10 degree Celsius and -25 degree Celsius respectively (Arctic LTER website; <http://ecosystems.mbl.edu/ARC/>).

A portion of this research was funded by the National Science Foundation for GIS and remote sensing applications in geomorphic-trophic hypothesis (GTH) research for benthic-pelagic coupling in arctic lakes study. The image area contains many of the GTH lakes where scientific research related with vegetation, topographic factors and water composition (e.g. primary productivity) of the arctic lakes are being carried out (Figure 1).

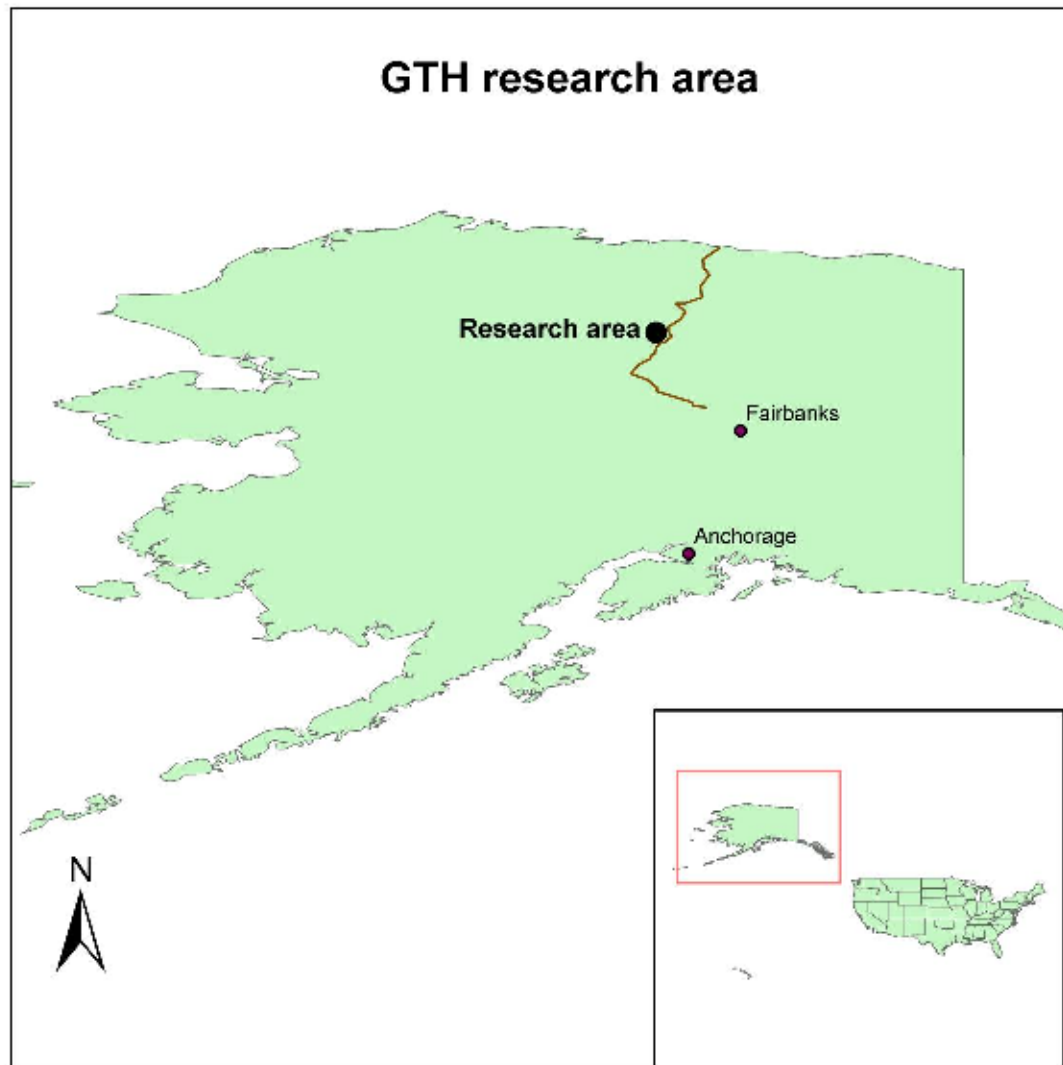


Figure 1: General location of the research area around Lake Toolik, in Alaska.

The SPOT image in consideration has a spatial resolution of 5m X 5m which is originally resampled from the original (10m X 10m) multi-spectral Band 1, Band 2 , Band 3 of SPOT-5 by a resolution merge technique with the panchromatic band (2.5m X 2.5m). Lastly, the radiometric resolution of the

imagery is 8 bit and it was geo-referenced to Universal Transverse Mercator (UTM) projection in zone 6.

Table 1: Description of the image data properties.

Acquisition date	July-25th- 2005
Radiometric resolution	8 bits
Spatial resolution	5m X 5m, resampled
Spectral resolution	Band1(Green): 500-590 nm
	Band2(Red): 610-680 nm
	Band3(NIR): 780-890 nm
Projection	UTM, Zone 6

Table 1 above describes the relevant sensor characteristics of the image data and following sections describes ancillary data used and the pre-classification image processing techniques applied. However, due to the lesser extent of the DEM data and also to restrict the ground sample collection sites and number of accuracy assessment sites to be used later, instead of using the entire available SPOT-5 image, a subset of the image (Figure 2) was used to test the proposed classification technique.

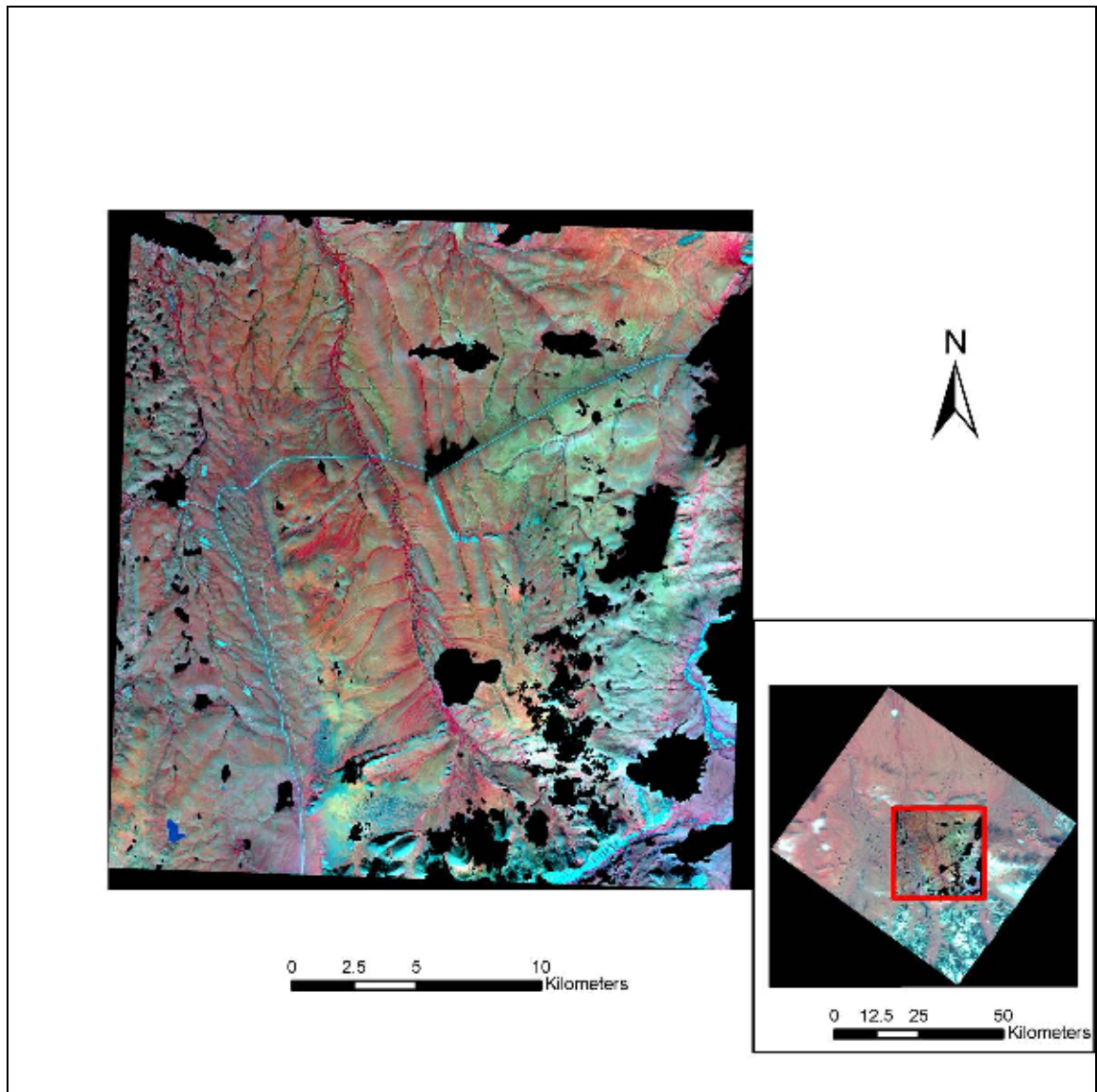


Figure 2: The subset of SPOT image (cloud and shadow pixels removed) used and the subset on the entire image in the inset (RGB 3, 2, 1).

3.1.1. Ancillary Data

Remotely sensed data are often used essentially for vegetation classification along with the integration of ancillary data into classification because classes are not always separable in the spectral feature space. The ancillary data used are Normalized difference vegetation index (NDVI), Slope, Aspect, and texture (variance) layers for each of the three spectral bands used as well as the NDVI layer.

3.1.2. Normalized Difference Vegetation Index (NDVI)

The normalized difference of the vegetation index (NDVI) is a non-linear transformation of the visible (red) and near-infrared bands of satellite information. NDVI is defined as the ratio of the difference between the visible (red) and near-infrared (NIR) bands, to the sum of the visible (red) and near-infrared (NIR) bands (Huete et al. 2002). The NDVI ratioing reduces many forms of multiplicative noise (e.g. Sun illumination, cloud shadows, topographic variation) and is an alternative measure of vegetation amount and condition (Jensen, 2005). It is associated with vegetation canopy characteristics such as biomass, leaf area index and percentage of vegetation cover. For SPOT data, it is given by:

$$NDVI = \frac{NIR - Red}{NIR + Red} OR \frac{Band3 - Band2}{Band3 + Band2}$$

NDVI values varies between -1.0 and +1.0 and the brighter pixels represent higher NDVI values which in turn is a representative of higher biomass under normal circumstances (Figure 3). Higher NDVI values also indicate greater and healthier plant cover (vegetation density) in an area. NDVI values are also used as spatial context for measurements of carbon flux.

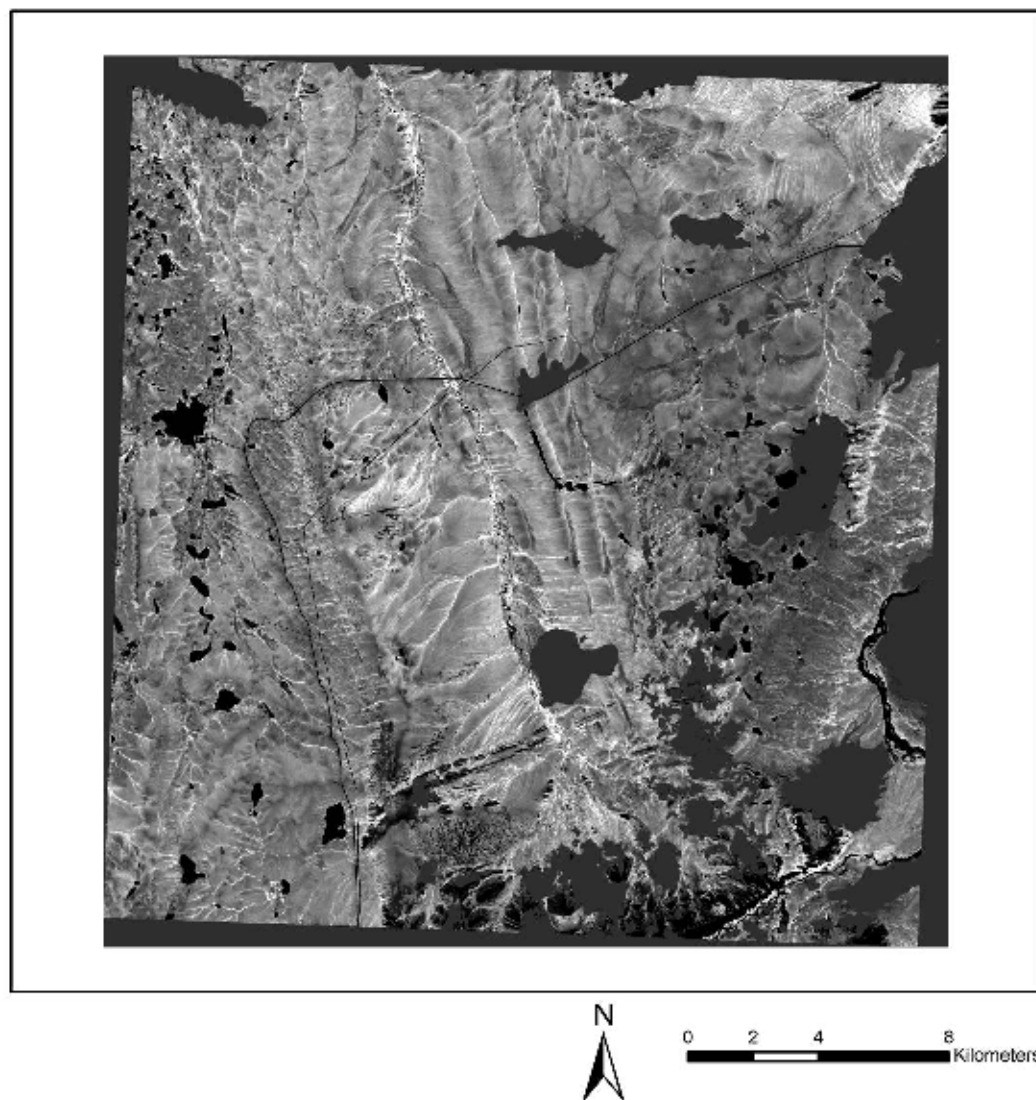


Figure 3: NDVI Image used (cloud and shadow pixels removed).

3.1.3. Slope and Aspect

Slope is expressed as the change in elevation (rise) over a certain distance (run). In this case of the raster data, the certain distance is the size of the pixel. The slope function in ArcGIS 9.2 uses a 3 x 3 cell neighborhood around the processing (center) cell in elevation raster applying the average maximum technique (Burrough et al. 1998; ESRI, 2007). The lower the slope value, the flatter the terrain; the higher the slope value, the steeper the terrain. Slope is most often expressed as a percentage, but can also be calculated in degrees. In this research, slope data (Figure 4) was calculated from the DEM (5m X 5m) data available and was expressed in degrees (ERDAS, 2005). Here higher (brighter) values of pixels represents higher slope.

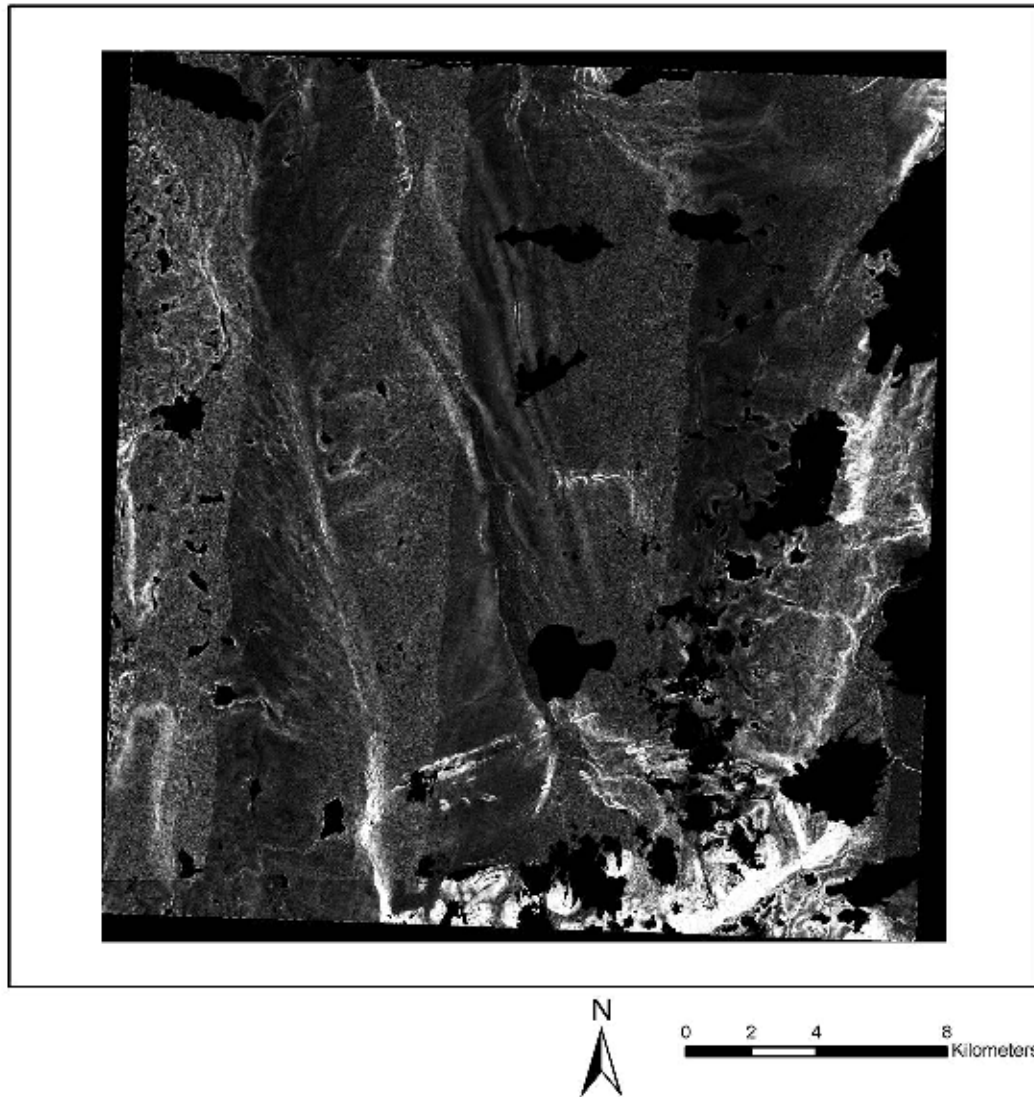


Figure 4: Slope map for the image area (cloud and shadow pixels removed).

An aspect image data (Figure 5) which is gray scale coded according to the prevailing direction of the slope at each pixel. Aspect can be considered as the slope direction and is expressed in degrees from north, clockwise, from 0 to 360. Due north is 0 degrees. A value of 90 degrees is due east, 180 degrees is

due south, and 270 degrees is due west. Aspect was calculated in ArcGIS 9.2 which uses the aspect function to fit a plane to the elevation values in a 3 x 3 cell neighborhood around the processing cell (ESRI, 2007). The direction the plane faces is the aspect for the processing cell. A value of 361 degrees is used to identify flat surfaces such as water bodies (ERDAS, 2005).

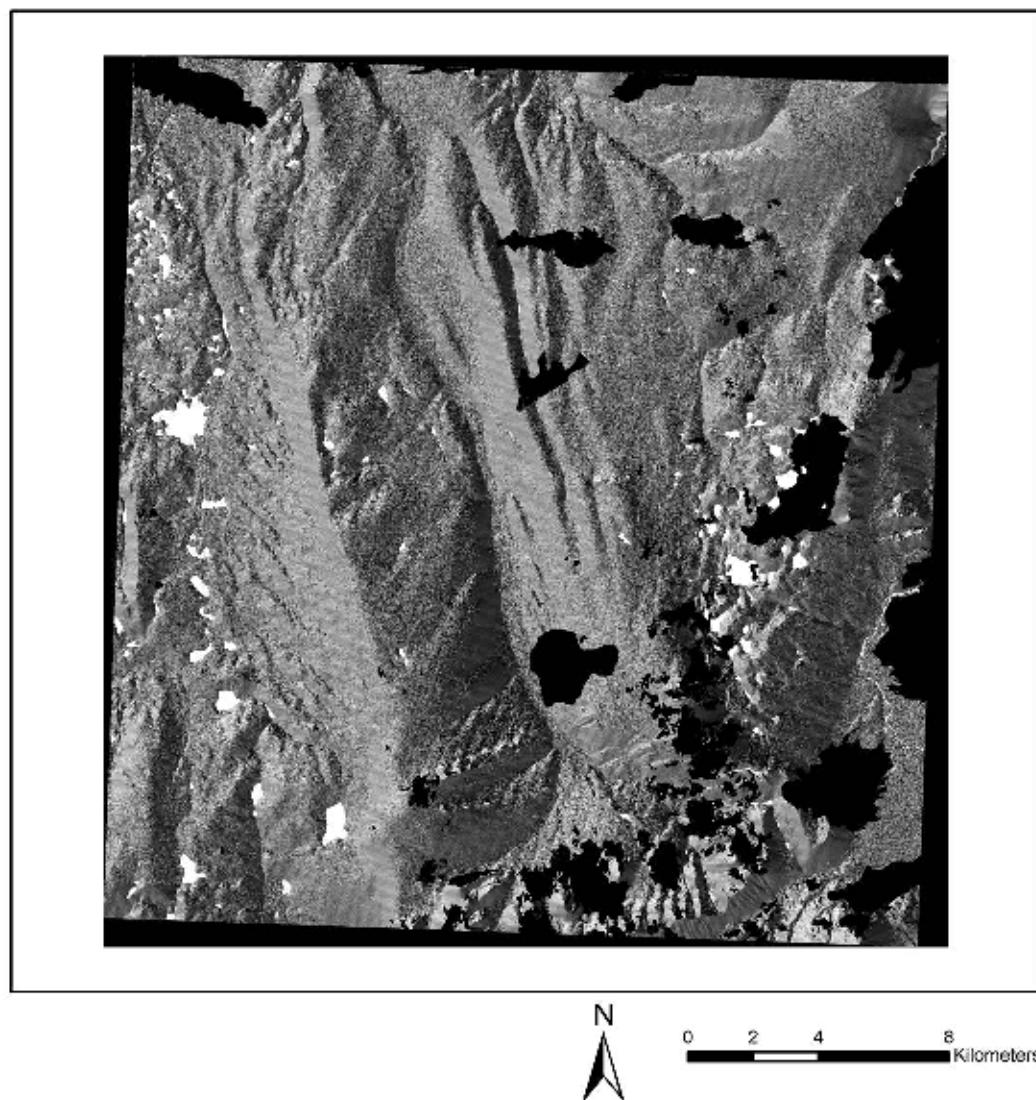


Figure 5: Aspect map for the image area (cloud and shadow pixels removed).

3.1.4. Texture

Texture is a simple contextual measure that may be extracted from an $n \times n$ window and incorporated in the classification process. When a small area of the image (e.g. 3×3 pixel area) has wide variation of discrete tonal features, the dominant property of that area is texture. Among the several different approaches, variance, a first order statistics in the spatial domain is chosen to make the texture layers for each of the three spectral bands and the NDVI layer (Figures 6 - 9).

$$\text{Variance} = \frac{\sum (BV_{ik} - \text{mean})^2}{n}$$

Where BV_{ik} is the brightness value of a pixel at i^{th} row and k^{th} column of the raster value table and mean is the overall mean of brightness value and n is the number of pixels for the particular band under consideration (Jensen, 2005).

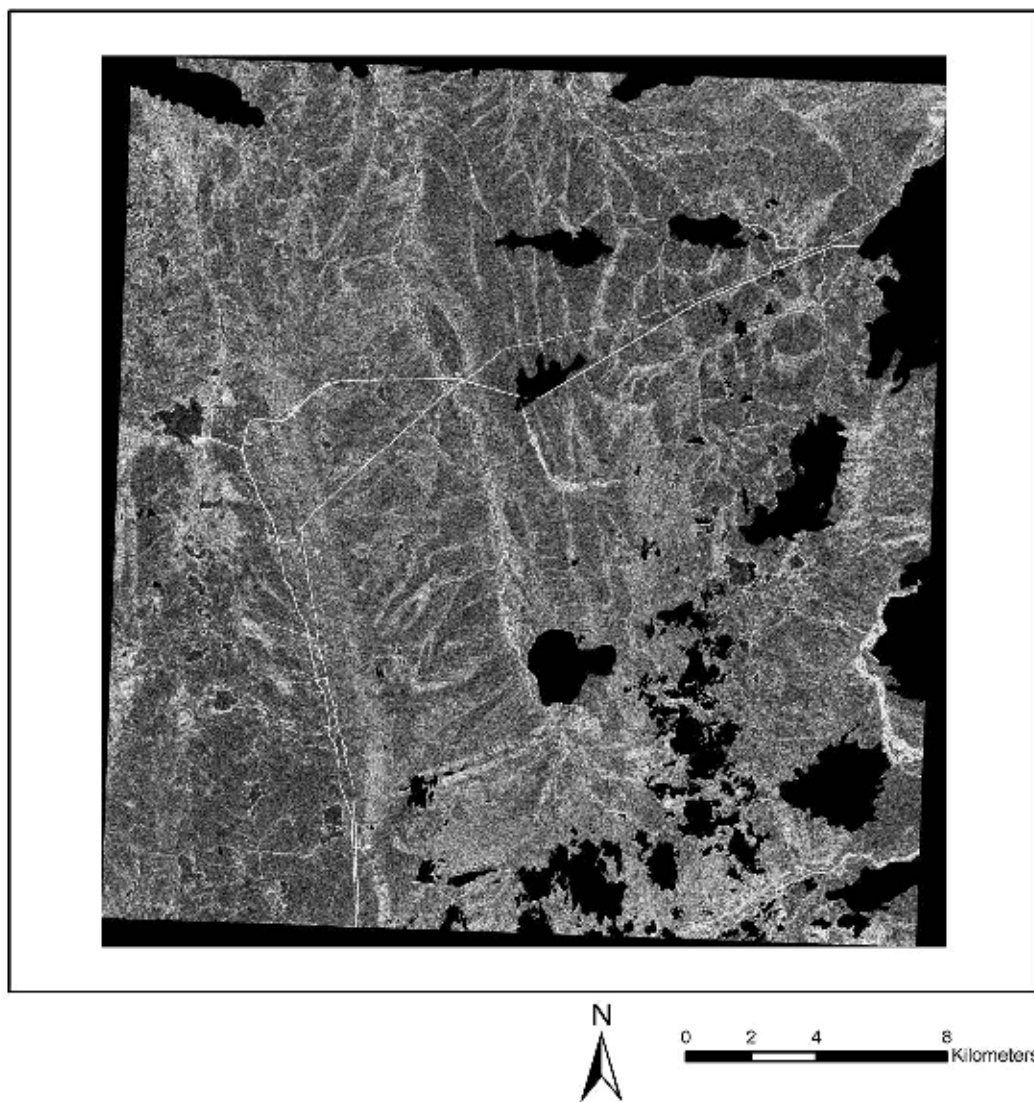


Figure 6: Band 1 texture (cloud and shadow pixels removed).

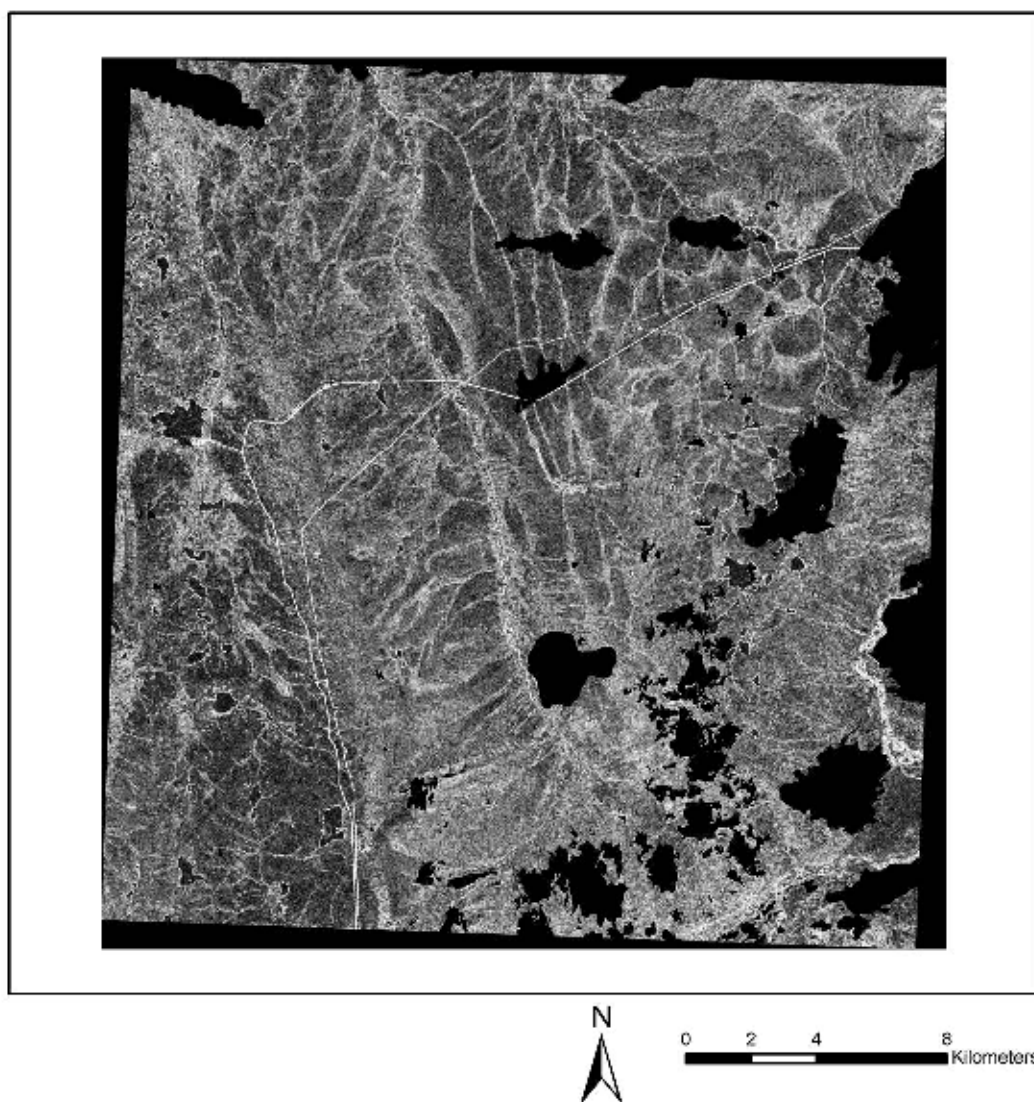


Figure 7: Band 2 texture (cloud and shadow pixels removed).

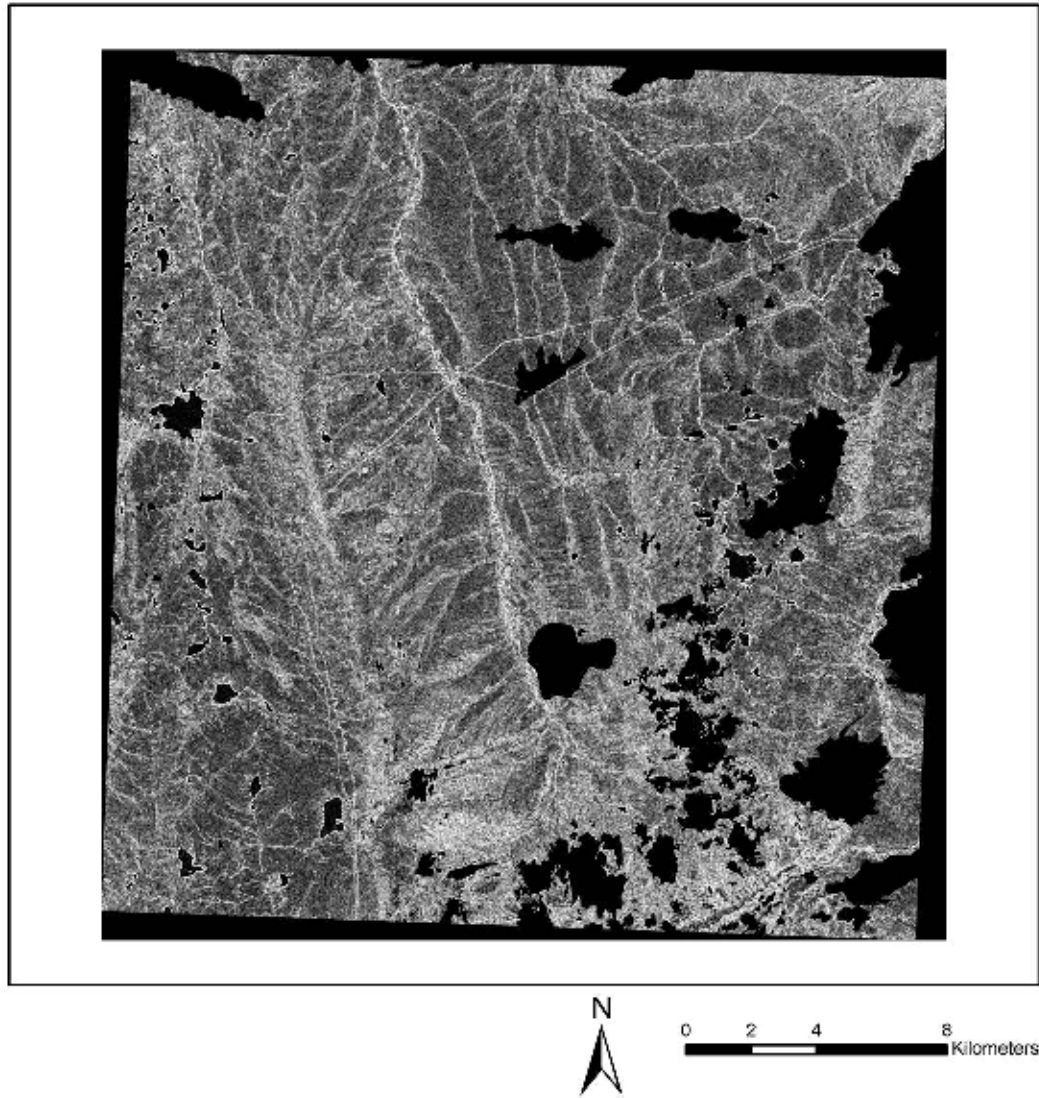


Figure 8: Band 3 texture (cloud and shadow pixels removed).

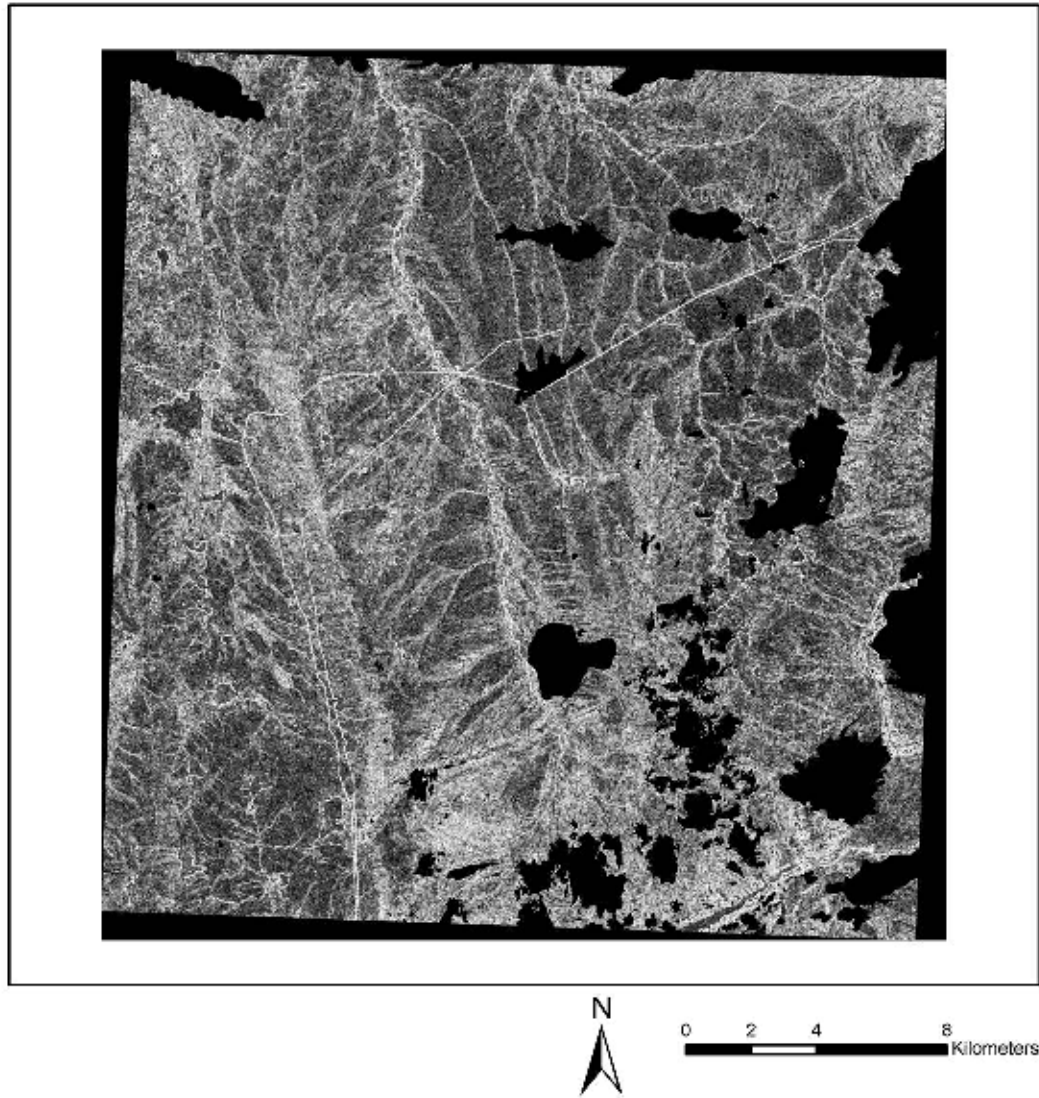


Figure 9: NDVI texture (cloud and shadow pixels removed).

3.2. Software Used

The different software used in this research are: ArcGIS 9.2, ERDAS Imagine 9.1, IDRISI Andes, GPS Pathfinder Office 3.1

3.3. Geometric Correction

Remotely sensed image data are representations of the irregular surface of the Earth and have both systematic and non-systematic geometric errors. Systematic errors are well documented, and are primarily related to the sensor functionality e.g. scan skew, panoramic distortion, platform velocity, earth rotation etc. while non-systematic errors are caused by the position and attitude angles of the satellite platform (Jenson, 2005). Non-systematic errors can be corrected in image-to-map rectification or image-to-image registration with the use of Ground Control Points (GCP). Geometric correction is applied to remove the geometric distortion so that the individual pixels of the corrected imagery will have the correct positions (x, y) on a planimetric map. The GPS locations of the training sites collected in summer 2006 showed that the SPOT-5 image used here is no exception and did have some non-systematic errors in the form of distortions or shifts (Figure 10). Hence there was a need for geometric correction that would minimize the positional errors in the image pixels in order to use the training data as well as to allow meaningful incorporation of accuracy assessment.

3.3.1. Collection of Ground Control Points

GCPs were collected in summer of 2007 in order to carry out the geometric correction of the image in the study area. The SPOT imagery of the field site was used for preliminary on site assessment of the ability to choose and interpret the position of each GCP on the image and surrounding ground features. Collection of GCPs was done using a Trimble Geo-Explorer CE GPS unit (using WGS 84 ellipsoid and the Universal Transverse Mercator (UTM)) projection system which provided a spatial accuracy of less than 1 meter. A minimum of 5 coordinates for each point were logged to achieve a confident level of accuracy (low Position Dilution of Precision (PDOP)) which depends on factors such as atmospheric conditions, satellite coverage, line of sight and geometry of the visible satellites. Several points, lines and polygons (Figure 10) were collected from ground features that were available as ground references for check points having large spectral difference or a target with high contrast, mostly in the form of man made features such as road intersections in the Dalton Highway, airstrip runways, and the Alaskan pipeline. More than 70 GCPs were collected.

This is worth mentioning here that the SPOT scene used in this region lacks enough ground references that can be used for collecting GCPs because of obvious less anthropogenic developments. This was a limitation for collecting enough GCPs which lead to a non-uniform geometry or spatial distribution of the

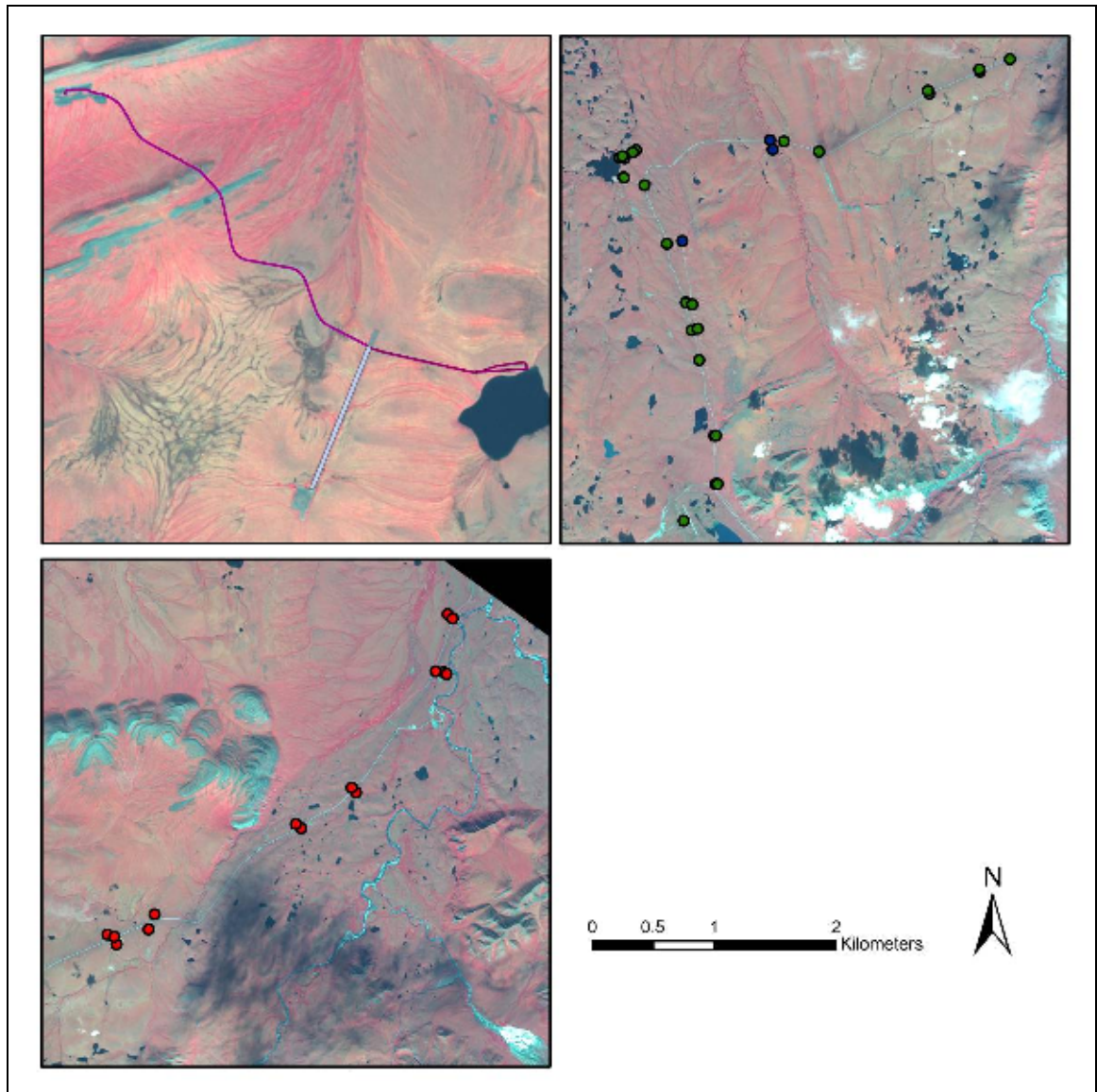


Figure 10: Collection of ground control points in the study area

GCPs used. Most of the GCPs were clustered along the Dalton Highway and around Lake Toolik in the central part of the image while the northwest and southern corners lacked GCPs. Also several GCPs had to be discarded due to high RMS errors (greater than 12m) to keep the total Root Mean Square (RMS) error below 5m (1 pixel).

RMS error is given by:

$$RMSE = \sqrt{((X_i - X_j)^2 + (Y_i - Y_j)^2)}$$

where i, j are the position of the point on the image and the ground respectively; X and Y are the easting and northing of the points.

3.3.2. Post-processed Differential Correction

Differential correction is used to increase the accuracy of collected GPS positions by reducing errors in GPS data by using a base station receiver whose position is accurately known. In postprocessed differential GPS, the base station receiver records the correction for each satellite directly to a file which is used and processed in the GPS processing software and the output is differentially corrected roving GPS data.

Toolik field station GPS base is a Trimble NetRS (survey grade) which supports mapping and survey grade corrections. Base data were available in the Toolik Field Station website

(http://www.uaf.edu/toolik/gis/TFS_GIS_gps_base.html) and was downloaded for postprocessed differential correction using GPS Pathfinder office software, Version 3.1. However, poor accuracy results due to high PDOP and excessively weak signals could not be improved with differential correction and were then manually discarded. A total of 64 GCPs were used in the image correction process.

3.3.3. Geometric Correction model

Geometric correction models are based on the empirical positional relationship between points on a satellite image and GCPs using conventional polynomials. Polynomial transformation approach applies separate single equations for x and y coordinates respectively across the whole image to adjust pixel locations. First order polynomial transformation allows translation, rotation, and scaling correction in both the x and y axes of an image. Higher order polynomials keep in consideration the correction of larger and non-linear distortions (Bannari, 1995).

A first order polynomial transformation (affine) was applied for the geometric correction. The equation for the transformation can be written as follows(Jensen, 2005):

$$x_{\text{predict}} = a_0 + a_1x + a_2y; y_{\text{predict}} = b_0 + b_1x + b_2y$$

where a_0 and b_0 are coefficients that control shifting (translation), a_1 and b_2 are coefficients that control scale changes, and a_2 and b_1 are coefficients that control

rotation (shear). A total of 64 collected GCPs were used in the correction in order to get a total RMS error of 4.99969 meters which is less than a pixel (5m). Ideally an RMS error of 0.5 pixels is preferred but due to lack enough GCPs with uniform spatial distribution as discussed in the previous section this was taken as the acceptable accuracy for geometric correction. A Nearest Neighborhood algorithm resampling was executed as it preserves the spectral integrity of the image pixels (Lillisand, 2000). Table 2 below shows the descriptive statistics for the X residuals ($\sqrt{(X_{\text{predicted}} - X_{\text{original}})^2}$) and Y residuals ($\sqrt{(Y_{\text{predicted}} - Y_{\text{original}})^2}$) and the RMS errors (RMSE).

Table 2: Statistics for residual and RMS errors

	X Residual	Y Residual	RMSE
Min	-8.12	-8.17	1.09
Max	10.24	10.32	11.68
Mean	0.00	0.00	4.18
S.D.	3.98	3.91	2.76

Figure 11 shows the comparison between the image before and after correction.

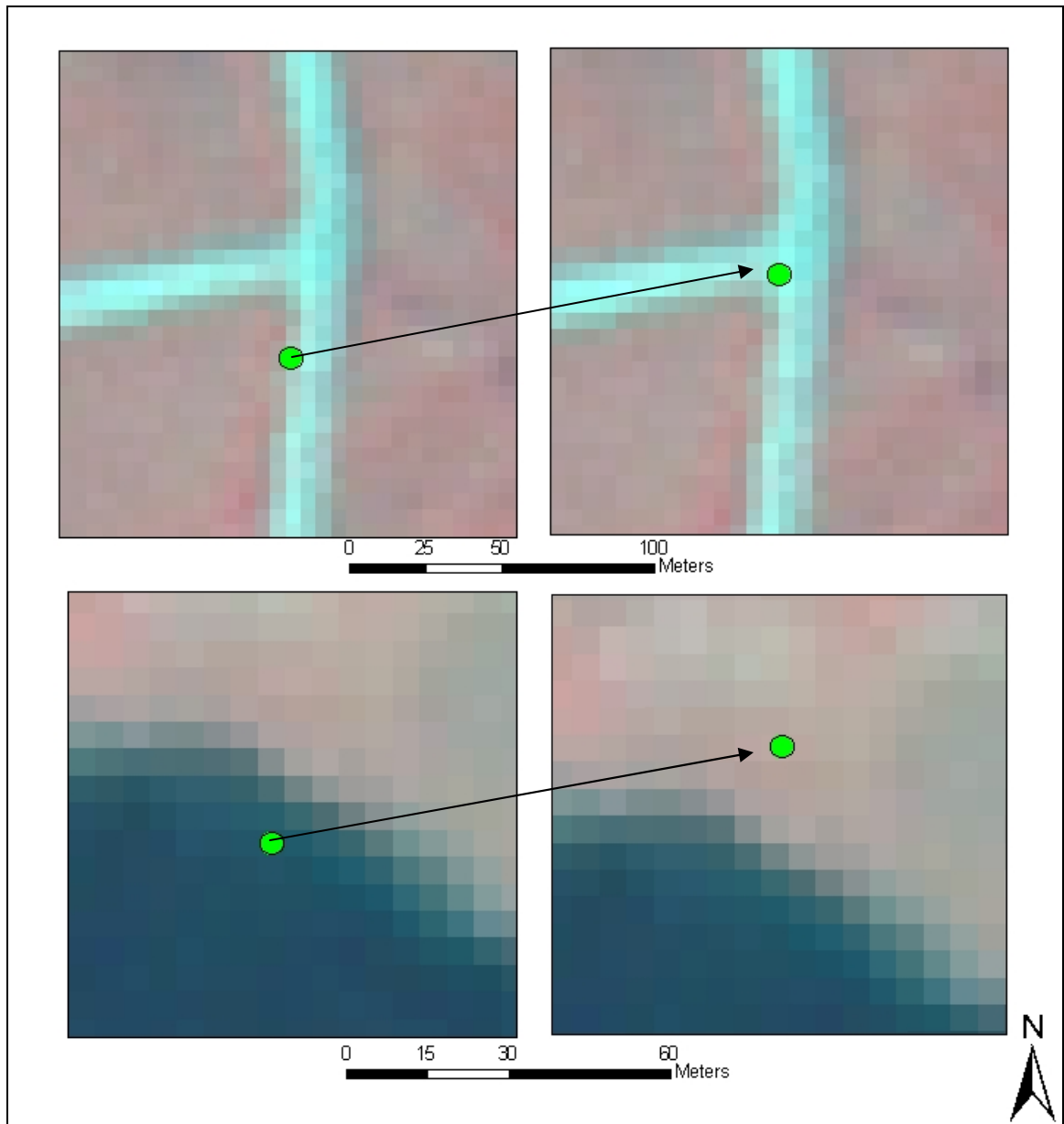


Figure 11: Example of geometric distortion of the SPOT-5 image.

Note: Image to the left shows location before rectification and to the right shows location after rectification of two points. Above is a road intersection and below a sample point which was shown in the lake before rectification.

3.4. Cloud and Shadow pixels removal

Clouds and shadow pixel contamination is a continuous problem for remote sensing studies. There were several cloud patches, their corresponding shadows on the ground, and shadows of the uplifted land patches due to the low sun angle (22 degrees). Cloud pixels in several cases had spectral similarity with barren, scarcely vegetated land-cover while shadow pixels had similarity with water pixels (Figure 12 - 13).

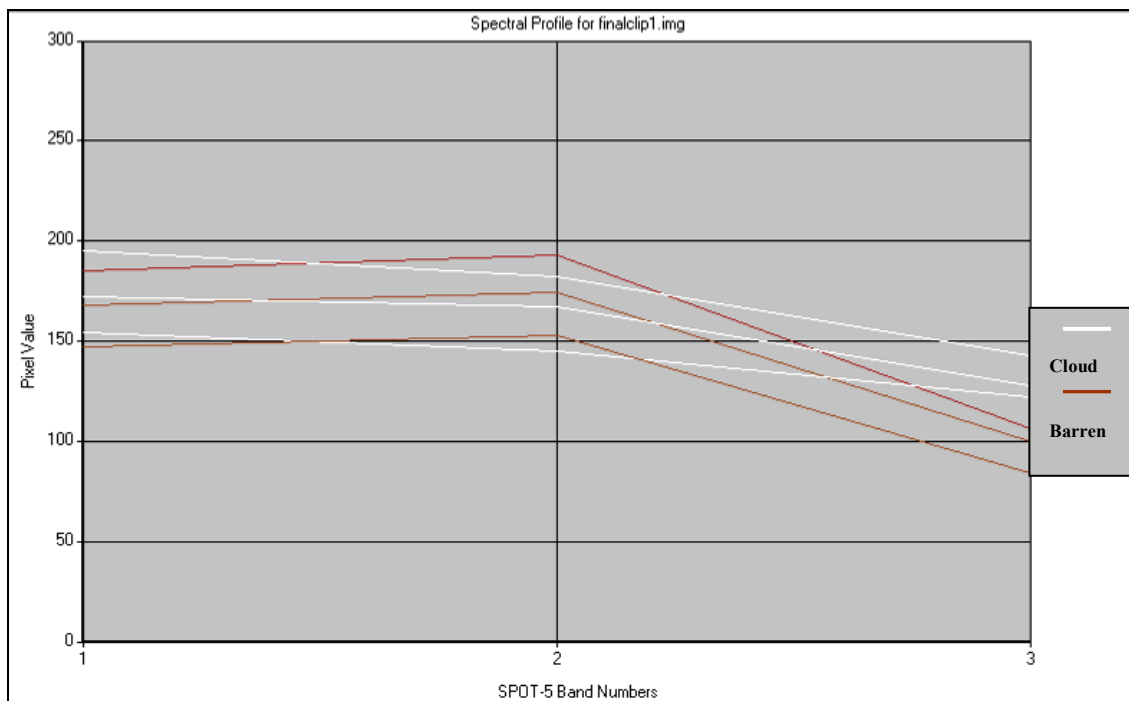


Figure 12: Spectral profile for barren (brown) and cloud (white) pixels showing the spectral similarity between the two.

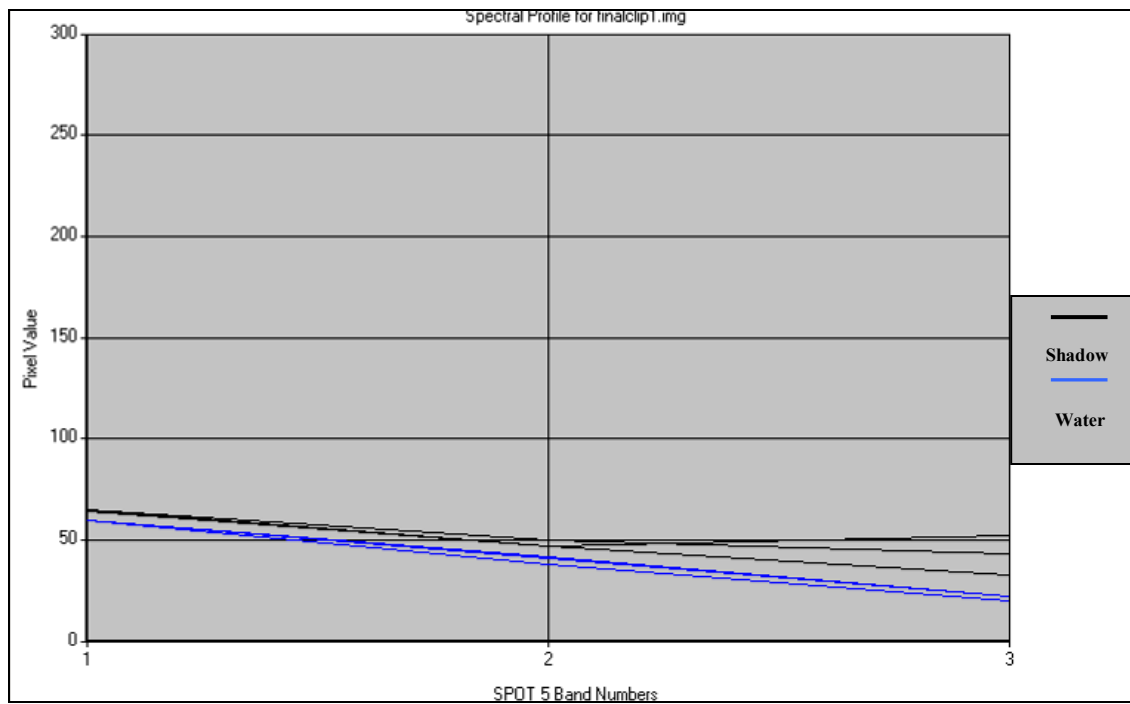


Figure 13: Spectral profile for water (blue) and shadow (black) pixels showing the spectral similarity between the two.

Since the clouds boundaries are diffused and so are the corresponding shadow boundaries it is difficult to detect all cloud pixels and shadow pixels and differentiate them from the similar barren pixels or water pixels respectively. Creation of the most approximate land surface reflectance from multi- temporal observations is a practice among remote sensing researchers to solve this problem (Holben, 1986). However due to the absence of a second SPOT image from a different date (temporal resolution) approximation of surface reflectance could not be achieved in this study. Instead, the cloud and shadow pixels were identified manually and then removed by masking. Then the DEM shadow map

(Figure 14) was prepared in ATCOR 3.0, an extension in ERDAS Imagine 9.1 in order to identify the shadow pixels formed by the landform and low sun angle. These shadow pixels were also removed by masking. This process of removing clouds and shadow pixels reduce the amount and spectral variations of pixels in the classification process, thus optimizing the satellite dataset to only include pixels of interest for obtaining the land-cover classes.

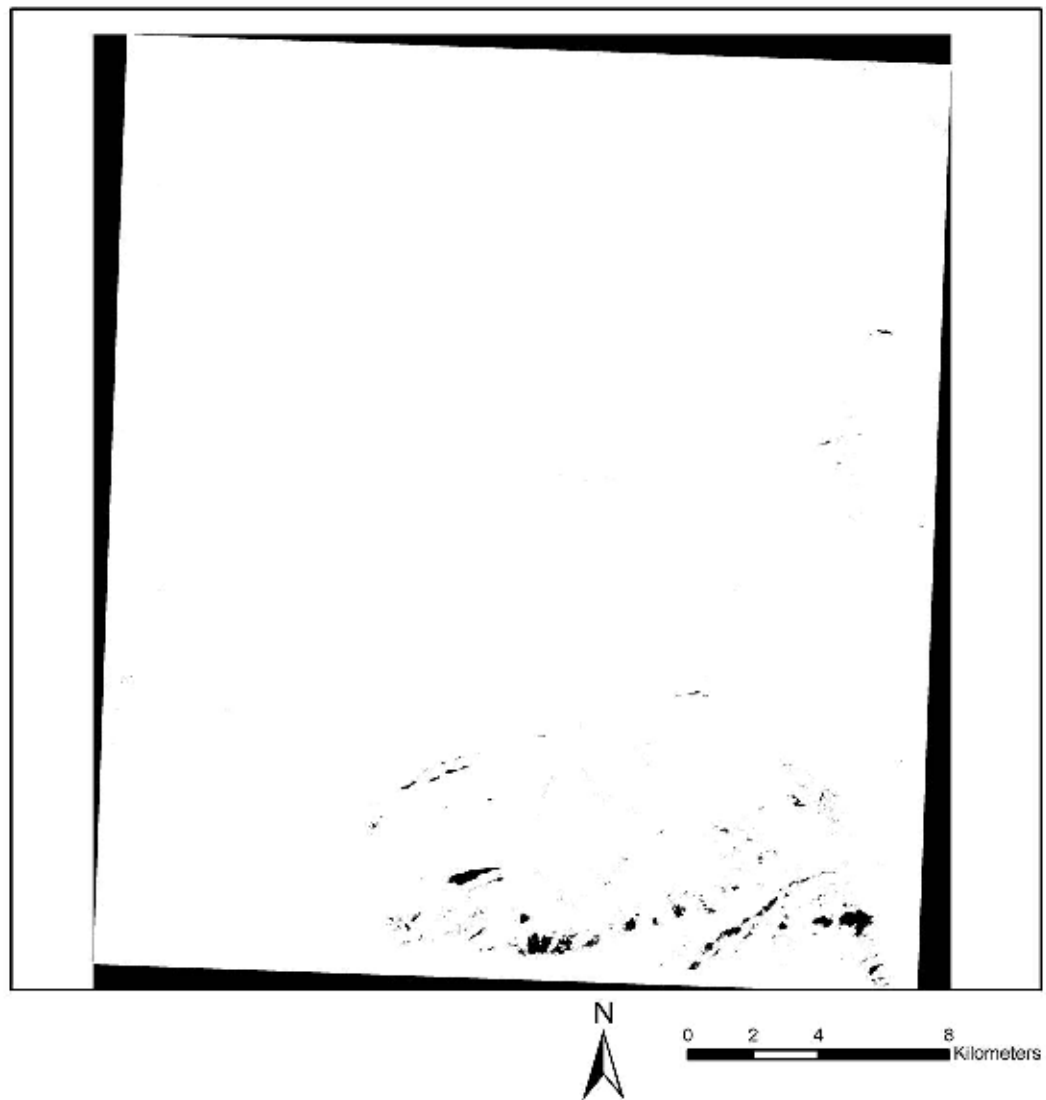


Figure 14: Figure showing the shadow cast (black) by the terrain due to the sun angle estimated by ATCOR in ERDAS Imagine 9.1

3.5. Classification scheme

Classes developed by Auerbach et al. (1997b), Muller et al. (1998) and Walker et al. (1994), were followed for the overall vegetation class guidelines following the Braun-Blanquet approach, which is a worldwide used standard

hierarchical system of vegetation classification based on plant-community floristics. (Westhoff et al. 1978). The data were classified into the following seven vegetation complexes:

1. Moist Low-Shrub Tundra and other Shrublands: Tussock tundra dominated by low shrubs (more than 50%). Willow dominated uplands areas dominated by dwarf and low shrubs mainly on interfluvial areas with well developed moss carpets. Common on lower hillslopes, in association with water-track complexes, and some floodplain areas. Areas dominated by willows along watertracks, streams and rivers (riparian shrubs) which includes willow communities in water tracks: *Eriophorum angustifolium*-*Salix pulchra*, *Salix alaxensis*-*Salix richardsoni*, *Salix glauca*-*Alnus crispa*, *Salix lanata*- *Betula nana*, *Salix pulchra*-*Calamagrostis canadensis*. (Appendix A, 1)

2. Water and aquatic complex: Marshes and aquatic vegetation with *Carex aquatilis* , *Hippuris vulgaris* with > than 50% standing water *Arctophila fulva*-*Eriophorum angustifolium*. (Appendix A, 2)

3. Barren complex: Roads, disturbed (anthropogenic) and re-vegetated gravel mines, construction pads, lichen-covered *Cetraria nigricans*-*Rhizocarpon geographicum*, and partially vegetated (<50%) exposed rocks in foothills and mountains, barren and partially vegetated river alluvium) (Appendix A, 3).

4. **Snowbed complex:** Generally, north facing areas with gentle slope where the snow cover stays longer than the adjoining areas, dominated by *Cassiope tetragona*, and other dwarf shrubs (*Ledum decumbens* and *Diapensia lapponica* in acidic sites and *Dryas integrifolia*, *Salix reticulata* and *Salix rotundifolia* in nonacidic sites) and fruticose lichens (*Cladina* spp. *Cetraria* spp., *Nephroma arctica*); *Carici microchaetae-Cassiopetum*, *Dryas integrifolia-Cassiope tetragona* (Appendix A, 4).

5. **Moist Dwarf-shrub, Tussock-Graminoid Tundra complex:** This the typical tussock tundra also known as moist acidic tundra (MAT) found in moist acidic hillslopes and moderately drained terrain with (pH<5.5) dominated by tussock-sedges, nontussock sedges, dwarf shrubs and mosses; *Sphagno-Eriophoretum vaginati* (AppendixA, 5).

6. **Wet Graminoid Tundra:** Rich fens on wetland areas with organic soils (pH > 4.5) dominated by sedges and mosses. Poor fens in wetland areas with organic soils (pH < 4.5) and dominated by sedges. Lawns of *Sphagnum* ssp. and sedges are common around the margins of basins of poor fens and some watertracks and foothills: *Carex chordorrhiza*, *Carex aquatilis*, *Carex aquatilis-Eriophorum angustifolium*, with *Carex. aquatilis-Carex chordorrhiza*, *Dryado integrifolia* - *Caricetum bigelowii* (Appendix A, 6).

7. Moist Graminoid, Prostrate-shrub Tundra complex:

- a. Typically also known as moist non-acidic tundra (MNT) found in moist nonacidic hillslopes and moderately well-drained surfaces (pH> 5.5) dominated by non-tussock sedges, prostrate and dwarf shrubs mosses; *Dryado integrifolia*-*Caricetum bigelowii* , *Astragalus umbellatus*-*Dryas integrifolia*.
- b. Dry nonacidic river terraces and gravelly well-drained slopes (pH> 5.5) dominated by *Dryas integrifolia* and other prostrate and dwarf shrubs, mat and cushion plants and lichens; *Oxytropis bryophila*-*Dryas integrifolia*
- c. Dry acidic tundra on hill crests, moraines and kames with (pH< 5.5), typically found on dry glacial tills and outwash deposits, steep south facing slopes and alpine areas on the mountains, dominated by prostrate and dwarf shrubs; *Astragalus umbellatus*-*Dryas integrifolia* -*Dryadetum octopetalae*, *Salici phlybophyllae*-*Arctoetum alpinae*, *Hierochloe alpina*-*Betula nana*, *Juncus biglumis*-*Saxifraga oppositifolia*.

The last two types (b, and c) are although compositionally different from MNT are spectrally similar to MNT and not separable by spectral means. Thus these two classes were merged into the class MNT as it was done by other researchers (Auerbach et al. 1997b; Muller et al. 1998; Walker et al. 1994). (Appendix A, 7)

3.6. Collection of sample data for training and accuracy (testing) assessment.

The SPOT image was first classified with an ISODATA clustering method into 60 classes. The classes were preliminarily identified with the help using the classified map by Walker et al. 1994 as a reference. Then, a stratified random sampling method was used to create more than 500 points on the SPOT image around Lake Toolik and in selected watersheds as required by the GTH researchers. This was done to optimize helicopter-hours cost, since most of the areas in the image are not accessible without a helicopter which in turn was extremely costly in terms of research budget. In summer 2006 and 2007, ground truth data were collected after visiting the created random points as well as additional training samples were collected by digitizing homogenous patches of pixels in the form of points, polygons and line, using Trimble Geo-Explorer CE GPS unit (Figure 15). Appendix B shows example of two field forms used to document the data collection. These collected data were corrected by post-processed differential correction and then used for making the training and testing data sets. Although, all the points created could not be visited due to shortage of helicopter hours and bad weather, a total of 349 points were visited and finally selected as representative pixels for the seven classes concerned. The pixels collected by random sampling were divided into two subsets, one of which was used for training (128 points) along with the field collected sample data, and the other for testing (221 points) the classification accuracy to avoid

any bias resulting from the use of the same set of pixels for both training and testing (Figure 16).

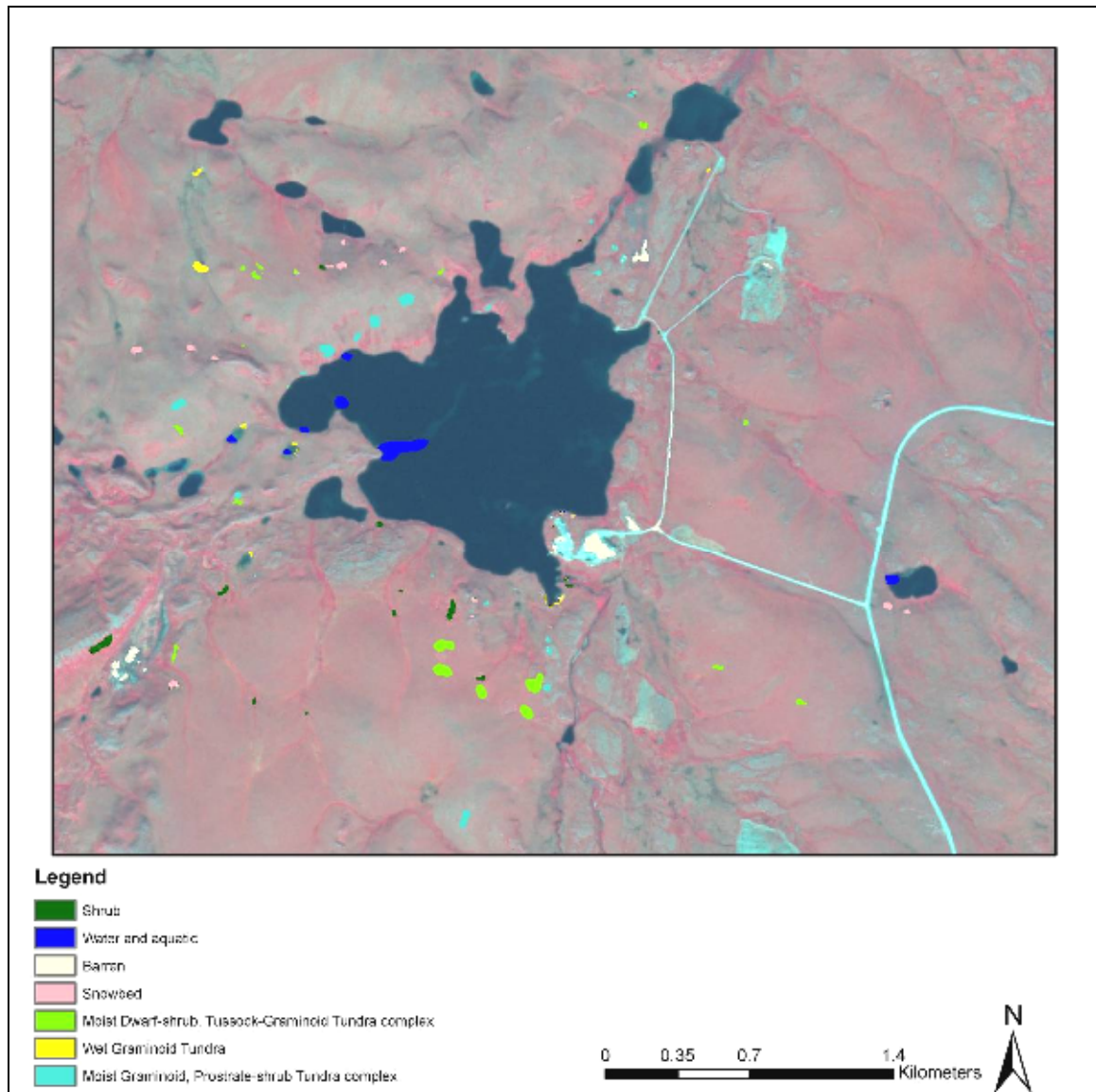


Figure 15: Training sites from field knowledge and sample sites collected in the form of lines and polygons.

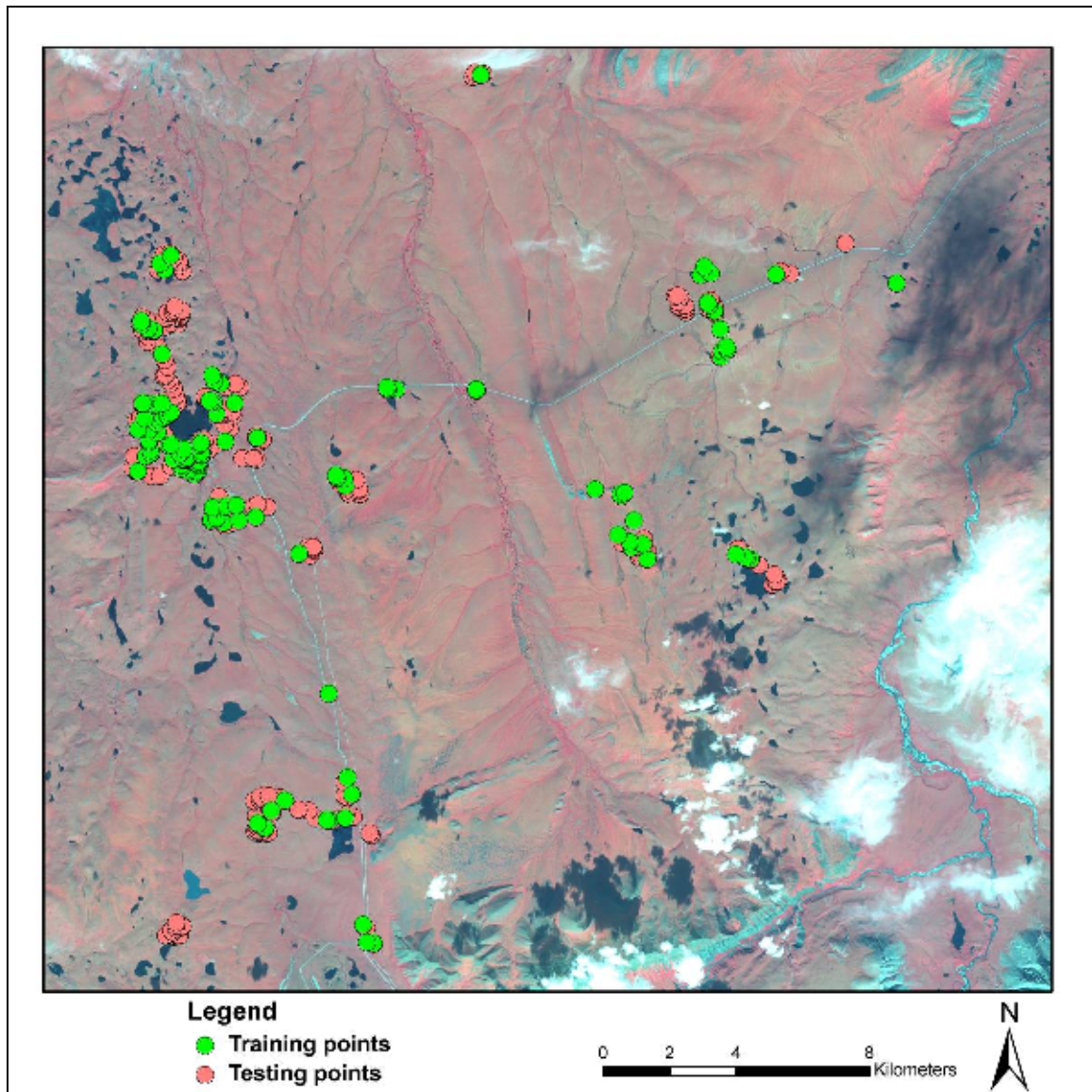


Figure 16: Overall distribution of training and testing (accuracy) sample points in the study area.

3.7. Unsupervised Classification (ISODATA clustering)

Unsupervised classification, also known as clustering, is a classification technique where the analyst needs no prior knowledge before performing the

classification. The computer arrange data into clusters by grouping similar spectral characteristics into unique clusters based on some statistically determined criteria (Jensen, 2005) which minimizes interclass similarity and maximizes intraclass similarity. Later on the clusters are relabeled and combined into information classes of interest.

Iterative Self-Organizing Data Analysis Technique (ISODATA) is one of the most common unsupervised classification techniques. It is iterative because it repeatedly performs the entire classification and recalculates statistics. It is "Self-Organizing" as it refers to the locating the clusters that are inherent in the data. The ISODATA clustering method uses the minimum spectral euclidean distance formula to form clusters and it begins with either arbitrary cluster means or means of an existing signature set, and each time the clustering repeats, the means of these clusters are shifted which are then used for the next iteration. The two stopping criterions that are set by the analyst are: 1) a maximum number of iterations have been performed; 2) a maximum percentage of unchanged pixels have been reached between two iterations (ERDAS, 2005).

The 3 bands of the SPOT image and the NDVI band (with clouds and shadow pixels removed) were stacked together to form a single image data for the unsupervised classification. First 60 unique clusters were formed using the four bands of the image data, each of which were assigned to one of the seven land-cover classes concerned using the training data set. The number of

iterations was set to 20 and the convergence threshold to 0.95. Pixels with zero values (no data, clouds or shadow) are excluded from the classification process. Each cluster was identified by using the maximum number of the training pixels belonging to the corresponding majority class. For few clusters that did not have any reference data, were assigned class values using the field knowledge of the area with careful inspection. Some clusters representing barren in the Brooks Range or representing shallow water were obvious but for other clusters spatially adjacent clusters were considered and help of areal photos, Landsat Image from 2000 and classified map by Walker et al. (1994) were considered before assigning the class values. Interestingly, for these clusters there were no testing data points (for accuracy) since these clusters represented relatively inaccessible areas in the far eastern and south-eastern corners of the study area. Finally all these clusters were grouped into the seven concerned land-cover classes and recoded in order to have the final image (Figure 17).

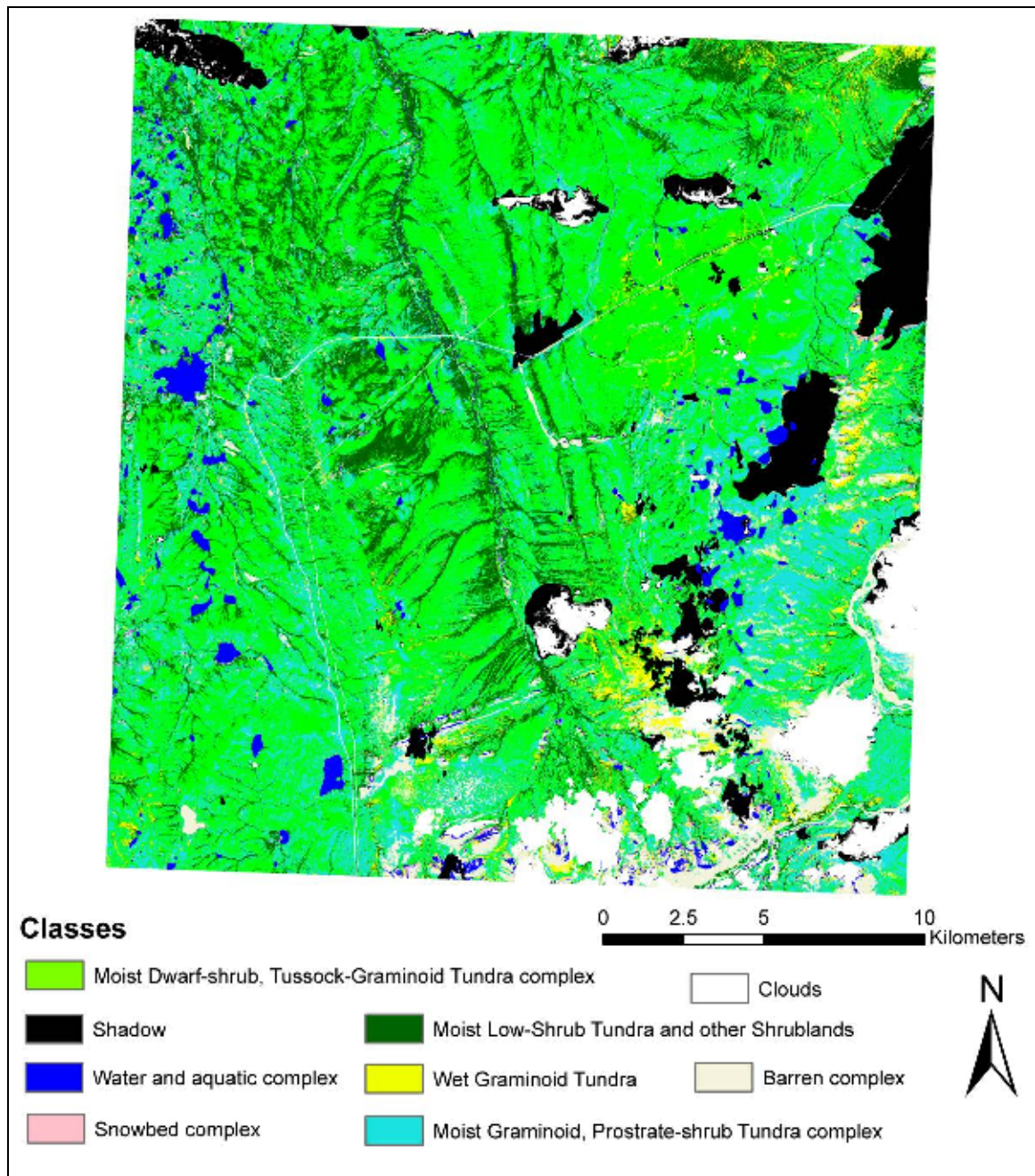


Figure 17: Classified Image for unsupervised ISOADATA clustering.

3.8. Classification (supervised) with Feature Analyst

Both traditional *unsupervised* and *supervised* spectral based approaches are routinely applied to remotely sensed data relying entirely upon the spectral information in an image, while neglecting the spatial arrangement of the pixels. Feature Analyst (a commercial software application by Visual Learning Systems, Inc. (VLS), of Missoula, Montana) utilizes multiple spatial attributes (size, shape, texture, pattern, spatial association) with spectral information to collect geospatial features and incorporates advanced inductive machine learning techniques to classify high resolution satellite imagery (O' Brien, 2003; Al-AbdulKader, et al. , 2002). Previous studies showed that Feature analyst works better than standard spectral based classification techniques in extracting urban features such as roads, houses etc from high resolution imagery (O'Brien, 2003; Jackson et al. 2005). Feature Analyst was used in this research to classify the SPOT image and indentify the typical arctic tundra type vegetation in Alaska and the results were compared to the unsupervised clustering and the proposed hybrid classification technique.

Feature Analyst is similar to a standard supervised classification in the sense that the analyst needs to provide training sites of each feature of interest which the software uses to find pixels in the image that are similar. After the first pass, in order to assist in refinement, Feature Analyst allows the user to define examples of "correct", "incorrect", and "missed" areas for each map produced

(Figure 18). These new examples are then used in the next pass to produce a new output which is, in most cases, more refined than the previous one. This process can be repeated until the analyst is satisfied to achieve the best results.

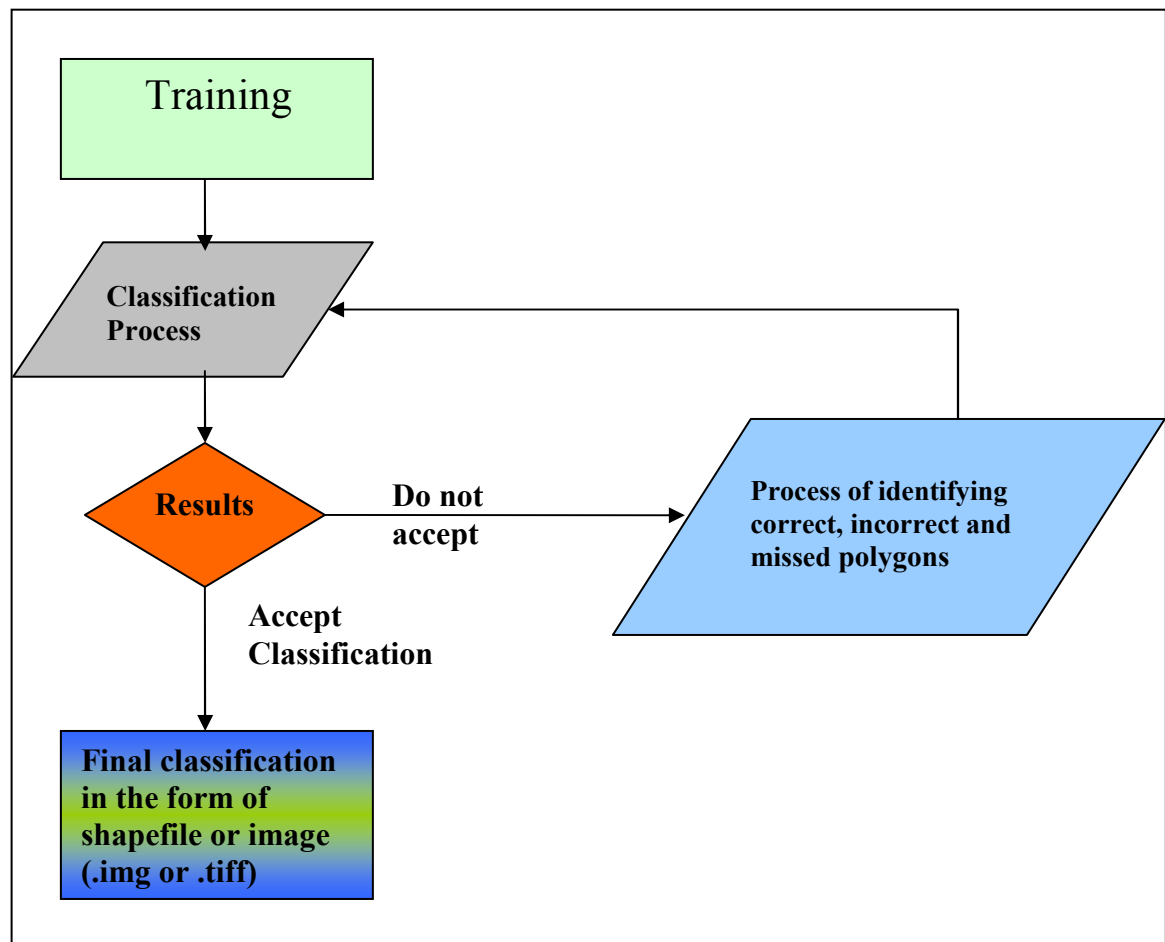


Figure 18: Showing the general work flow of the Feature Analyst

There are several steps involved in using Feature Analyst software. For the learner settings, appropriate input feature (land-cover) and spatial resolution (5m) was selected. One important characteristic of the Feature Analyst software

is the 'Input Representation'. Input Representation is the pattern that Feature Analyst uses to classify each pixel in the image to determine if it is part of the target feature. It is possible to uniquely define the shape of the area to be sampled for feature extraction via the 'Input Representation' input representation' as the learner not only looks for the single pixel but also surrounding pixels selected by the user in order to more accurately extract features. For this project the 'Foveal' representation (with pattern width 3) seems to give the most consistent results for extracting features in the lake (shoals or shallow aquatic) from all other reflective surfaces (Figure 19 a - c). However for the other classes the manhattan representation gave more or less similar results.

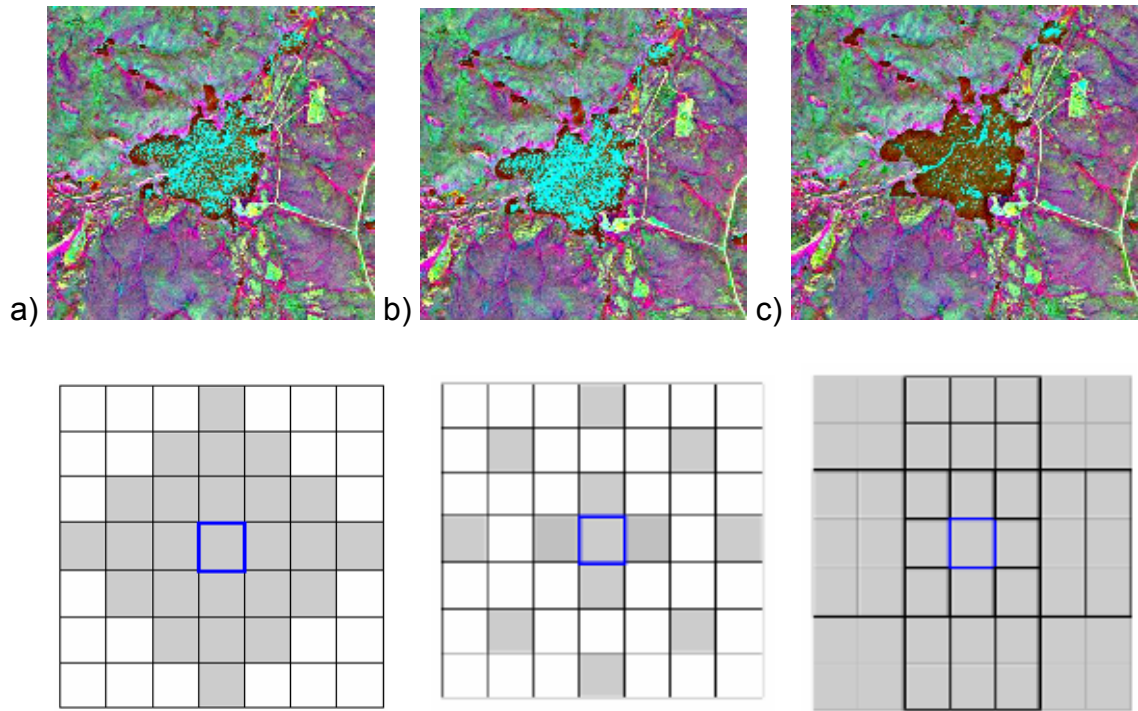


Figure 19: Comparing extraction of the features in the Toolik Lake with the different input representations.

Note: a) Manhattan, b) Bull's Eye 4, c) Foveal 3x3. The graphical representations of the three different input representations are given below each extraction.

The same image data consisting the three bands of the SPOT image, stacked with the NDVI as the fourth band was used for classification. Numerous polygons were digitized for each class represented in the area using the training data set. Careful attention was needed to digitizing features of interest (i.e. different vegetation classes) in a variety of locations and with a variety of spectral reflectance. Then a multilayered classification scheme was prepared and all the selected classes were used as input. To create a wall-to-wall classification, training sites for each feature need to be selected—multiple examples were

selected for each feature and then combined for use in classifying the whole image. After the first pass, each of the classes was split out for further clutter removals and corrections (Figure 20). Then after the corrections, they were combined to be used for the final supervised classification (Figure 21) and the desired output format was chosen as raster (.img) (Figure22).

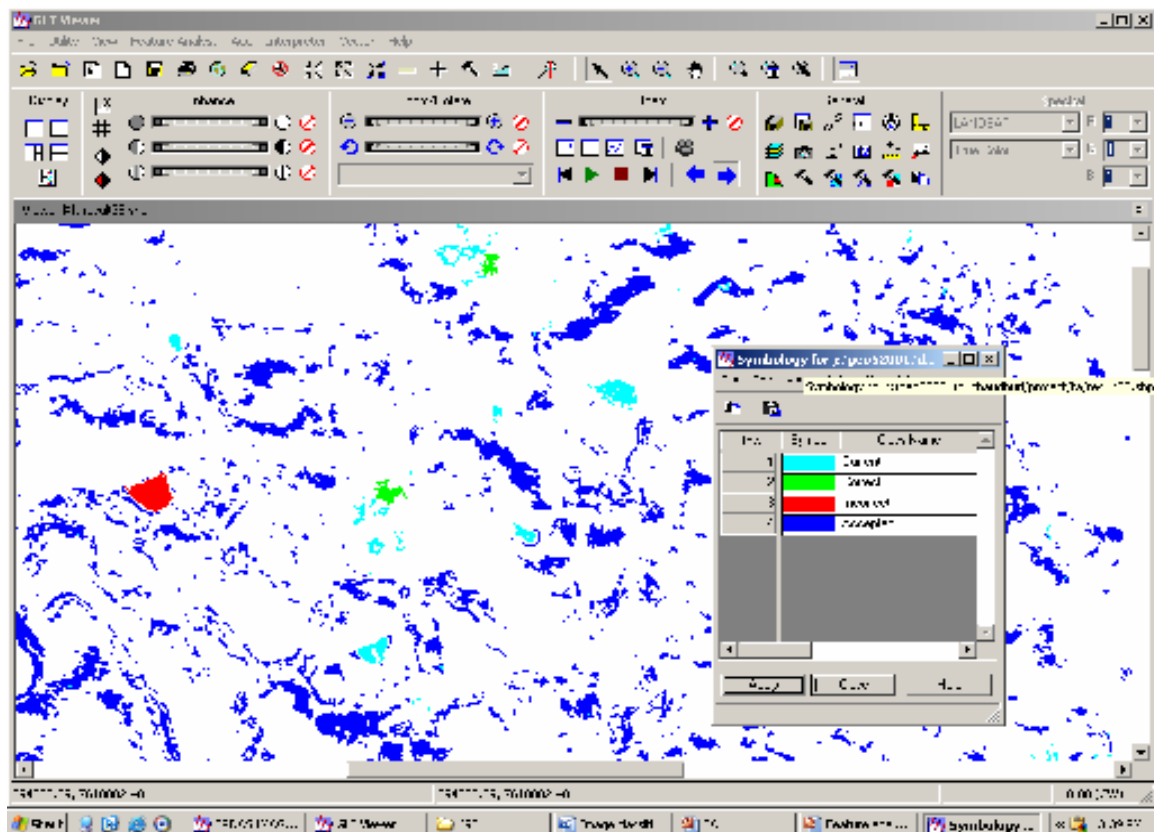


Figure 20: Figure showing the accepted, incorrect, correct and current signatures for snowbed complex.

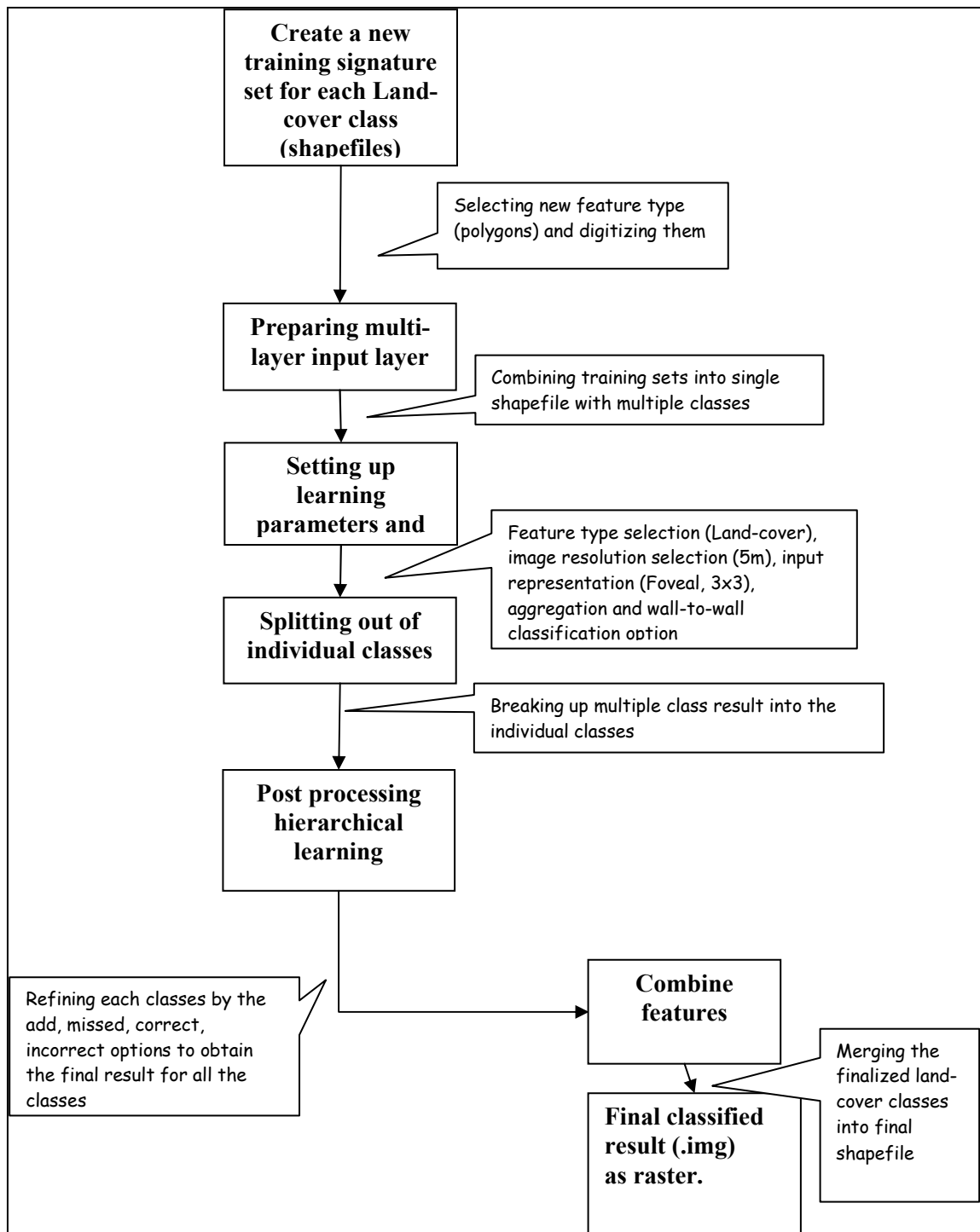


Figure 21: Flow chart showing the work flow for supervised classified image with Feature Analyst.

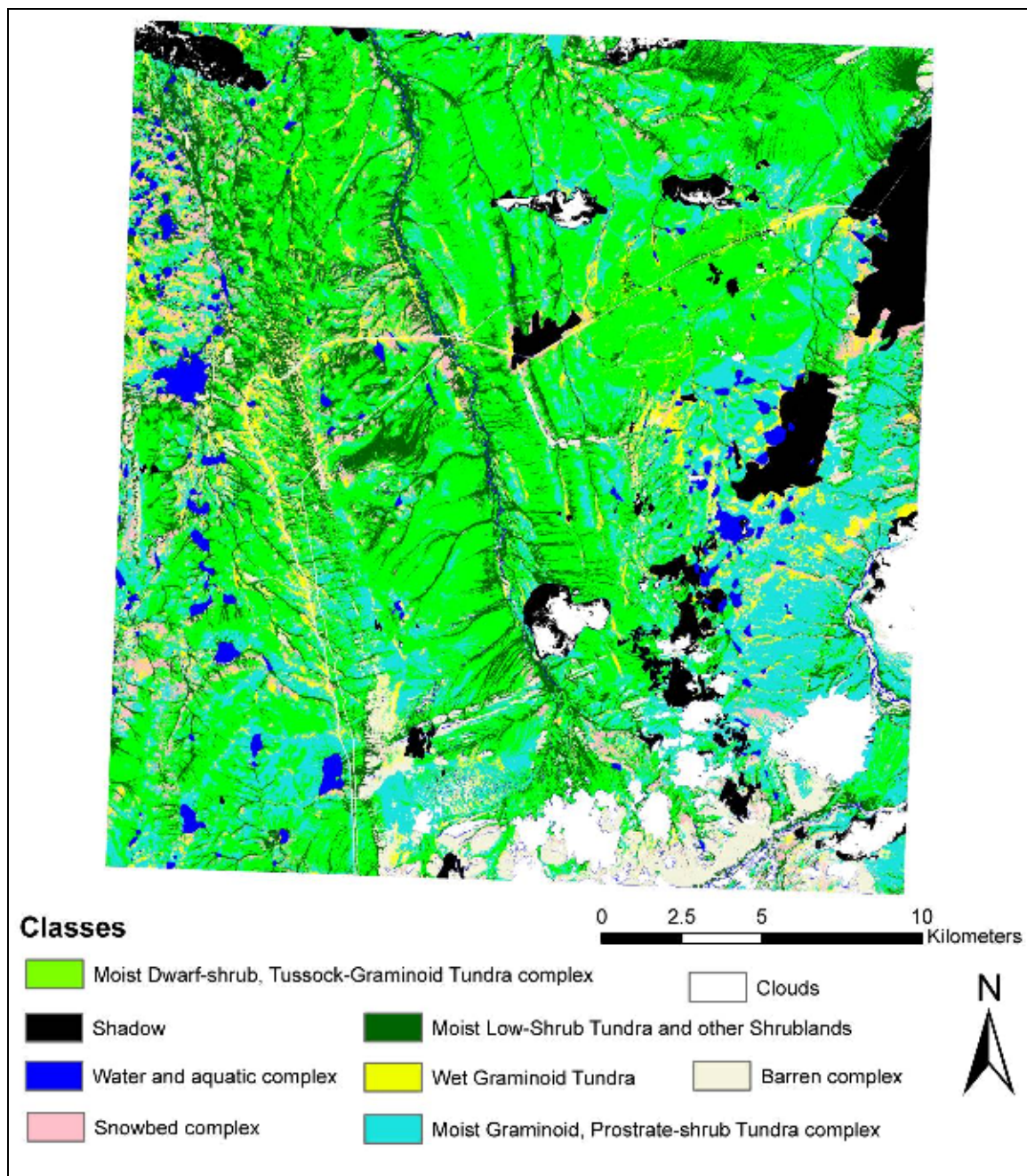


Figure 22: Classified Image (Supervised) with Feature Analyst.

3.9. Hybrid Classification

3.9.1. Hybrid Methodology

The methodology for the hybrid classification considered three different classifiers that are commonly used in remote sensing image classification i.e. knowledge based (expert) classification, unsupervised clustering (ISODATA), and a non- parametric classifier, multi-layer perceptron (MLP) that utilizes artificial neural network. The approach was to combine the three classifiers to classify the SPOT image data with ancillary geo-spatial data and form a multiple classifier system which combined the relative strengths from the different classifiers and applied them in a sequence in such a way that the overall accuracy was the maximized. First, a knowledge based classification was applied after gathering knowledge in the form of heirchical rule set. Then, the pixels that were not being classified by the rules were masked out and an ISODATA clustering was applied to classify those pixels and then they were merged with the previous classified pixels. A MLP classifier with two hidden layers was applied which took the three bands of SPOT image, NDVI, slope, aspect, and the output of the rule-based classifier as input layers. Finally those classes of each classifier having higher kappa values were merged together to get the final classified image (Figure 23).

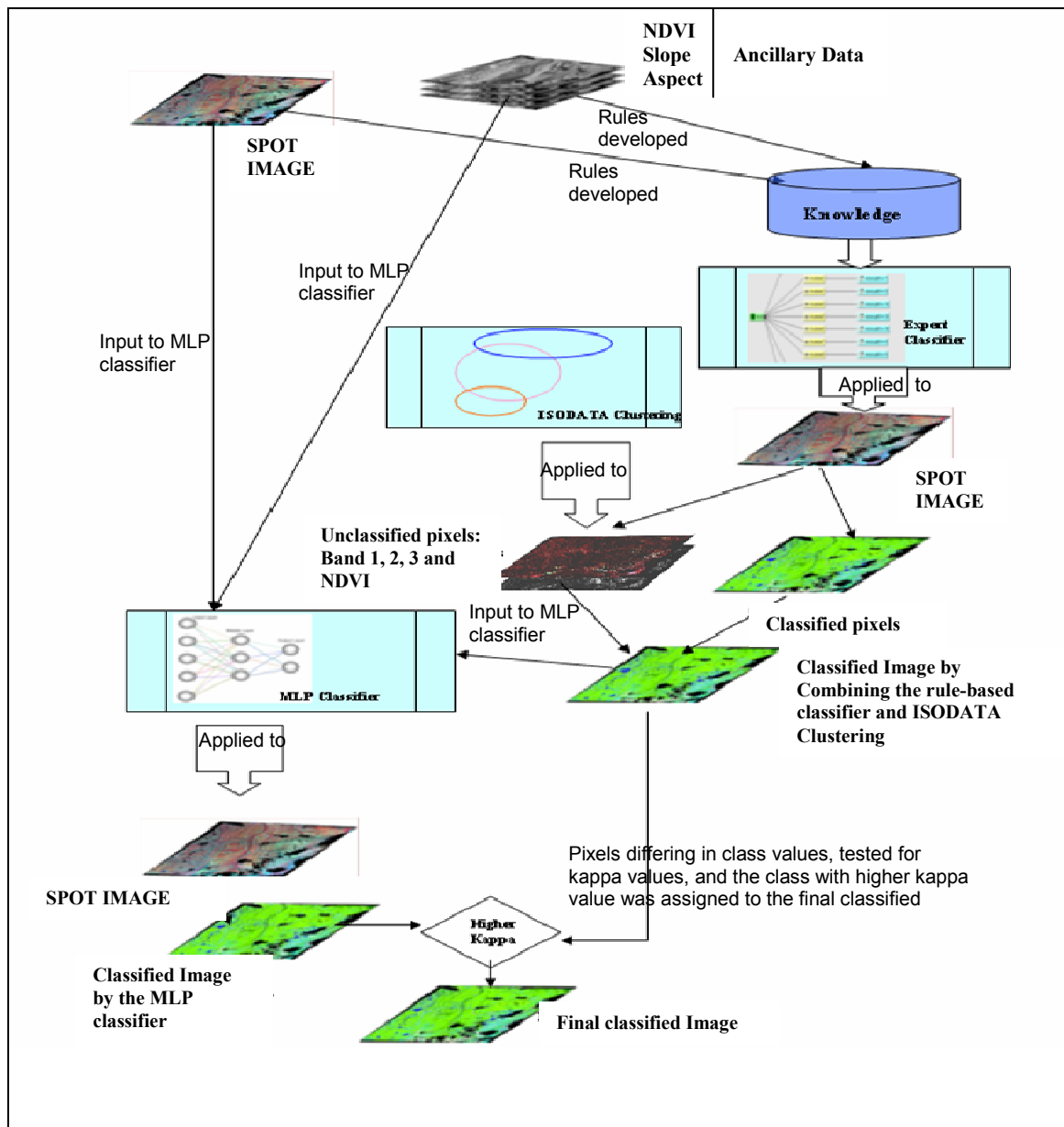


Figure 23: Flowchart for proposed hybrid classification methodology.

3.9.2. Knowledge Based (Expert) Classification

3.9.2.1. Expert Classifier in ERDAS Imagine

A rule can be defined as a list of conditional statements that determine the informational contents of a scientific hypothesis. The Expert Classifier in ERDAS Imagine implements multiple rules and hypothesis that are linked together into a hierarchy that describes a final set of target informational class. An Expert Classifier has two major components, the Knowledge Engineer and the Knowledge Classifier. The Knowledge Engineer (Figure 24) in the Expert Classifier provides a graphical user interface to build a knowledge base which is represented as a tree diagram consisting of final and intermediate class definitions (hypotheses), rules (conditional statements concerning variables), and variables (raster, vector, or scalar). The Knowledge Classifier provides an interface to implement the developed knowledge base in classifying an image.

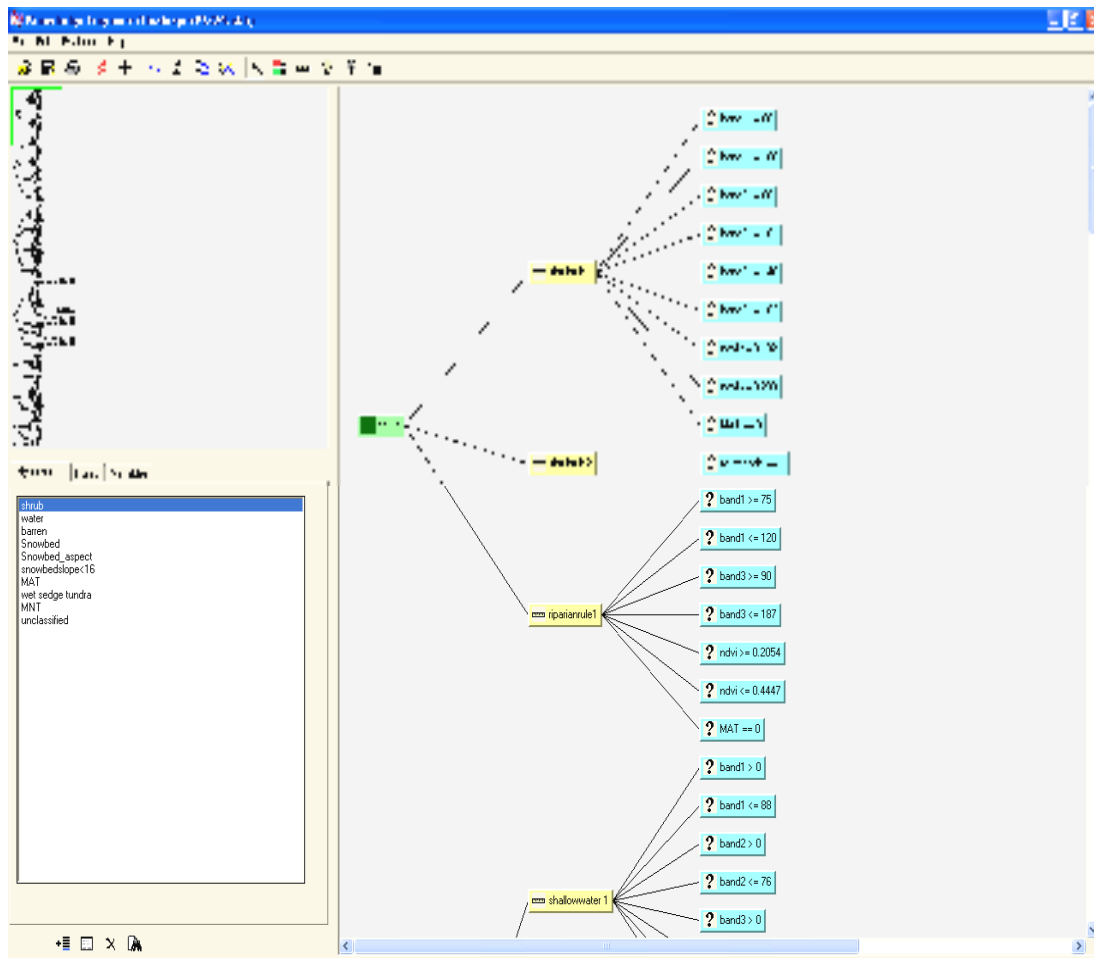


Figure 24: Snapshot of the Expert Classifier of the Expert Classifier in ERDAS Imagine.

3.9.2.2. Advantages and Disadvantages of Expert Classification

Knowledge may be defined as a deterministic collection of related and useful information. The set of rules formed in the knowledge engineering process are stored and turned into information which in turn is used as “knowledge” and thus can be reused for any similar data set for an unseen geographical region. A knowledge based (expert system) classifier is very useful in classifying high-

resolution imagery since traditional classifiers (such as Maximum Likelihood supervised classification or ISODATA unsupervised classification) cannot incorporate spatial association among the pixels of interest. Uncertainties is handled by placing confidence in each rule and as multiple rules are triggered within a tree, the Knowledge Classifier combines the confidences and when several rules are true at a particular pixel – the rule with the highest confidence is assigned to be the class for that pixel (ERDAS, 2005).

Acquisition of knowledge is a biggest disadvantage for knowledge based classifier. A vivid knowledge about the study area and the class composition is required for the analyst before assigning the rules for the individual classes in the Knowledge Engineer. The second problem is known as the “knowledge acquisition bottleneck” which refers to the problem of inefficient formulating of the gathered knowledge in a systematic, correct, and completely usable format for quantitative analysis (Huang et al 1997).

3.9.2.3. Preliminary Classification Rules

Three bands of the SPOT image, and the ancillary data consisting of NDVI, slope, and aspect were used for the process of making rules.

Representative sites for each class from the training data set were selected on the image. Hypothesis for each class was derived based on spectral properties of these sites and the related secondary data. Output of each class was saved and then whenever necessary was used as a constraint in deriving the final output of

the informational classes. A buffer of 2 pixels were created around water class, and was called shallow fen since in most obvious cases these pixels represented shallow fen type vegetation and those pixels were assigned to the Wet Graminoid Tundra complex. Classes Snowbed complex and Wet Graminoid Tundra complex had severe interclass spectral mixing and had to be masked out. The masked out pixels were stacked with NDVI and the result was classified into 8 clusters, in a file named snowbed-WST. These clusters were identified and used in the rules. There were total 7 rules formed as follows:

Rule for Moist Low-Shrub Tundra and other Shrublands:

(IF Band1 >= 86 AND Band1 <= 106, Band2 >= 86 AND Band2<= 101, Band3>= 140 AND Band3<= 163, NDVI >= 0.182 AND NDVI <= 0.293, NOT in Moist Dwarf-shrub, Tussock-Graminoid Tundra complex)

OR

(IF Band1 >= 75 AND Band1 <= 120, Band3>= 90 AND Band3<= 187, NDVI >= 0.205 AND NDVI <= 0.4447, NOT IN class Moist Dwarf-shrub, Tussock-Graminoid Tundra complex)

→ Moist Low-Shrub Tundra and other Shrublands complex

Rule for water and aquatic complex:

(IF Band1 > 0 AND Band1 <= 88, Band2 > 0 AND Band2 <= 76, Band3 > 0 AND Band3 <= 82, NDVI >= -0.111 AND NDVI <= 0.101)

OR

(IF Band1 > 0 AND Band1 <= 88, Band2 > 0 AND Band2 <= 76, Band3 > 0 AND Band3 <= 82, NDVI <= -0.314902)

OR

(IF Band1 >= 97 AND Band1 <= 127, Band2 >= 75 AND Band2 <= 106, Band3 >= 22 AND Band3 <= 50, NDVI <= -0.0314902)

OR

(IF Band1 < 88, Band2 <= 95, Band3 >= 18 AND Band3 <= 64, NDVI <= -0.111, NOT IN Mountainshadow region)

→ Water and aquatic complex.

Rule for Barren complex:

(IF Band1 >= 105, Band2 >= 90, Band3 >= 45 and Band 3 <= 145, NDVI >= -0.349087 AND NDVI <= 0.0725263, NOT IN shallowwater, NOT IN class Moist Graminoid, Prostrate-shrub Tundra complex)

OR

(IF NDVI > -0.328 AND NDVI < -0.033826, DEM > 1034, NOT IN water and aquatic complex)

OR

(IF Band1 >= 90 AND Band1 <= 111, Band2>= 80, Band3 >= 45 and Band 3 <= 161, NDVI >= -0.349087 AND NDVI <= 0.0725263, NOT IN shallowwater, DEM >= 947.569 AND DEM <= 1024.41, IN lowelevation region)

OR

(IF Band1 >= 88, Band2> 75, Band3 >= 54, NDVI <= -0.025, NOT IN water and aquatic complex)

→ Barren complex

Rule for Snowbed complex:

(If Band1 >= 85 AND Band1 <= 97, Band2 >= 72 AND Band2 <= 88, Band3 >= 78 AND Band 3 <= 100, NDVI <= 0.064 AND NDVI <= 0.101, (aspect > 0 AND aspect < 90) or (aspect > 270 AND aspect < 360), slope < 16, NOT IN water and aquatic complex, NOT IN shallow fen, Slope < 16)

OR

(snowbedWST = 4, (aspect > 0 AND aspect < 90) or (aspect > 270 AND aspect < 360), slope < 16, NOT IN water and aquatic complex, NOT IN shallow fen, Slope < 16)

OR

(snowbedWST = 5, (aspect > 0 AND aspect < 90) or (aspect > 270 AND aspect < 360), slope < 16, NOT IN water and aquatic complex, NOT IN shallow fen, Slope < 16)

→ Snowbed complex.

Rule for Moist Dwarf-shrub, Tussock-Graminoid Tundra complex:

(If Band1 >= 85 AND Band1 <=100, Band2 >=77 AND Band2<=97, Band3 >=107 and Band3 <=147, NDVI >= 0.077 AND NDVI <=0.244, NOT IN Moist Graminoid, Prostrate-shrub Tundra complex)

→ Moist Dwarf-shrub, Tussock-Graminoid Tundra complex

Rule for Wet Graminoid Tundra

(If Band1 >= 74 AND Band1 <= 97, Band2 >= 62 AND Band2 <= 99, Band3 >= 60 AND Band3 <= 114, NDVI >= -0.105 AND NDVI <= 0.062, NOT IN water and aquatic complex, NOT IN snowbed complex)

→ Wet Graminoid Tundra

Rule for Moist Graminoid, Prostrate-shrub Tundra complex

(If Band1 >= 89 AND Band1 <= 114, Band2 >= 89 AND Band2 <= 106, Band3 >= 96 AND Band3 <= 126, NDVI >= 0.004 AND NDVI <= 0.126)

→ Moist Graminoid, Prostrate-shrub Tundra complex

The rules were run all together from the knowledge classifier and it was found that the rules could not exhaustively classify all the pixels in the image and the 19.2% of the pixels remained unclassified. The pixels that were classified

were separated from those that were unclassified (Figure 25) and a partially classified map was formed (Figure 26).

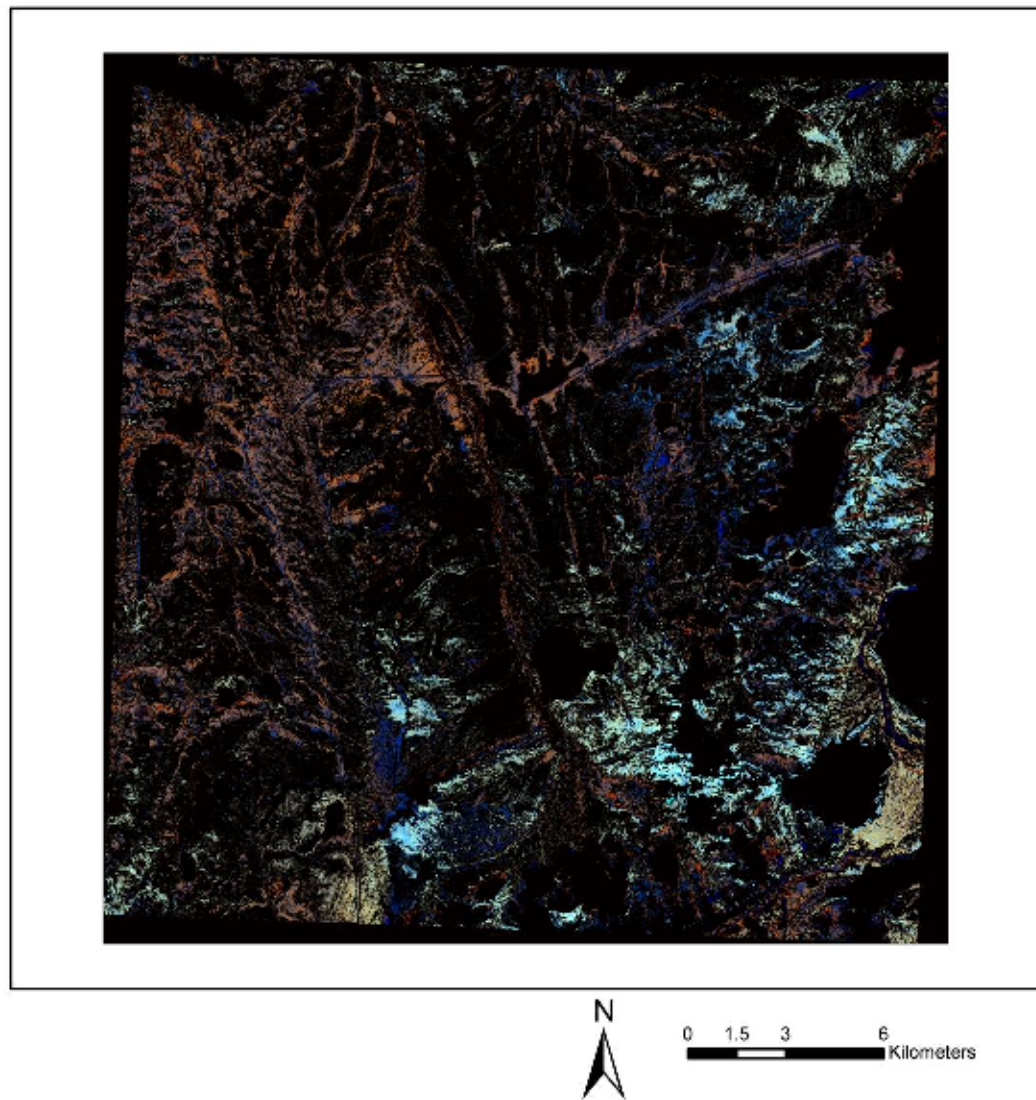


Figure 25: Pixels unclassified from the Expert classifier.

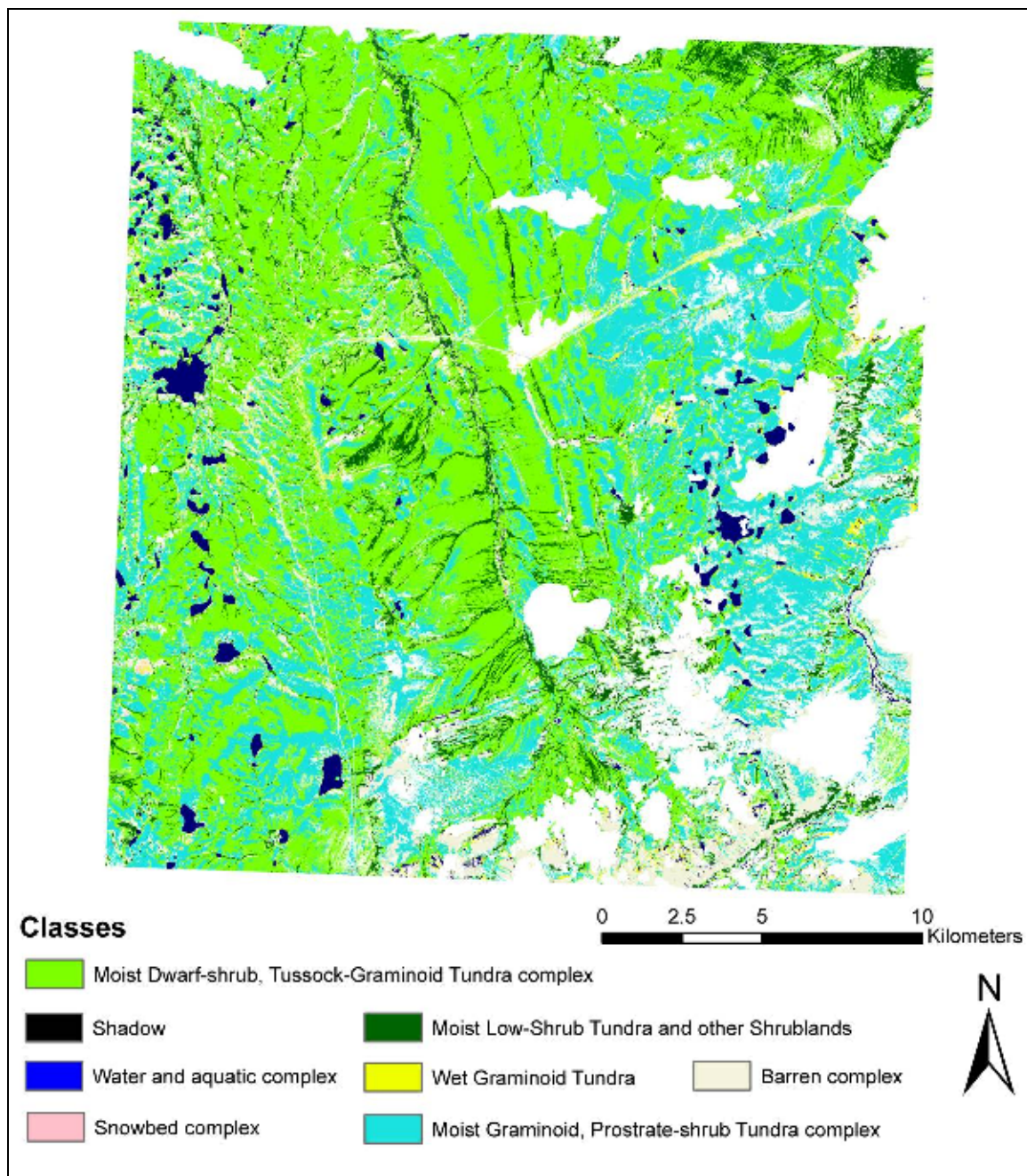


Figure 26: Partially classified image by Expert classifier.

3.9.3. ISODATA clustering of the remaining pixels

Interestingly, the pixels that were not classified comprised mostly of the same areas in that were left out in the ISODATA clustering as clusters that lacked reference data located relatively in accessible areas in the far eastern and south-eastern corners of the study area separately using unsupervised ISODATA clustering method. The remaining pixels were layer stacked with the NDVI band, added as a fourth band and then they were classified into 60 spectral classes which were identified into one of the seven classes (Figure 27) with the help of field knowledge, classes represented by spatially adjacent clusters, aerial photos, Landsat TM Image from year 2000, and previously classified map by Walker et al. (1994) as done in the ISODATA clustering for the entire image. Then these identified clusters were recoded into the seven concerned land-cover classes.

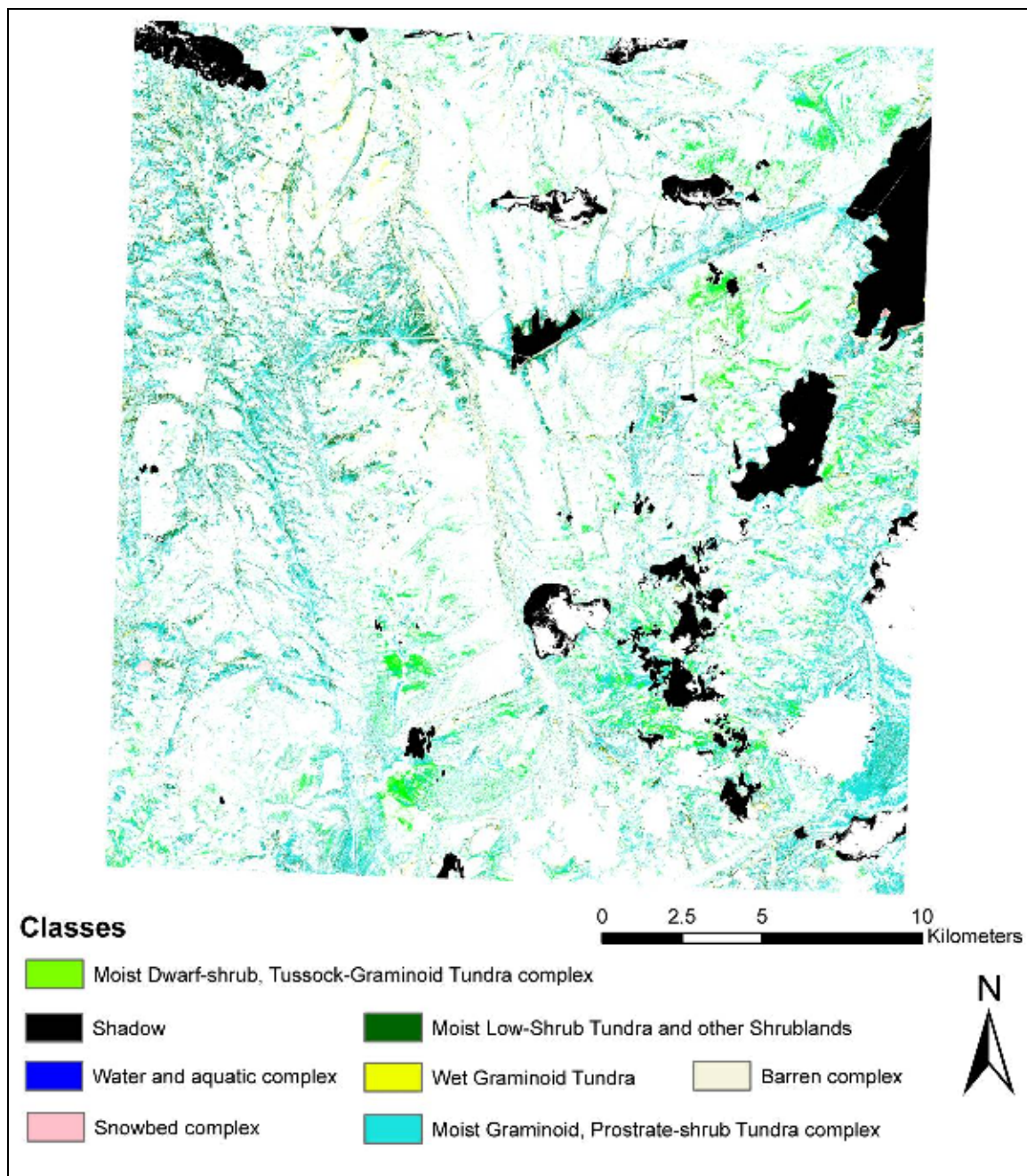


Figure 27: Unclassified pixels from Expert Classifier, classified and recoded

3.9.4. Final Classification rules

The results were merged with the partially classified map resulting from the knowledge based classifier by adding new rules to the knowledge base that would now consider these recoded pixels to derive the complete classified image (Figure 28). The modified set of rules (See Appendix C for rule diagrams) that classifies the whole image is as follows:

Rule for Moist Low-Shrub Tundra and other Shrublands:

(IF Band1 >= 86 AND Band1 <= 106, Band2 >= 86 AND Band2<= 101, Band3>= 140 AND Band3<= 163, NDVI >= 0.182 AND NDVI <= 0.293, NOT IN Moist Dwarf-shrub, Tussock-Graminoid Tundra complex)

OR

(IF Band1 >= 75 AND Band1 <= 120, Band3>= 90 AND Band3<= 187, NDVI >= 0.205 AND NDVI <= 0.4447, NOT IN class Moist Dwarf-shrub, Tussock-Graminoid Tundra complex)

OR

(Restclassified = 1)

→ Moist Low-Shrub Tundra and other Shrublands complex

Rule for water and aquatic complex:

(IF Band1 > 0 AND Band1 <= 88, Band2 > 0 AND Band2 <= 76, Band3 > 0 AND Band3 <= 82, NDVI >= -0.111 AND NDVI <= 0.101)

OR

(IF Band1 > 0 AND Band1 <= 88, Band2 > 0 AND Band2 <= 76, Band3 > 0 AND Band3 <= 82, NDVI <= -0.314902)

OR

(IF Band1 >= 97 AND Band1 <= 127, Band2 >= 75 AND Band2 <= 106, Band3 >= 22 AND Band3 <= 50, NDVI <= -0.0314902)

OR

(IF Band1 < 88, Band2 <= 95, Band3 >= 18 AND Band3 <= 64, NDVI <= -0.111, NOT IN Mountainshadow region)

→ Water and aquatic complex.

Rule for Barren complex:

(IF Band1 >= 105, Band2 >= 90, Band3 >= 45 and Band 3 <= 145, NDVI >= -0.349087 AND NDVI <= 0.0725263, NOT IN shallowwater, NOT IN class Moist Graminoid, Prostrate-shrub Tundra complex)

OR

(IF NDVI > -0.328 AND NDVI < -0.033826, DEM > 1034, NOT IN water and aquatic complex)

OR

(IF Band1 >= 90 AND Band1 <= 111, Band2>= 80, Band3 >= 45 and Band 3 <= 161, NDVI >= -0.349087 AND NDVI <= 0.0725263, NOT IN shallowwater, DEM >= 947.569 AND DEM <= 1024.41, IN lowelevation region)

OR

(IF Band1 >= 88, Band2> 75, Band3 >= 54, NDVI <= -0.025, NOT IN water and aquatic complex) OR (Restclassified = 3)

→ Barren complex

Rule for Snowbed complex:

(If Band1 >= 85 AND Band1 <= 97, Band2 >= 72 AND Band2 <= 88, Band3 >= 78 AND Band 3 <= 100, NDVI <= 0.064 AND NDVI <= 0.101, (aspect > 0 AND aspect < 90) or (aspect > 270 AND aspect < 360), slope < 16, NOT IN water and aquatic complex, NOT IN shallow fen, Slope < 16)

OR

(snowbedWST = 4, (aspect > 0 AND aspect < 90) or (aspect > 270 AND aspect < 360), slope < 16, NOT IN water and aquatic complex, NOT IN shallow fen, Slope < 16)

OR

(snowbedWST = 5, (aspect > 0 AND aspect < 90) or (aspect > 270 AND aspect < 360), slope < 16, NOT IN water and aquatic complex, NOT IN shallow fen, Slope < 16)

OR

(Restclassified = 4, NOT IN shallowfen)

→ Snowbed complex.

Rule for Moist Dwarf-shrub, Tussock-Graminoid Tundra complex:

(If Band1 >= 85 AND Band1 <=100, Band2 >=77 AND Band2<=97, Band3 >=107 and Band3 <=147, NDVI >= 0.077 AND NDVI <=0.244, NOT IN Moist Graminoid, Prostrate-shrub Tundra complex)

OR

(Restclassified = 5, NOT IN Moist Graminoid, Prostrate-shrub Tundra complex)

→ Moist Dwarf-shrub, Tussock-Graminoid Tundra complex

Rule for Wet Graminoid Tundra

(If Band1 >= 74 AND Band1 <= 97, Band2 >= 62 AND Band2 <= 99, Band3 >= 60 AND Band3 <= 114, NDVI >= -0.105 AND NDVI <= 0.062, NOT IN water and aquatic complex, NOT IN snowbed complex)

OR

(Restclassified = 6)

OR

(Restclassified = 4, shallowfen = 1)

→ Wet Graminoid Tundra

Rule for Moist Graminoid, Prostrate-shrub Tundra complex

(If Band1 >= 89 AND Band1 <= 114, Band2 >= 89 AND Band2 <= 106, Band3 >= 96 AND Band3 <= 126, NDVI >= 0.004 AND NDVI <= 0.126)

OR

(Restclassified = 7)

→ Moist Graminoid, Prostrate-shrub Tundra complex.

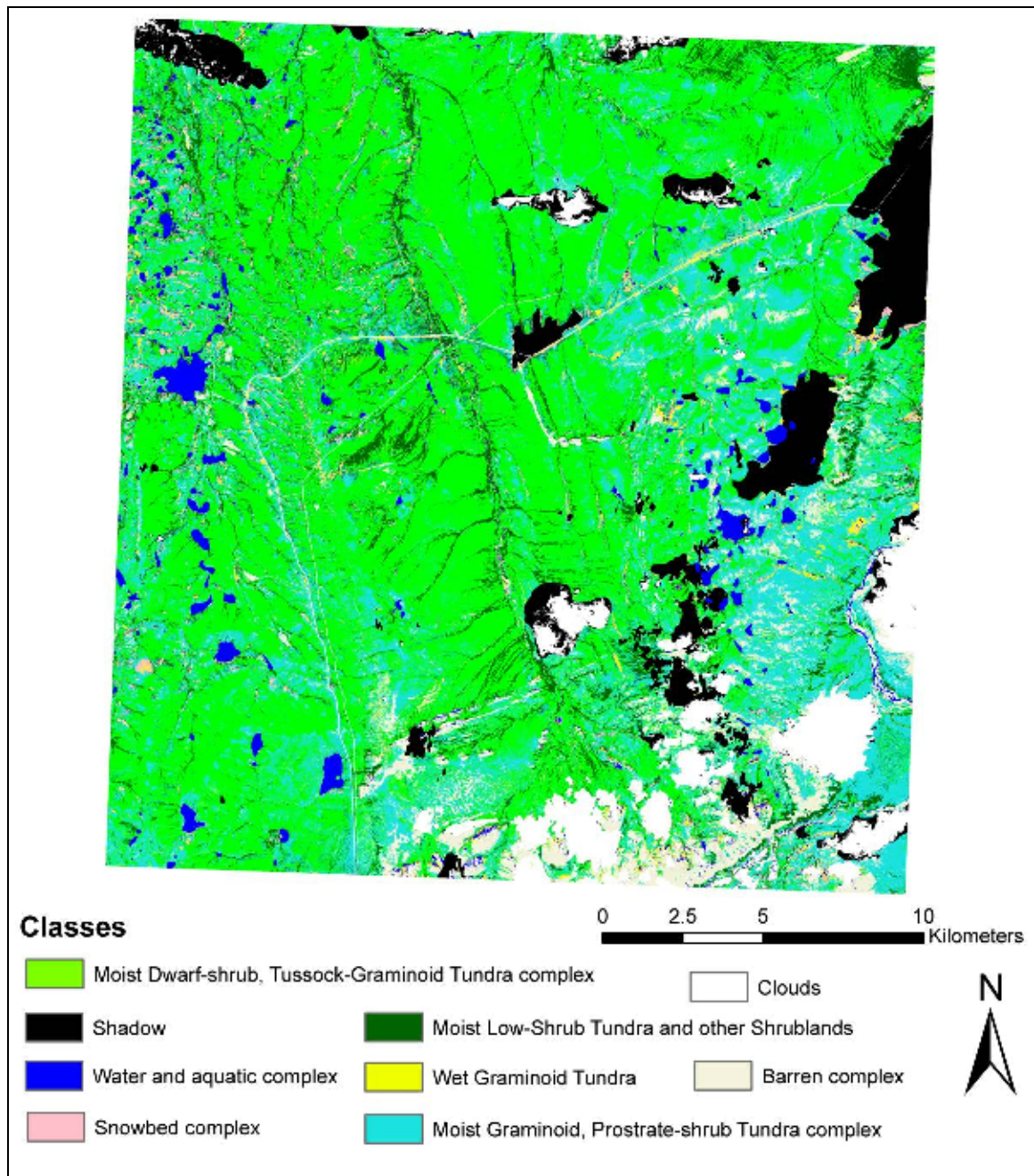


Figure 28: Classified Image from Expert Classifier.

3.9.5. MLP Classifier

3.9.5.1. Artificial Neural Network and its advantages

Artificial neural network is a data mining tool, developed to emulate the brain's interconnected system of neurons to imitate the brain's ability to sort patterns and acquire knowledge through from trial and error learning process. Neural nets typically consist of many simple processing units, called nodes which are connected together in a complex communication network. Interneuron connection strengths known as synaptic weights are used to store the knowledge.

Neural networks are different from statistical or algorithm based models in several respects (Skapura, 1996). Firstly, neural networks do not require formal mathematical specification and the weights derived between inputs, hidden nodes and the output(s) through the iterative processes performed by the computer are not directly interpretable. Secondly, unlike statistical models, neural networks are not highly sensitive to noise in data; statistical or mathematical algorithms treat noise in data similar to data of high quality. Thirdly, information developed by neural networks can be transferable by saving the weight files and implement them on the unseen data set (Pijanowski et al. 2001).

ANNs have been employed to process classification of multispectral remote sensing imagery and often achieve improved accuracies (Benediktsson

et al. 1990, Pijanowski et al. 1997, Jensen et al. 1999). The advantages of an ANN as depicted by (Jensen, 2005) are:

1. ANNs can readily accommodate ancillary data in the classification technique.

2. ANNs makes no a priori assumptions of normal and linear distribution due to its nonparametric operation.

3. ANNs are able to learn from existing non-linear empirical examples adaptively instead of “prespecified” by an analyst which makes the classification objective.

4. Neural network can handle noisy information inevitably included in the examples with the ability to generalize thus making it more robust than other mining methods. Individual bias in training and incorrect or incomplete information are excluded from the knowledge acquisition process unlike decision trees.

5. In decision trees knowledge is represented by logical rules made up of binary predicates. Numerical attributes have to be converted to binary true/false statements which may cause a loss of large amount of information. Whereas, ANN can accept all data formats as long as data is converted to a numeric representation.

6. ANNs are good at generalizing both discrete and continuous data and have a capability to interpolate or adapt to the patterns never seen in training process and attempts to find the best fit for input patterns. Decision trees on the

other hand, fail to generalize a predictable inference if an appropriate match with the perfect rules cannot be found.

7. Neural networks continuously adjust the weights as more training data are provided and capable to learn continuously.

3.9.5.2. The Multilayer Perceptron

The most widely used neural network model in classifying remotely sensed imagery is the multilayer perceptron (MLP), a feed-forward artificial neural network model. An MLP, consists of three types layers, the input, hidden and output layers, each consisting of processing nodes that are interconnected to each other, but there are no interconnections between nodes within the same layer. An MLP in general comprises one input layer, one or two hidden layers and one output layer. The input layer nodes correspond to individual data sources, such as the different bands of imagery, ancillary data etc. Hidden layers are used for computations, and the values associated with each node are calculated from the input node values and weights of the links connected to that node. The output layer includes a set of codes to represent the informational classes to be classified by the analyst (Figure 29).

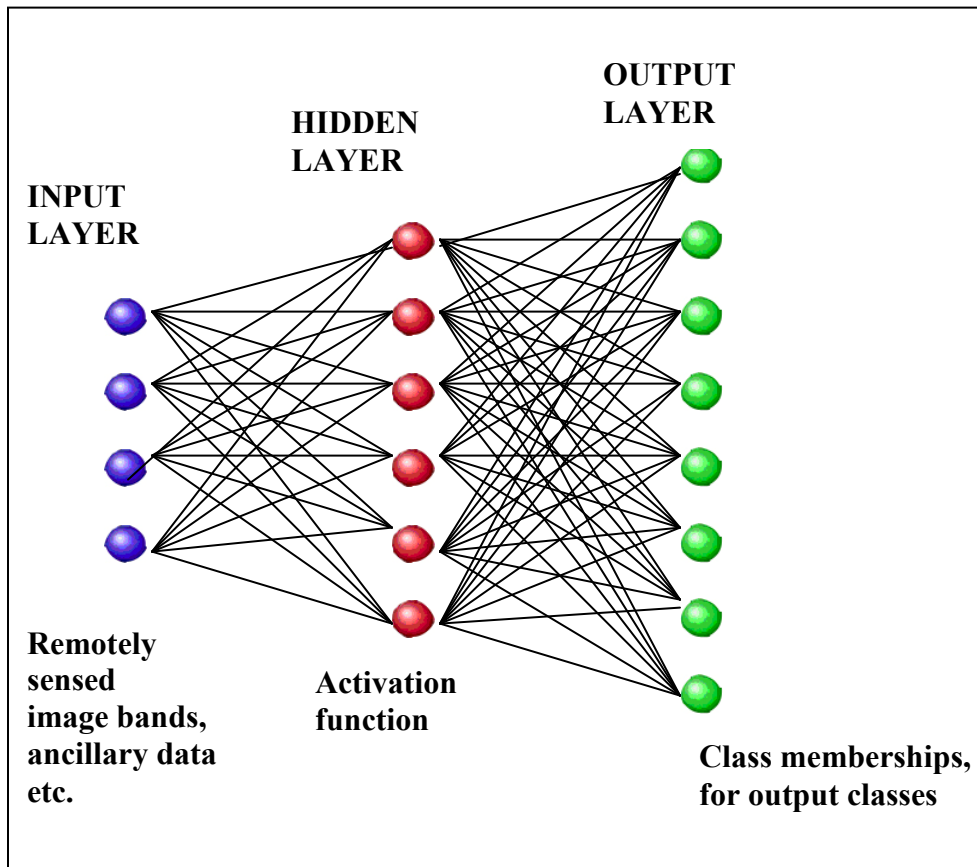


Figure 29: An Artificial Neural Network with single hidden layer.

MLP uses an algorithm, called the generalized delta rule, which is an iterative gradient descent training procedure based on error-correction learning rule that is carried out in two stages. First, once the random network weights have been initialized, the input data are propagated forward to estimate the output value for each training pattern set. In the next stage, the difference (error) between known and estimated output is fed backward through the network, and the weights associated with the nodes are changed in such a way that the

differences between the actual and the desired outputs is minimized. The whole process is repeated, with weights being recalculated at every iteration, until the total error (RMSE) is minimal, or else either the number of iterations or the RMSE is lower than a given threshold value provided by the analyst (Kavzolgu et al.2003).

Each node in the network may include a nonlinear transfer function at the output end. Being initialized with all the synaptic weights and thresholds set to small arbitrary numbers, the network is provided with the training sample patterns. In the forward pass, the input parameter to a node in an MLP network is the weighted sum of the outputs from the layer below or previous layer. The net input to the j^{th} neuron, net_j , is given by

$$net_j = \sum_i w_{ij} o_i$$

where w_{ij} is the weight between node i and node j , o_i being the output from node i . The corresponding output from node j is given by

$$o_j = \frac{1}{1 + \exp(-net_j + \theta_j)} \text{ (using sigmoidal transfer function)}$$

or

$$o_j = m \tan h(k(net_j)) \text{ (using hyperbolic tangent)}$$

where θ_j , m and k are constants. The difference between the output activation

and the desired response of node j at the output layer is called the error which is then propagated backward with weights for relevant connections corrected by the delta rule in the backward pass and the weights are updated during training. The delta rule can be stated as the following equation:

$$\Delta w_{ji}(n+1) = \eta(\partial_j o_i) + \alpha \Delta w_{ji}(n)$$

where η is the learning rate, α is the momentum factor, and ∂ is the computed error with respect to the output from node j , $\Delta w_{ji}(n+1)$ is the change of a weight connecting nodes i and j , in two successive layers, at the $(n+1)^{\text{th}}$ iteration (Jensen, 2005; Kanellopoulos et al. 1997; IDRISI Andes Manual, 2007).

3.9.5.3. Neural Network Architecture and Parameters

A trial-and-error strategy was employed to find out appropriate values for the different important parameters in the MLP that influences the performance of the learning algorithm and produces highest classification accuracy. The basic architecture to start with was constructed according to the guidelines suggested by Kavzoglu et al. (2003) and Kanellopoulos et al. (1997) using IDRISI Andes image processing software (Figure 31). The important parameters are discussed below.

Number of input nodes

Various combinations of input layers were run in the trial runs (Appendix D). It was found that the inclusions of the texture layers for the SPOT-5 bands and the NDVI did not improve the classification accuracies. Also various other transformations of the SPOT-5 bands were applied as trial inputs. The final input layer in the final neural network contained 8 nodes, each representing the different input layers consisting of the 3 bands (green, red and IR) of SPOT image, slope, aspect, NDVI, tundra index (discussed later in this chapter in section 3.8.6), and the output of the rule-based classifier.

Number of output nodes

The output layer consisted of eight nodes, seven of them representing the different tundra land-cover classes and the eighth one being the unclassified or undefined pixels that consisted of the background pixels and the masked out pixels in shadow and cloud regions in the image.

Input Image Normalization or Scaling

Values of the individual pixels of all the eight different input layers were normalized or scaled from 0.0 to 1.0 by using a script in ERDAS Imagine (Figure 30) to ensure that the network's iterative weight adjustment fits within the numerical range of the activation function's range thus preventing early saturation effects that causes the network to 'stall' (Kanellopoulos et al. 1997).

```

COMMENT "Script for normalization of raster layers";
#
# set cell size for the model
#
SET CELLSIZE MIN;
#
# set window for the model
#
SET WINDOW UNION;
#
# set area of interest for the model
#
SET AOI NONE;
#
# declarations
#
Float RASTER n1_SPOTband1 FILE OLD NEAREST NEIGHBOR AOI NONE
"f:/work/research/final data/SPOTband1.img";
Float RASTER n4_dem0 FILE NEW IGNORE 0 ATHEMATIC FLOAT SINGLE
"f:/work/research/idrisi/Band10-1.img";
#
# function definitions
#
#define n2_memory Float ($n1_SPOTband1)
n4_dem0 = (($n2_memory - GLOBAL MIN ($n2_memory)) / (GLOBAL MAX
($n2_memory) - GLOBAL MIN ($n2_memory))) * 1;
QUIT;

```

Figure 30: Script in ERDAS Imagine used for scaling the input raster.

Training and Validation Pixels

In each iteration, for the seven land-cover classes concerned, 300 pixels were randomly selected for training the network, and 100 pixels were randomly selected for testing the trained networks. Thus, from the training site layer, for each iteration, number of training and test datasets was 2400 and 800 random pixels in total, respectively. As suggested by Kavzolgu et al. (2003) the sample sizes should range between $30 * N_i * (N_i + 1)$ and $60 * N_i * (N_i + 1)$ depending on

the difficulty of the problem under consideration, where N_i is the number of input features or nodes. In this case, the minimum number of training pixels required was 2160 while the optimum number is 4320.

Number of hidden layers and hidden nodes

Determination of the optimum number of hidden nodes in a neural network classification technique is a serious concern in order to avoid overfitting or underfitting and produce lower classification accuracies. Overfitting is a state which occurs when the network is too large and become overspecific to the training data while underfitting occurs when the network is too small and thus is unable to identify the internal structure of the data. The optimum number of nodes in a hidden layer is between $2N_i$ to $3N_i$ (Kanellopoulos et al. 1997; Hush, 1989; Hecht-Nielsen, 1987).

In this research, the thumb rule followed for the number of nodes in the first hidden layer was $n_1 = \text{ceiling}(2.5 * N_i)$, where N_i is number of nodes in the input layer. For example, N_i being 8, the number of nodes in the first hidden layer was calculated be 20. It was found from the trial runs (Appendix D) that the use of two hidden layers did have significant effect on network classification performance. The thumb rule followed for the number of nodes in the second hidden layer was $n_2 = \text{ceiling}(2.5 * N_o)$ where N_o is the number of nodes in the output layer, being a constant = 8 in this research. As suggested by Kanellopoulos et al. (1997), for the trial runs, if a single hidden layer was used,

the higher of the two cases (n_1 and n_2) considered was applied for the number of nodes in the hidden layer.

Learning rate, momentum and sigmoid constant

In order to optimize the speed and efficiency of the learning process, the learning rate and momentum term are very important considerations. The momentum term determines the direction of search for the global minimum of the error using the previous weight configurations. The learning rate is used to update the inter-node weights (Kavzolgu et al. 2003).

An automatic learning with dynamic learning rate was opted for the trial runs in IDRISI Andes image processing software with learning rate being 0.01 and the end learning rate being 0.001. With the automatic learning option, the MLP automatically adjust the learning rate, and if adjustments are made to the learning rate, the iteration process starts again. The dynamic learning option enables the MLP, to decrease the learning rate towards the minimum learning rate (IDRISI Andes Manual, 2007). All these adjustments and reiterations occur until the learning process becomes stable. The momentum factor used was 0.5 and the sigmoid constant chosen was 1.0.

Stopping criterion

A stopping criterion for the MLP learning process has to be established as it is generally impossible to train neural networks at an accuracy rate of 100%.

Three different stopping criteria were used in this research are, the RMS error (0.0001) which is the error associated with the learning of the network, the number of iterations (10000), and the accuracy rate percent (100%). Any of these criteria reached first would terminate the learning process.

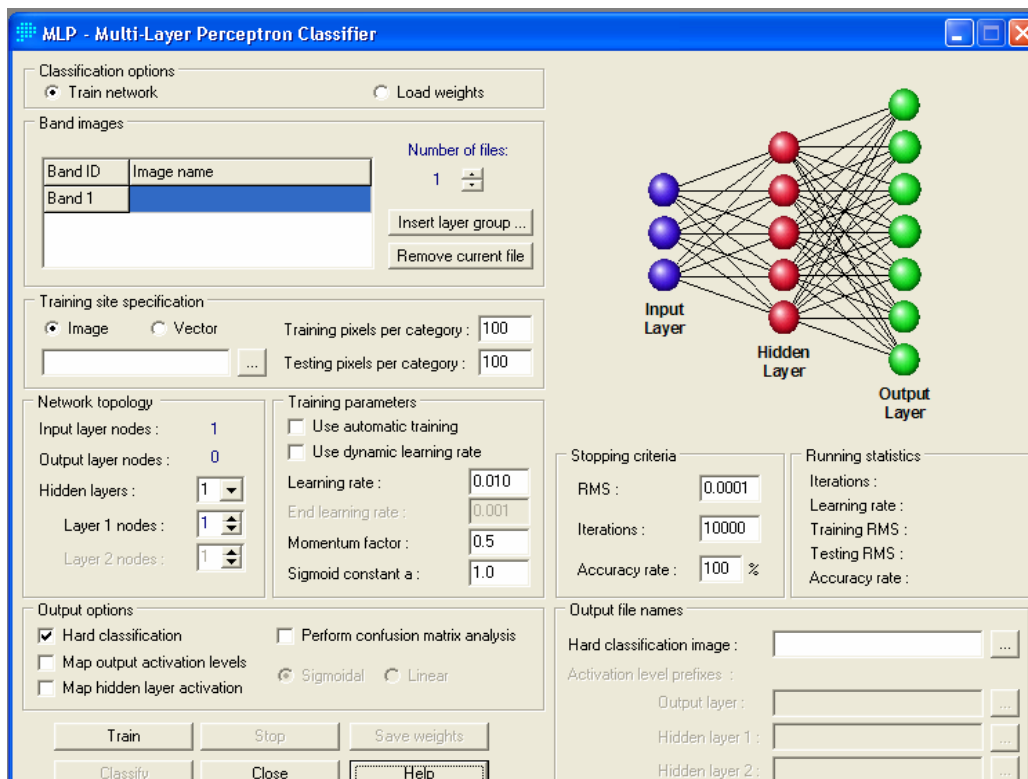


Figure 31: The MLP classifier interface in IDRISI Andes.

Appendix D shows a table with the trial runs, and the different parameters used for the neural network classifier. The best combination of input layers consists of Bands 1, 2, 3 from the SPOT image, NDVI, slope, aspect, tundra

index, and the output classified image of the rule-based classifier. The average training accuracy and corresponding kappa value for this combination was 96.0433 and .9901 respectively. Addition of the NDVI layer to the SPOT bands improved the training accuracy from 83.54% to 92.50% for a double layered inclusion of the texture layers for SPOT bands did not have any improvement in the training process in terms of accuracy and kappa. Also two different transformations of the tundra index were tried as input: square of tundra index and log of tundra index; both of these transformations did not contribute to any further improvement in the training accuracy.

3.9.6. The Tundra Index

The tundra index is a new calculated spectral index (Figure 34) which is a non-linear transformation of the spectral values of the SPOT bands given by:

$$\text{Tundra index} = \frac{2 * \text{IR} - (\text{Green} + \text{Red})}{2 * \text{IR} + (\text{Green} + \text{Red})}$$

It was found that the two major classes, Moist Dwarf-shrub, Tussock-Graminoid Tundra complex and Moist Graminoid, Prostrate-shrub Tundra complex which consisted of the major land-cover for the study area had spectral mixing effect in them and it was hard to separate the two classes when they had heterogeneous occurrence in contiguous pixels. The spectral profile (Figure 33) created from homogenous occurrence of five sample pixels for each of the two

classes explains that for Moist Dwarf-shrub, Tussock-Graminoid Tundra complex,

```
COMMENT "Model for tundra index";
#
# set cell size for the model
#
SET CELLSIZE MIN;
#
# set window for the model
#
SET WINDOW UNION;
#
# set area of interest for the model
#
SET AOI NONE;
#
# declarations
#
Integer RASTER n1_SPOTimage FILE OLD NEAREST NEIGHBOR AOI NONE
"f:/work/research/final data/SPOTimage.img";
Float RASTER n2_myindex FILE DELETE_IF_EXISTING IGNORE 0
ATHEMATIC FLOAT SINGLE "f:/work/research/final
data/tundraindex.img";
#
# function definitions
#
#define n10_memory Float($n1_SPOTimage(3) + $n1_SPOTimage(1))
#define n9_memory Float($n1_SPOTimage(3) + $n1_SPOTimage(2))
#define n8_memory Float($n1_SPOTimage(3) - $n1_SPOTimage(1))
#define n7_memory Float($n1_SPOTimage(3) - $n1_SPOTimage(2))
n2_myindex = EITHER 0.0 IF ( $n9_memory == 0.0 AND $n10_memory
== 0.0 ) OR ($n7_memory + $n8_memory) / ($n9_memory +
$n10_memory) OTHERWISE;
QUIT;
```

Figure 32: Script in ERDAS Imagine for modeling the tundra index.

(referred to as MAT in the figure) the values for Band1 and Band2 of the SPOT-5 image is lower than Moist Graminoid, Prostrate-shrub Tundra complex (referred

to as MNT in the figure), while it is the vice versa for Band3. Thus a non-linear transformation was applied to the SPOT-5 raster bands, which involved all the three bands, instead of only Band3 and Band2 in NDVI, which would magnify the differences for the two classes. This transformation was applied along with the NDVI, as another input node in the input layer of the MLP classifier.

The inclusion of the tundra index in the MLP layer improved the average training accuracy from 90.79% to 93.75%, average training kappa values from 0.95 to 0.97 in single hidden layer architecture. While in a two hidden layer scenario, the improvements were not so significant (Table 3). Figure 35 shows the classified image from the MLP classifier.

Table 3: Iterations showing the improvements in training accuracies and kappa values with the inclusion of the tundra index in the MLP classifier

Input Layers**	Iterations	SLR*	ELR*	n ₁ *	n ₂ *	Training RMS*	Testing RMS*	Training Accuracy	Training Kappa	Average Accuracy	Average Kappa
	1	0.00015	0.0003			0.00103	0.0019	90.75	0.954		
A	2	0.00138	0.0003	20	0	0.00109	0.00189	90.25	0.952	90.7933	0.9544
	3	0.00238	0.0003			0.00097	0.00168	91.38	0.957		
	1	0.00488	0.0005			0.00085	0.00147	93.88	0.964		
B	2	0.00488	0.0005	20	0	0.00082	0.00144	94.25	0.974	93.7533	0.9663
	3	0.00181	0.0003			0.00099	0.00171	93.13	0.96		
	1	0.00203	0.0003			0.00078	0.00135	93.75	0.976		
A	2	0.00213	0.0003	16	20	0.0008	0.00137	94.13	0.976	93.7100	0.9741
	3	0.00198	0.0003			0.00082	0.0014	93.25	0.971		
	1	0.00085	0.0001			0.00083	0.00144	92.88	0.976		
B	2	0.00238	0.0003	18	20	0.00072	0.00123	95.38	0.98	93.7533	0.9752
	3	0.00075	0.0001			0.00085	0.00148	93.00	0.97		

Notes:

- ** Input Layer A: SPOT-5 Bands (1, 2, 3), NDVI, Slope, Aspect
Input Layer B: SPOT-5 Bands (1, 2, 3), NDVI, Slope, Aspect, tundra index
- * n₁: No. of nodes in Hidden Layer1
n₂: No. of nodes in Hidden Layer2
RMS: Root mean square error

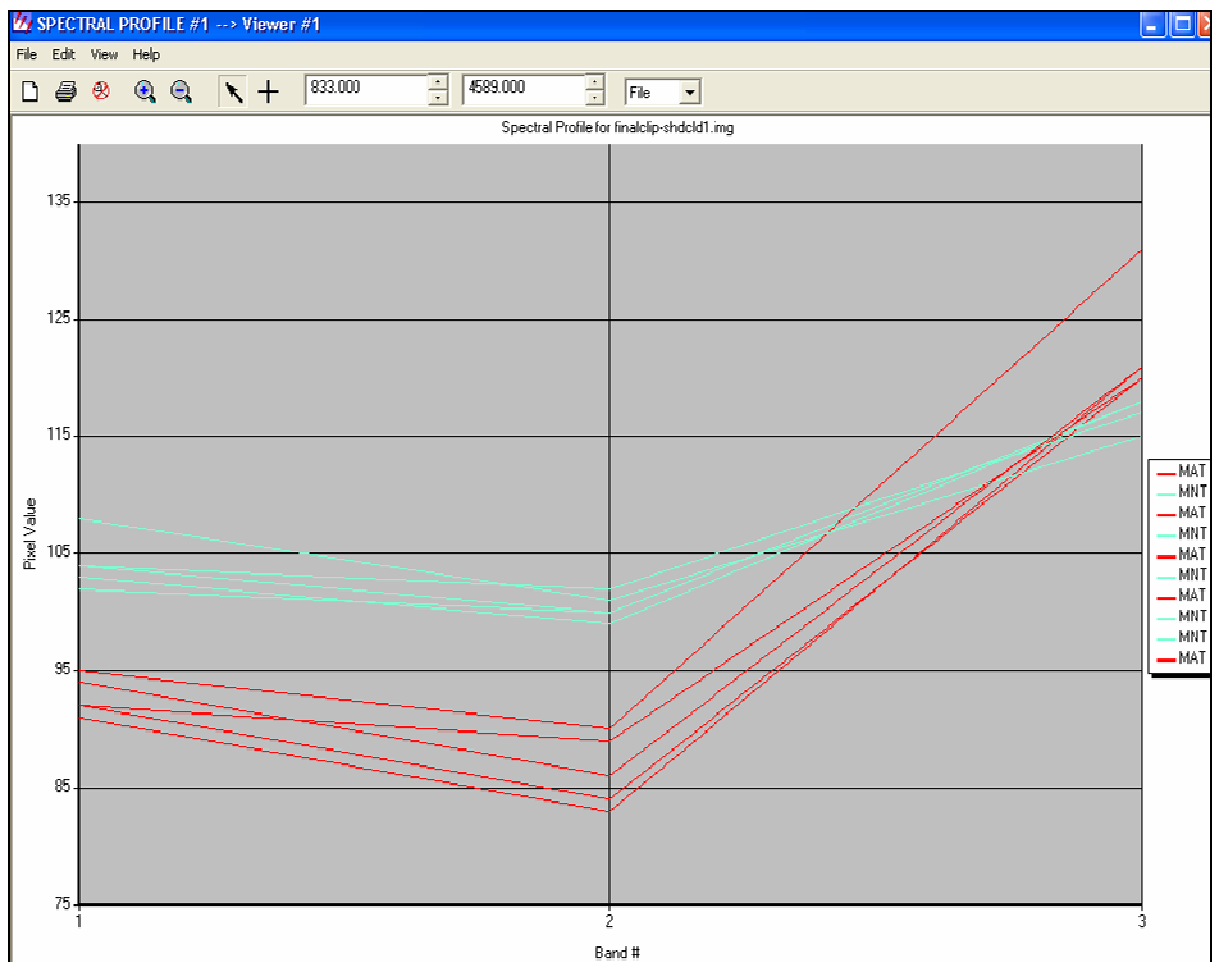


Figure 33: Spectral profile for MAT and MNT in the SPOT-5 image.

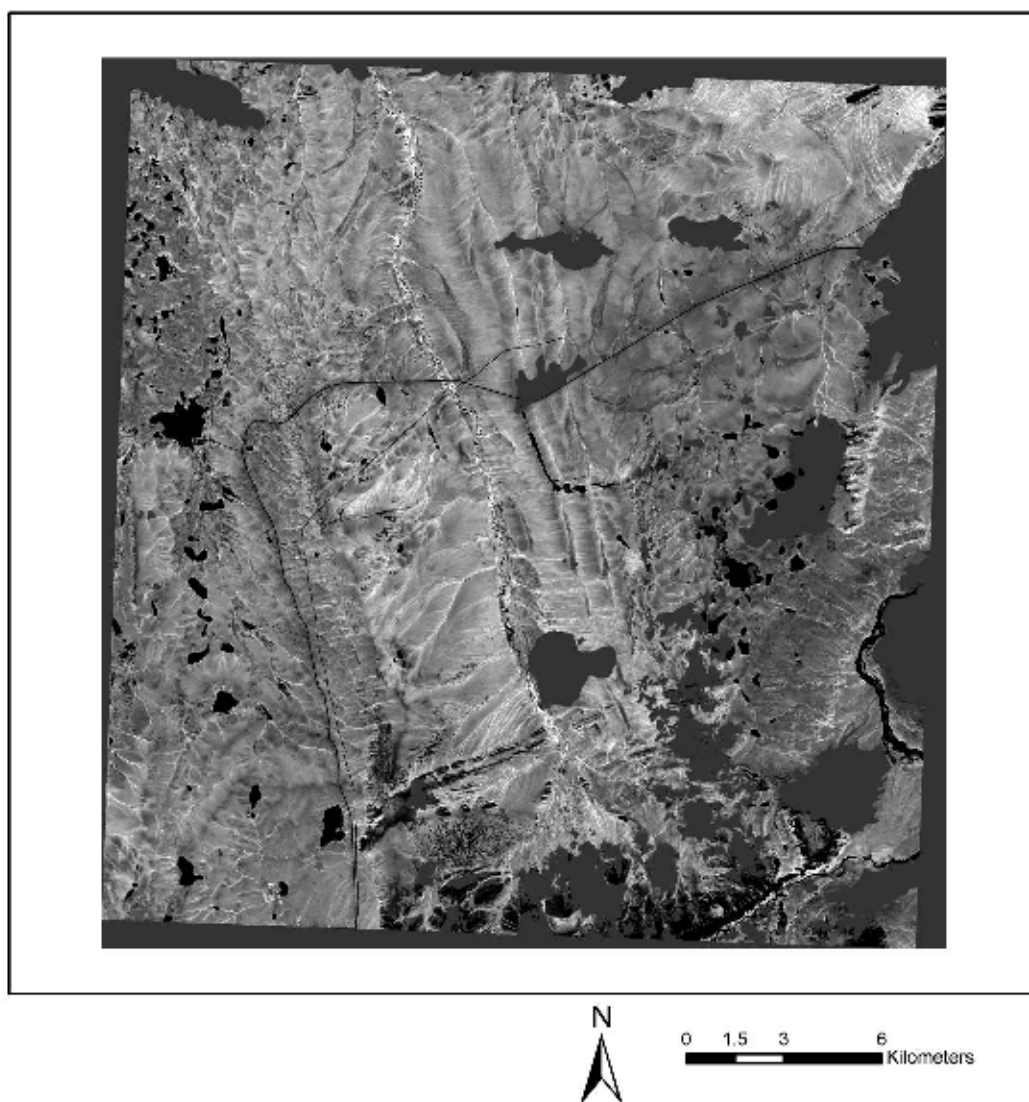


Figure 34: The tundra index layer (with cloud and shadow pixels removed).

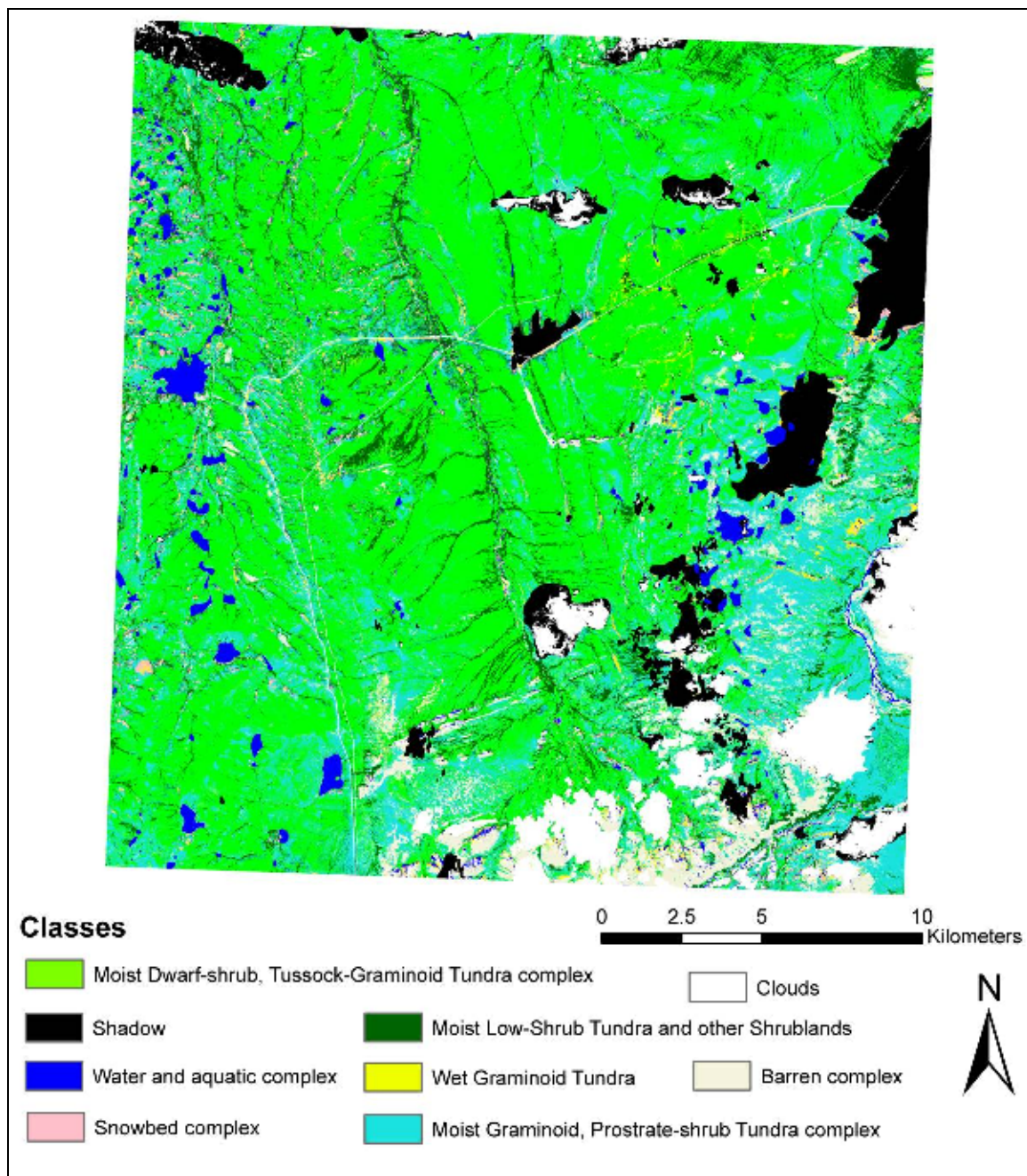


Figure 35: The classified image from the MLP classifier

3.9.7. Integration of the two classifiers: rule-based and MLP

Liu et al. (2002) suggested that different classifiers have complementary capabilities and integrating the results from individual classifiers improves classification accuracy. In their study, Liu et al. (2002) used a consensus builder approach to adjust classification output in the case of disagreement in classification between three different classifiers where, if the output classes for each individual pixel differed, the producer accuracies for each class were compared and the class with the highest producer accuracy was assigned to the pixel of the final classified image (map) output.

Following the same principle (Liu et al. 2002), in this research, the output classified images from the two classifiers, namely the rule-based and MLP, were compared. For pixels where there was a disagreement for class values, instead of the producer accuracies as suggested by (Liu et al. 2002), kappa values for the classes of each individual pixel were compared and the class with the highest kappa value was assigned to the pixel for the final classified map output. The classes having higher kappa values were extracted in ArcGIS 9.2 with the extraction (by attributes) tool in the spatial analyst extension and saved as a raster (.img format). This raster was used as a decision zone (Figure 36) to assign the pixel values (classes) of the final classified map output for the hybrid classifier, using the outputs of the rule-based and MLP classifiers.


```

COMMENT "Model for integrating the two classifiers; rule-based and
MLP";
#
# set cell size for the model
#
SET CELLSIZE MIN;
#
# set window for the model
#
SET WINDOW UNION;
#
# set area of interest for the model
#
SET AOI NONE;
#
# declarations
#
Integer RASTER n1_rule FILE OLD NEAREST NEIGHBOR AOI NONE
"e:/research/idrisi/finaltest/rule-based.img";
Integer RASTER n2_MLP FILE OLD NEAREST NEIGHBOR AOI NONE
"e:/research/idrisi/finaltest/MLP.img";
Integer RASTER n3_higher FILE OLD NEAREST NEIGHBOR AOI NONE
"e:/research/idrisi/finaltest/higherkappa.img";
Integer RASTER n7_hybrid FILE NEW IGNORE 0 THEMATIC BIN DIRECT
DEFAULT 8 BIT UNSIGNED INTEGER
"e:/research/idrisi/finaltest/hybrid.img";
#
# function definitions
#
#define n5_memory Integer(EITHER $n1_expert IF ( $n3_nn == 0) OR
$n2_nn02232 OTHERWISE )
n7_hybrid = $n5_memory;
QUIT;

```

Figure 36: Script in ERDAS Imagine for modeling the tundra index

3.10. Post-processing of the classified images for the area under haze

A section of the image towards the north-west corner was covered by a haze probably caused by a very low cloud or mist that could not be identified during the cloud pixels removal due to its amorphous presence in the image. Due to this, the area had unusual spectral reflectance from the land-cover and the

classes produced by the different classification techniques in this haze area were identified to be incorrect from field knowledge and observation. For example most of the area in this region was assigned to the class Moist Graminoid, Prostrate-shrub Tundra complex where in reality it should be Moist Dwarf-shrub, Tussock-Graminoid Tundra complex. This can be explained by the brighter appearance of the pixels due to the haze, especially in band1 and band 2 (Figure 37, RGB 3, 2, 1).

A specific field inspection for ground truth data for this area was done in this area in order to ascertain the classes properly. The area was clipped out and an ISODATA clustering was applied to obtain 20 clusters. These clusters were identified using the ground truth data and field knowledge into the seven land-cover classes concerned. The clustered image was recoded and merged (Figure 37) into each of the three classified images obtained previously from the three classifiers.

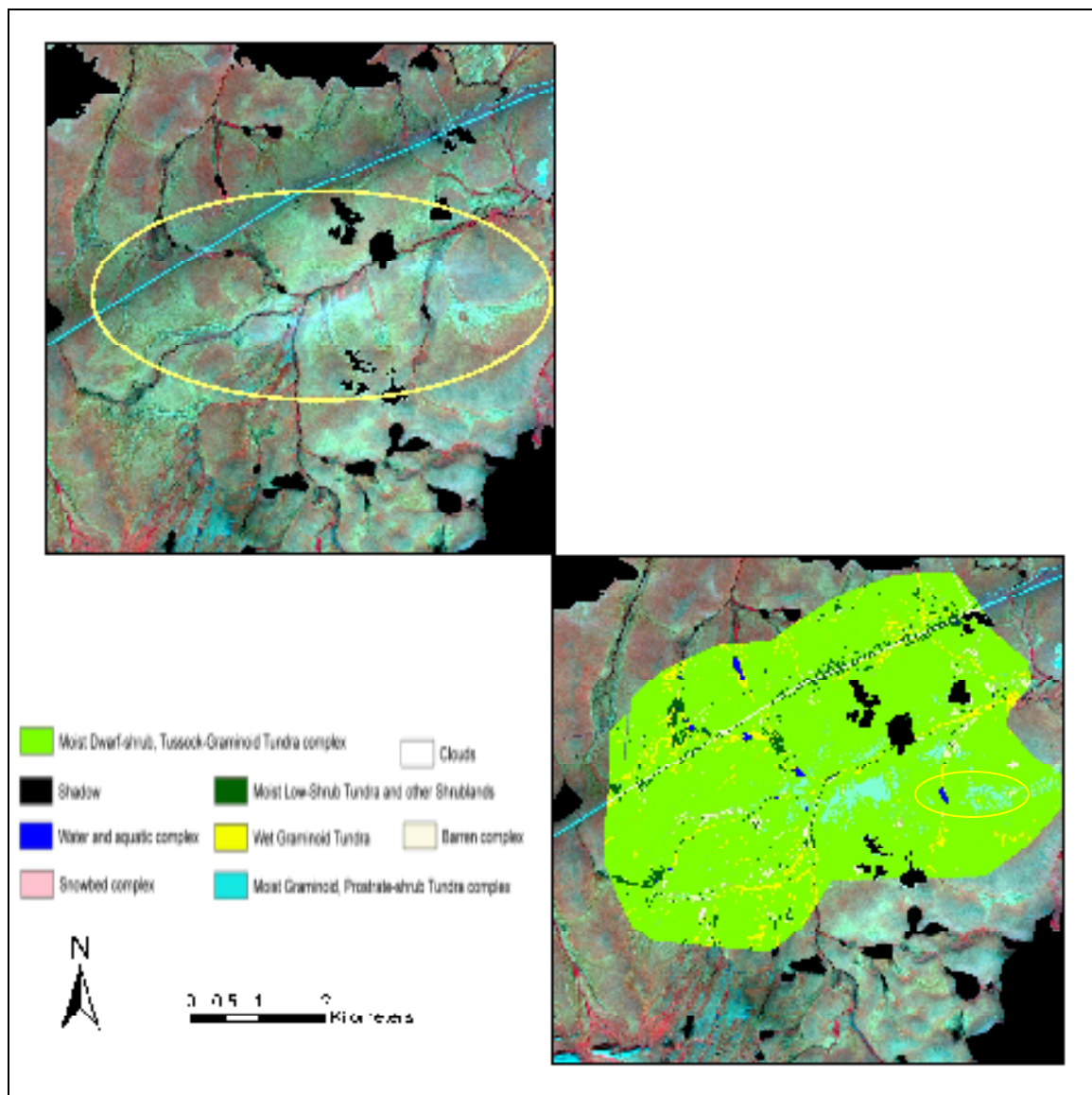


Figure 37: The area under haze (top) and the clipped area reclassified and recoded (bottom).

```

COMMENT "Model used to merge the recoded haze area classification into
the output of the MLP classifier";
#
# set cell size for the model
#
SET CELLSIZE MIN;
#
# set window for the model
#
SET WINDOW UNION;
#
# set area of interest for the model
#
SET AOI NONE;
#
# declarations
#
Integer RASTER n1_MLP FILE OLD NEAREST NEIGHBOR AOI NONE
"e:/research/idrisi/finaltest/MLP.img";
Integer RASTER n3_recodehaze20 FILE OLD NEAREST NEIGHBOR AOI NONE
"f:/work/research/final data/test/recodehaze20.img";
Integer RASTER n7_finalMLP FILE DELETE_IF_EXISTING IGNORE 0 THEMATIC
BIN DIRECT DEFAULT 8 BIT UNSIGNED INTEGER
"e:/research/idrisi/finaltest/finalMLP.img";
#
# function definitions
#
#define n5_memory Integer(EITHER $n1_MLP IF ( $n3_recodehaze20 == 0 )
OR $n3_recodehaze20 OTHERWISE )
n7_finalMLP = $n5_memory;
QUIT;

```

Figure 38: Script in ERDAS Imagine for model merging of the recoded haze area classification into the output of the MLP classifier.

CHAPTER IV

RESULTS AND DISCUSSIONS

4.1. Results

4.1.1. Accuracy Assessment

It is necessary to check the accuracy of the land-cover classification with ground truth data before it can be used in scientific investigations and decision making policies (Jensen, 2005). Errors in a thematic map from a classification process can be introduced from several sources like data acquisition (sensor error), radiometric resampling, geometric registration, data conversion and misclassification by the analyst or by the classification model. A design-based inference which is most commonly applied in remote sensing studies involving statistical measurements including overall accuracy, producer's error, consumer's error and the Kappa coefficient was applied in this study (Congalton et al.1999). The results of the accuracy assessments were used to compare the results of the different classification techniques.

A stratified random sampling technique was applied for collecting the ground truth data for accuracy assessment. In this technique a minimum number

of samples are selected from each class and samples are allocated to all the classes for accuracy assessment without depending on the proportion of each individual class in the entire study area. As mentioned in section 3.5, the data (pixels) collected were divided into two subsets, one of which was used for training (128 points) and the other for testing (221 points) the classification accuracy to avoid any bias resulting from the use of the same set of pixels for both training and testing.

4.1.1.1. Error Matrix

An error matrix is an effective technique involving a square matrix that presents the overall accuracy of the classification, the producer and user accuracy of each class. The columns contain the reference data (from the ground truth data), while the rows represent data from the classifications. Values of each row and column represent a unique combination of classified data and accuracy assessment data. The major diagonal of the matrix reflect the sites correctly classified while the cell values in the off-diagonal positions express disagreement between the classified and the reference data. From the error matrix (Table 4 - 6) various descriptive evaluations for accuracies can be derived as explained in following sections.

Table 4: Error Matrix for unsupervised ISODATA clustering

Reference Data								
Classified Data	shrub	water	barren	Snowbed	MAT**	WST**	MNT**	Row Total
shrub	17	0	0	2	5	2	2	28
water	0	30	0	0	0	0	0	30
barren	0	0	11	0	0	1	1	13
Snowbed	0	0	0	2	1	2	0	5
MAT**	2	0	0	0	55	5	21	83
WST**	1	0	1	0	0	3	1	6
MNT**	1	0	1	4	11	5	34	56
Column Total	21	30	13	8	72	18	59	221

Table 5: Error Matrix for supervised classification with Feature Analyst

Reference Data								
Classified Data	shrub	water	barren	Snowbed	MAT**	WST**	MNT**	Row Total
shrub	14	0	0	0	4	4	2	24
water	0	30	1	0	1	1	2	35
barren	1	0	11	0	1	1	4	18
Snowbed	4	0	0	8	5	3	9	29
MAT**	1	0	0	0	45	1	10	57
WST**	1	0	1	0	0	7	9	18
MNT**	0	0	0	0	16	1	23	40
Column Total	21	30	13	8	72	18	59	221

Table 6: Error Matrix for hybrid classification

Reference Data								
Classified Data	shrub	water	barren	Snowbed	MAT**	WST**	MNT**	Row Total
shrub	12	0	0	0	2	2	2	18
water	0	30	0	0	0	0	0	30
barren	0	0	11	0	1	1	0	13
Snowbed	0	0	0	8	1	2	2	13
MAT**	8	0	0	0	62	6	15	91
WST**	1	0	1	0	0	5	1	8
MNT**	0	0	1	0	6	2	39	48
Column Total	21	30	13	8	72	18	59	221

Notes (Table 4, 5, 6)

MAT** referred to as Moist Dwarf-shrub, Tussock-graminoid Tundra complex

WST** referred to as Wet graminoid Tundra complex

MNT** referred to as Moist Graminoid, Prostrate-shrub Tundra complex

4.1.1. 2. Overall Accuracy

The overall accuracy is determined by the sum of all samples on the diagonal (total correct pixels) divided by the total number of samples. However, the overall accuracy (or percentage classified correctly) gives no insight into how well the classifier is performing for each of the individual classes and also it does not consider the case of particular classes that covers large proportion of the test data and offers a bias to the overall accuracy. Table 7 shows the overall accuracies for each of the three different classifiers.

Table 7: Comparison of overall accuracies for the three different classifiers

Classifiers	Total number of pixels	Number of correct pixels	Overall accuracy%
Unsupervised ISODATA clustering	221	151	68.33
Supervised classification with Feature Analyst	221	138	62.44
Proposed Hybrid Classifier	221	167	75.57

4.1.1. 3. User's and Producer's Accuracy

The two most important descriptive accuracy measures derived from the error matrix are: user's accuracy (also known as error of commission or reliability) and producer's accuracy (also known as error of omission).

User's accuracy is the probability for a pixel classified as a particular information class on the classified map actually represents that particular class on the ground. It is obtained by dividing the total number of correct pixels in a category by the total number of pixels actually classified in that category (Jensen, 2005).

Producer's accuracy is the probability of a reference pixel being correctly classified as a particular information class on the classified map. It is obtained by dividing the total number of correct pixels in a category by the total number of pixels of that category (reference data column total) (Jensen, 2005).

Tables 8 - 10, show the user's and producer's accuracies for the three different classifiers.

Table 8: User's and Producer's accuracies for unsupervised ISODATA clustering

Class names	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Shrub complex	21	28	17	80.95%	60.71%
Aquatic complex	30	30	30	100.00%	100.00%
Barren complex	13	12	10	76.92%	83.33%
Snowbed complex	8	5	2	25.00%	40.00%
MAT** complex	72	83	55	76.39%	66.27%
WST** complex	18	6	3	16.67%	50.00%
MNT** complex	59	56	34	57.63%	60.71%

Table 9: User's and Producer's accuracies for supervised classification with Feature Analyst

Class names	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Shrub complex	21	24	14	66.67%	58.33%
Aquatic complex	30	35	30	100.00%	85.71%
Barren complex	13	18	11	84.62%	61.11%
Snowbed complex	8	29	8	100.00%	27.59%
MAT** complex	72	57	45	62.50%	78.95%
WST** complex	18	18	7	38.89%	38.89%
MNT** complex	59	40	23	38.98%	57.50%

Table 10: User's and Producer's accuracies for proposed hybrid classifier

Class names	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Shrub complex	21	18	12	57.14%	66.67%
Aquatic complex	30	30	30	100.00%	100.00%
Barren complex	13	13	11	84.62%	84.62%
Snowbed complex	8	13	8	100.00%	61.54%
MAT** complex	72	91	62	86.11%	68.13%
WST** complex	18	8	5	27.78%	62.50%
MNT** complex	59	48	39	66.10%	81.25%

Notes (Table 8, 9, 10)

MAT** referred to as Moist Dwarf-shrub, Tussock-graminoid Tundra complex

WST** referred to as Wet graminoid Tundra complex

MNT** referred to as Moist Graminoid, Prostrate-shrub Tundra complex

4.1.1.4. The Kappa Statistic

Kappa analysis is a discrete multivariate technique used in accuracy assessment which uses K_{hat} statistic as a measure of agreement or accuracy between classified map and reference data (Jensen, 2005). K_{hat} uses the major diagonal elements of the error matrix and the chance agreement indicated by the row and column totals (marginals), thus considering interclass agreement. The Kappa analysis tests if a land-use or land-cover map is significantly better than if the map had been generated by (random) chance (Congalton, 1996). K_{hat} statistic is computed as (Jensen, 2005, Congalton, 1991):

$$K_{hat} = \frac{N \sum_{i=1}^k x_{ii} - \sum_{i=1}^k (x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^k (x_{i+} * x_{+i})}$$

where k is the number of rows in the matrix, x_{ij} is the number of observation in row i and column j , and x_{i+} and x_{+i} are the marginal totals for row i and column i , respectively, and N is the total number of observations.

Typically, Kappa coefficient lies between 0 and 1 scale, where 1 indicates complete agreement. Kappa values are also characterized into 3 groups in general: a value > 0.80 (80%) represents strong agreement, a value between

0.40 and 0.80 (40 to 80%) represents moderate agreement, and a value < 0.40 (40%) represents poor agreement (Congalton, 1996; Jensen, 2005).

Table 11 shows the comparison of the different kappa values for each individual class for the three classifiers in consideration and the corresponding overall kappa values.

Table 11: Comparison of the kappa values for the three classifiers

Class names	Unsupervised ISODATA clustering	Supervised classification with Feature Analyst	Proposed Hybrid Classification
Shrub complex	0.5659	0.5396	0.6317
Aquatic complex	1.0000	0.8347	1.0000
Barren complex	0.8229	0.5868	0.8365
Snowbed complex	0.3775	0.2487	0.6009
MAT^{**} complex	0.4996	0.6877	0.5273
WST^{**} complex	0.4557	0.3347	0.5917
MNT^{**} complex	0.4641	0.4202	0.7442
OVERALL	0.5904	0.5418	0.6840

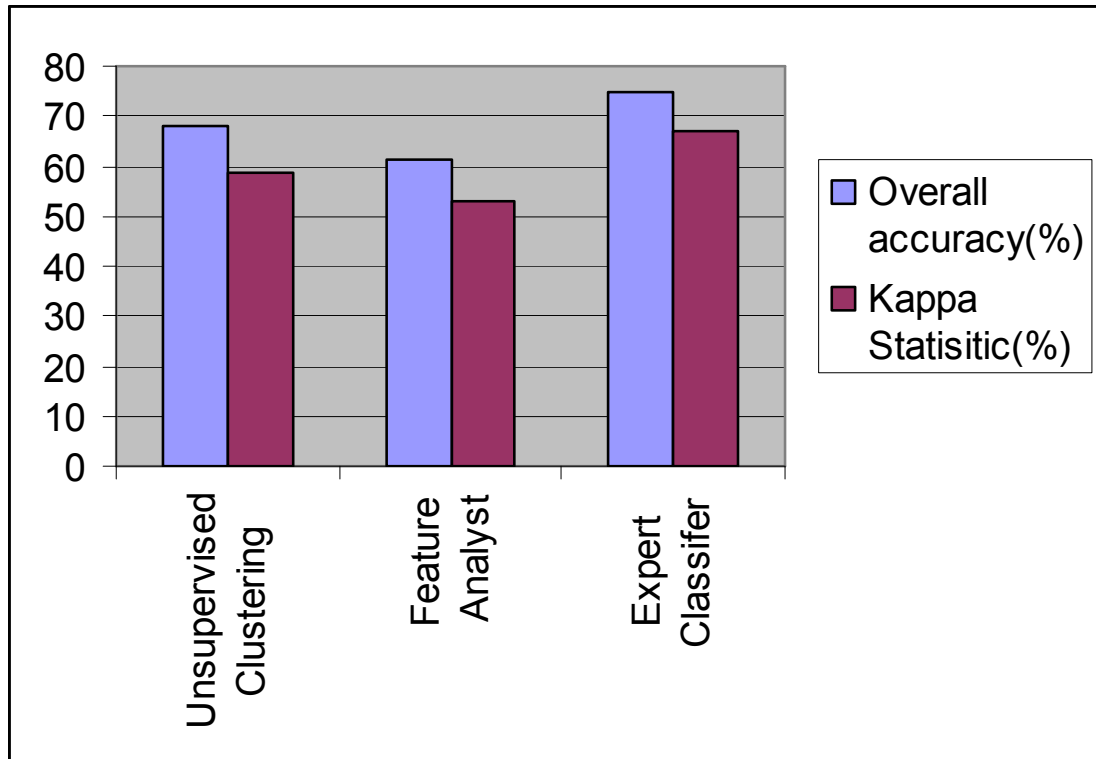


Figure 39: Showing the comparison of the accuracy values for the three classifiers

4.1.2. Statistical Significance of the Accuracy Assessment

To compare the performance of the proposed hybrid classifier with two other standard classification techniques: unsupervised ISODATA clustering and supervised classification with Feature Analyst, a comparison between three different kappa values was performed. Estimated variance for the three different kappa values from their respective error matrices was calculated using SAS 9.1 (Figure 40) (Congalton 1999; Fleiss 1969; Shine et al. 1999).

```

data accuracy;
input observed $ predicted $ count;
datalines;
shrub shrub 12
shrub water 0
shrub barren 0
shrub snowbed 0
shrub MAT 2
shrub wst 2
shrub mnt 2
water shrub 0
water water 30
water barren 0
water snowbed 0
water MAT 0
water wst 0
water mnt 0
barren shrub 0
barren water 0
barren barren 11
barren snowbed 0
barren MAT 1
barren wst 1
barren mnt 0
snowbed shrub 0
snowbed water 0
snowbed barren 0
snowbed snowbed 8
snowbed MAT 1
snowbed wst 2
snowbed mnt 2
MAT shrub 8
MAT water 0
MAT barren 0
MAT snowbed 0
MAT MAT 62
MAT wst 6
MAT mnt 15
wst shrub 1
wst water 0
wst barren 1
wst snowbed 0
wst MAT 0
wst wst 5
wst mnt 1
mnt shrub 0
mnt water 0
mnt barren 1
mnt snowbed 0
mnt MAT 6
mnt wst 2
mnt mnt 39
;
proc freq data=accuracy order=data;
weight count;
tables observed*predicted / agree norow nocol;
test kappa wtkap ;
run;

```

Figure 40: SAS Program to calculate kappa variance from error matrix for hybrid classifier

Let K_1 is the estimated kappa for unsupervised ISODATA clustering, $\text{var}(K_1)$ is its estimated variance, K_2 is the estimated kappa for supervised classification with Feature Analyst and $\text{var}(K_2)$ is its estimated variance, K_3 is the estimated kappa for proposed hybrid classification and $\text{var}(K_3)$ is its estimated variance. To test the hybrid classifier with each of the standard classification procedures, a standardized normal variable Z was used, and the null hypothesis (H_0) that the two kappa values concerned are equal versus the alternative (H_A) that they are not equal. Z is given by:

$$Z = \frac{K_3 - K_1}{\sqrt{(\text{var}(K_3) + \text{var}(K_1))}}, \text{ for comparing unsupervised ISODATA clustering}$$

and the proposed hybrid classification (Table 12) and,

$$Z = \frac{K_3 - K_2}{\sqrt{(\text{var}(K_3) + \text{var}(K_2))}}, \text{ for comparing supervised classification with}$$

Feature Analyst and the proposed hybrid classification (Table 13).

Z was compared against normal distribution functions and rejected if $|Z|$ was greater than 1.96 at a 95% significance level (Shine et al. 1999).

Table 12: Statistical significance between hybrid classifier and unsupervised classification with ISODATA clustering

	Proposed Hybrid Classifier	Unsupervised Classification with ISODATA clustering
KAPPA	0.6840	0.5904
KAPPA VARIANCE	0.00059049	0.00064009
KAPPA STD ERROR	0.0243	0.0253
Z VALUE	2.6682	**SIGNIFICANTLY DIFFERENT

Table 13: Statistical significance between hybrid classifier and supervised classification with Feature Analyst

	Proposed Hybrid Classifier	Supervised Classification with Feature Analyst
KAPPA	0.6840	0.5418
KAPPA VARIANCE	0.00059049	0.00071824
KAPPA STD ERROR	0.0243	0.0268
Z VALUE	3.9307	**SIGNIFICANTLY DIFFERENT

4.2. Discussions

Eight land-cover classes were identified from image classification process from the SPOT-5 data using the three different techniques. Figure 41 shows the comparison of the three final maps obtained from the three different techniques.

4.2.1. Comparison of classification performance for each class: error matrices, producer's and user's accuracies, and kappa values.

The results obtained for the different classes and their interpretations in context of the three different classifiers are described as follows:

Moist Low-Shrub Tundra and other Shrublands complex had a highest producer's accuracy (80.95%) in the unsupervised classification ISODATA clustering but the user's accuracy was low (60.71%) due to the fact that it was over estimating the class and most of the confusion was with *Moist Dwarf-shrub, Tussock-graminoid Tundra complex* class. Hybrid classification had the highest user's accuracy (66.67%), producer's accuracy (57.14%) lower than ISODATA clustering and highest kappa value (0.6317) for this class. Supervised classification with Feature Analyst had a higher producer's accuracy (66.67%) than the hybrid classifier but lower than ISODATA clustering, lowest user's accuracy (58.33%) and lowest kappa value (0.5396).

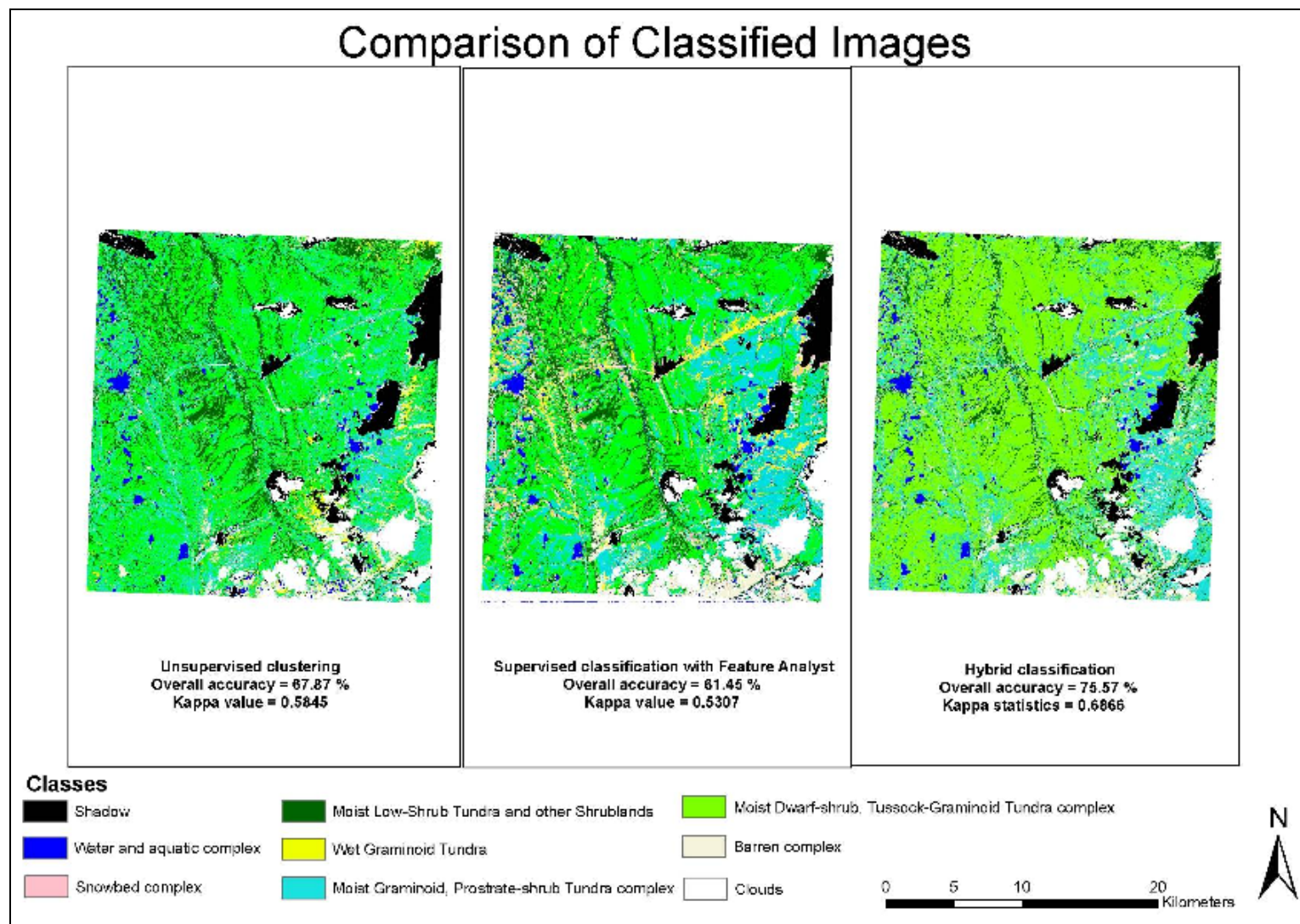


Figure 41: Comparison of the three classified images.

Water and aquatic complex was classified with an accuracy of 100%, (both producer's and user's) by both ISODATA clustering and the hybrid classification technique. Supervised classification with Feature Analyst had a producer's accuracy of 100% but the user's accuracy was lower than the other two classifiers (85.71%).

Barren complex had a highest producer's accuracy (84.62%) and user's accuracy (84.62%) by the hybrid classifier, where as supervised classification with Feature Analyst had a higher producer's accuracy (84.62%) than unsupervised ISODATA clustering (76.92%), but a lower user's accuracy (61.11%) compared to unsupervised ISODATA clustering (83.33%). Hybrid classifier had the highest kappa value of 0.8365, followed by ISODATA clustering (0.8229), supervised classification with Feature Analyst being the lowest (0.5868).

Snowbed complex had a highest producer's accuracy (100.00%) and user's accuracy (61.54%) by the hybrid classifier, where as supervised classification with Feature Analyst also had a producer's accuracy (100.00%) but the user's accuracy was very low (27.59%). Unsupervised ISODATA clustering had a poor user's (40.00%) and producer's (25.00%) accuracy for this class. Hybrid classifier had the highest kappa value of 0.6009, followed by ISODATA clustering (0.3775), supervised classification with Feature Analyst being the

lowest (0.2487). This class had the most confusion with *Wet Graminoid Tundra* complex especially those around the lake and shallow water margins.

Moist Dwarf-shrub, Tussock-graminoid Tundra complex is the class that is most extensive in coverage area and had a highest producer's accuracy (86.11%) by the hybrid classifier, followed by unsupervised ISODATA clustering (76.39%) and supervised classification with Feature Analyst (62.50%). In terms of user's accuracy, supervised classification with Feature Analyst (78.95%) was the highest for this class, followed by the hybrid classifier (68.13%) and unsupervised ISODATA clustering (66.27%). The most confusion for this class was with *Moist Graminoid, Prostrate-shrub Tundra* complex. Supervised classification with Feature Analyst had the highest kappa value of 0.6877, followed by the hybrid classifier (0.5273), unsupervised ISODATA clustering being the lowest (0.4996).

Wet Graminoid Tundra complex was the class that a very low accuracy rate for all the classifiers. The highest producer's accuracy (38.89%) was achieved by supervised classification with Feature Analyst, where as the hybrid classifier had a producer's accuracy of 27.78% and unsupervised ISODATA clustering had 16.67% for the same. The hybrid classifier had the highest user's accuracy (62.50%), followed by unsupervised ISODATA clustering (50.00%), and supervised classification with Feature Analyst (38.89%). This class had major confusions with *Moist Graminoid, Prostrate-shrub Tundra* complex and *Snowbed*

complex. In terms of kappa values, the hybrid classifier was the highest with a value of 0.5917, followed by ISODATA clustering (0.4557), and supervised classification with Feature Analyst (0.3347).

Moist Graminoid, Prostrate-shrub Tundra complex had both the highest producer's (66.10%) and user's (81.25%) accuracy achieved by the hybrid classifier. Unsupervised ISODATA clustering had the next highest value for both producer's (57.63%) and user's (60.71%) accuracy, followed by supervised classification with Feature Analyst with producer's (38.98%) and user's (57.50%) accuracy (38.98%) being the lowest for this class. This class being the second biggest class in terms of area, had major confusions with *Moist Dwarf-shrub, Tussock-graminoid Tundra* complex. In terms of kappa values, the hybrid classifier was the highest with a value of 0.7442, followed by unsupervised ISODATA clustering (0.4641), and supervised classification with Feature Analyst (0.4202).

4.2.2. Overall Accuracy and Statistical Significance

As evident from table 6, the hybrid classifier had the highest overall accuracy rate of 75.57% (167 correct pixels out of 221) in comparison to unsupervised ISODATA clustering (68.33%; 151 out of 221) technique, which is the most used image classification technique used in the area so far, and supervised classification with Feature Analyst (62.44%; 138 out of 221), having the lowest overall accuracy.

Table 9(i-ii), shows that the improvement of the hybrid classifier in terms of accuracy is significantly different statistically from the other two classifiers: unsupervised ISODATA clustering and supervised classification with Feature Analyst. A Z statistic value of 2.6682 (> 1.96) indicates that the null hypothesis (H_0) stating that the results from the hybrid classifier and the unsupervised ISODATA clustering are the same, can be rejected at a significance level of 95%. Similarly, with a Z-value of 3.9307 (> 1.96) indicates that results from the hybrid classifier and supervised classification with Feature Analyst are different, at a significance level of 95%.

4.2.3. Confusion between vegetation classes

Except for the water class, there was confusion between the vegetation types on the classified maps. This section specifically discusses the interclass confusions in the classified map made by the hybrid classifier. As evident from the error matrix, (Table 5-iii), most of the confusions involved *Moist Dwarf-shrub, Tussock-graminoid Tundra* and *Moist Graminoid, Prostrate-shrub Tundra* complex, *Moist Dwarf-shrub, Tussock-graminoid Tundra* and *Moist Low-Shrub Tundra* and other *Shrublands* complex.

Moist Dwarf-shrub, Tussock-graminoid Tundra and *Moist Graminoid, Prostrate-shrub Tundra* complexes were the two most extensive vegetation classes in terms of area of coverage. About 16.5% of the pixels that were, classified as *Moist Dwarf-shrub, Tussock-graminoid Tundra* (typical moist acidic

tundra) were misclassified as *Moist Graminoid, Prostrate-shrub Tundra* (mostly moist non-acidic tundra), where as, about 13% of the pixels that were classified as *Moist Graminoid, Prostrate-shrub Tundra* were misclassified as *Moist Dwarf-shrub, Tussock-graminoid Tundra*. The transition zone between moist acidic tundra and moist non-acidic tundra where spatial extent of one class ends and the other begins, represents the shift between non-tussock sedge, erect dwarf-shrub, moss tundra and tussock sedge, dwarf-shrub, moss tundra. The moist acidic side of the transition is dominated by vegetation species like *Betula nana*, *Salix* spp., *Eriophorum vaginatum*, *Carex bigelowii*, and *Sphagnum* spp. etc., while, the non-acidic side rarely has *Betula nana*, more *Carex bigelowii* and *Sphagnum* spp. are absent. (Jia et al. 2002). This suggests that there was a lot of spectral intermixing between the two classes and most of these intermixing were present in the transition region.

Research by Gough et al. (2000) suggest that there are no strict spatial boundary between acidic ($\text{pH} < 5.5$) soil and non-acidic ($\text{pH} > 5.5$) soil due to old and new geologic sites respectively in a local scale at 5m pixel spatial resolution (as opposed to in a regional scale and 50m pixel resolution (Walker et al. 1995; Jia et al. 2002)), species found in acidic sites were also found in non-acidic sites. Thus there were ample number of sites found in this research where occurrence of moist acidic tundra being dominated by moist non-acidic tundra and vice versa in the transition zones. In these situations it was not impossible to differentiate between the two at a 5m pixel level and spectral resolution of SPOT-5 satellite.

Moist Dwarf-shrub, Tussock-graminoid Tundra and Moist Low-Shrub Tundra and other Shrublands complex were spectrally similar in many instances when the domination of low shrub canopy (*Salix pulchra*, *Betula nana*, *Salix glauca*) over tussock tundra (*Sphagno-Eriophoretum vaginati*) changes in height and complexity in some occasions. The boundary between these two classes occurred as irregular and intermittent patches that represents the shift from tussock sedge, dwarf-shrub, moss tundra to either erect dwarf-shrub tundra or low shrub tundra shows gradual increases in the abundance and biomass of deciduous dwarf birch (*Betula nana* L.) and willow (*Salix pulchra* Cham.) shrubs (Jia et al. 2002). This gradual transition might have caused spectral mixing that might have contributed to low accuracy results. *Moist Low-Shrub Tundra and other Shrublands* complex is more accurately identified in cases of willow dominated uplands areas dominated by dwarf and low shrubs mainly on interfluvial areas, water-track complexes, streams and rivers (riparian shrubs), and some floodplain areas.

Besides, the classes discussed above, the most problematic vegetation class to identify, evidently having lowest accuracy rate and kappa value was *Wet graminoid tundra* complex. Other researchers also found low accuracy for similar species combination (wet sedge tundra) (Markon, 1994; Stow et al. 2000; Hope et al. 1999). This class consisting of fens, dominated by sedges and mosses is mostly found in wetland areas, margins of lake basins, watertracks and foothills. In most cases misrepresentation occurred in isolated occurrences of low centre

polygon areas, where the centers containing sedges in wet to flooded conditions and the wide rims were dominated by *Moist Low-Shrub Tundra and other Shrublands* (*Eriophorum angustifolium*-*Salix pulchra* and *Salix pulchra*-*Calamagrostis Canadensis*), or *Moist Dwarf-shrub, Tussock-graminoid Tundra* (typical moist acidic tundra) type vegetation. In some cases, *Wet graminoid tundra* complex (specifically rich fen; *Carex chordorrhiza*, *Carex aquatilis*, *Eriophorum angustifolium*) was misidentified as *Moist Low-Shrub Tundra and other Shrublands* (riparian shrubs, specifically) due to the high spectral reflectance in band 3 of the SPOT image (Figure 42). In some cases when the occurrences were less wet than average, it was often confused with *Moist Dwarf-shrub, Tussock-graminoid Tundra* (dry acidic/non-acidic tundra; occurrences of *Dryado integrifolia*-*Caricetum bigelowii* and *Juncus biglumis*-*Saxifraga oppositifolia*). Also, in few cases, there were some confusion with *snowbed* complex, specially occurring in gentle slopes and shadowy areas. This class was found to be more accurate when located adjacent to lake basins and extensive gently sloping low lands draining into lakes.

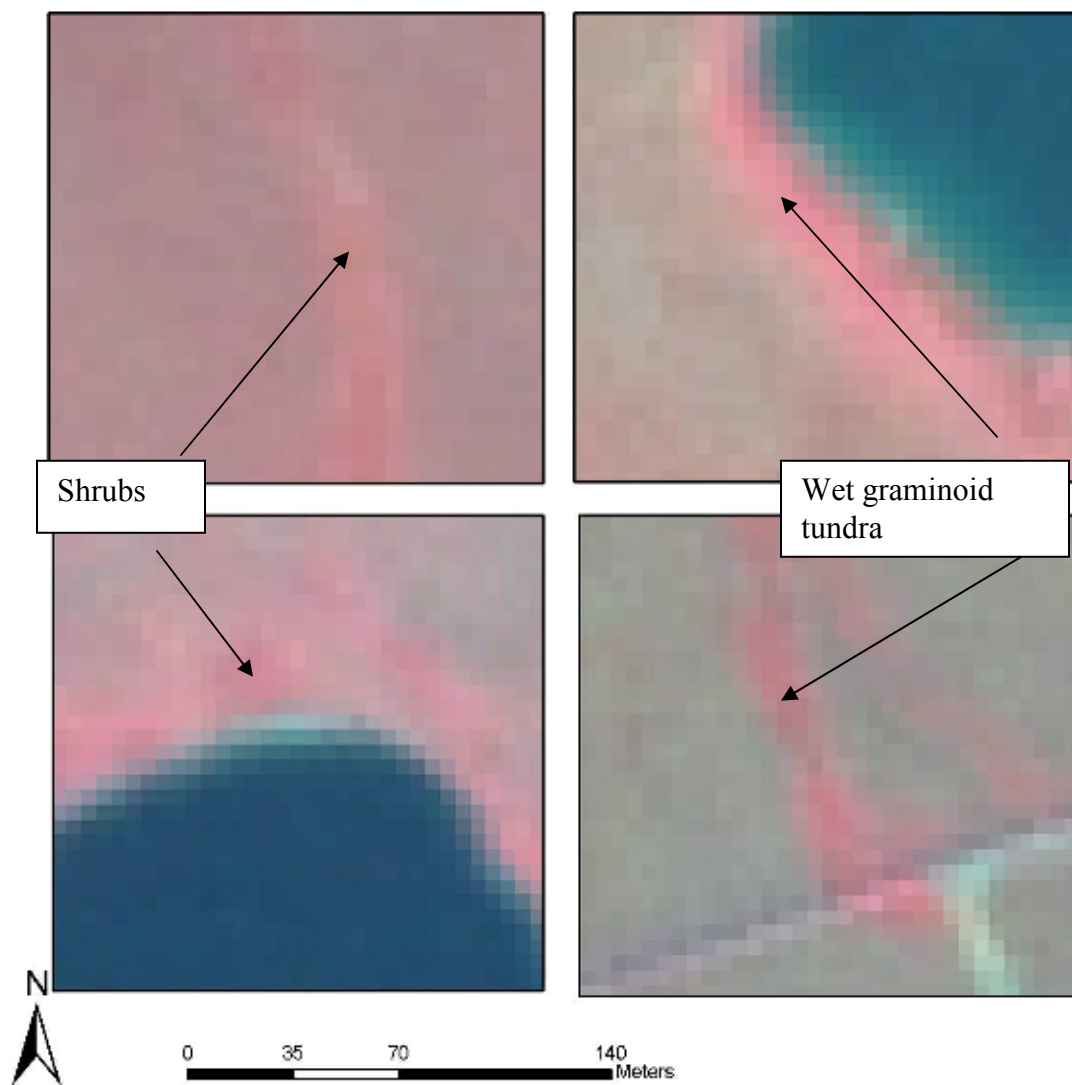


Figure 42: Spectral similarity between shrubs (riparian) on the left and wet graminoid tundra complex (rich fen) on the right.

4.2.4. Solutions to the research questions

This research, SPOT-5 image classification of the Toolik area, is first time when a SPOT image resampled to 5 meter pixel have been used in land-cover classification of this area. It shows more recent and detailed vegetation cover than the existing Walker et al (1994) map made from Landsat MSS data. This section answers the specific research questions framed towards the end of the first chapter in this dissertation.

- How to extract spatial and spectral knowledge for the unique arctic tundra vegetation type that can be utilized for expert classification?

The set of rules that were developed to build the knowledge base for the rule-based classification, as a part of the hybrid classification process showed that both spatial and spectral knowledge can be efficiently extracted and be used in expert classification process. The set of rules developed can be generalized and used in other parts of the SPOT image that were outside in the study area. Use of spatial (slope and aspect) and spectral (SPOT bands, NDVI etc.) knowledge together, in the form of if-then rules in the rule-based classifier enabled the extraction and classification of relatively complex classes like *snowbed* complex with considerable better efficiency than the other traditional classifiers used.

- How can a hybrid classifier be used to classify SPOT-5 data (resampled to 5 meter pixel) to achieve higher classification accuracy than traditional classification techniques used using actual ground truth data?

This research showed that using the proposed hybrid classification technique to classify SPOT-5 data to achieve higher classification accuracy than traditional classification techniques used in this area. The hybrid classifier produced an overall accuracy of 75.57%, in comparison of 68.33% from unsupervised ISODATA clustering and 62.44% from supervised classification with Feature Analyst (Table 6). The overall kappa values (Table 8) also suggest that the proposed hybrid classifier (0.6840) is superior to unsupervised ISODATA clustering (0.5904) and supervised classification with Feature Analyst (0.5418). For the individual classes, except for *Moist Dwarf-shrub*, *Tussock-graminoid* *Tundra* all the other classes have higher individual kappa values in the classified map by the hybrid classifier than the other two classification techniques (Figure 43). The overall accuracy obtained was found to be greater than the existing classification of the arctic tundra from a SPOT-4 satellite by Stow et al. (1989), having overall accuracy of about 56%.

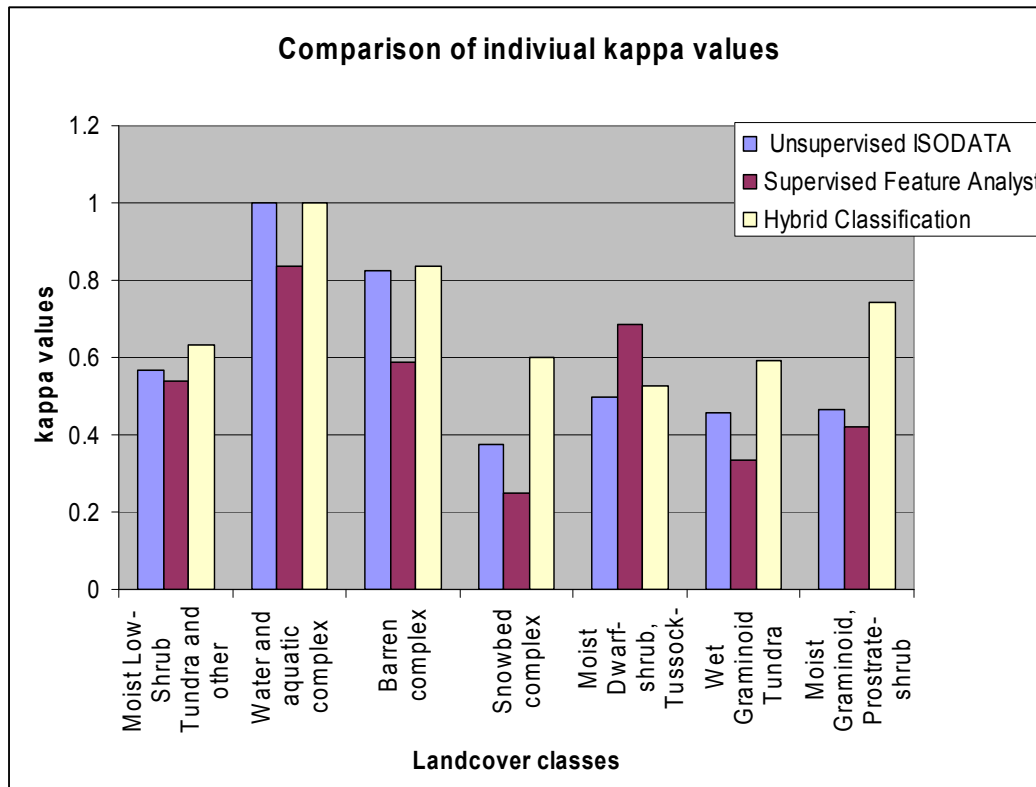


Figure 43: Comparison of the three classifiers in terms of individual kappa values of the classes.

- What are the statistical significances of the classification accuracy obtained from the proposed method as compared to traditional spectral classifiers used in that area?

The error matrices from the three different classifiers were compared statistically to determine whether results of the hybrid classifier was significantly different of the results obtained from the other two classification techniques. A Z-statistic value greater than > 1.96 in both the cases, indicates that results from the hybrid classifier and the two other traditional classification techniques are different, at a significance level of 95%.

This research meets the stated research objectives to develop a hybrid image classification approach that effectively integrated ancillary information (slope, aspect, NDVI etc) into the classification process and combined unsupervised ISODATA clustering, rule-based classifier and the Multilayer Perceptron (MLP) classifier in a sequence that typically classified arctic tundra type vegetation in SPOT-5 (resampled at 5m spatial resolution) satellite image that yields higher accuracy than the existing classification techniques applied on SPOT data in this region. Application of data mining tool (MLP classifier that implements artificial neural network) with geo-spatial and spectral knowledge (developed knowledge base for the rule-based classifier) for hybrid classification technique produced better results in classifying unique arctic tundra vegetation in Alaska. Comparing and contrasting the proposed hybrid image classification with the standard (spectral) classification techniques: unsupervised ISODATA clustering and supervised classification with Feature Analyst, showed that the hybrid classification technique produces better results.

Of particular importance was the identification of the *snowbed* complex that was not identifiable on the existing resampled 50 m Landsat data classification (Auerbach et al. 1997b; Muller et al. 1998). Instead of five vegetation classes in the classified map from Landsat data, this research extracted six major vegetation classes. The 5m spatial resolution of SPOT-5 data allowed identification of smaller and isolated parcels of typical tundra vegetation and the unique *snowbed* which was not detectable on the Landsat data

resampled to 50 m resolution. Also this research depicted the formulation of a new spectral index: the tundra index. The use of the tundra index with commonly used NDVI, produced better results in the training of the neural network of the MLP classifier (Table 3).

4.2.5. Research Concerns and Limitations

According to Anderson (1976) and Congalton et al. (1999), a classified map with an 85% overall accuracy and kappa value greater than 0.70 is considered valid and acceptable to be applied toward other scientific applications. The classified map from the proposed hybrid classifier (75.57% overall accuracy and kappa 0.6840) in this research, still needs improvements to meet this goal of using this classified map in other scientific investigations and ecological applications. Several factors may have had a significant influence that affected the accuracy results of the hybrid classification are discussed below.

The use of SPOT-5 imagery may be an important factor that influenced the accuracy results of the classified map. Although the spatial resolution (5m resampled) was very inviting to discriminate the small heterogeneous vegetation parcels of typical arctic tundra vegetations, SPOT-5 offers very limited spectral resolution in the green, red and near-infrared portion of the electromagnetic spectrum which limits separation of vegetation in this very heterogeneous landscape. The existing Landsat MSS classification by Auerbach et al. (1997b) was based on homogeneity of 3X3 pixels at 50m pixel level which depicts a

generalized classification of the area because it is hard to find a homogenous extent of a vegetation cover even at a 5m pixel level except for the two classes: *Moist Dwarf-shrub, Tussock-graminoid Tundra* complex and *Moist Graminoid, Prostrate-shrub Tundra* complex. The use of SPOT-5 imagery at 5m pixel level in this research focused on the heterogeneity of the tundra landscape in general.

DEM creation is an inherently inaccurate process that always inevitably incorporates some errors. The ancillary data used, specifically slope and aspect in the hybrid classification was derived from a DEM data (Intermap Technologies Corporation STAR-3i) at a 5m resolution level with RMSE of 2.5m horizontal accuracy. Although slope and aspect data being used as ancillary data helped the classification process, since the DEM data was not fully accurate, the derived information slope and aspect were having some inaccuracies. The NDVI data used was derived from the resampled SPOT-5 (band 3 and band2) data, which might not have reflected the true NDVI values for the vegetation under consideration. These factors might have contributed towards the lower accuracy than expected for the hybrid classification.

The SPOT-5 to begin with had severe georectification errors (Figure 10). The image was georectified with 64 ground control points with an RMSE error of 4.99m. Ideally the rectification RMSE should be less than 2.5m (half a pixel) in order to be successfully used in a scientific image classification (Jensen, 2005). The major problem faced in this research with georectification is the absence of

proper locations like road intersections, urban features that are prominently and precisely identifiable on the image during the image registration process. The Dalton Highway, the pipeline and the Toolik field station provides some GCP locations but that covers only the central and NE portion of the image. In order to have proper georectification process there is a need of uniform distribution of the GCPs through out the image, which is a major limitation in this study that might have contributed towards the low accuracy results.

Considering the heterogeneity of the typical arctic tundra vegetations there was a major problem regarding having several sample sites and testing sites on transition edges or boundaries between one vegetation class and another. These sample locations had spectral mixing effect in the pixels which could have contributed in assignment of different class. Also while classifying the two major classes *Moist Dwarf-shrub*, *Tussock-graminoid Tundra* and *Moist Graminoid*, *Prostrate-shrub Tundra* complex, it was very difficult to differentiate and interpret the dominant one in the transition boundary zone (Figure 44). Also *Wet graminoid tundra* complex (specifically rich fen; *Carex chordorrhiza*, *Carex aquatilis*, *Eriophorum angustifolium*) was misidentified as *Moist Low-Shrub Tundra and other Shrublands* (riparian shrubs, specifically) due to the high spectral reflectance in band 3 of the SPOT image (Figure 42). Overall, it was hard to find homogenous sample locations for the classes in the heterogeneous landscape of the study area. Thus it could be the case when the training sample

sites had spectral mixing or the identification of the clusters from ISODATA clustering of the hybrid classifier was wrong, using those sample points.

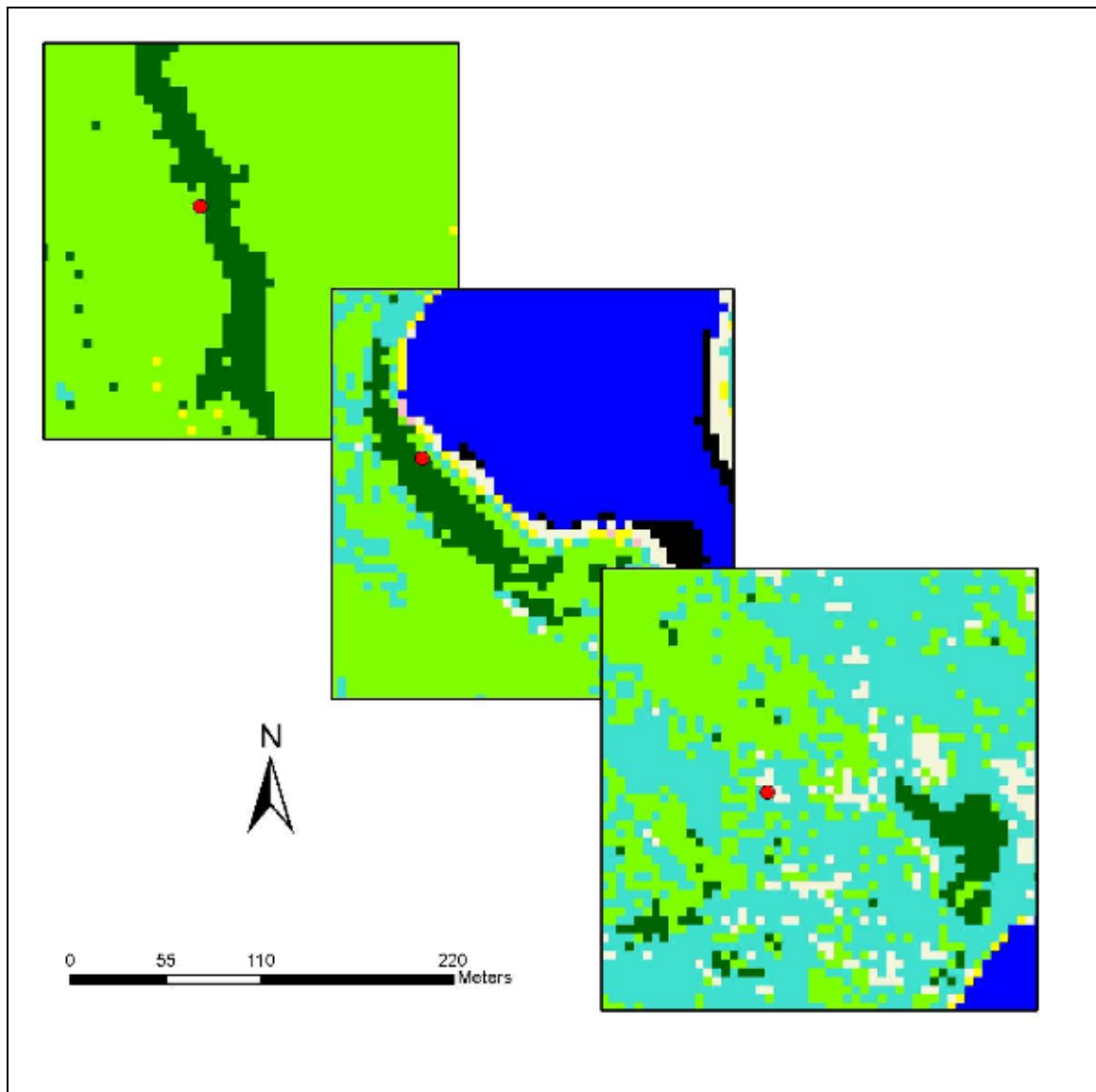


Figure 44: Spectral mixing in the transition zones: the effect of heterogeneity in the area.

According to Congalton et al. (1999), for less than, 10 land-cover classes, a good rule of thumb is to collect a minimum of 50 samples for each land-cover class to reach a confidence of 95% in the accuracy assessment. This means the total number of sample points should be 350, for the seven land-cover classes for this study. But the total number of points collected through the stratified random sampling technique and used for the accuracy assessment in this study was only 221 which is a limitation of this research. Only two classes, *Moist Dwarf-shrub*, *Tussock-graminoid Tundra* and *Moist Low-Shrub Tundra and other Shrublands* complex which had greatest spatial extent in the study area had more than 50 sample points for the accuracy assessment. The overall distribution of the sample points were more clustered around Lake Toolik and the Dalton Highway in the central and north eastern portion of the image. Also, in the MLP classifier the number of training pixels used was 2400, which is more than the minimum requirement, but less than the optimum number of training pixels which is 4320. All of the above mentioned limitation can be explained by the inaccessibility of the terrain, limited helicopter hours and research budget, and bad weather in which sometimes it was not possible to fly. In addition to this, 45 sample points were discarded since they were within the 300m buffer around the Dalton Highway (Figure 45). Previous research suggests that the ecological effects of road and road dust disturbances due to the Dalton highway, causes pronounced effects on the substrate and vegetation properties in the arctic tundra (Auerbach et. al., 1997a; Walker et. al. 1987; Forbes et. al. 1999). As a

result of this, the spectral properties of the vegetation are affected very prominently (overall accuracy of the classification of this section was lowered to 50% and thus the collected points within 300m around the Dalton Highway had to be discarded.

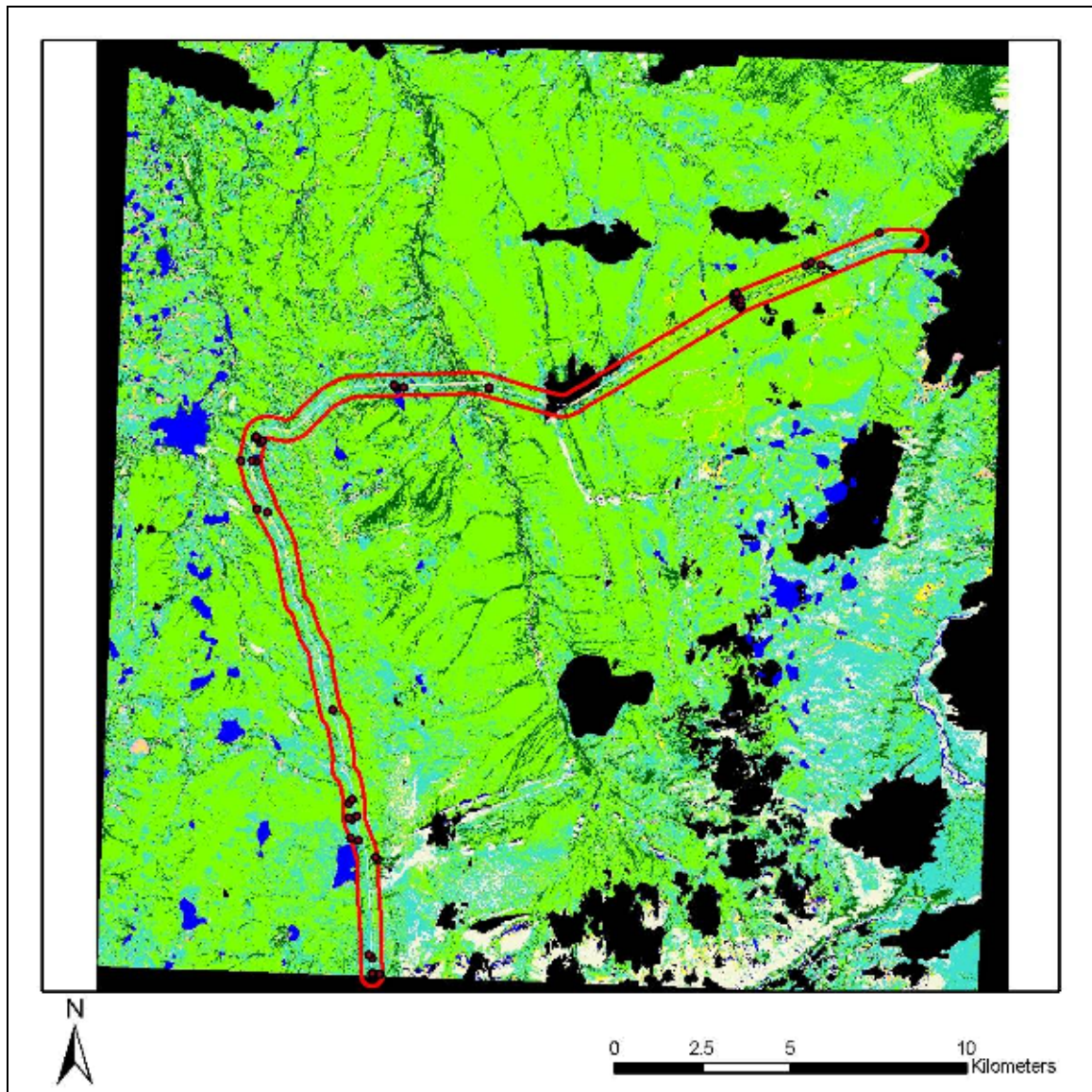


Figure 45: Sample points discarded which were collected within 300m (buffer) from the Dalton road.

Texture (variance) of the different bands of the SPOT image and the NDVI layer were used as input layers of the MLP classifier during the trial runs. It was observed that the inclusions of the texture did not improve the training accuracy

of the neural network. This might be explained by the fact that the SPOT data being pan-sharpened (resampled with the help of high resolution panchromatic band) to 5m pixel in order to increase the spatial resolution, the original reflectance (spectral) values of the pixels was not retained. As a result the texture layers might not be representing the true texture values of the pixels. Thus addition of the texture layers as input did not improve the training accuracy of the neural network classifier and hence were not considered in the final set of input layers.

It was found that there was no strict spatial boundary between moist acidic tundra and non-acidic tundra at a 5m pixel level in this heterogeneous landscape. However it was evident at a larger spatial resolution (50m pixel level) and regional scale as suggested by Walker et al. (1995) and Jia et al. (2002). An additional geology layer in this study would have been very useful in the data mining analysis of the MLP classifier used as part of the hybrid classifier. Unfortunately lack of geological data with suitable resolution spatial for this research, prevented the use of geology as an additional ancillary data layer to acquire better results in terms of accuracy.

CHAPTER V

CONCLUSIONS

Detailed and accurately classified map for land-cover map of the arctic tundra landscape in this region is a research need for many scientists all over the world. Use of SPOT-5 satellite imagery resampled at 5m pixel spatial resolution in this research has been a unique venture in order to extract more detailed classification of the heterogeneous landscape. The previous satellite classifications in this area mostly used Landsat data (at a spatial resolution of 80m to 50m, resampled), thus neglecting the detail and heterogeneity of the typical arctic tundra landscape.

The general objective of this research was to develop a hybrid image classification technique combining ISODATA clustering (unsupervised), expert classifier (rule-based) and the Multilayer Perceptron (MLP) classifier that effectively integrates data mining techniques with geo-spatial and spectral knowledge, and ancillary information into the classification process. The main goal was to find out the best possible combination or sequence of these classifiers for typically classifying tundra type vegetation in the SPOT-5 satellite image that yields higher accuracy than the existing classified vegetation map at the 5 meter spatial resolution level. The objectives also included comparing and

contrasting the image classification performance between the proposed hybrid classification and the standard (spectral) classification techniques: unsupervised ISODATA clustering and supervised classification with Feature Analyst. The main emphasis here was to explain the detailed heterogeneity of the tundra vegetation cover at a 5m scale, rather than the homogeneity of landscape as emphasized by the earlier studies. The conclusions of this dissertation and the recommendations for future research studies are given in this section.

5.1. Conclusions

This research showed the efficient use of a set of rules extracting both spatial and spectral knowledge to develop a knowledge base for the rule-based classification of arctic vegetation, as a part of the hybrid classification process. Use of spatial (slope and aspect) and spectral (SPOT bands, NDVI etc.) knowledge together, in the form of if-then rules in the rule-based classifier enabled the extraction and classification of relatively complex classes like *snowbed* complex with considerable better efficiency than the other traditional classifiers used.

The methodology uniquely combines together the two different approaches stated by Liu et al. (2002) to produce the hybrid classifier, and used the kappa values, in the consensus building step in order to improve the classification accuracies. It can be concluded from this research that hybrid classification technique to classify SPOT-5 data is more appropriate than

traditional classification techniques used in this area, in previous research studies. The hybrid classifier that combined ISODATA clustering (unsupervised), rule-based classifier and the Multilayer Perceptron (MLP) classifier, was found to be superior both in terms of overall accuracy and kappa analysis to the use of only unsupervised ISODATA clustering or supervised classification with Feature Analyst. The proposed hybrid classifier was able to extract all the individual classes, except for *Moist Dwarf-shrub*, *Tussock-graminoid Tundra* better than other two classifiers. The overall accuracy assessed was found to be greater than existing classification of the arctic tundra from a SPOT-4 satellite by Stow et al. (1988), which had an overall accuracy of about 56%.

It can also be concluded from this research that results of the accuracy assessment of the hybrid classifier were significantly better than that of the two other classifiers, at a confidence level of 95%.

Besides the above conclusions there are several other important outcomes of this research. Firstly, the hybrid classifier, using if-then rules and the artificial neural network, is nonparametric nature and it is easy to add ancillary layers to it than. It also doesn't require any statistical assumption about the distribution of the training sets such as normal distribution as required by standard ISODATA clustering. It requires much more effort and time to build the set of rules and train the neural network of the hybrid classifier. However, once the rules are made finally, the knowledge base of the rule-based classifier and

the weights obtained from the neural network learning iterations can be saved and used in other parts of the SPOT image that the classifier have not seen; in other words the knowledge can be transferred.

This research showed the use of a new spectral index called tundra index which was specifically formulated to help the differentiation of moist acidic and moist acidic tundra. It is a linear transformation of the spectral values of the SPOT bands given by

$$\frac{2 * IR - (Green + Red)}{2 * IR + (Green + Red)}$$

It was observed that the inclusion of the tundra index as one of input layers in the MLP classifier improved the average training accuracy from 90.79% to 93.75%, and average training kappa values from 0.95 to 0.97 in a single hidden layer architecture.

None of the satellite image classifications done so far in this region stated in the literature have used this hybrid technique. Although the final classified still map needs some improvement in terms of accuracy measures before it can be used in other scientific researches, the proposed hybrid classification technique showed a major improvement than the traditional classifiers. With those improvements, the classified map will be important to studies pertaining to biogeochemical cycling, landscape ecology, wildlife habitat, ecosystems,

hydrology, and general resource management applications in the arctic. An improved understanding of the vegetation species distribution may therefore provide insight into the effect of a changing environment on vegetation and ecosystem change. With the use of SPOT satellite data, this research provided yet another great opportunity to utilize the scope of remote sensing toward an enhanced and updated understanding of arctic tundra vegetation distribution in study area with constantly changing landforms and land-cover types.

5.2. Future Research Recommendations

In future, application of hyperspectral data (e.g. Advanced Land Imager (ALI), Hyperion) instead of SPOT imagery may be a very interesting research that might offset the low spectral resolution of SPOT. The classes having spectral mixing effect like *Moist Dwarf-shrub, Tussock-graminoid Tundra* complex and *Moist Graminoid, Prostrate-shrub Tundra* complex may be extracted with better accuracy. Although a resampling technique might be applied to obtain optimum spatial resolution to discriminate heterogeneous parcels of tundra vegetation. Also the greater number of spectral bands from the hyperspectral data, as input layers to the artificial neural network could provide a better data mining opportunity for MLP classifier used in the hybrid classification.

Using DEM data with better horizontal and vertical accuracy might be good improvisation from this research. As result of more accurate DEM data, the derived slope and aspect would be expected to be more accurate than that used

in this research. A better accuracy estimate can be obtained with these improved set of ancillary data. NDVI data produced from hyperspectral data might also have a better use as ancillary data to produce better accuracy results.

There is a severe research need of creation of ground control points for georectification of satellite imagery in this region of Alaska. Since, as stated in this research, there is lack of proper locations like road intersections, urban features that are prominently and precisely identifiable on the image and used as ground control points during the image registration process. The Dalton road, the pipeline and the Toolik field station provides some GCP locations but they are clustered and located only in the central and eastern portion of the area. To have accurate georectification process there is a need of uniform distribution of the GCPs through out the image area.

Future additions to this research will involve collection of more ground truth samples in the form of points, lines and polygons from areas that were not possible to visit during the last two summers. The collected points would be used to increase both the number of training points as well as the number of points used for accuracy assessments. Also the new sample points collected in homogenous pixels will be used to replace the training points having spectral mixing effects in the transition zones between vegetation classes to ensure better training. Addition of more points for accuracy assessment, distributed evenly in the study area would offset the research limitation of not having optimum number

of points for accuracy assessment and thus provide better insight of the accuracy assessment of the proposed hybrid classification technique.

In future a glacial geology map with appropriate spatial resolution could be added as an input layer to neural network. Incorporation of geology could be very useful to differentiate acidic and non-acidic vegetation. Also using a geomorphology layer with hyperspectral data might be proved useful to differentiate and extract moist non-acidic tundra, dry non-acidic tundra and dry acidic tundra classes which in this research were categorized as a single class: *Moist Graminoid, Prostrate-shrub Tundra* complex.

For the hybrid technique applied, it can be experimented with some or all of the following research alternatives: 1) kNN clustering technique replacing the ISODATA (C-means) clustering in the first stage of the hybrid classification, 2) the neural network (MLP) can be replaced with a decision tree classifier (C 5.0), 3) incorporation of NHD plus hydrographic data (stream flowline, waterbodies) to extract out the riparian shrubs as a separate class from the shrub complex, and 4) fuzzifying the output of the MLP classifier and bringing in another rule-base classifier from fuzzy output.

REFERENCES

- Al-AbdulKader, K. A., Blundell, J.S., & Farrand, W.H.** (2002). Marine Habitat Mapping Using High Spatial Resolution Multi-spectral Satellite Data. *Saudi Aramco Journal of Technology*. Retrieved on September 23, 2005, from <http://www.vls-inc.com/reviews/Reviews.html>
- Aksoy, S., Koperski, K., Tusk, C., et al.** (2005). Learning Bayesian classifiers for scene classification with a visual grammar. *IEEE Transactions on Geoscience and Remote Sensing*. 43, 581-589
- Anderson J. R., Hardy E. E., Roach J. T. & Witmer R. E.** (1976). A Land Use and Land Cover Classification System for the Use with Remote Sensor Data. *Geological Survey Professional Paper*, 964, p. 28. U.S. Government Printing Office, Washington D.C..
- ATCOR** for ERDAS Imagine Geosystems. Retrieved on July 21, 2006 from <http://www.geosystems.de/atcor/>.
- Auerbach, N.A., Walker, M.D., & Walker, D.A.** (1997a). Effects of roadside disturbance on substrate and vegetation properties in arctic tundra. *Ecological Applications*, 7, 218-235.
- Auerbach, N.A., Walker, D.A. and Bockheim, J.G.** (1997b). Land Cover Map of the Kuparuk River Basin, Alaska. *Tundra Ecosystem Analysis and Mapping Laboratory*. Institute of Arctic and Alpine Research, University of Colorado, Boulder.
- Avci, M., & Akyurek, Z.** (2004). A hierarchical classification of Landsat TM imagery for land-cover mapping. Retrieved August 3, 2007, from www.cartesia.org/geodoc/isprs2004/comm4/papers/402.pdf
- Bannari, A., Morin, D., Benie, G.B. et al..** (1995). A theoretical review of different mathematical models of geometric corrections applied to remote sensing images. *Remote Sensing Reviews*, 13, 27-47.
- Benediktsson, J. A., Swain, P. H., & Ersoy, O. K.** (1993). Conjugate-gradient neural networks in classification of multisource and very-high dimensional remote sensing data. *International Journal of Remote Sensing*, 14, 2883 - 2903.

- Benediktsson, J. A., Swain, P. H. & Ersoy, O. K.** (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE transactions on Geoscience and Remote Sensing*, 28, 540-551.
- Bolstad, P. V. & Lillesand, T. M.** (1991). Rapid maximum likelihood classification. *Photogrammetric Engineering and Remote Sensing*, 57, 67-74.
- Brown, D.G., Lusch, D.P., & Duda, K.A.** (1998). Supervised classification of types of glaciated landscapes using digital elevation data. *Geomorphology* 21, 233–250.
- Burrough, P. A. & McDonell, R.A.** (1998). *Principles of Geographical Information Systems*. Oxford University Press, New York, p. 190.
- Calef, M.P., H.E. Epstein, A.D. McGuire, T.S. Rupp, & H.H. Shugart** (2005). Analysis of vegetation distribution in Interior Alaska and sensitivity to climate change using a logistic regression approach. *Journal of Biogeography*, 32, 863-878.
- Canada Centre for Remote Sensing**, (n.d.). Retrieved December 23, 2007, from http://www.ccrs.nrcan.gc.ca/resource/tutor/fundam/chapter4/07_e.php
- Carpenter, G. A., Gjaja, M. N., Gopal, S., & Woodcock, C. E.** (1997). ART neural networks for remote sensing: vegetation classification from Landsat TM and terrain data. *IEEE Transactions on Geoscience and Remote Sensing*, 35, 308-325.
- Caudill, M.** (1990). Expert networks. In Eberhart, R.C., & Dobbins, R.W., (Ed.), *Neural Network PC Tools: A Practical Guide* (pp.189–214), Academic Press, San Diego.
- Chapin III, F. S., Sturm M., Serreze M. C. et al.** (2005). Role of terrestrial ecosystem changes in arctic summer warming. *Science*, 310, 657–660.
- Congalton, R.G.** (1996). Accuracy Assessment: A Critical Component of Land Cover Mapping. In Scott, J.M., Tear, T.H., & Davis, F., (Eds.), *Gap analysis: a landscape approach to biodiversity planning* (pp. 119–131). American Society for Photogrammetry and Remote Sensing.
- Congalton, R.G.** (1991). A review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, 37, 35-46.

- Congalton, R. & Green, K.** (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. CRC/Lewis Press, Boca Raton, FL. 137 p
- Di, K, Li, D., & Li, D.** (2000). A study of remote sensing image classification based on spatial data mining techniques. *Journal of Wuhan Technical University of Surveying and Mapping*, 25(1).
- Duda, R.O., Hart, P.E., & Stork, D.G.** (2001). *Pattern Classification*. New York, John Wiley & Sons, pp. 654
- Eklund, P. W., Kirkby, S.D., Salim, A.** (1998). Data mining and soil salinity analysis. *International Journal of Geographical Information Science*, 12, 247-268.
- Epstein, H.E., Beringer, J., Gould, W.A. et al.** (2004). The nature of spatial transitions in the arctic. *Journal of Biogeography*, 31, 1917–1933.
- ERDAS** Imagine Field Guide. (2005).
- ESRI**, ArcGIS Desktop Help. (2007).
- 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.
- Fleiss, J. L., Cohen, J., & Everitt, B. S.** (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 72, 323-327
- Forbes, B.C., & Sumina O.I.** (1999). Comparative Ordination of Low Arctic Vegetation Recovering from Disturbance: Reconciling Two Contrasting Approaches for Field Data Collection. *Arctic, Antarctic and Alpine Research*, 31, 389 -399
- Foody, D. M., McCulloch, M. B., & Yates, W. B.** (1995). Classification of remotely sensed data by an artificial neural network: Issues related to training data characteristics. *Photogrammetric Engineering and Remote Sensing*, 61, 391–401.

- Foody, G.** (1992). On the Compensation for Chance Agreement in Image Classification Accuracy Assessment. *Photogrammetric Engineering & Remote Sensing*, 58, 1459-1460.
- Gerçek, D.** (2002). Improvement of Image Classification with the integration of Topographical data. *M.S. Thesis*. The Graduate School of Natural and Applied Sciences, Middle East Technical University.
- Gonzalez, A., J., & Dankel, D., D.** (1993). *The engineering of knowledge-based systems: theory and practice*. Prentice-Hall, Inc., Upper Saddle River, NJ.
- Gough, L., Shaver, G.S., Carroll, J., Royer, D. L., Laundre, J.A.** (2000). Vascular plant species richness in Alaskan arctic tundra: the importance of soil pH. *Journal of Ecology*, 88, 54-66.
- Gould, W. A.** (2000). Remote Sensing of Vegetation, Plant Species Richness and Regional Biodiversity Hotspots. *Ecological Applications*, 10, 1861–1870.
- Gould, W. A. & Walker, M.A.** (1999). Plant communities and landscape diversity along a Canadian Arctic river. *Journal of Vegetation Science*, 10, 537–548.
- Green, K., Congalton, R.G.** (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Lewis Publishers, New York.
- Grossberg, S.,** (1976). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23, 121-134.
- Hand, D.J.** (1998). Data Mining: Statistics and More? *The American Statistician*, 52, 112-118.
- Hansen, M.C., Defries, R.S., Townshend, J.R.G., Sohlberg, R.** (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, 21, 1331-1364.
- Hazarika, R., Saikia, A.** (n.d.) The rift in the lute: Rhino habitat in the Kaziranga National Park, India. Retrieved on August 23, 2007, from http://www.gisdevelopment.net/application/application/natural_hazards/landslides/ma06_96.htm
- Hengl, T.** (2002). Neural Network Fundamentals: A Neural Computing Primer. *Personal Computing Artificial Intelligence*, 16, 32 - 43.

- Heute, A., Didan, K., Miura, T. et al.**, (2002). Overview of the radiometric performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, 195–213.
- Holben, B. N.** (1986). Characterization of maximum value composites from temporal AVHRR data. *International Journal of Remote Sensing*, 7, 1417-1434.
- Hope, A. S., Kimball, J. S., & Stow, D. A.** (1993). Relationship between tussock tundra spectral reflectance properties and biomass and vegetation composition. *International Journal of Remote Sensing*, 14, 1861– 1874.
- Hope, A., & Stow, D.** (1995). Shortwave reflectance properties of Arctic tundra. In Reynolds, J., & Tenhunen, J. (Eds.), *Landscape function and disturbance in Arctic tundra. Ecological Studies*, 120, 155– 164. Heidelberg: Springer.
- Hope, A., Engstrom, R., Stow, D.** (2005). Relationship between AVHRR surface temperature and NDVI in Arctic tundra ecosystems. *International Journal of Remote Sensing*, 26, 1771– 1776.
- Huang X. & Jensen J. R.** (1997). A machine-learning approach to automated knowledge-base building for remote sensing image analysis with GIS data. *Engineering & Remote Sensing*, 63, 1185-1194.
- Hush, D. R.** (1989). Classification with neural networks: a performance analysis. *Proceedings of the IEEE International Conference on Systems Engineering*, Dayton, Ohio, USA, 277–280.
- IDRISI Andes User Manual** (2007) *Guide to GIS and Image Processing*.
- Jackson, S. S., & Bourne, S. G.** (2005). Using Feature Extraction To Monitor Urban Encroachment. *Earth Observation Magazine*, XIV. Retrieved on September 10, 2006, from http://www.eonline.com/EOM_Apr05/features01.html
- Jensen, J. R.** (1978). Digital land cover mapping using layered classification logic and physical composition attributes. *American Cartographer*, 5, 121-132.
- Jensen, J. R.** (2005). *Introductory Digital Image Processing: A Remote Sensing Perspective*. NJ: Prentice Hall.
- Jensen, J.R., Qiu, F. & Patterson, K.** (2001). A Neural Network Image Interpretation System to Extract Rural and Urban Land Use and Land Cover Information from Remote Sensor Data. *GeoCarto International, A Multi-disciplinary Journal of Remote Sensing and GIS*, 16, 19-28.

- Jensen, J.R., Qiu, F. & M. Ji.** (1999). Predictive Modelling of Coniferous Forest Age Using Statistical and Artificial Neural Network Approaches Applied to Remote Sensing Data. *International Journal of Remote Sensing*, 20, 2805-2822.
- 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.
- Jia, G. J., Epstein, H. E., & Walker, D. A.** (2004). Controls over intra-seasonal dynamics of AVHRR-NDVI for the Arctic Tundra in Northern Alaska. *International Journal of Remote Sensing*, 25, 1547-1564.
- Kanellopoulos, I., Wilkinson, G.G., & Megier, J.** (1993). Integration of neural network and statistical image classification for land cover mapping. *Proceedings of the 1993 International Geoscience and Remote Sensing Symposium*, 511 - 513.
- Kanellopoulos, I., & Wilkinson, G.G.** (1997). Strategies and best practice for neural network image classification. *International Journal of Remote Sensing*, 18, 711- 725.
- Kavzoglu, T., & Mather, P.M.** (2003). The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, 24, 4907- 4938.
- Kempka, R.G., Macleod, R. D., Payne, J. et al.** (1995). National Petroleum Reserve; Alaska Land-cover Inventory: Exploring Arctic Coastal Plain Using Remote Sensing. *9th Annual Symposium on Geographic Information Systems, GIS World, Inc.*, 788 – 793.
- Koperski, K., Marchisio, G., Aksoy, S., & Tusk, C.** (2001). Applications of Terrain and Sensor Data Fusion in Image Mining. *IEEE International Geoscience and Remote Sensing Symposium (2002)*, 2, 1026-1028.
- Landis, J., & Koch, G.** (1997). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33, 159-174.
- Lillesand, T. M., & Kiefer, R. W.** (2000). *Remote sensing and Image Interpretation*. New York: Wiley.

- Li, J., & Narayanan, R.M.** (2004). Integrated Spectral and Spatial Information Mining in Remote Sensing Imagery. *IEEE Transaction on Geoscience and Remote Sensing*, 42, 673-685.
- Li, D., Di, K., & Li, D.** (2000). Feature based model verification (fbmv): A new concept for hypothesis validation in building reconstruction. *International Archives of Photogrammetry and Remote Sensing*, XXXIII- B3, 24-35.
- Liu, X.H., Skidmore, A.K. & Oosten, H.V.** (2002). Integration of Classification Methods for Improvements of Land-cover Map Accuracy, *ISPRS Journal of Photogrammetry & Remote Sensing*, 56, 257-268.
- Liu, W., Seto, K., Wu, E., Gopal, S., & Woodcock, C.** (2004). ART-MMAP: A neural network approach to subpixel classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1976–1983.
- McMichael, C., Hope, A., Stow, D. et al.** (1999). Estimating CO2 exchange in Arctic Tundra ecosystems using a spectral vegetation index. *International Journal of Remote Sensing*, 20, 683– 698.
- Markon C.J.** (1992). Land cover mapping of the upper Kuskokwim Resource Management Area, Alaska, using Landsat and a digital data base approach. *Canadian Journal of Remote Sensing* 18, 62–71.
- Markon, C. J., Fleming, M.D., & Binnian, E.F.** (1995). Characteristics of vegetation phenology over the Alaskan landscape using AVHRR time-series data. *Polar Record*, 31, 179-190.
- Markon, C. J., & Derksen, D. V.** (1994). Identification of Tundra Land Cover near Teshekpuk Lake, Alaska Using SPOT Satellite Data. *Arctic*, 47, 222-231.
- Morrissey, L.A., & Ennis, R.A.** (1981). Vegetation mapping of the National Petroleum Reserve in Alaska using Landsat digital data. *U.S. Geological Survey Open-File Report* 81– 315.
- 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.
- Muller, S.V., & Walker, D.A (1998).** Land-Cover Map of the North Slope of Alaska. Retrieved on September 23, 2006, from ftp://sidads.colorado.edu/pub/DATASETS/ARCSS/data/arcss020/ns_landcov.meta.txt

- Narumalani, S., Hladý, J.T., & Jensen, J.R.** (2002). Information Extraction from Remote Sensed Data. In Bossler, J.D., Jensen, J.R., McMaster, R.B. & C. Rizos (Eds.), *Manual of Geospatial Science and Technology* (298- 324). New York: Taylor and Francis.
- National Resource Research Institute** (1997). Land cover characteristics based on Landsat Imagery. Retrieved on July 19, 2005, from www.nrri.umn.edu/toolkit/vegetation.html.
- Noyle, B. M.** (1999). The Vegetation and a Landsat Assisted Vegetation Map of the Barrow Peninsula, Northern Alaska. *M. S. Thesis*, (pp.183), Michigan State University.
- O'Brien, M. A.** (2003). Feature Extraction with the VLS Feature Analyst System. *Proceedings of the American Society for Photogrammetry and Remote Sensing*. Retrieved on October 11, 2005, from http://featureanalyst.com/feature_analyst/publications/reviews/obrien_asprs.pdf.
- Pacific Meridian Resources** (1996). *National Petroleum Reserve Alaska land-cover inventory: Phase 2 Eastern NPR-A*. Interim Report.
- Pijanowski, B.C., Shellito, B.A., Bauer, M.E. et al.** (2001). Using GIS, Artificial Neural Networks and Remote Sensing to Model Urban Change in the Minneapolis-St. Paul and Detroit Metropolitan Areas. *ASPRS Proceedings*. St Louis, MO,.
- Pijanowski, B.C., D.T. Long, S.H. Gage et al.** (1997). A Land Transformation Model: Conceptual Elements, Spatial Object Class Hierarchies, GIS Command Syntax and an Application to Michigan's Saginaw Bay Watershed. *Land-use Modeling Workshop*. Sioux Falls, South Dakota.
- Quinlan, J. R.** (1993). C4.5: Programs for Machine Learning. *Machine Learning*, 16, 235-240.
- Quinlan, J. R.** (2000). Data Mining Tools See5 and C5.0. Retrieved on September 29, 2007, from <http://www.rulequest.com/see5-info.html>
- Redmond, R., & Winne, J. C.** (2000). Classifying & Mapping Wildlife Severity. *University of Montana Wildlife Spatial Analyst Lab*. Retrieved on September 29, 2005, from http://www.vls-inc.com/feature_analyst/publications/reviews/wildlife_severity.pdf.

- Riedel, S. M., Epstein, H. E, Walker, D. A. et al.** (2005). Spatial and temporal heterogeneity of vegetation properties among four tundra plant communities at Ivotuk, Alaska, USA. *Arctic, Antarctic, and Alpine Research*, 37, 25-33.
- Roli,F., Serpico, S.B., & Vernazza, G.** (1996). Neural Networks for classification of Remotely Sensed Images. In C.H. Chen (Ed.), *Fuzzy Logic and Neural Network Handbook*, New York: McGraw-Hill.
- Grossberg, S.,** (1976). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23, 87-202.
- Shasby, M., Cameggie, D.** (1986). Vegetation and terrain mapping in Alaska using Landsat MSS and digital terrain data. *Photogrammetric Engineering and Remote Sensing*, 52, 779-786.
- Shekhar, S., & Chawla, S.** (2003). *Spatial Databases: A Tour*. New Jersey: Prentice Hall.
- Shine, J.A., & Wakefield G.I.** (1999). A comparison of supervised imagery classification using analyst-chosen and geostatistically-chosen training sets. *Geocomputation*, 99.Retrieved on May 29, 2007, from http://www.geovista.psu.edu/sites/geocomp99/Gc99/044/gc_044.htm
- Shippert, M.M., Walker, D.A., Auerbach, N.A., & Lewis, B.E.** (1995). Biomass and leaf area index maps derived from SPOT images for the Toolik Lake and Imnavait Creek Area, Alaska. *Polar Record*, 31,147-154.
- Skapura, D.M.,** (1996). *Building neural networks*. ACM Press/Addison-Wesley Publishing Co., New York, NY.
- Skidmore, A. K., & Turner, B. J.** (1988). Forest mapping accuracies are improved using a supervised nonparametric classifier with SPOT data. *Photogrammetric Engineering and Remote Sensing*, 54, 1415–1421.
- Skidmore, A.K.** (1989). An expert system classifies Eucalypt forest types using Thematic Mapper Data and a Digital terrain model. *Photogrammetric Engineering and remote sensing*, 55, 1449–1464.
- Skidmore, A.K., Turner, B.J., Brinkhof, W., & Knowles, E.** (1997). Performance of a neural network: mapping forests using GIS and remote sensing data. *Photogrammetric Engineering and Remote Sensing*, 63, 501-514.

- Soh, L.K., Tsatsoulis, C.** (1998). Data Mining in Remotely Sensed Images: A general model and an application. In *Proceedings of the 1998 International Geoscience and Remote Sensing Symposium*, Seattle, Washington, 798-800.
- 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.
- Stow, D., Hope, A., et al.** (2004). Remote Sensing of vegetation and Land-Cover Change in Arctic Tundra Ecosystems. *Remote Sensing of Environment*, 89, 281-308.
- 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., 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., Coulter, L., Kaiser, J., et al.** (2003). Irrigated Vegetation Assessment for Urban Environments. *Photogrammetric Engineering and Remote Sensing*, 69, 381- 390.
- 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., Boynton, W., et al.** (1998). Satellite-derived vegetation index and cover type maps for estimating carbon dioxide flux for arctic tundra regions. *Geomorphology*, 21, 313-327.
- Truax, D.D., Repaka, S. R.** (2003). Comparing Spectral and Object Based Approaches for Classification and Transportation Feature Extraction from High resolution Multi-spectral Imagery. *Proceedings of the ASPRS 2004 Annual Conference*, Denver, USA.
- Vanderzanden, D., Morrison, M.** (2004). High Resolution Image Classification: A Forest Service Test of Visual Learning Systems. USDA Forest Service. Retrieved on August 15, 2006, from http://www.vls-nc.com/feature_analyst/publications.htm

- Vatsavai, R.R., Burk, T.E., Shekhar, S.** (2001). Multi-Spectral Image Classification using Spectral and Spatial Knowledge. Retrieved on January 13, 2007, from <http://www-users.cs.umn.edu/~vatsavai/papers/cisst01-raju.ps>.
- Vatsavai, R.R., & Shekhar, S.** (2002). An Efficient Hybrid Classification System for Mining Multi-spectral Remote Sensing Imagery Guided by Spatial Database. Retrieved on January 13, 2007, from <http://www-users.cs.umn.edu/~vatsavai/papers/prrs02f-vatsavai.ps>.
- Vatsavai, R.R., & Shekhar, S.** (2005). A Semi-Supervised Learning Method for Remote Sensing Data Mining. *17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'05)*, 207-211.
- Vierling, L.A., Deering, D. W., & Eck, T. E.** (1997). Differences in arctic tundra vegetation type and phenology as seen using bidirectional radiometry in the early growing season. *Remote Sensing of Environment*, 60, 71-82.
- Visual Learning Systems** (2002). User Manual, *Feature Analyst Extension for Imagine 8.x*, Visual Learning Systems, Inc., Missoula, MT.
- Visual Learning Systems, Inc. White Paper.** The Feature Analyst Extension for ArcView and ArcGIS: Assisted Feature Extraction Software for GIS database maintenance. 2001.
- Vourlitis, G., Oechel, W., Hope, A., et al.** (2000). Physiological models for scaling plot measurements of CO₂ flux across an arctic tundra landscape. *Ecological Applications*, 10, 60-72.
- Walker, D.A.** (1985). Terrain and vegetation types of the Sagavanirktok Quadrangle, Alaska. *NASA-Ames consortium agreement no. NCA2-OR170-303*, final report.
- Walker, D.A., & Everett, K.R.** (1987). Road Dust and its environmental impact on Alaskan Taiga and Tundra. *Arctic and Alpine Research*, 19, 479 -489.
- Walker, D. A.** (1996). Distribution and recovery of arctic Alaskan vegetation. In Reynolds, J. F., & Tenhunen, J. D. (Eds.), *Landscape function and disturbance in arctic tundra. Ecological Studies*, 120, 35-71. Heidelberg: Springer.
- Walker, D.A.** (1999). An integrated vegetation mapping approach for northern Alaska (4,000,000 scale). *International Journal of Remote Sensing*, 20, 2895-2920.

- Walker, D.A., & Acevedo, W.** (1987). Vegetation and a Landsat-derived land cover map of the Beechey Point Quadrangle, Arctic Coastal Plain, Alaska. *U.S. Army Cold Regions Engineering and Research Laboratory*, Hanover, NH, CRREL Report 87-5, 63
- Walker, D. A., Bay, C., Daniels, F. J. A., et al.** (1995). Toward a new arctic vegetation map: A review of existing maps. *Journal of Vegetation Science*, 6, 427-436.
- 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.
- Walker, D. A., & Walker, M. D.** (1991). History and Pattern of Disturbance in Alaskan Arctic Terrestrial Ecosystems: A Hierarchical Approach to Analysing Landscape Change. *The Journal of Applied Ecology*, 28, 244-276 .
- Westhoff, V., & Van der Maarel, E.** (1978). The Braun-Blanquet approach. In Whittaker R. H. (Ed.). *Classification of Plant Communities*, 287- 399.
- Weiguo Liu., Karen, C. Seto., Elaine, Y., et al.** (2004). ART-MMAP: A Neural Network Approach to Subpixel Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 2004.
- Wilkinson, G.G., Kanellpoulos, I., Kontoes, C., et al.**(1992). A comparison of neural network and expert system methods for analysis of remotely-sensed imagery. In *Proceedings of the 1998 International Geoscience and Remote Sensing Symposium*, Houston, USA, 62–64.
- Williams, M., Rastetter, E. B., Shaver, G. R. et al.** (2001). Primary Production of an Arctic Watershed: An Uncertainty Analysis. *Ecological Applications*, 11, 1800-1816
- Zhou, L., Tucker, C. J., Kaufmann, R. K. et al.** (2001). Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999. *Journal of Geophysical Research*, 106, 20069– 20083.
- Zhang, S., & Liu, X.** (2005). Realization of Data Mining Model for Expert Classification using multi-scale spatial data. *ISPRS Workshop on Service and Application of Spatial Data Infrastructure*, 36, Hangzhou, China.

APPENDIX A: FIELD PHOTOS COLLECTED ILLUSTRATING THE DIFFERENT LAND-COVER CLASSES

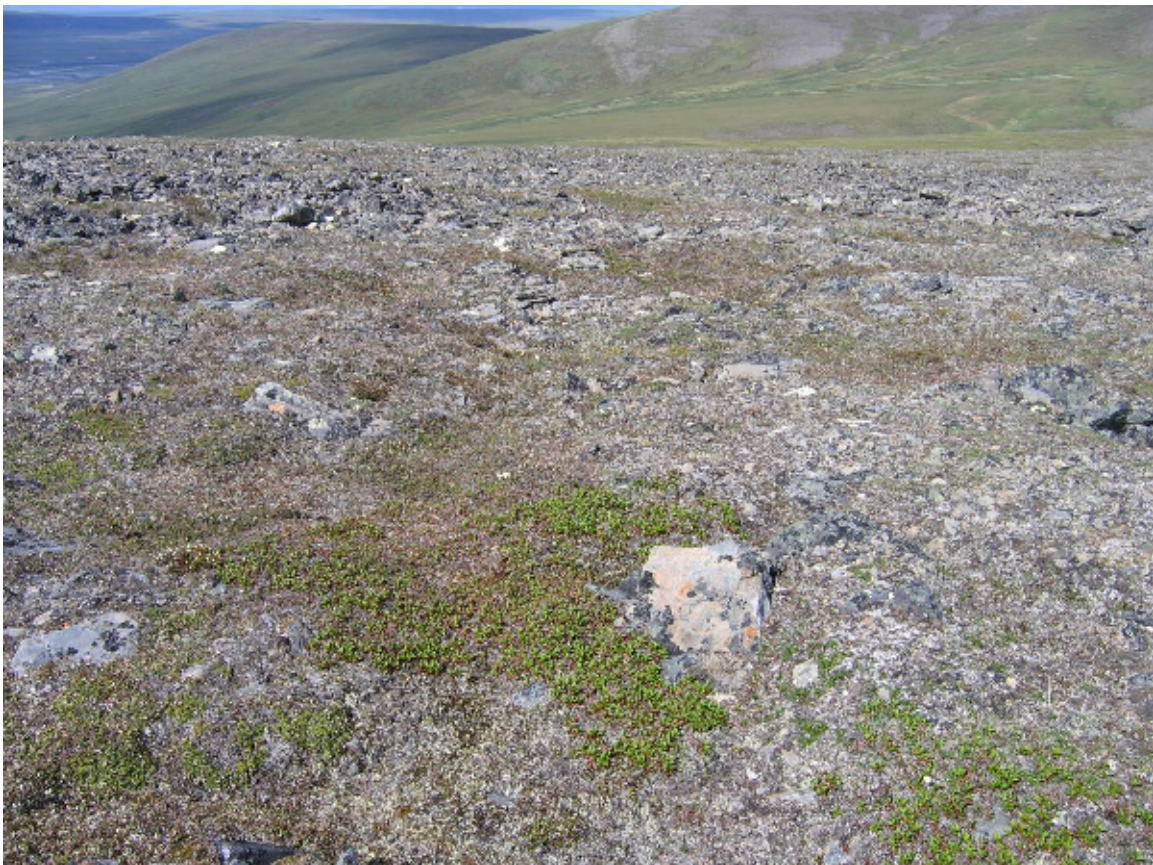
1: Moist Low-Shrub Tundra and other Shrublands



2: Water and aquatic complex



3: Barren complex



4: Snowbed complex



5: Moist Dwarf-shrub, Tussock-graminoid Tundra complex



6: Wet Graminoid Tundra



7: Moist Graminoid, Prostrate-shrub Tundra complex



APPENDIX B: EXAMPLE OF FIELD FORMS AND FIELD DATA COLLECTION METHODOLOGY

266 - Sphagnum + Pink bells
267 → Tussac + Betula
+ Cotton grass 268 - General

~~266 - Sphagnum~~
~~267 - Tussac~~

Field Work for Arctic Tundra Vegetation classification With DT and Spatial Data Mining Techniques, North Slope Alaska Summer 2007

Date 06/24 Time 3:15 PM Surveyors: P
Field Form # Grid #
Location Details
1 UTM Reading -
2 Shapefile - Name Poly-possible - MAT-24 Type ☐ Point ☐ Line ☒ Poly
3 Photo #
Classification Check
1. Preliminary Class MAT
2. Estimated Class
Vegetation Details
1. Dominant Species Tussac, moss (Sphagnum spp.)
2. Secondary Species Betula nana, Pink bells
3. Tertiary Species
4. Others
Surface Form Description
Dry / Wet ☐ D ☐ M ☐ W
General Description / Notes 1/2 m deep water all over
Taken near Sag. River

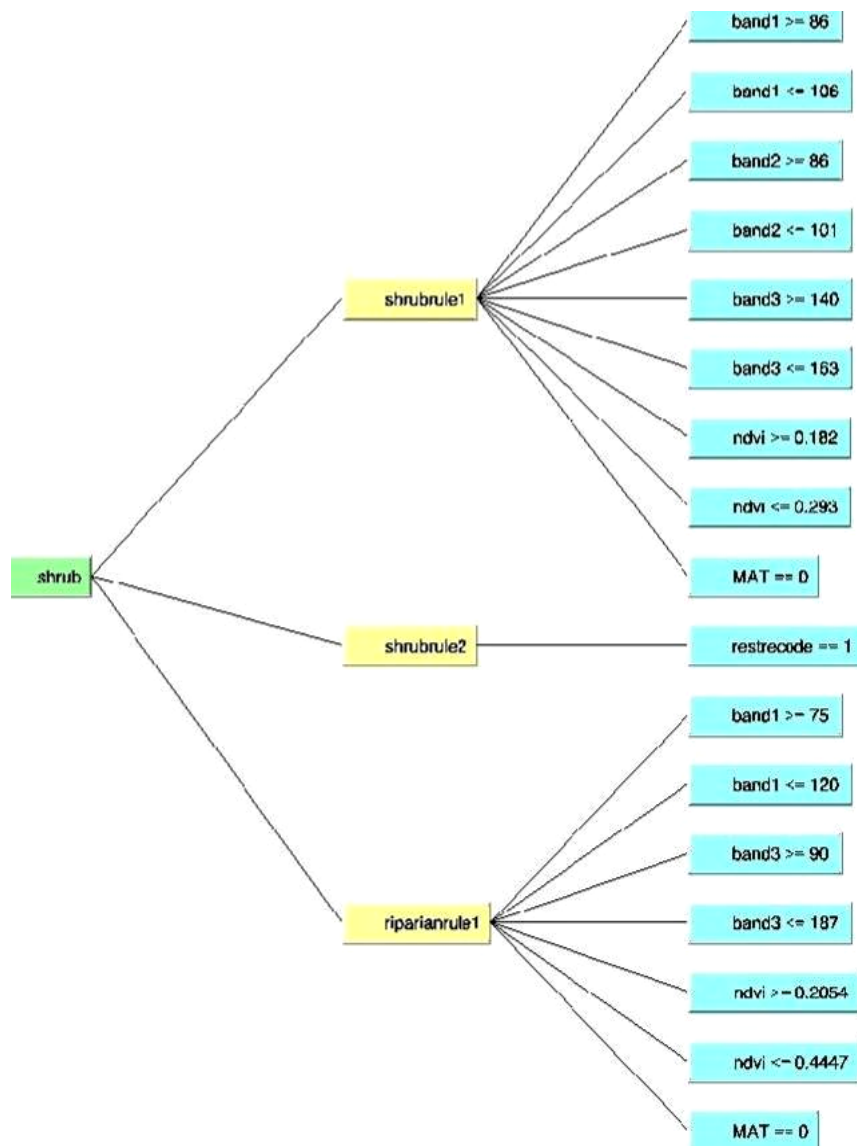
212

Photos
220
General photo
226
General photo from truck up toward slope

Field Work for Arctic Tundra Vegetation classification With DT and Spatial Data Mining Techniques, North Slope Alaska Summer 2007

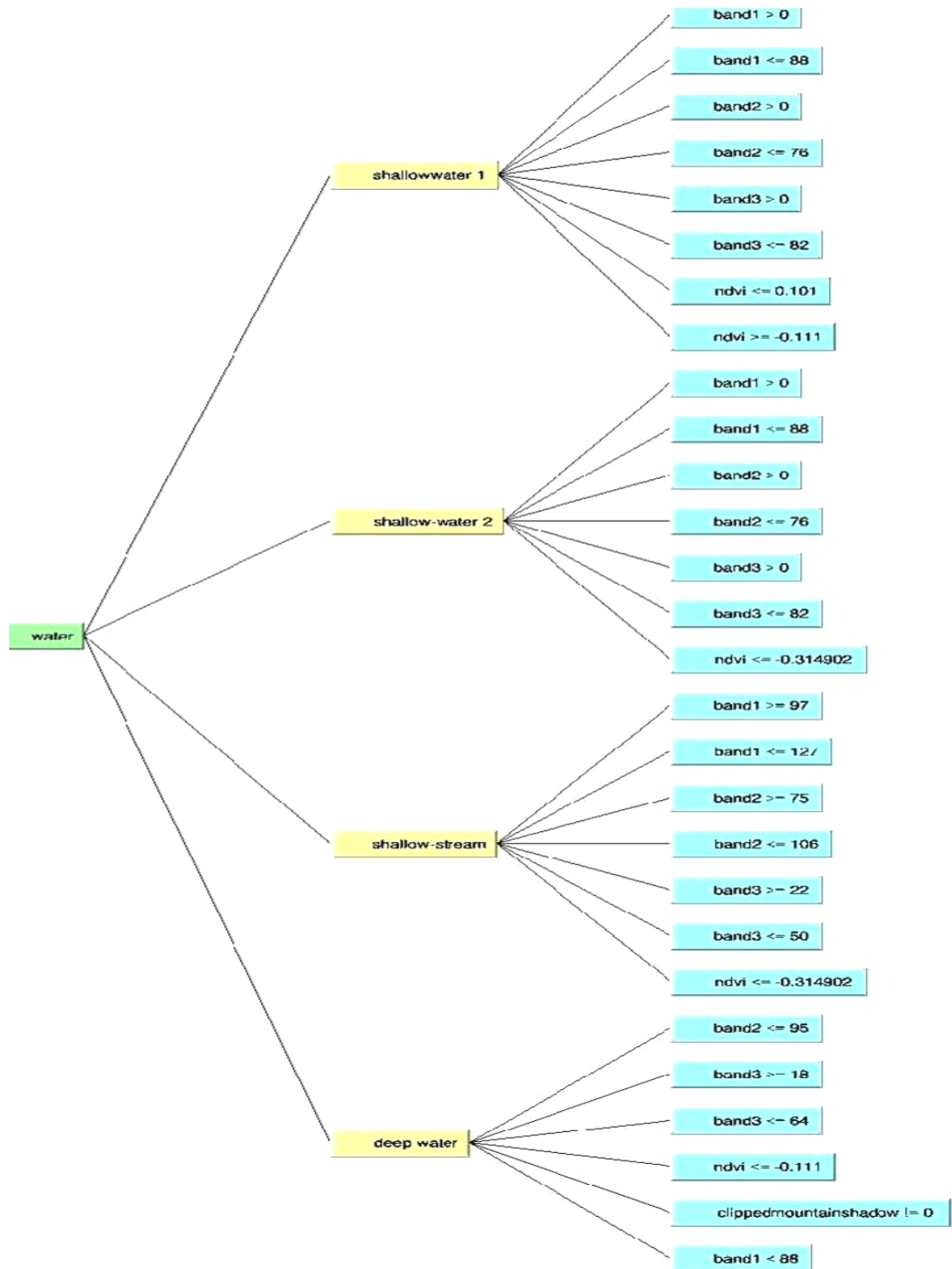
Date 06/24/07 Time 7:00 PM Surveyors: P/BB
Field Form # 8 Grid #
Location Details
1. UTM Reading - 7616817.85N 401698.92E 3.4 Pdp
2. Shapefile - Name - Not possible to get accurate Type ☐ Point ☐ Line ☐ Poly
3. Photo #
Classification Check
1. Preliminary Class Shrub
2. Estimated Class Shrub
Vegetation Details
1. Dominant Species Salix alaxensis (221)
2. Secondary Species Moss spp. → photo 222 bud like appearance moss
3. Tertiary Species → photo 223 lush green to red moss (w/o buds)
→ photo 224 dark green standing spp. (224)
→ photo 225 white-gray moss
4. Others Betula nana
Surface Form Description
Dry / Wet ☐ D ☒ M ☐ W
General Description / Notes On a gentle slope - NW (45°) / Right in front of Kupuk river bridge
Good site for shrub class
Salix in high density near road

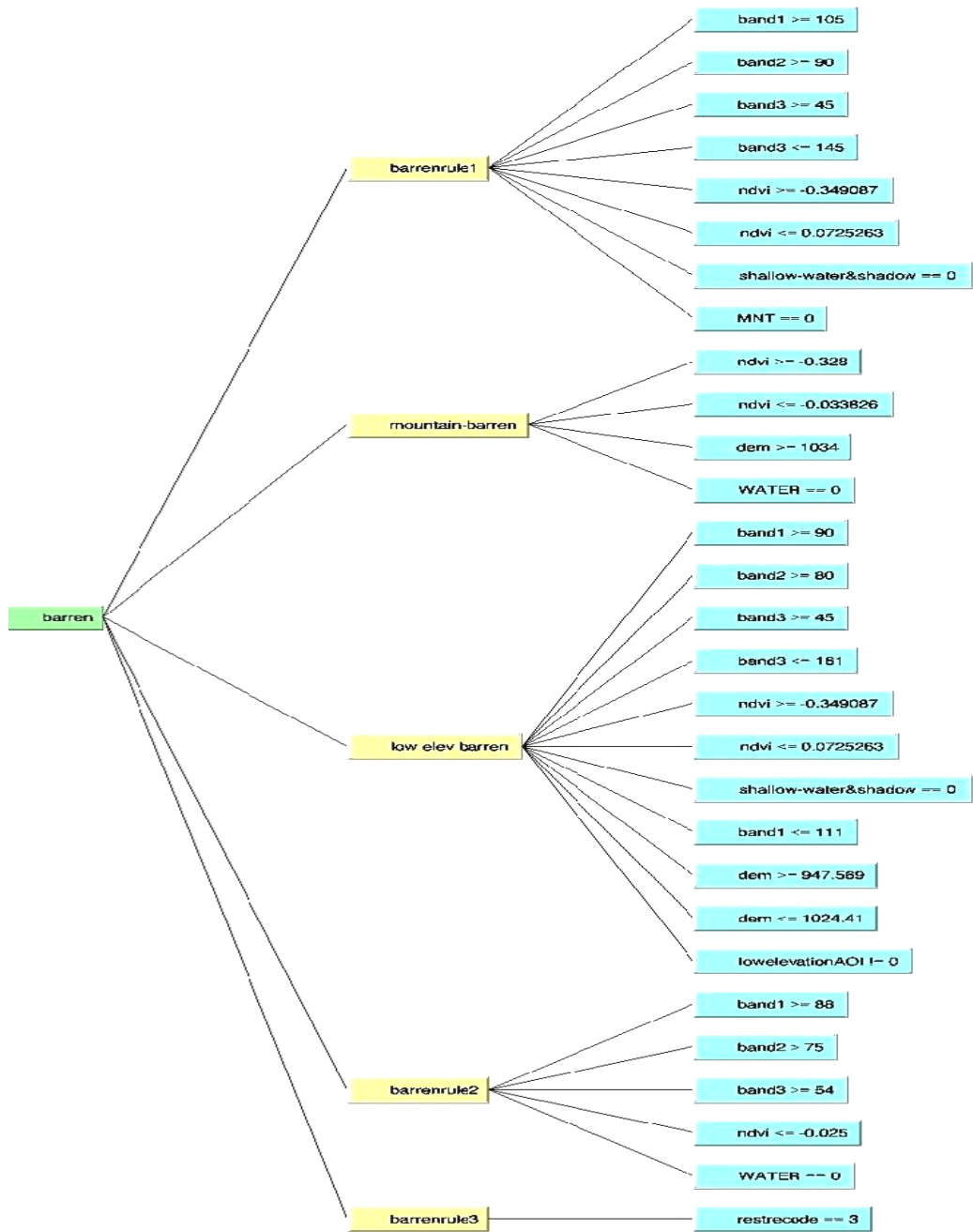
APPENDIX C: FIGURE SHOWING THE DIFFERENT RULES OF THE RULE-BASED CLASSIFIER



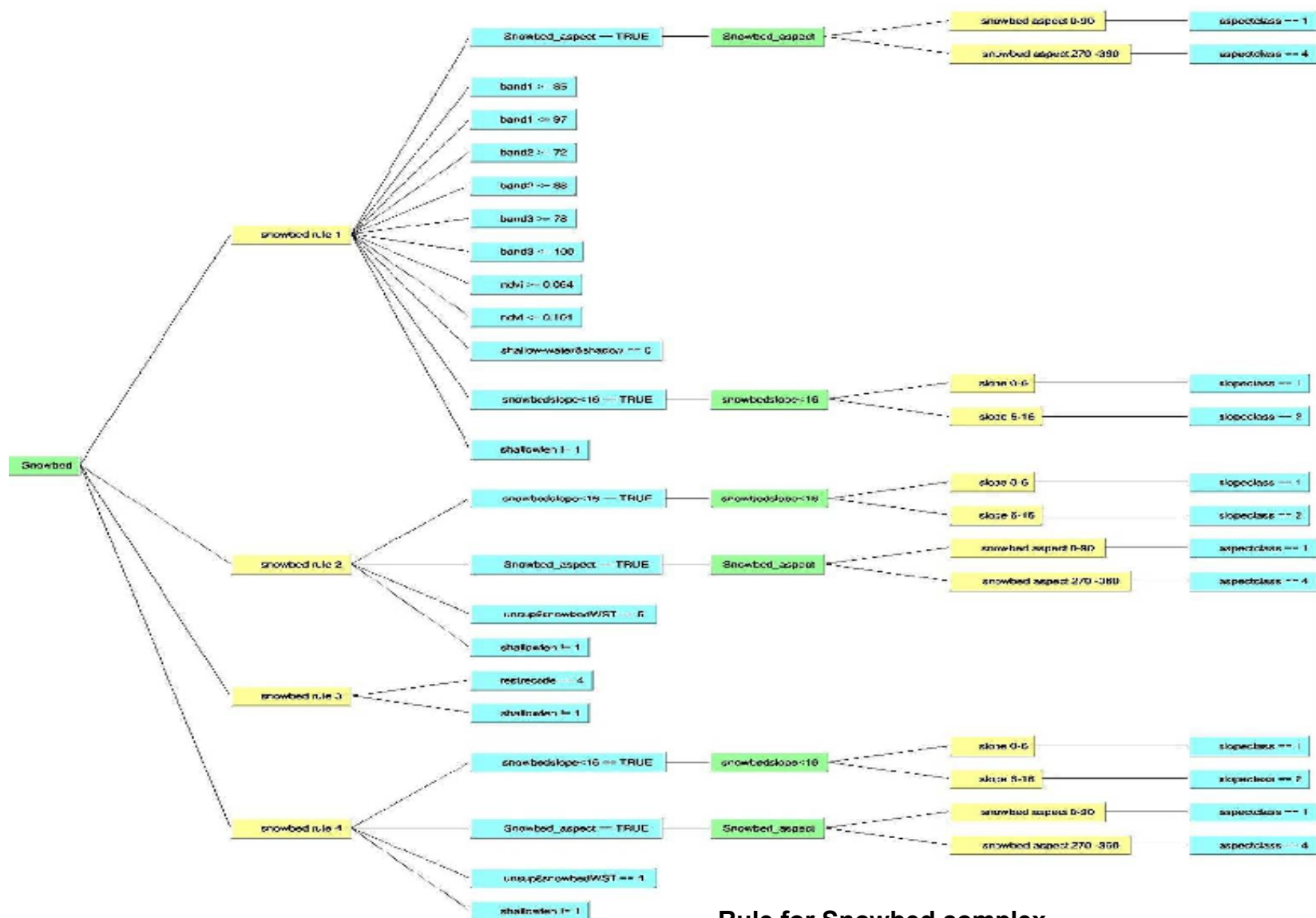
Rule for Moist Low-Shrub Tundra and other Shrublands

Rule for Water and aquatic complex

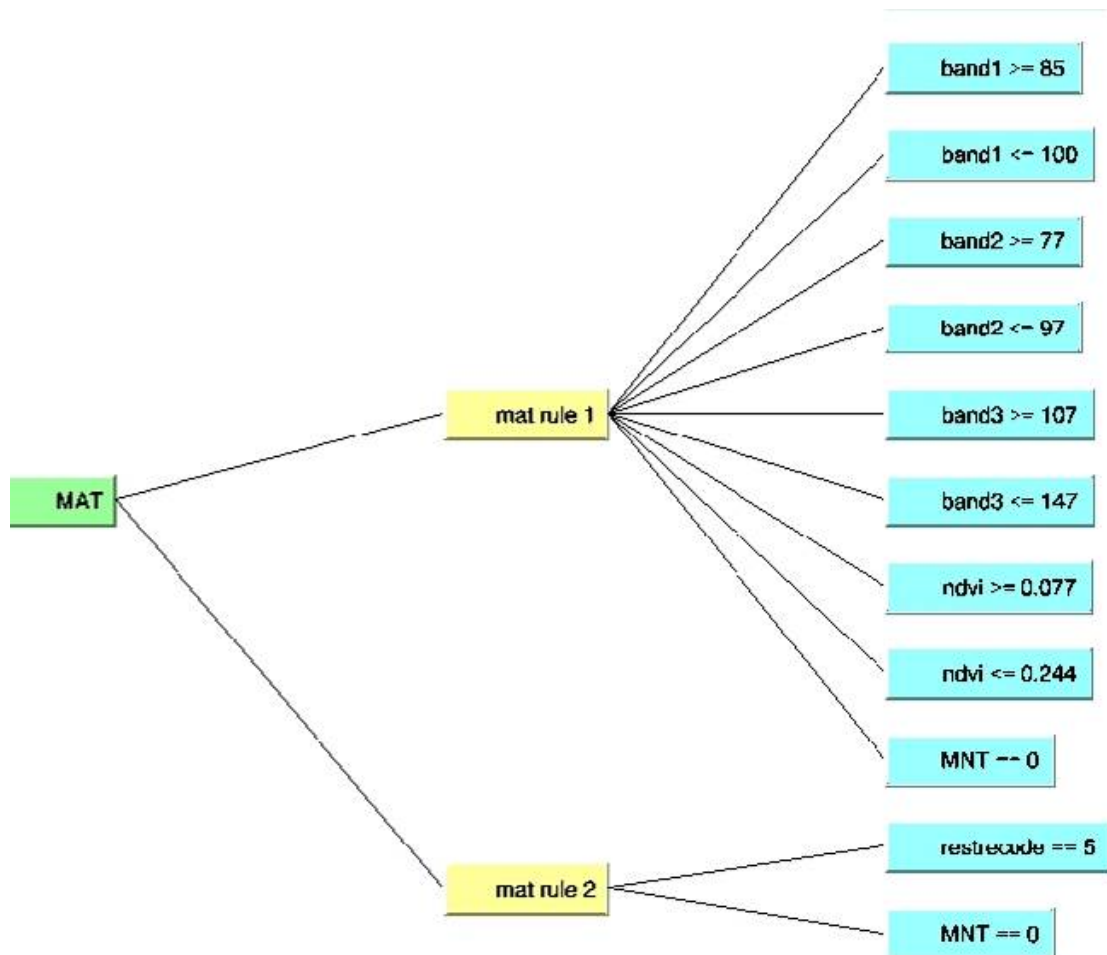




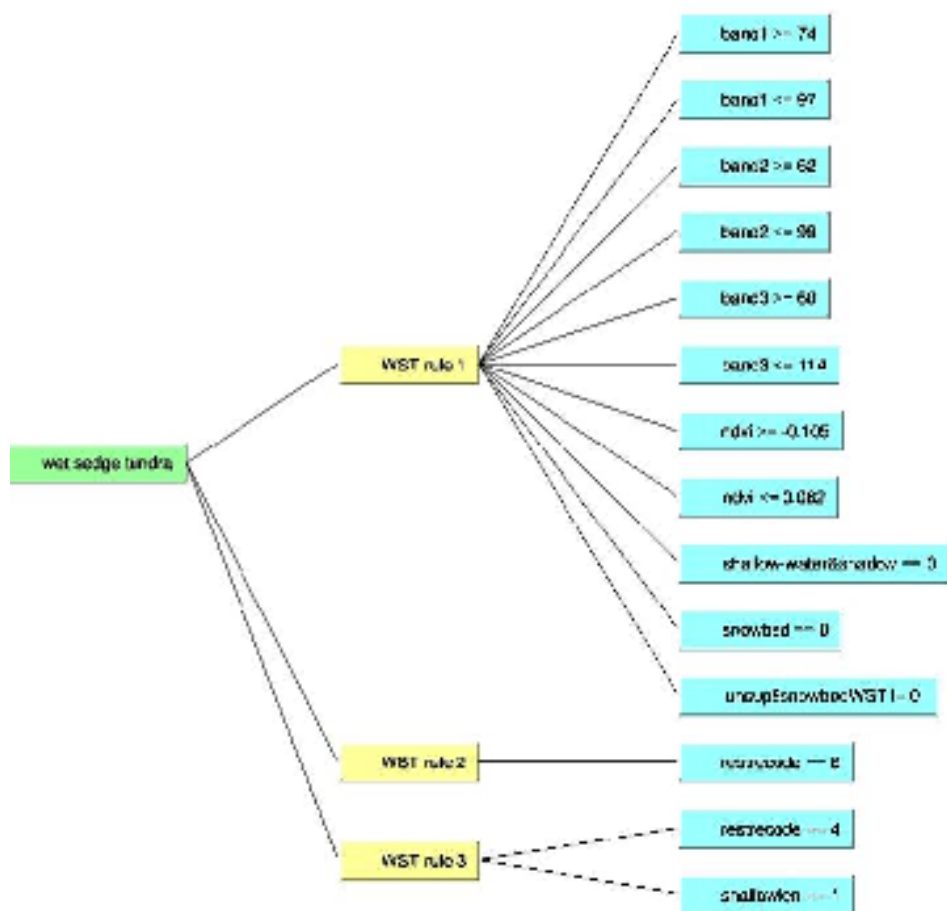
Rule for Barren complex



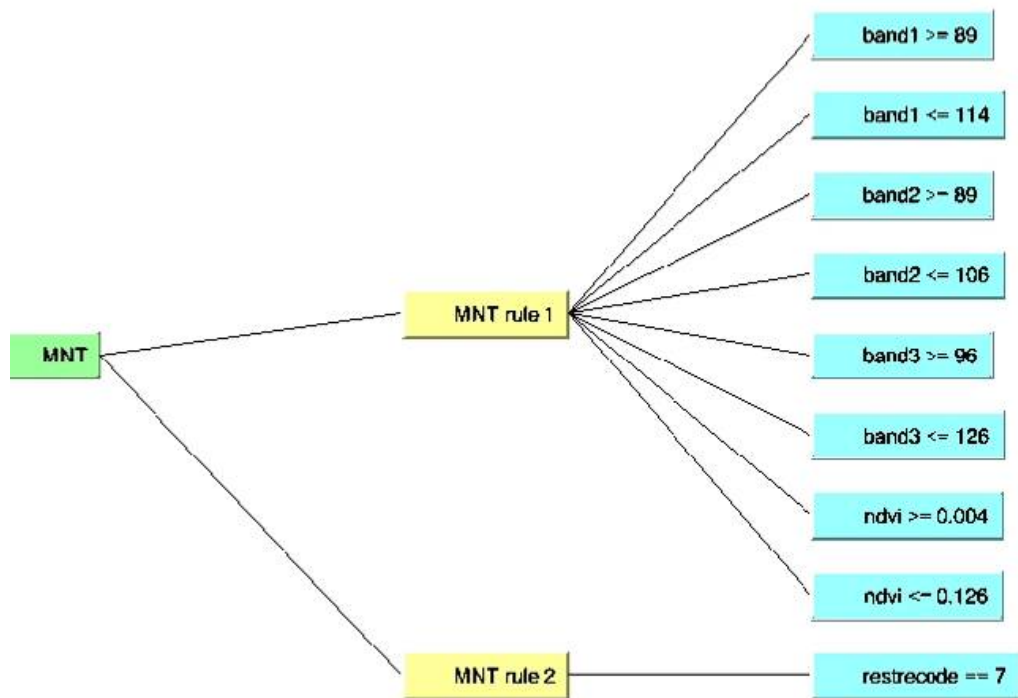
Rule for Snowbed complex



Rule for Moist Dwarf-shrub, Tussock-Graminoid Tundra complex



Rule for Wet Graminoid Tundra



Rule for Moist Graminoid, Prostrate-shrub Tundra complex

APPENDIX D: TABLE SHOWING THE TRIAL RUNS FOR THE MLP CLASSIFIER WITH DIFFERENT INPUT LAYERS AND NUMBER OF NODES IN THE HIDDEN LAYERS VALUES TO FINALIZE THE NETWORK ARCHITECTURE. The most optimum values are marked in bold (Iteration 6).

Iterations	InputBands	SLR	ELR	No. of nodes in Hidden Layer1	No. of nodes in Hidden Layer2	Training RMS	Testing RMS	Training Accuracy	Training Kappa	Average Accuracy	Average Kappa
1ai	Band1, Band2, Band3	0.00238	0.000235			0.001038	0.003103	86.13	0.9523		
1aii	Band1, Band2, Band3	0.00233	0.00025	20	0	0.001031	0.003082	90.13	0.9555	81.4633	0.5734
1aiii	Band1, Band2, Band3	0.00488	0.0005			0.001112	0.003346	68.13	0.1945		
1bi	Band1, Band2, Band3	0.00238	0.00025			0.000914	0.002729	78.88	0.2463		
1bii	Band1, Band2, Band3	0.00194	0.00194	8	20	0.000898	0.001548	92.75	0.965	83.5433	0.4861
1biii	Band1, Band2, Band3	0.00238	0.00025			0.000948	0.002841	79.00	0.247		
2ai		0.00158	0.00025			0.001024	0.001897	90.63	0.9603		
2aii	1+NDVI	0.00145	0.00025	20	0	0.001033	0.001899	90.63	0.9608	91.2133	0.9639
2aiii		0.00488	0.0005			0.000489	0.001605	92.38	0.9705		
2bi		0.00074	0.00012			0.000911	0.001713	91.38	0.9647		
2bii	1+NDVI	0.00208	0.00025	10	20	0.000771	0.001478	93.25	0.9768	92.5033	0.9724
2biii		0.00208	0.00025			0.000784	0.001505	92.88	0.9757		
3ai		0.000154	0.00025			0.001028	0.001901	90.75	0.9544		
3aii	2+Slope+Aspect	0.00138	0.00025	20	0	0.001088	0.001885	90.25	0.9522	90.7933	0.9544
3aiii		0.00238	0.00025			0.000968	0.001679	91.38	0.9566		
3bi		0.00203	0.00025			0.000782	0.001351	93.75	0.9757		
3bii	2+Slope+Aspect	0.00213	0.00025	16	20	0.000797	0.001366	94.13	0.9759	93.7100	0.9741
3biii		0.00198	0.00025			0.000823	0.001399	93.25	0.9707		
4ai		0.00181	0.00025			0.000952	0.001671	92.75	0.9689		
4aii	3+textures band 1,2,3, NDVI	0.00099	0.00025	25	0	0.001088	0.001892	91.63	0.9613	91.0433	0.9617
4aiii		0.00072	0.00012			0.001241	0.002141	88.75	0.9548		
4bi		0.00082	0.00012			0.00087	0.001586	91.63	0.9709		
4bii	3+textures band 1,2,3, NDVI	0.00082	0.00012	25	20	0.000875	0.001588	91.38	0.9709	91.5033	0.9713
4biii		0.00087	0.00012			0.000865	0.001581	91.50	0.9722		
5ai		0.00488	0.0005			0.000847	0.001465	93.88	0.9644		
5aii	SPOT+NDVI+Slope+Aspect+tundra index	0.00488	0.0005	20	0	0.000823	0.001437	94.25	0.974	93.7533	0.9663
5aiii		0.00181	0.00025			0.000986	0.001711	93.13	0.9604		
5bi		0.00085	0.00012			0.000828	0.00144	92.88	0.976		

5bii	SPOT+NDVI+Slope+Aspect+tundra index	0.00238	0.00025	18	20	0.00072	0.001231	95.38	0.9798	93.7533	0.9752
Iterations	InputBands	SLR	ELR	No. of nodes in Hidden Layer1	No. of nodes in Hidden Layer2	Training RMS	Testing RMS	Training Accuracy	Training Kappa	Avg Accuracy	Avg Kappa
5biii		0.00075	0.00012			0.000848	0.001478	93.00	0.9697		
6ai		0.00036	0.00025			0.000888	0.001543	94.63	0.9702		
6aai	5+irule**	0.00046	0.00025	20	0	0.00087	0.001508	95.34	0.9731	95.2833	0.9779
6aiii		0.00185	0.00025			0.000677	0.001288	95.88	0.9903		
6bi		0.00238	0.00025			0.000623	0.001201	96.00	0.9917		
6bii	5+irule**	0.00079	0.00012	20	20	0.000702	0.001285	95.63	0.9855	96.0433	0.9901
6biii		0.00406	0.0005			0.000637	0.001113	96.50	0.9932		
7ai		0.00076	0.000125			0.000741	0.001366	94.88	0.9811		
7aai	3+logmyindex+irule**	0.00046	0.00025	20	20	0.00087	0.001508	95.50	0.9916	95.2100	0.9887
7aiii		0.0008	0.000125			0.000701	0.001317	95.25	0.9934		
8bi		0.00488	0.0005			0.000626	0.001329	94.63	0.9913		
8bii	3+sqrmindex+irule**	0.00488	0.0005	20	20	0.000624	0.001315	94.88	0.9935	94.9200	0.9929
8biii		0.00488	0.0005			0.000627	0.001317	95.25	0.994		

irule** represents the classified image from the rule-based classifier.