Assessing and Improving Methods for the Effective Use of Landsat Imagery for Classification and Change Detection in Remote Canadian Regions

Juan Xia He

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate in Philosophy degree in Geography

Department of Geography
Faculty of Arts
University of Ottawa

Abstract

Canadian remote areas are characterized by a minimal human footprint, restricted accessibility, ubiquitous lichen/snow cover (e.g. Arctic) or continuous forest with water bodies (e.g. Sub-Arctic). Effective mapping of earth surface cover and land cover changes using free medium-resolution Landsat images in remote environments is a challenge due to the presence of spectrally mixed pixels, restricted field sampling and ground truthing, and the often relatively homogenous cover in some areas. This thesis investigates how remote sensing methods can be applied to improve the capability of Landsat images for mapping earth surface features and land cover changes in Canadian remote areas. The investigation is conducted from the following four perspectives: 1) determining the continuity of Landsat-8 images for mapping surficial materials, 2) selecting classification algorithms that best address challenges involving mixed pixels, 3) applying advanced image fusion algorithms to improve Landsat spatial resolution while maintaining spectral fidelity and reducing the effects of mixed pixels on image classification and change detection, and, 4) examining different change detection techniques, including post-classification comparisons and threshold-based methods employing PCA(Principal Components Analysis)-fused multi-temporal Landsat images to detect changes in Canadian remote areas. Three typical landscapes in Canadian remote areas are chosen in this research. The first is located in the Canadian Arctic and is characterized by ubiquitous lichen and snow cover. The second is located in the Canadian sub-Arctic and is characterized by well-defined land features such as highlands, ponds, and wetlands. The last is located in a forested highlands region with minimal built-environment features.

The thesis research demonstrates that the newly available Landsat-8 images can be a major data source for mapping Canadian geological information in Arctic areas when Landsat-7 is decommissioned. In addition, advanced classification techniques such as a Support-Vector-Machine (SVM) can generate satisfactory classification results in the context of mixed training data and minimal field sampling and truthing. This thesis research provides a systematic investigation on how geostatistical image fusion can be used to improve the performance of Landsat images in identifying surface features. Finally, SVM-based post-classified multi-temporal, and threshold-based PCA-fused bi-temporal Landsat images are shown to be effective in detecting different aspects of vegetation change in a remote forested region in Ontario. This research provides a comprehensive methodology to employ free Landsat images for image classification and change detection in Canadian remote regions.

Preface

This is a paper-based PhD thesis. Chapters 2 and 3 have been published in peer-reviewed journals. Chapter 4 has been submitted to a peer-reviewed journal and is conditionally accepted, pending revision. Chapter 5 will be submitted to an applied geomatics journal. I am intellectually responsible for Chapters 2, 4, and 5, for which I am the primary author formulating the research questions, designing the research scheme, undertaking the analysis, writing the manuscripts and generating the figures and tabular results. I am the second author on only one paper, Chapter 3, for which my contribution is in the methodology of the paper.

The full citation for Chapter 2 is:

He, J., J.R. Harris, M.Sawada, P. Behnia. 2015. "A comparison of classification algorithms using Landsat-7 and Landsat-8 data for mapping lithology in Canada's Arctic." ***International Journal of Remote Sensing*** 36(8): 2252-2276.

The full citation for Chapter 3 is:

Harris, J.R., **J.X. He**, R. Rainbird, P. Behnia. 2014. "A Comparison of Different Remotely Sensed Data for Classifying Bedrock Types in Canada's Arctic: Application of the Robust Classification Method and Random Forests." ***Geoscience Canada*** 41(4): 557-584.

The full citation for Chapter 4 is:

He, J., M. Sawada, J.R. Harris. 2015. "A Systematic Investigation of Geostatistical Image Fusion for the Improvement of the Spectral Fidelity and Spatial Detail in Landsat MS Imagery." ***International Journal of Remote Sensing*** (in revision).

Another paper for which I am the third author has been published in a peer-reviewed journal in the course of my PhD study. The full citation for this paper is:

Harris, J. R., E.C. Grunsky, **J.He,** D.Gorodetzky, and N.Brown. 2012. "A robust, cross-validation classification method (RCM) for improved mapping accuracy and confidence metrics." ***Canadian Journal of Remote Sensing*** 38 (1): 69-90.

Acknowledgements

This thesis would not have been possible without help and support from many people, not all of whom I can mention here.

I gratefully acknowledge the love and support from my husband Jian and my son Yiyang. Thanks for their love, encouragement and confidence in my ability to finish this thesis.

I want to thank my supervisor Dr. Michael Sawada for his guidance and support. Dr. Sawada has guided my academic career since 2007 and since then I am grateful to have been able to count on his guidance, advice, and support whenever I need it. I want to thank my co-supervisor Dr. Jeff Harris for his encouragement and support over the course of the study. Also, I want to express my gratitude to my committee members Dr. Huhua Cao, Dr. Luke Copland, and Dr. Konrad Gajewski for their advice and support.

I also want to thank my friends and lab mates with whom I have had the pleasure time in the Laboratory for Applied Geomatics and GIS Science. Thanks you everyone for all the great advice and encouragement regarding my thesis research. A special thanks to Kate Ploeger, Lazar Ilic, Jingle Liu, Ruibo Han and his wife Linna.

Table of Contents

List of Tables

List of Figures

List of Acronyms

AI	Aggregation Index
AR1	Lag-1 Autocorrelation
ASTER VNIR_ SWIR_r	Aster very near infrared and shortwave infrared (raw)
ASTER_rr	Aster VNIR and SWIR (raw and ratios)
ASTER–TIR	Aster thermal bands
Ave A	Average Overall Accuracy
Ave P	Average Producer's Accuracy
Ave U	Average User's Accuracy
Ave-PA	Average Producer's Accuracy
Ave-UA	Average User's Accuracy
BPNN	Back-propagation algorithm
C	Contrast
CA	Total class area
CC	Correlation Coefficient
COHESION	Patch cohesion index
DEM	Digital Elevation Model
DOS	Dark Object Subtraction
DRAPE	Digital Raster Acquisition Project for the East
ERGAS	Relative Non-dimensional Global Synthesis Error
ERTS	Earth Resources Technology Satellite Technology *Satellite*
GEM	Geo-mapping for Energy and Minerals
GEMS	The Geo-mapping for minerals and energy program
GIS	Geographic Information System
GLS	Generalized Linear Squares
GS	Gram-Schmidt
GSC	Geological Survey of Canada
HPCC	High-pass Correlation Coefficient
HSV	Hue, Saturation, Value
ICB	information-class-basis
IHS	Intensity, Hue, Saturation
L	Luminance
L7	Landsat-7
L7_MLC	Landsat-7 Maximum Likelihood classification generated majority map
L7_NN	Landsat-7 Neural Network generated majority map

L7_r	Landsat-7 (raw)
L7_RF	Landsat-7 Random Forest generated majority map
L7_rr	Landsat-7 (raw and ratios)
L7_rr_mag	Landsat-7 (raw, ratios and magnetics)
L7_SVM	Landsat-7 Support Vector Machine generated majority map
L8	Landsat-8
L8_NN	Landsat-8 Neural Network generated majority map
L8_r	Landsat-8 (raw)
L8_RF	Landsat-8 Random Forest generated majority map
L8_rr	Landsat-8 (raw and ratios)
L8_rr_mag	Landsat-8 (raw, ratios and magnetics)
L8_SVM	Landsat-8 Support Vector Machine generated majority map
L8-MLC	Landsat-8 Maximum Likelihood classification generated majority map
LPI	Largest patch index
MCS	Multiple Classification Systems
MLC	Maximum Likelihood classifier
MLP	Multi-layer perceptron
MS	Multispectral
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
NNs	Neural Networks
NP	Number of patches
NWT	Northwest Territories
OK	Ordinary Kriging
OLI	Operational Land Imager
OLS	Ordinary Linear Squares
oob	out-of-bag
PA	Producer’s accuracy
PAFRAC	Perimeter-area fractal dimension
PAN	Panchromatic
PC1	The first component of a PCA result
PCA	Principal Components Analysis
PCCD	post-classification change detection
PIF	pseudo-invariant feature
PLAND	Percentage of landscape
RCM	Robust Classification Method
RCM-MLC	Robust Classification Method-Maximum Likelihood Classification

RF	Random Forest
RK	Regression-kriging
RKF	Radial Kernel function
RPM	Remote Predictive Mapping
S	Structure
SCB	Spectral-class-basis
SPD	Spectral Discrepancy
SSIM	Structure Similarity
SVM	Support-Vector Machine
SWIR	Short Wavelength Infrared
TBCD	Threshold-based change detection
TD	Transformed Divergence
TIR	Thermal
TM	Thematic Mapper
TOA	Top-of-atmosphere
UA	User's accuracy
UK	Universal Kriging
USGS	US Geological Survey
V	visible
VNIR	Visible and Near Infrared
XS	multispectral

Chapter 1 Introduction

This thesis is a methodological contribution that focuses on ways of leveraging free Landsat images to effectively map land surface cover and detect changes in Canadian remote regions. Remote regions in the thesis are defined as areas that have minimal human footprints. The research is approached from the following four perspectives: 1) determining the continuity of Landsat-8 images for mapping surficial materials when Landsat-7 is decommissioned, 2) selecting classification algorithms that best address challenges involving mixed pixels, 3) applying advanced image fusion algorithms to improve Landsat spatial resolution while maintaining spectral fidelity and reducing the effects of mixed pixels on image classification and change detection, and, 4) examining different change detection techniques, including the post-classification change detection (PCCD) method and the threshold-based change detection (TBCD) method based on PCA(Principal Components Analysis)-fused bi-temporal Landsat images to detect changes in Canadian remote areas. These four approaches will form a comprehensive methodology to employ free Landsat images for image classification and change detection in Canadian remote regions.

Over the past three decades, remotely sensed images with various spatial, spectral, temporal, and radiometric resolution have become an important data source for earth surface feature identification (e.g. bedrock types, land cover types), classification and change detection. Landsat images are widely used due to their long record of continuous measurement (data archive back to 1972), spatial resolution, global coverage, near nadir observations, and most importantly, their free availability. The literature is proliferated with Landsat applications in various environments, such as foresty (e.g. Tottrup 2004; Margono et al. 2012; Zhu et al. 2012), geological mapping (e.g. Mwaniki et al. 2015; Yazdi et al. 2013), land cover mapping and

change monitoring (e.g. Haack et al. 1987; Vittek et al. 2014). In 2013, Landsat-8, with improved sensors and technology, started to provide continuous global coverage. Landsat 8 differs from Landsat 7 by including different wavelengths with narrower bandwidths, two new spectral bands, and 16-bit radiometric resolution. Some research has been conducted to investigate the Landsat-8 OLI (Operational Land Imager) sensor in aboveground biomass estimation (Dube and Mutanga 2015), land cover classification and change detection (Estoque and Murayama 2015; Mei et al. 2015), and geological mapping (Mwaniki et al. 2015). A question requiring further investigation is how the newly available Landsat-8 images together with the existing Landsat archive data can be used in consistently identifying earth surface cover and detecting land cover changes in remote areas – areas that are characterized by a lack of rapid landscape changes, such as the Canadian Arctic and sub-Arctic.

The first approach to this question is to determine if Landsat-8 images are capable of mapping geological information in Canada's remote areas. A very remote area on Victoria Island in Northwest Territories (NWT) of Canada is selected as the first study area in the thesis. Mapping geological information using medium-resolution images in such a remote area with ubiquitous lichen and snow cover and a short season for acquisition of suitable imagery, poses some challenges due to, for example, spectrally mixed pixels present in the images as well as restricted field sampling and ground truthing. Landsat images, mainly Landsat-7 images, together with ancillary data (e.g. DEM (digital elevation model)), have been used to successfully generate geological maps in Canadian northern territories including Schultz Lake (Grunsky et al. 2009), Wager Bay (Campbell et al. 2013), Hall Peninsula and Foxe Basin (Harris et al. 2012c), as well as Hearne Lake (Stevens et al. 2013). However, there are few studies dealing with the application of Landsat-8 in mapping geology in Canada's Arctic environments. Therefore, in

order to map and update geological information in such areas when Landsat-7 is decommissioned, it is necessary to investigate the continuity of Landsat-8 as a major data source for geological applications in Canada's North.

In addition to selecting suitable Landsat images, the application of classification algorithms for geological mapping in such a remote area is also a challenge. Many studies have extensively investigated the performance of conventional and advanced classifiers, including Neural Networks (NNs), Support Vector Machines (SVMs) and Random Forest (RF) (e.g. Akar and Güngör 2012; Nitze et al. 2012; Rodriguez-Galiano et al. 2014; Cracknell and Reading 2014). However, there are few studies that compare the performance of these conventional and advanced classifiers in the application of mapping surficial materials in Canada's North. With more Landsat 8 images becoming available, the capabilities of such advanced classification algorithms in generating geological maps in Arctic environments need to be explored.

The study area chosen for the investigation on the continuity of Landsat-8 images and appropriate classifiers is located in the western Minto Inlier on Victor Island, Northwest Territories. In addition to ubiquitous lichen and glacial cover, a variety of sedimentary and volcanic bedrocks are generally well exposed in the study area. This bedrock has sufficient differences in mineralogy to be seen spectrally. In concert with its coastal environment within Canadian Arctic regions, the Minto Inlier presents to be an ideal environment for mapping geological information based on remotely sensed satellite images.

Given the challenge of spectrally mixed pixels present within medium-resolution Landsat images, and with the exception of classifiers that are insensitive to mixed pixels, image fusion is an effective approach to reduce the effects of spectral mixing in Landsat images. Images from different sensors have different characteristics and capabilities in mapping land covers and

detecting change. The objective of image fusion techniques is to synthesize images with different spectral and/or spatial resolution to compensate for the insufficiency of a single-source image and to combine the complementary features within various images. Various efforts have been made to develop satisfactory fusion images (e.g. Choi 2006, Gangkofner et al. 2007). One under explored example is the regression-kriging based (RK) fusion method, a member of geostatistically-based image fusion approaches. RK was firstly proposed by Odeh et al. (1994) and has been used to successfully fuse images acquired at the same time (e.g. Hengl 2009; Meng et al. 2010). However, a systematic investigation of the RK-based fusion method is not available in the literature and this promising technique could hold significant potential for resultant image classification workflows in remote Canadian regions. The study area chosen to systematically explore that technique is a sub-Arctic remote area near Victor Mine in northern Ontario, Canada. Aside from open pits and an accommodation complex, this study area is characterised by well-defined patterning of natural features including highlands, creeks and ponds, wetlands, trees and shrubs, and therefore presents an ideal area to systematically investigate advanced image fusion techniques using the medium resolution imagery from the Landsat archive.

Given the time period covered by the imagery archive, change detection is another application of Landsat in areas that are considered remote and that exhibit only subtle changes in landscape characteristics on decadal time scales. Change detection can be accomplished in one of two ways, directly using raw satellite images or using fused satellite images. In the literature, change detection on fused imagery can be achieved by directly applying classification algorithms on fused results generated from multi-temporal images (e.g. Rokni et al. 2015). Alternatively, change detection can also be achieved by applying threshold-based (TBCD) methods on a differenced image that is developed by comparing fused images from two different times to one

of the input images. Du et al. (2012) demonstrated the capability of the TBCD method by fusing a Landsat image at Time 1 and the SPOT panchromatic image from Time 2. However, SPOT panchromatic images are not freely available in Canadian remote areas. Thus it is more valuable to explore how Landsat images acquired at different times can be fused and then used in a TBCD approach. Meanwhile, as a comparison to the TBCD method, a widely used PCCD method based on multi-temporal Landsat images is also examined concurrently. The study area chosen for change detection is a typical Canadian remote area characterized by highlands, forests, and lakes, with a minimum human footprint. This investigation will provide a new method on how to fuse Landsat images at different times to highlight changed features in remote areas.

1.1 Research Objectives

The overall research goal of this thesis is to investigate how remote sensing methods can be used to assess and improve the use of free Landsat images in identifying and classifying earth surface cover and detecting changes in Canadian remote areas. In order to pursue this overall goal, four specific research objectives are formulated:

(1) To examine the continuity of Landsat-8 images in mapping geological information in Arctic environments by comparing classification results using other available remotely sensed images, such as SPOT 5, Landsat-7 and ASTER;

(2) To investigate which classification algorithm is optimal for mapping geological information in Canadian remote areas that present spectrally mixed pixels, non-normally distributed training data, minimal field sampling and ground truthing, and relatively homogeneous land cover;

(3) To systematically investigate the effectiveness of the RK-based image fusion technique for improving spatial resolution of Landsat images while preserving spectral fidelity and reducing the influence of spectrally mixed pixels in Landsat images;

(4) To examine the performance of the post-classification detection method using multi-temporal Landsat images, and to explore how image fusion techniques can be applied to bi-temporal Landsat images to identify land cover changes in remote areas.

1.2 Organization of Thesis

The first chapter has briefly introduced the rationale and the research objectives addressed in this thesis. The fulfilment of these objectives is presented in four self-contained research articles that comprise this thesis.

Chapter 2 examines the continuity of Landsat-8 images in mapping geological information in Canadian Arctic environments by comparing Landsat-8 to Landsat-7 images. It also investigates which classification algorithms best address the challenges of using Landsat images to map land cover in the Canadian Arctic.

Chapter 3 compares Landsat images to SPOT 5 and ASTER images in mapping Canadian Arctic geological information. It also evaluates two different classification methods: Random Forest (RF) and the Robust Classification Method-Maximum Likelihood Classification (RCM-MLC). Finally a new classification method called Multiple Classification Systems (MCS) is proposed to leverage different characteristics of all remotely sensed images used for classification.

Chapter 4 provides a systematic investigation on how RK-based image fusion can be used to effectively improve the spatial resolution of Landsat images while preserving spectral fidelity. It is approached by qualitatively and quantitatively evaluating spectral and spatial

characteristics of the fused images developed by applying various combinations of regression algorithms and kriging methods and comparing these to existing common fusion approaches such as Brovey, Gram-Schmidt, and Wavelet-based fusion methods.

Chapter 5 examines the capability of Landsat archive images in identifying land cover changes using both post-classification (PCCD) and threshold-based detection (TBCD) methods. The PCCD method using multi-temporal Landsat images is first examined. The TBCD method is then used to explore how a PCA-based fusion method can be used to merge multi-temporal Landsat multispectral (MS) images and a panchromatic image to identify vegetation changes in remote areas.

Chapter 6 summarizes the conclusion of this thesis research, outlines the main contributions, and points out potential directions for future research.

Chapter 2 A Comparison of Classification Algorithms Using Landsat-7 and Landsat-8 Data for Mapping Lithology in Canada's Arctic

Abstract: To map Arctic lithology in central Victoria Island, Canada, the relative performance of advanced classifiers (Neural Network (NN), Support Vector Machine (SVM), and Random Forest (RF)) were compared to maximum likelihood classifier (MLC) results using Landsat-7 and Landsat-8 imagery. A ten-repetition cross-validation classification approach was applied. Classification performance was evaluated visually and statistically using the global classification accuracy, producer's and user's accuracy for each individual lithogical/spectral class, and cross-comparison agreement. The advanced classifiers outperformed MLC especially when training data was not normally distributed. The Landsat-8 classification results were comparable to Landsat-7 using the advanced classifiers but differences were more pronounced when using MLC. Re-scaling the Landsat-8 data from 16 bit to 8 bit substantially increased classification accuracy when MLC was applied but had little impact on results from the advanced classifiers.

Keywords: image classification, Neural Network, Support Vector Machine, Random Forest, lithology

2.1 Introduction

Remotely sensed images have been increasingly used for the creation and updating of geological maps since the advent of Landsat-1 in 1972. Numerous studies have demonstrated that remote sensing techniques, as a complement and an alternative to traditional fieldwork, provide an effective and time-saving approach to extract geological information, especially in isolated inaccessible areas, such as Canada's North (e.g., Behnia et al. 2012; Cracknell and Reading 2014; Harris et al. 2011). Among those studies, supervised image classification is the most commonly used approach for geological information extraction.

Supervised classification involves a number of stages including selection of appropriate data types, identification of representative training areas for the features to be classified and the selection of an image classification algorithm. Over the decades, traditional supervised classifiers, for example, the Maximum likelihood classifier (MLC), have been extensively applied to generate geological maps (e.g., Patel and Rampal 1993; Loizzo et al. 1995; Chen, Warner, and Campagna 2007; Wang, Niu, and Wu 2011; Li, Frei, and Altermann 2011; LaRocque et al. 2012; EI et al. 2013; Hadigheh and Ranjbar 2013). MLC labels the pixels in an image according to user defined classes based on posterior probabilities computed from the statistics of the training data (e.g., mean, covariance) (Richards and Jia 2006). MLC is an optimal classifier if the training data is normally distributed (i.e., a parametric classifier). However, obtaining normally distributed training data is often difficult in practice.

In the last two decades, various non-parametric classification algorithms have been developed in order to obtain more accurate and more reliable classifications. The most intensively-investigated and widely-used of these algorithms include Neural Networks (NNs), Support Vector Machines (SVMs), and Random Forest (RF). Many studies have compared the performance of these advanced classifiers to MLC. For example, Akar and Güngör (2012) compared RF, SVM and MLC for land cover mapping and found that RF outperformed SVM and MLC using IKONOS and QuickBird data. Nitze, Schulthess, and Asche (2012) compared NN, RF, SVM and MLC for classifying agricultural crop types using Landsat TM data, and the comparative study showed that SVM provided higher classification accuracy and lower classification variability when compared to NN, RF and MLC. In addition, some studies have investigated the capabilities of these advanced classifiers for geological mapping (e.g. Leverington 2010; Leverington and Moon 2012; Yu et al. 2012; Abedi, Norouzi, and Bahroudi

2012; Cracknell and Reading 2014; Waske et al. 2009; Harris et al. 2014). In general, previous comparative studies on the performance of these advanced classification approaches have demonstrated that NNs, SVMs, and RF outperformed conventional classification methods like MLC. Such advanced classifiers offer advantages that are not provided by conventional classifiers. For example, NNs, RF, and SVMs make no statistical assumptions in terms of the distribution of training data, and can produce satisfactory classification results even with a fewer number of training data or with spectrally-mixed training data (Waske and Benediktsson 2010).

A NN includes three main components: a set of inputs, hidden layers, and outputs. The most widely used NN model is the multi-layer perceptron (MLP) neural network based on a back-propagation algorithm (BPNN). Some comparative studies (e.g., Paola and Schowengerdt 1995; Mas and Flores 2008) have demonstrated the superiority of NNs because 1) NNs do not assume that the training data are normally distributed, and 2) the nature of NNs, for example, a BPNN classification process, involves a repeated feed-forward and back-propagation process in order to minimize the root-mean-square error between the model's predicted and target values (Paola and Schowengerdt 1995). Neural networks are also able to process large amounts of noisy data. However, NNs run slowly during the training phase and setting parameters during the training phase is difficult because of the many input parameters required (Arora, Tiwari, and Mohanty 2000).

SVMs are based on statistical learning theory and employ optimization algorithms to generate boundaries (hyperplanes) that separate classes in multi-dimensional feature space. The optimal procedure aims to maximize the margin between the closest training data and the hyperplane, therefore minimizing the confusion between classes. In practice, SVMs use a subset of training data (called support vector) to partition the feature space without assumptions on the

statistical distribution of the training data. SVM is increasingly used in the remote sensing community due to its ability to generate good classification results even with limited and/or spectral-mixed training data, a common limitation for traditional digital image classification (Fauvel, Chanussot, and Benediktsson 2006; Foody and Mathur 2006; Mountrakis, Im, and Ogole 2011). When SVM is applied, a set of user-defined parameters (i.e., choice of kernel, kernel specific parameters, regularization parameter) is required to obtain optimal results. However, the choice of the required parameters can significantly influence the classification accuracy achieved by SVMs (Huang, Davis, and Townshend 2002; Pal 2005; Mountrakis, Im, and Ogole 2011; Mondal et al. 2012). Thus, an optimum parameter search must be performed to achieve the best classification result for a given dataset (Kavzoglu and Colkesen 2009).

RF, unlike NNs or SVMs, is an ensemble classifier that breaks a complex classification process down into a series of decision processes (i.e., trees). Each tree (classifier) is generated by applying a subset of variables (i.e. features) that is randomly sampled from the total number of variables on a bootstrapped sample of training data (generally 2/3 used for classification and 1/3 for validation). The most frequently used feature selection measures used by RF are the Information Gain Ration criterion (Quinlan 1993) and the Gini Index (Breiman et al. 1984). Each tree is grown to the maximum extent with no pruning so that an ensemble of trees is created. Each pixel in the final classification map is assigned to the majority class across all the tree predictions in the forest by employing a voting procedure (most popular). The RF classifier is gaining increasing attention because of fast computation, few parameters requirements, few statistical assumptions on training data, less sensitivity to noise or over-fitting, and the capability to determine variable importance (Pal 2005; Rodriguez-Galiano et al. 2012; Harris et al. 2014).

A wide range of remotely sensed data are now available for mapping geological information, ranging from free coarse and medium resolution imagery (e.g. free Landsat and SPOT imagery for the Canada area available from www.geobase.ca) to expensive high resolution images like QuickBird (Harris et al. 2011). Landsat-7 data, due to its advantages (free-cost, large coverage, and regional spatial resolution), have been used as a major data source for mapping land cover, bedrock and surficial cover for the past three decades. Landsat-8 offers some technical advancements over Landsat-7, mainly different wavelengths of narrower bandwidths and two new spectral bands. The narrower optical bandwidths are expected to increase capabilities for feature discrimination and change detection. Thus there is a need to investigate the capability of Landsat- 8 data for geological mapping, and to compare the results to Landsat-7 data. Harris et al. (2014) compared various sensors, including Landsat- 7 and Landsat-8 for classifying lithologies in central Victoria Island, Northwest Territories of Canada using the Robust Classification Method (RCM). They found that Landsat -7 and Landsat-8 provided similar classification accuracies.

This study compared three state-of-the-art advanced classifiers, namely NNs, SVMs, and RF to the conventional and well-used MLC classifier for mapping different lithologies (rock types) in Canada's North using Landsat-7 and Landsat-8 data. The objectives of the study were two-fold: the first was to examine the performance of the aforementioned four supervised classification methods, and the second was to investigate if Landsat-8 imagery could provide improved classification results when compared to Landsat-7 imagery.

2.2 Study Area

The study area is located within the western Minto Inlier on Victoria Island, Northwest Territories of Canada (Figure 2.1). The lithological units were first mapped by Thorsteinsson and

Tozer (1962), and more recently revised by Rainbird et al. 2013a, b, c, d. The main lithological units in the study area (Figure 2.1) include the Minto Inlet and Wynniatt, Killian and Kuujjua Formations of the Shaler Supergroup and Cambro-Ordovician succession of the Palaeozoic. A more detailed lithological description can be found in Behnia et al. (2012).

In this study, both the Landsat-7 and Landsat-8 data were downloaded from the US Geological Survey (USGS) Global Visualization Viewer (http://glovis.usgs.gov). The Landsat-7 orthorectified imagery (in 8-bit format) was collected on 14 July 2002 whereas the Landsat-8 imagery (in 16-bit format) was acquired on 4 July 2013. Atmospheric effects on the imagery were removed by using the dark subtraction method (Gupta 1991). Figure 2.2 shows the contrast enhanced Landsat-7 (Band 742 RGB composite) and Landsat-8 images (Band 753 RGB composite) of the Victoria Island study area.

In order to better discriminate various lithologies, areas covered by water, ice and vegetation were masked using SPOT-5 images acquired on 22 August 2009, purchased from Blackbridge (http://www.blackbridge.com) as well as Landsat-7. First, the vegetation cover within the study area was distinguished by applying a NDVI (Normalized Difference Vegetation Index, Richard & Jia 2006) derived from the Landsat-7 and Landsat-8 imagery that were acquired in July when biomass was the most extensive. Second, the delineation of snow and ice cover used SPOT-5 imagery because the 2009 SPOT-5 images appeared to show the largest extent of snow and ice.

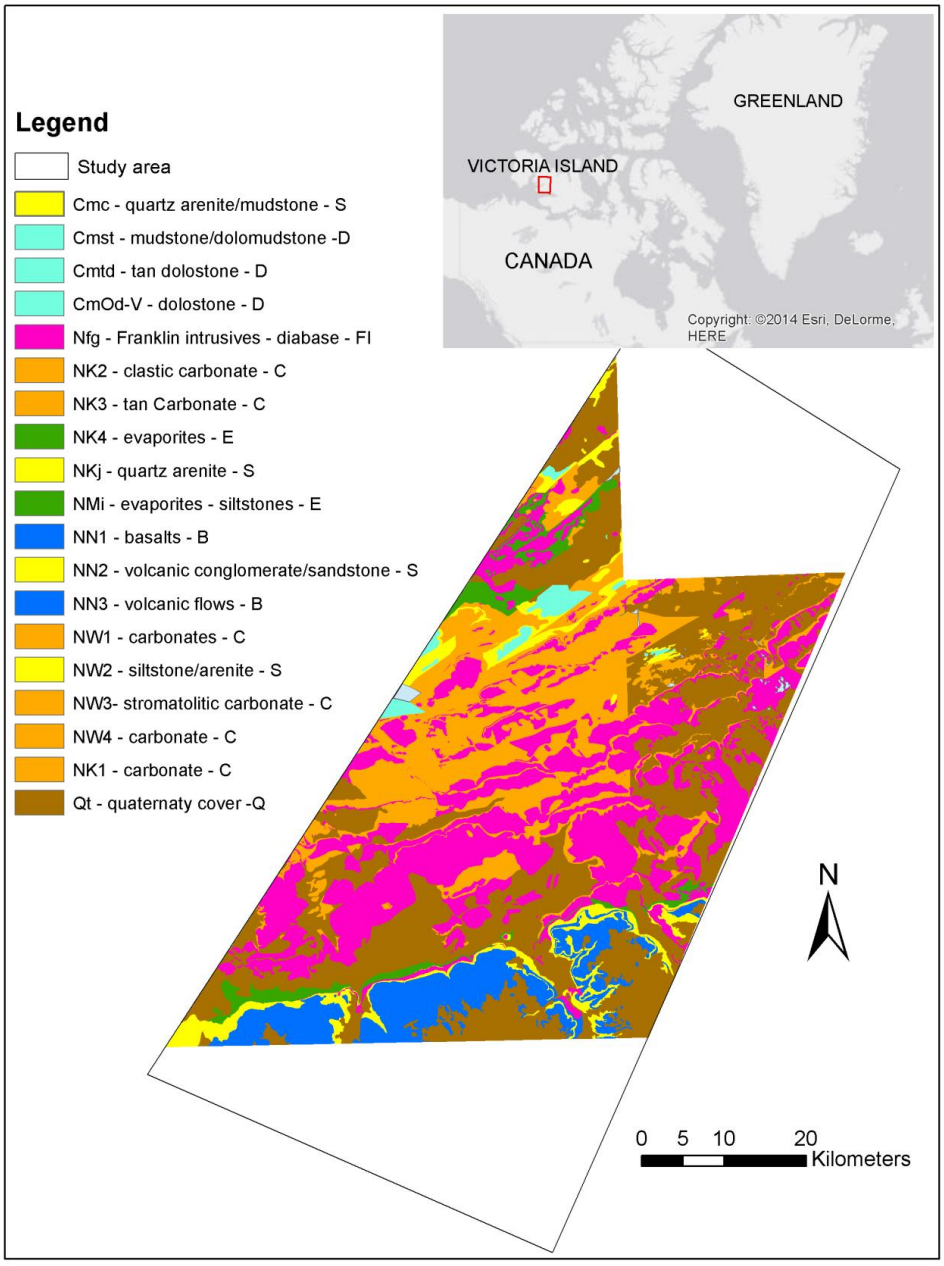

Figure 2.1: Location of the study area (on the top): Minto Inlier, Victoria Island, NWT, Canada. The new detailed geology map of the study area (Rainbird et al. 2013a, b, c, d) is shown in the centre. See Table 2.1 for description of the lithological unit. The symbols represent the generalized and grouped lithologies and one surficial material (quaternary cover, tills) used for comparing the various classifications produced (see text) (S = sandstone, D = dolostone, FI = Franklin sill, C = carbonate, E = evaporate, B = basalts, Q = quaternary cover).

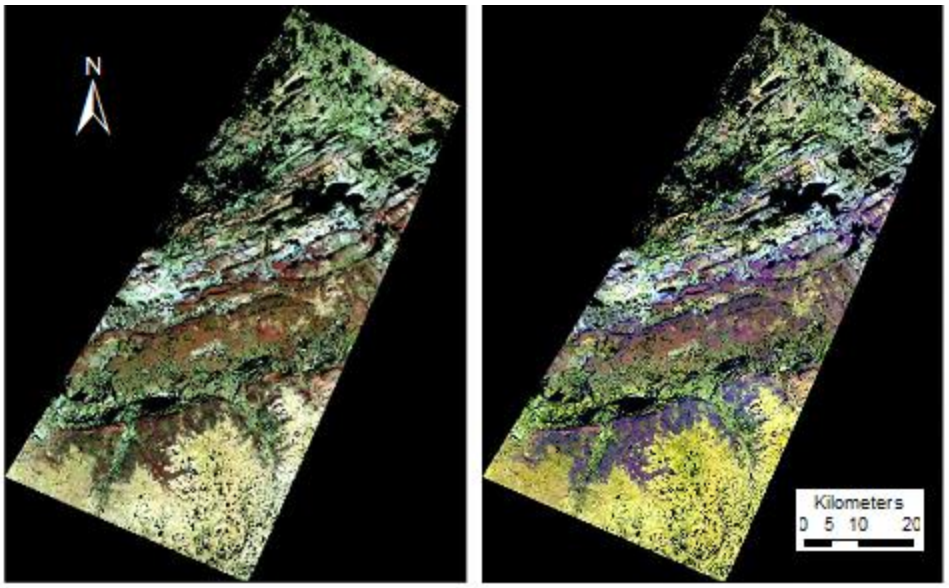

Figure 2.2: The Landsat-7 (left, bands 742-RGB) and Landsat-8 (right, bands 753-RGB) images for the study area – contrast enhanced using a 2% linear stretch. Black areas within the image are a mask used to exclude vegetation with high-chlorophyll, snow, ice and water.

2.3 Methods

2.3.1 Training data

Based on a combination of field observations, visual interpretation of SPOT-5 stereo imagery, and the spectral diversity observable in the Landsat-7 imagery, six main lithological units were defined in this study, including the Natkusiak Formation basalt, Franklin diabase, carbonate of the Wynniatt Formation and of the Cambro-Ordovician succession, Paleozoic dolostones, quartz-arenite of the Kuujjua Formation, and evaporate rocks of the Minto Inlet and Killian Formations (shown in Table 2.1).

Vegetation (with high chlorophyll) was excluded from the analysis by applying a mask as described above, however some small areas covered with dry (low chlorophyll) vegetation were still present in the images, which may cause spectral interference with other bedrock spectra. Thus, one class was created named dry vegetation in order to minimize spectral

interference. A given lithological feature may exhibit variable spectral responses in different parts of the study area due to the varying overlying glacial sediments and/or overburden materials, weathering process, moisture content, or vegetation cover. For example, basalt in some parts of the study area exhibits lower spectral reflectance in the Visible and Near Infrared (VNIR) and Short Wavelength Infrared (SWIR) bands of the Landsat-7 image. To address such variability, three basalt classes were created, two in the non-vegetated and one delineating basalt covered by thin and dry vegetation. Similarly, two spectral classes were used for Franklin gabbro/diorites sills and dykes, one non-vegetated and one covered by thin vegetation. The carbonate features with different formations (i.e., the Wynniatt Formation and the Cambro-Ordovician succession) were combined into one spectral class because of their indistinguishable spectral signatures. In addition, three surficial classes were developed for different formation-types of tills. In total, 13 spectral classes were used in the classification, details of which are listed in Table 2.1.

Table 2.1: Description of the lithological units and surficial materials (till and dry vegetation), together with the spectral units and the training areas used for classification of the Landsat-7 and Landsat-8 data using MLC, NN, SVM, and RF. In total, 13 spectral classes were classified in the study.

Lithology units and surfical materials	Spectral/lithological class and description	Training areas (13 units)
Natkusiak Formation basalt	Natkusiak Formation basalt extending SE-NW in southern part, having low reflectance in VNIR and moderate reflectance in SWIR of the Landsat-7 data	nfb1 (4 polygons, 184 pixels)
	Natkusiak Formation basalt having low reflectance in VNIR and SWIR of the Landsat-7 data	nfb2 (3 polygons, 144 pixels)
	Natkusiak Formation basalt covered by thin vegetation	nfbv (7 polygons, 87 pixels)
Franklin Diabase	Franklin Diabase extending in the central and northern parts of the area	nfg (9 polygons, 420 pixels)
	Franklin diabase covered by thin vegetation	nfgv (4 polygons,187 pixels)

Carbonates of the Wynniatt Formation Cambro/Ordovician (carbonates)	Carbonates	Carbonates (9 polygons, 293 pixels)
Paleozoic dolostones	Paleozoic dolostones	Paleozoic Dolostones (6 polygons, 82 pixels)
Quartz-arenite of the Kuujjua Formation (Occur in contact with the basalts)	Quartz-arenite of the Kuujjua Formation.	Sandstones (4 polygons, 178 pixels)
Evaporite rocks (Includes evaporates of Minto and Killian Formations)	Evaporite rocks	Evaporites (8 polygons, 140 pixels)
Quaternary cover	Till1	Till1 (5 polygons, 191 pixels)
	Till2	Till2 (9 polygons, 156 pixels)
	Till3	Till3 (8 polygons, 224 pixels)
Dry Vegetation (missed by vegetation mask)	Dry Vegetation/low chlorophyll	Dry Vegetation (13 polygons, 255 pixels)

All the training data for the assigned lithological/spectral classes were identified by geological scientists with the aid of aerial photo interpretation, Landsat-7 images and field work. Figure 2.3 plots the spectral signatures of the 13 spectral classes investigated in the study. Due to spatial heterogeneity of the lithological classes, the training data were identified using small polygons instead of individual pixels. In addition, spectral separabilities (here measured using Transformed Divergence (TD) (Richard and Jia 2006)) between pairs of the training data were computed and the TD values show that all the training data display good spectral separability. Thus, classification of the data is a viable processing method.

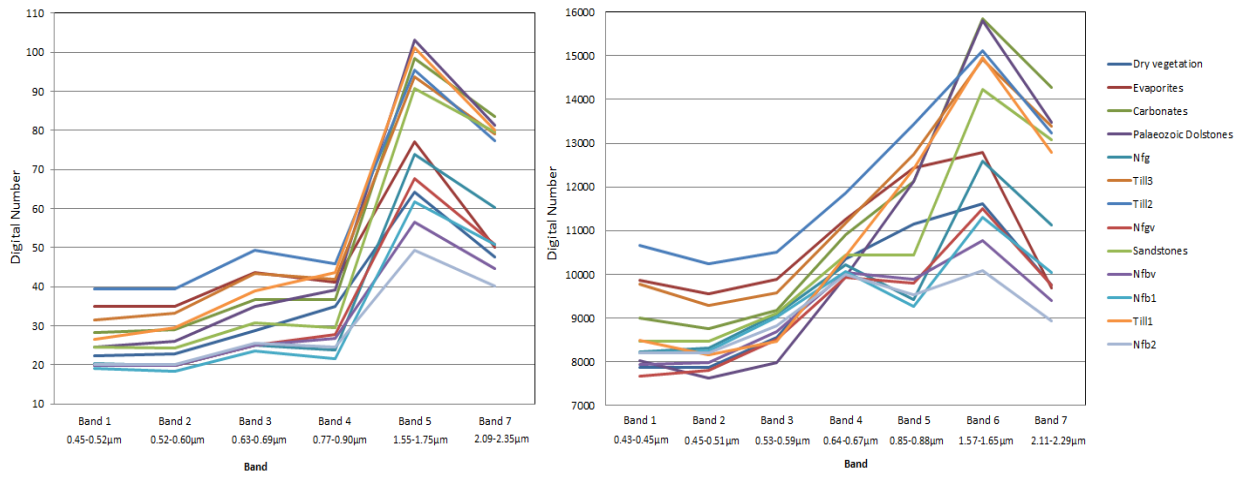

Figure 2.3: The spectral signatures of the 13 spectral classes investigated in the study. The left plot is based on the Landsat-7 data, and the right plot is based on the Landsat-8 data.

2.3.2. Classification

Supervised classification labels pixels to one of the user-defined classes based on a prior knowledge of the features to be classified. In this study, MLC (Maximum Likelihood classification), NN (Neural Network), SVM (Support-Vector Machine), and RF (Random Forest) classifiers were applied to label each pixel within the images. The original training data were split into three random subsets using the RCM (Robust Classification Method) tool (Harris et al. 2012b). One-third of the original training areas (independent test dataset) were randomly selected and were used for an independent evaluation of the classification accuracies derived from the four classification methods. The remaining two-thirds of the original training areas were then used to perform a 10-repetition cross-validation classification in order to reduce training data variability (i.e., bagging). In each repetition, the remaining two-thirds of the original training areas were randomly partitioned into two subsets with equal number of polygons, since 50%-50% splitting of the training areas for classification and evaluation tends to yield higher overall accuracy classification results (Harris et al., 2012). One subset (training dataset) was used

to train the classifier and the other subset (validating dataset) to evaluate classification accuracy. Each training dataset was evaluated using TD values which demonstrated good spectral separability (i.e., the TD values for each pair of the training dataset are generally larger than 1.9) and thus could be used for classification (Richard and Jia 2006). In addition, the normality of the distribution of each training dataset was examined before classification by visual examination of histograms and applying the Anderson-Darling Normality test (Stephens 1974) in Minitab (www.minitab.com). The examination revealed that all the training datasets were not normally distributed and were skewed to either right or left. After the 10-repetition classification, a majority classification map was then generated by aggregating the 10 classification results through majority voting. In the study, MLC, NN, SVM, and RF used the same randomly sampled training dataset and validating dataset in each of the 10 repetitions. The intention of this classification approach was to reduce the variance in the classification results caused by sampling variability of the training areas.

The standard backpropagation NN classifier (Richard and Jia 2006) was used in the study. The number of hidden layers was set to 1, and the training rate and training momentum values were adjusted to achieve optimal results. In the SVM classification in the study, the parameters of radial and polynomial kernels, two commonly used kernel functions in remote sensing applications, were tested, and the classification result giving the highest overall accuracy value was adopted as the final classification result for each repetition. In most of the classifications in the study, the radial kernel function (RKF) generated better classification results. The selection of RKF was also recommended as a good kernel function in some studies (e.g., Pal and Mather 2004; Kavzoglu and Colkesen 2009; Otukei and Blaschke 2010). The

MLC, NN, and SVM classifications were carried out using ENVI (*Exelis Visual Information Solutions, Boulder, Colorado*).

In RF classification, the *Gini index* which measures the impurity of each node in a decision tree was used to construct decision trees. In order to identify an optimal combination of the number of trees and the number of variables from which the out-of-bag (oob) error converges and the prediction (i.e., classification map) has the highest classification accuracy (calculated against the validating dataset), RF models with a different number of sampled variables and different number of trees (from 100 to 1500) were constructed and validated for each training dataset for both the Landsat-7 and Landsat-8 data. Figure 2.4, for example, shows the RF classification accuracies (i.e., overall accuracy values derived from the validating dataset) using the Landsat-7 data and the training dataset used in Repetition 1 classification (Table 2.2) with different number of trees and number of sampled variables. Figure 2.4 suggests that the square root of the total number of variables (i.e., 2 bands of the Landsat-7 data) consistently yielded the highest accuracy classification result across any number of trees tested in the study. Selection of the square root of the number of variables for each tree was also recommended by Breiman (2001) and Gislason, Benediktsson, and Sveinsson (2006). As a result, the number of randomly sampled variables in all the RF classifications in the study was set to the square root of the input bands, which is 2 bands for the Landsat-7 data and 3 bands for the Landsat-8 data. In terms of the optimal number of trees for each training dataset using either the Landsat-7 or Landsat-8 data, results vary from 100 to 500. For example, Figure 2.4 suggests that for the training dataset used in Repetition 1, 200 trees generated the highest classification accuracy when using the Landsat-7 data; however, 100 trees provided the highest classification accuracy when using Landsat-8 data (Figure 2.5). Thus, different numbers of trees ranging from 100 to 1500 were

examined for each training dataset and the classification result with the highest overall accuracy was chosen for further analysis. The RF classification was implemented by using EnMap-Box (Rabe et al. 2014).

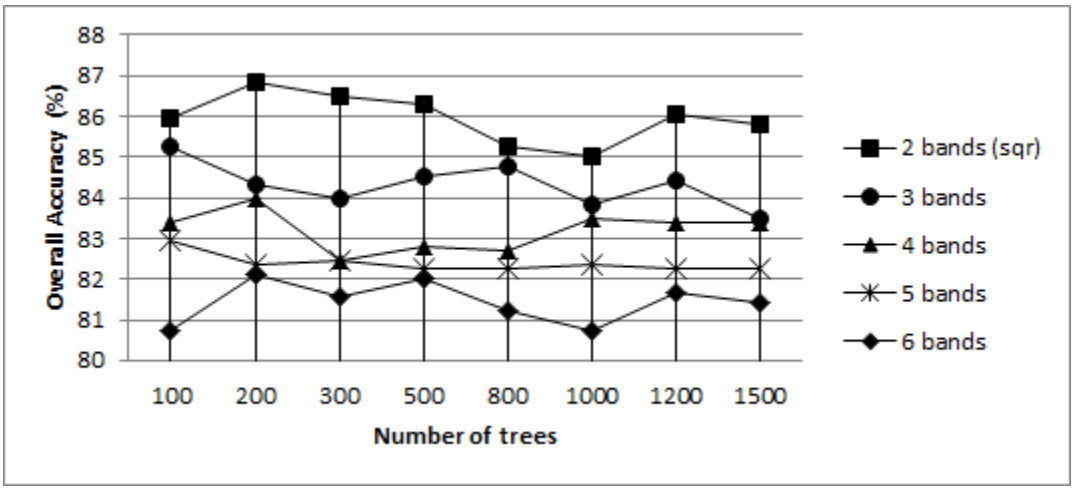

Figure 2.4: Variation of overall accuracy values for the RF classification result by using varying number of trees from 100 to 1500 and varying number of sampled variables from the Landsat-7 image. The training and validating datasets used in Repetition 1 classification (see Table 2.2) were used.

In addition to classification maps, NN, SVM, and RF also produce probability maps whereby pixel values range from 0 to 1 and indicate the probability that a given pixel belongs to a specific class (i.e. certainty of class membership). Unlike NN, SVM, and RF, MLC creates rule images whose values are discriminant functions for each pixel, rather than probability values; so that, in this study, the rule images generated from MLC were first converted to class membership probability images based on the formula provided by Richard and Jia (2006), and then the converted probability images were standardized according to the method of Chen (2008). The probability images were used to label tie pixels in the process of generating majority classification maps.

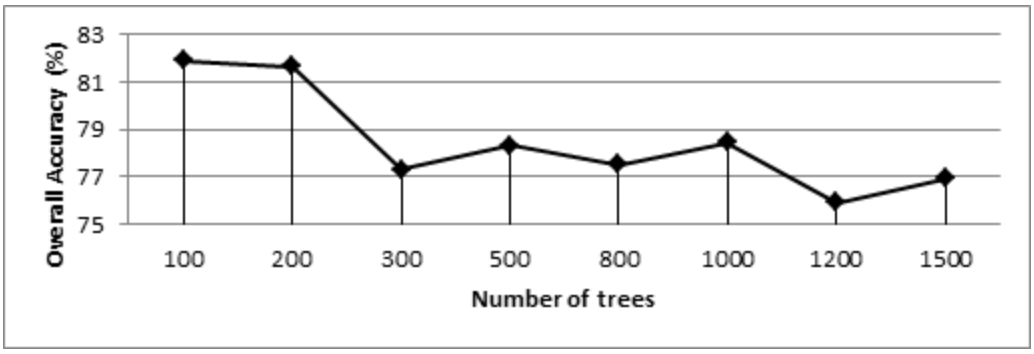

Figure 2.5: Variation of overall accuracy values for the RF classification result by using a different number of trees and using the Landsat-8 data and the training dataset and validating dataset used in Repetition 1 classification (Table 2.2). The number of sampled variables at each node is 3, the square root of the total number of bands of the Landsat-8 image.

In the study, each of the four classification methods was run 10 times on the Landsat-7 and Landsat-8 data. In each repetition, the classifier was trained and validated by a paired training and validating dataset discussed above. A majority classification map for each classifier using the Landsat-7 and Landsat-8 data was then created, where each pixel was populated by the class that occurred most frequently (i.e. majority voting) over the 10 repetitions, as well as an uncertainty map was generated based on the variability in the classification on a pixel-to-pixel basis. The tied pixels (i.e., the pixels were assigned the same number of different classes throughout the process) were assigned to the class having the highest probability value across the 10- repetition classification. The accuracy of the majority maps was evaluated by using the independent test dataset.

2.4 Results

2.4.1. The majority and uncertainty maps from the 10-repetition classification

Figure 2.6 presents the majority classification maps for each of the 4 classification algorithms, generated over the 10 repetitions, applied to the Landsat-7 and Landsat-8 data. Overall, all the

majority maps exhibit visually similar patterns of the regional distribution of bedrock lithological units. For example, Franklin diabase units (i.e. nfg and nfgv, see Table 2.1 for symbols, purple and dark purple pixels in the classification maps) extend from west to east in the central part of the study area; most basalts (i.e. nfb1,nfb2, and nfbv, blue pixels in the classification maps) present in south part of the study area above carbonate features (i.e. brown pixels in the classification maps), and some basalts present in contact with diabase features; till features present in the north part; areas covered by dry vegetation are mostly in the north part of the area. However, there are some appreciable differences between the Landsat-8 MLC generated (L8-MLC) majority map and the other majority classification maps. For example, in the L8-MLC majority map (Figure 2.6*b*), pixels in the center part of the study area were assigned to basalts (blue pixels) rather than Franklin diabase (purple pixels) when compared to the other majority maps because diabase (nfg) exhibits very similar spectral reflectance to basalts (nfb1) in the VNIR bands on the Landsat-8 data due to similar mineralogy. Moreover, a greater percentage of the study area, particularly in the south of the study area, was classified as dry vegetation (light green pixels) on the Landsat-8 MLC map rather than carbonates (brown pixels) as in the other maps. The different radiometric resolutions of the Landsat-8 (16 bit) and Landsat-7 (8 bit) data may contribute to these differences (see discussion below).

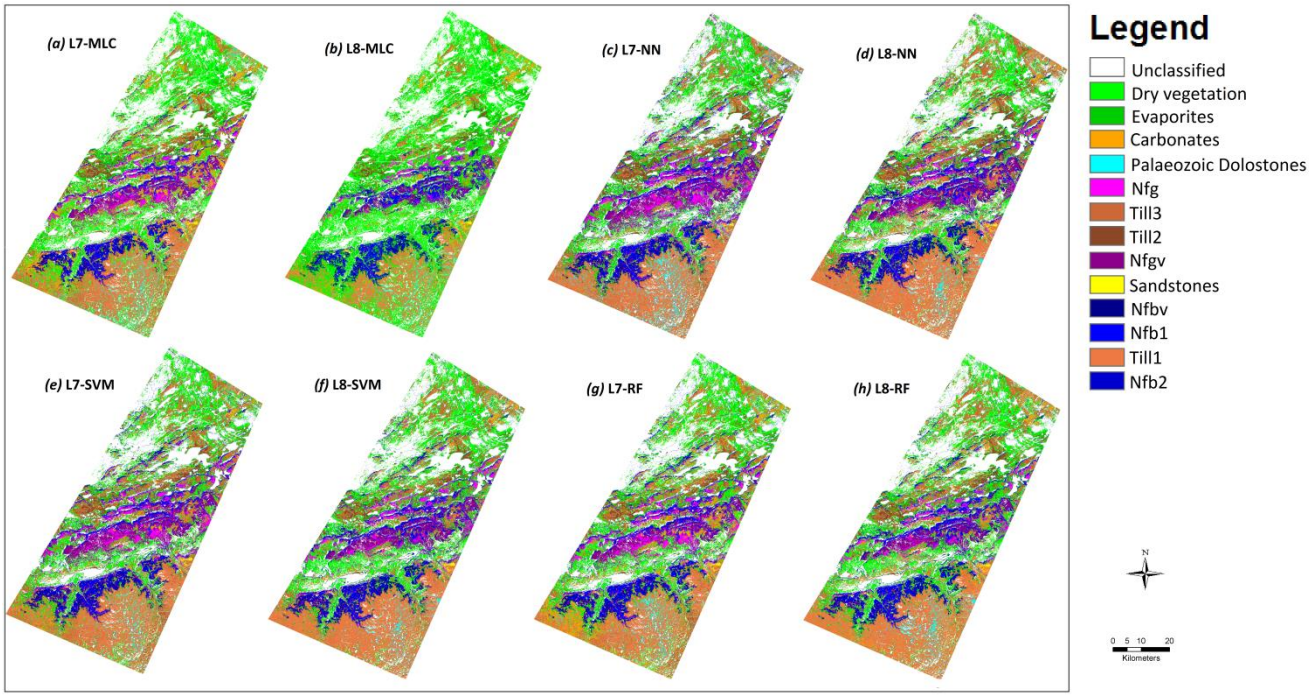

Figure 2.6: The majority classification maps generated by using MLC, NN, SVM, and RF on Lnadsat-7 and Landsat-8. (a) the majority map from MLC and Landat-7 (L7-MLC), (b) the majority map from MLC and the Landsat-8 data (L8-MLC), (c) the majority map from NN and the data (L7-NN), (*d*) the majority map from NN and the Landsat-8 data(L8-NN),(*e*) the majority map from SVM and the Landsat-7 data (L7-SVM), (*f*) the majority map from SVM and the Landsat-8 data(L8-SVM), (*g*) the majority map from RF and the Landsat-7 data(L7-RF), (*h*) the majority map from RF and the Landsat-8 data(L8-RF).

The majority classification uncertainty maps (Figure 2.7) exhibit the classification variability across the 10-repetition of the classification algorithms. The pixel values comprising the uncertainty maps range from 1 to 8 (pixels values in some uncertainty maps range from 1 to 6), where value 1(dark blue) means no variability (i.e., the pixels with value 1 were classified as only 1 class type throughout the 10 repetitions) and the highest value (red), for example 8, indicates the highest variability since these pixels were classified to 8 different spectral/lithological classes across the 10 repetitions. If a classification uncertainty map contains a larger amount of pixels with smaller values then it indicates that the classifier is more stable with areas of less classification uncertainty. The uncertainty maps, together with quantitative

analysis of the uncertainty values (Figure 2.8) suggest that the SVM classification applied to the Landsat-7 data performed most consistently compared to the other classifiers, as more of the pixels were classified with more certainty (39% of the pixels were assigned to 1 class and 42% to the next 2 different classes). The most uncertain classification is MLC applied to the Landsat-8 data, where only 12% pixels were assigned to 1 class type, 27.5% to 2 classes and 41% to 3 classes.

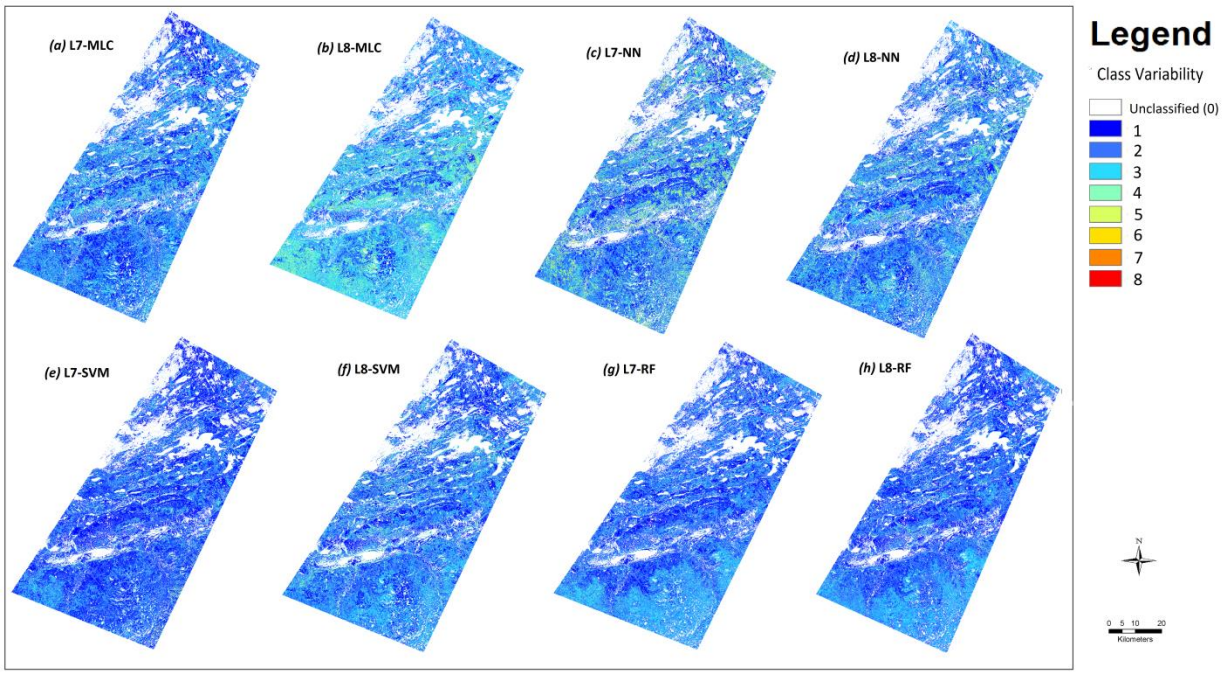

Figure 2.7: The majority classification uncertainty maps showing the spatial variability in classification results across the 10-repetition classification process. See Figure 2.6 for order of maps and abbreviations.

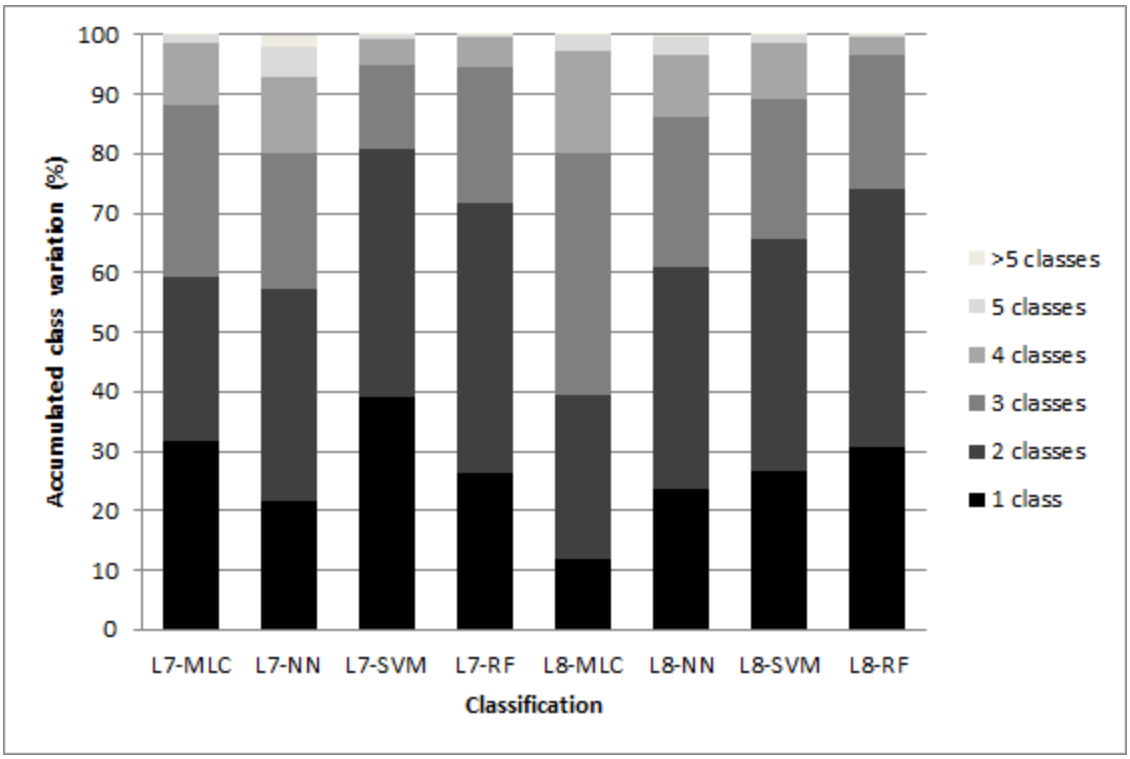

Figure 2.8: Class variation of pixels over the 10-repetition classification for each classifier using Landsat-7 and Landsat-8 image. The Y values show the accumulated percentages across all the class variation.

2.4.2. Evaluation of the 10-repetition classifications and the majority classification maps

For each classification over the 10 repetitions, a confusion matrix was computed by comparing the test (validation) dataset to the classification map class labels. Table 2.2 lists the individual overall classification accuracies for the 10 repetitions of the classification using the four classifiers applied to the Landsat-7 and Landsat-8 data. These overall accuracy values were then used to quantitatively compare the relative performance of the four classification algorithms and the two Landsat datasets for mapping lithology.

Table 2.2: Overall accuracy values (unit: %) across the 10 classification repetitions using MLC, NN, SVM, and RF applied to the Landsat 7 and Landsat 8 data. The bold numbers indicate the lowest overall accuracies, and the bold and italic numbers are the highest overall accuracies over the 10-repetition classification. ^

Repetition	L7-MLC	L8-MLC	L7-NN	L8-NN	L7-SVM	L8-SVM	L7-RF	L8-RF
1	**75.32**	72.32	**75.43**	79.82	**78.55**	**78.78**	87.08	81.89
2	82.74	81.14	76.85	81.24	84.35	80.92	81.35	***86.92***
3	***89.16***	**71.68**	***90.7***	86.81	***91.41***	***89.26***	***87.83***	84.87
4	79.78	82.78	83.89	83.22	85.44	81.33	81.22	81.67
5	78.13	86.14	77.45	***87.91***	83.97	82.61	80.3	79.62
6	80.85	76.17	85.85	87.66	90.21	84.89	85.21	85.64
7	82.12	**87.61**	83.22	87.11	79.32	82.82	**75.32**	85.51
8	82.18	76.85	85.7	**79.32**	86.48	81.14	80.23	79.71
9	82.31	73.53	86.87	84.32	84.87	81.87	75.86	**77.53**
10	83.45	84.6	87.82	83.1	83.91	85.52	80.8	79.31
Average	81.60	79.28	83.38	84.05	84.85	82.91	81.52	82.27
Average*	80.44		83.72		83.88		81.89	

^ (Average-the average accuracy values for each classifier applied on the Landsat-7 and Landsat-8 data respectively; Average*-the average accuracy values for each classifier applied on both the Landsat-7 and Landsat-8 data).

Overall there is not a great difference between the classifiers: a range of approximately 3% when using the Landsat-7 data and 5% when using the Landsat-8 data. With respect to the performance of each classifier applied on the Landsat-7 data, the highest mean overall accuracy value comes from SVM (84.85%), followed by NN (83.38%), MLC (81.60%), and RF (81.52%) (shown in the second last row of Table 2.2-Average). The Wilcoxon signed-rank test, an alternative non-parametric method to the paired-T test when the sample size is too small (less than 30) (Demšar 2006), was used to compare the performance difference on a pair-wise basis between each classification. The Wilcoxon signed-rank test at the 95% confidence level on the 10 repetitions of the classifications shows that the median of the overall accuracy value derived

from SVM is significantly higher than that derived from MLC, suggesting that SVM significantly outperformed the MLC when applied to the Landsat-7 data. The Wilcoxon test also indicates that the difference between the median of the overall classification accuracy values between RF applied to Landsat 7 and the other classifiers is not statistically significant, though the average overall accuracy value from RF (Landsat-7) is lower than all the other classifiers.

Regarding the performance of each classifier on the Landsat-8 image, NN produced the highest averaged overall accuracy value (84.05%), followed by SVM (82.91%) and RF (82.27%), and MLC produced the lowest accuracy (79.28%) (shown in the second last row of Table 2.2-Average). The Wilcoxon signed-rank test indicates that NN significantly outperformed compared to MLC if the Landat-8 images used, but there is no significant difference between NN, SVM, and RF.

The evaluation of the performance of each classifier, regardless which imagery was used, showed that, SVM produced the highest average overall classification accuracy(83.88%), followed by NN (83.72%), RF (81.89%) and MLC (80.44%) (shown in the last row of Table 2.2-Average*). The subtle difference between NN and SVM is obviously not statistically significant, indicating that NN and SVM offered similar performance. The average overall accuracy values from NN and SVM are higher than those from MLC, perhaps due to the advantages offered by NN and SVM mentioned in the *Introduction* section. However, it is surprising to observe that the average overall classification accuracy value from RF is only slightly higher than that from MLC. The difference between these two average overall accuracies is not statistically significant (by using paired-T test), nor is the difference between the medians of the overall accuracies from RF and MLC (by using the Wilcoxon signed-rank test). This observation suggests that the

performance of MLC is similar to RF even though MLC is not ideal for classification with non-normally distributed training datasets.

To reduce training data sampling variability, majority classification maps for each classifier using the Landsat-7 and Landsat-8 data were generated (shown in Figure 2.6) as previously mentioned. To evaluate how well the majority classification maps agree with the independent check (validation) data, an error matrix was constructed by comparing class labels on each majority classification map to the independent test dataset. Figure 2.9 shows the overall classification accuracy values for the majority maps derived by comparison to the test dataset (diamonds). For comparison purposes, the average overall map accuracy values for the 10-repetitions of the classifications derived from the four classifiers applied to both Landsat-7 and Landsat-8 data are also shown in Figure 2.9 (squares). Figure 2.9 shows that for the Landsat-7 data, SVM offered the best results and NN offered the best results for the Landsat-8 data. Moreover, a significance test (Z-statistic) based on Khat (Cohen's Kappa) values (Congalton and Green 2008) derived from the majority maps and the test data indicates that among the four classifiers using either the Landsat-7 or Landsat-8 data, MLC performed significantly differently from NN and SVM. In addition, when applied to the Landsat-8 data, NN, SVM, and RF did not exhibit significantly different classification accuracies, supporting test results derived from the 10-repetition overall accuracy values.

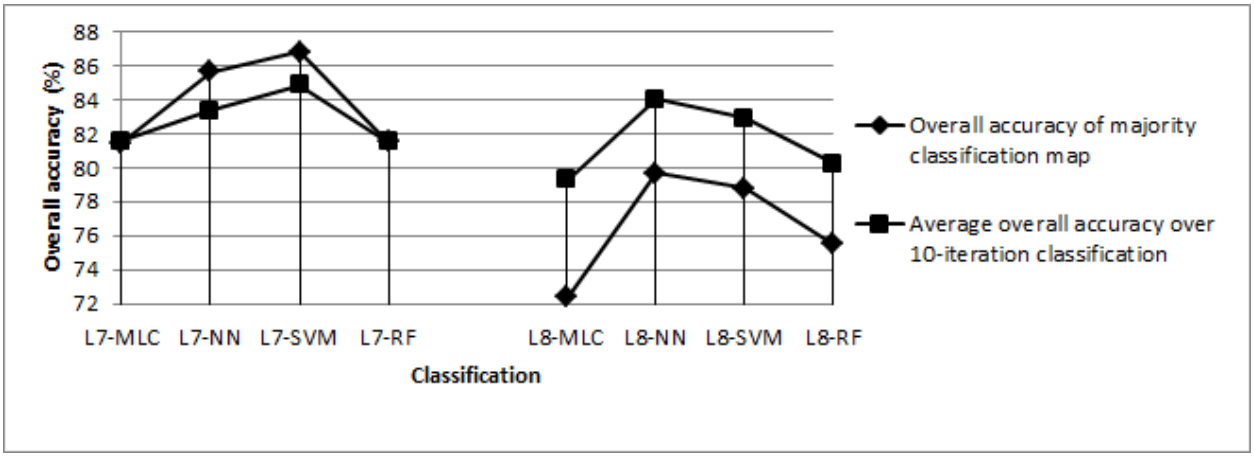

Figure 2.9: The overall classification accuracy values for the majority maps compared to the independent test dataset (diamonds) and the average overall accuracy values over the 10-repetition classification using each classifier on the Landsat-7 and Landsat-8 images respectively (squares).

In addition, all the overall accuracy values for the Landsat-8 generated majority classification maps derived from the test datasets are lower than the corresponding average overall accuracy values derived from the 10 repetitions of the classification algorithms, but the patterns (i.e., the order) of these two sets of overall accuracy values are the same, which is depicted in the right part of Figure 2.9. Furthermore, all the classifiers applied to the Landsat-7 data outperformed the classifiers when applied to the Landsat-8 data. The Landsat-8 derived accuracies (diamonds on the right side of Figure 2.9) are significantly lower than the Landsat-7 derived accuracies (diamonds on the left portion of Figure 2.9) when the overall accuracies of majority classification maps derived from the independent test dataset are considered. The reasons for the lower overall accuracies when using Landsa-8 data may be because: 1) the training and the independent test data were identified based on the Landsat-7 image which may upward bias the classification results using the Landsat-7 data; 2) the Landsat-7 and Landsat-8 data were acquired in different years (2002 vs. 2013) and during the 11-year period some environmental factors like moisture and vegetation cover may have changed, so that the training

and test data may not be entirely representative for the lithological classes present in the Landsat-8 image; 3) the narrower bandwidth of the Landsat-8 data may lead to the loss of spectral response captured by the Landat-8 data; and 4) the Landsat-8 data have a different radiometric resolution (16-bit).

Quantitative analyses on the overall classification accuracies derived from the 10-repetition classification and from the majority classification maps in concert with visual comparison of the majority maps, suggested that: 1) using either Landsat-7 or Landsat-8 data, SVM performed very similarly to NN, and both of them performed better than MLC and RF given that the training data were not normally distributed; 2) RF performed better than MLC when applied to the Landsat-8 data, and RF performed similarly to MLC if the Landsat-7 data were used; and 3) the Landsat-7 data generated better bedrock lithological maps than the Landsat-8 data if the overall accuracies from the majority maps were used as a performance measurement tool, but both Landsat images performed comparably if the average overall accuracy values across the 10-repetition classification were used as an evaluation tool. This suggests that Landsat-7 data are still valuable for geological mapping and Landsat-8 data offer similar (in this case slightly lower) accuracies for bedrock mapping.

2.4.3 Classification accuracy for individual classes

Table 2.3 and Table 2.4 list the average producer's accuracy (Ave-PA) and user's accuracy (Ave-UA) for individual classes over the 10 repetitions of the four classification algorithms. Landsat-7 and Landsat-8 performed differently in terms of discriminating the 13 spectral classes. Landsat-7 discriminated dry vegetation, basalts (nfb1,nfb2,nfbv), and till1 better than Landsat-8 across all the classifiers; but Landsat-8 performed better in distinguishing carbonates, nfgv, and

sandstones compared to Landsat-7 for all the classifiers. Both Landsat-7 and Landsat-8 accurately classified evaporate rocks, Paleozoic dolostones, nfg, till2, and till3.

Table 2.3: Average Producer's Accuracy values (Ave-PA) from the 10-repetition classification ((unit: %).

	Class Name	L7-MLC	L7-NN	L7-SVM	L7-RF	L8-MLC	L8-NN	L8-SVM	L8-RF
1	Dry vegetation	99.2	99.0	98.0	97.9	76.8	81.0	83.5	88.1
2	Evaporites	90.9	77.1	82.7	84.7	88.4	82.6	78.3	83.7
3	Carbonates	73.6	72.1	71.1	77.9	95.9	81.7	77.8	78.7
4	Paleozoic Dolostones	88.7	90.0	90.2	78.1	91.0	87.6	91.6	89.5
5	Nfg	81.7	85.0	85.5	88.7	82.7	87.1	84.8	86.1
6	Till3	93.8	95.9	99.1	90.4	92.2	96.7	99.8	91.7
7	Till2	63.0	75.4	68.9	77.6	52.9	74.5	62.4	86.5
8	Nfgv	78.1	86.4	87.3	79.5	92.6	98.1	96.6	94.5
9	Sandstones	78.6	88.0	80.4	65.6	94.3	84.9	83.3	76.8
10	Nfbv	73.3	69.0	78.3	71.6	53.0	67.0	62.9	70.3
11	Nfb1	84.7	91.3	92.0	88.8	73.7	68.6	80.0	78.3
12	Till1	83.0	77.2	82.9	52.3	69.1	56.0	56.6	47.8
13	Nfb2	76.4	78.5	89.4	68.5	23.3	81.5	86.8	47.1

Table 2.4: Average User's Accuracy values (Ave-UA) from the 10-repetition classification (unit: %).

	Class Name	L7-MLC	L7-NN	L7-SVM	L7-RF	L8-MLC	L8-NN	L8-SVM	L8-RF
1	Dry vegetation	96.2	94.4	93.0	96.0	75.8	91.0	90.8	88.7
2	Evaporites	80.3	84.5	71.6	82.3	54.6	87.2	88.3	90.2
3	Carbonates	76.3	88.0	92.8	79.3	81.5	94.1	92.2	87.3
4	Paleozoic Dolostones	81.9	68.7	78.3	60.8	91.9	63.0	64.8	61.8
5	Nfg	88.3	90.8	90.0	85.3	93.6	87.9	89.8	86.1
6	Till3	90.1	88.5	94.2	89.0	90.2	84.0	79.5	84.7
7	Till2	94.7	88.6	94.0	89.0	95.1	92.2	93.1	95.0
8	Nfgv	95.4	90.4	86.0	86.1	98.2	91.3	91.7	95.5
9	Sandstones	62.6	70.8	62.0	65.7	94.4	78.3	77.9	77.9
10	Nfbv	64.9	59.8	69.2	54.3	37.7	70.6	68.6	61.7
11	Nfb1	56.6	70.2	71.2	67.7	57.1	56.6	61.8	52.7
12	Till1	99.7	94.5	95.9	80.7	100.0	51.8	61.2	35.5

13	Nfb2	92.9	90.0	97.8	94.2	100.0	90.9	92.9	90.5

MLC applied to the Landsat-8 data performed differently compared to other classifiers from the perspective of the average PA and UA. Both the Ave-PA and Ave-UA from MLC and the Landsat-8 for sandstones are much higher than NN, SVM, and RF. However, the Ave-PA and Ave-UA for dry vegetation and nfbv are lower than NN, SVM, and RF. Moreover, the Ave-PA of nfb2 is much lower than other classifiers because most of the test (validating) dataset was assigned to nfbv, evaporite rocks, and carbonates in the classification process. In addition, the Ave-UA for evaporite rocks is lower than the others. The confusion matrixes reveal that in the Landsat-8 classification using MLC, dry vegetation was confused with evaporite rocks and nfbv was confused with nfb1, nfgv, carbonates, and evaporite rocks, even though all of these classes have very good spectral separability as measured by using TD. However, the confusion of nfbv with nfgv and evaporite rocks is lower on the MLC classification of the Landsat-7 data, indicating that Landsa-7 MLC classification more accurately identifies Nfbv than the Landsat-8 MLC classification. The lower Ave-PA and/or Ave-UA values when using the Landsat-8 imagery may, as previously mentioned, be caused by the different acquisition time for the Landsat-7 and Landsat-8 data (2002 vs. 2013, respectively). The Ave-PA and Ave-UA for Paleozoic dolostones, sandstones, and till1 are lower when RF was applied to the Landsat-7 data, because sandstones in the test dataset were classified as Franklin Diabase (nfg), and the classified sandstones were actually carbonates in the test dataset. Till1 was confused with carbonates and Paleozoic dolostones, though they both have very high TD separability of 1.99. For example, in Repetition 2, all sandstones pixels in the validating dataset were classified into nfg, while the labeled sandstones actually are carbonates. The TD value for the pair of sandstones and nfg in the validating dataset of Repetition 2 is 1.83, much lower than the other

pairs, which may cause confusion of sandstones and nfg in the reference dataset. In addition, the very low Ave-PA and Ave-UA values for the till1 class in the classification using Landsat-8 and RF were caused by the confusion of till1 with dry vegetation, carbonates and paleozoic dolostones, though they also have good separability of 1.99. The confusion indicates obvious variability in the training dataset.

2.4.4. Cross-comparison agreement

Cross-comparison analysis was employed to quantitatively compare the classification maps on a class-by-class basis. Higher agreement between two classification maps indicates similar performance of the two classifiers. The agreement analysis was conducted using two different approaches. The first approach was to compare agreement on an individual class basis; the second approach was based on the generalized lithological classes and two surficial material classes (these classes are called information classes in the remote sensing community). In the generalized comparison process, the lithogical/spectral classes in the majority maps were firstly aggregated to information classes (see Figure 2.1 to see how the 13 classes were generalized). That is: Class Nfb1, Nfb2 and Nfbv were merged together to form one basalt unit so that a pixel classified as Nfb1 in one classification and Nfb2 or Nfbv in another would be considered as a correct match. Similarly, Nfg and Nfgv were combined into one Franklin diabase sill class, and Till 1, Till 2, and Till 3 were merged into one till unit. Figure 2.10 shows the agreement values (in percentages) between pairs of the majority maps on a class-by-class basis, including the spectral-class-basis (SCB- original 13 classes) agreement for the majority maps and the information-class-basis (ICB – generalized classes) agreement for the aggregated majority maps. Figure 2.10 shows the agreement level based on information classes is higher than that based on lithological/spectral classes due to generalization of the training areas. Figure 2.10 also shows

that generally the agreements between the classification maps produced by the advanced classification algorithms are higher than that of the traditional MLC algorithm for both Landsat-7 and Landsa-8 data. The higher agreement(around 80%) between NN, SVM, and RF suggests that these three classifiers perform more similarly and are less sensitive to the noise (variability) in the training data (e.g., the training data is not normally distributed and may contain mixture spectral information). However, the agreement values between MLC with each advanced classifier are lower compared to the agreement among the advanced algorithms, especially if the Landsat-8 data were used. This suggests that MLC performs differently from NN, SVM, and RF, since the former depends on the normal distribution of the training data and more importantly, the quality of the training data. Mixed spectral pixels have a negative impact on the classification results when MLC is applied.

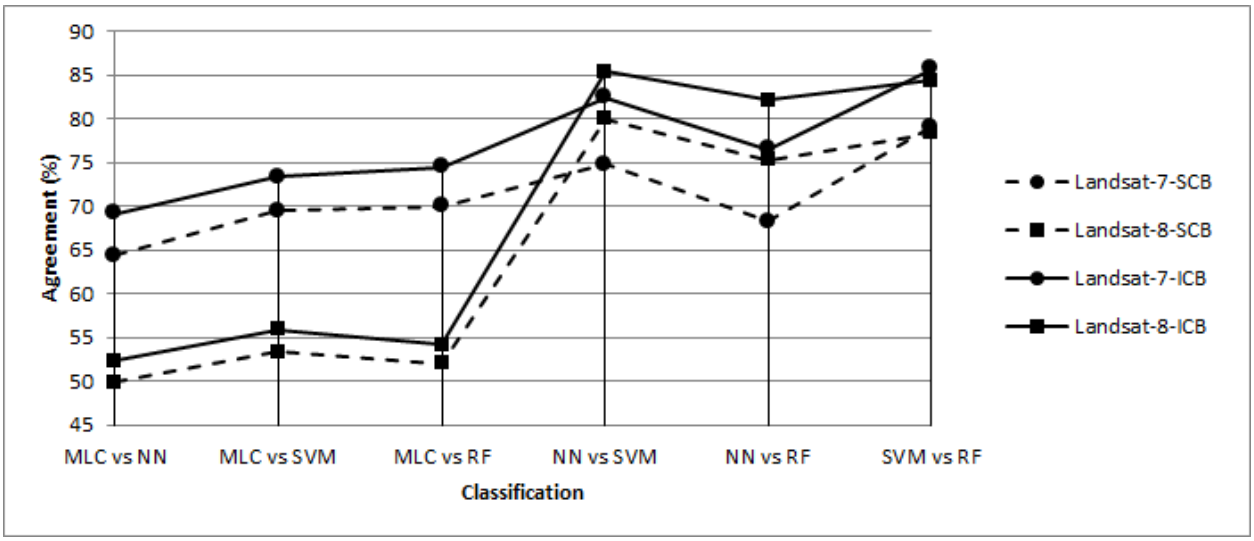

Figure 2.10: The cross-comparison agreement based on SCB-spectral classes (dashed lines) and ICB-information classes (solid lines) from the majority maps generated by applying each classification approach on the Landsat-7 (circles) and Landsat-8 (squares) images respectively.

The cross-comparison agreement (similarity) between the Landsat-7 and Landsat-8 classifications using the same classifier is presented in Figure 2.11. This figure indicates that the greatest similarity between the majority maps generated by these two image datasets using the same classifier is from using SVM, followed by NN, RF, and then MLC from both the perspective of lithogical/spectral classes and information classes. The lowest agreement is between the majority maps generated using MLC, although the agreements between the two datasets are very similar. The agreements between the maps produced by the advanced classifiers, although higher than MLC, are less similar when the two datasets are compared (Figure 2.11- there is a greater range in agreement between the Landsat- 7 and Landsat-8 data when using the advanced classification algorithms). Again this difference may be caused by either/or: 1) The training datasets were selected based on the Landsat-7 image, so that when these training data were applied to the Landsat-8 imagery, the statistics of these training datasets may have contained more noise; 2) The 16-bit radiometric resolution of the Landsat-8 image may have had an impact on the statistics of the training dataset. In addition, the higher agreement from SVM, NN and RF suggests that these three machine learning algorithms work more similarly on both Landsat images even when the training data are skewed.

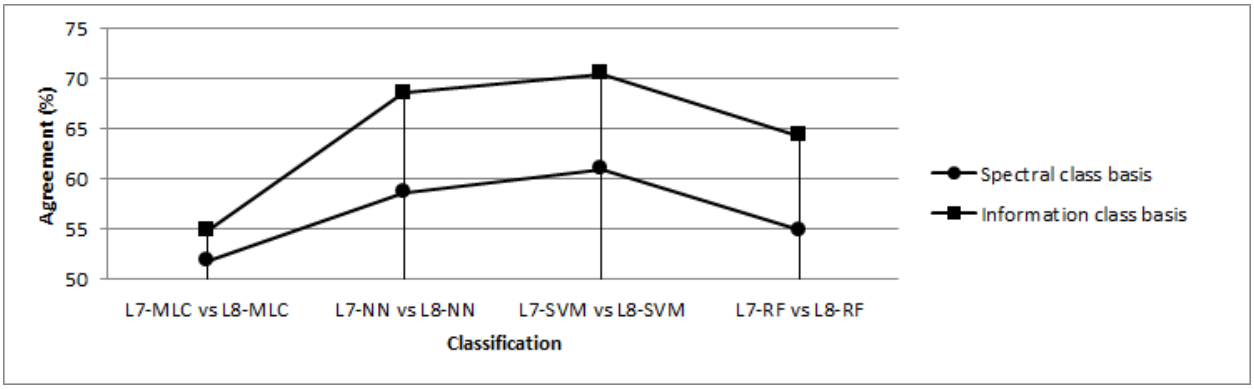

Figure 2.11: The cross-comparison agreement levels for the majority maps generated by using the same classification approach on Landsat-7 and Landsat-8 imagery respectively.

2.4.5. Important bands

One of the advantages offered by RF is that it can be used to rank the relative importance of each variable (i.e., the input bands) in the classification process. Table 2.5 lists the rank order of band importance based on the average normalized importance score (Breiman 2001) over the 10-repetitions of the RF classification applied to the Landsat-7 and Landsat-8 data, respectively. The variable importance rank shows that for both the Landsat-7 and Landsat-8 data, NIR (Near Infrared) and SWIR bands play a more important role in the classification process. The NIR band (Band 4 in the Landsat-7 data and Band 5 in the Landsat-8 data) contributes the most in the classification process for both the Landsat-7 and Landsat-8 images, followed by the SWIR bands (Band 7 and 5 in the Landsat-7 data, Band 7 in the Landsat-8 data), and then the visible bands. One exception is that for the Landsat-8 image, Band 6 (SWIR band, spectrum from 1.57 μm (micron) to 1.65 μm) is less important than Band 4 (Red band, spectrum from 0.64 μm to 0.67μm). A Principal Components Analysis (PCA) that was applied to the Landsat-8 image reveals the first and second components account for 65.87% and 26.82% of the variance in the entire Landsat-8 dataset respectively, and these two components have higher loadings in Band 4 than that in Band 6 of the Landsat-8 data. In addition, spectral separability based on Bands 4, 5, and 7 of the Landsat-8 data has no or subtle difference from that based on Bands 5, 6, and 7 for most of pairs of the classes. However, the spectral separability between dry vegetation and evaporite rocks on the Landsa-8 data decreased from good separation (>1.97) when Bands 4, 5 and 7 were evaluated to moderate-poor separability (< 1.5) when Bands 5, 6, 7 were used (Figure 2.12). The spectral separation between till1 and dry vegetation decreased from 1.99 when Bands 4, 5 and 7 were used to 1.81 when Bands 5, 6, 7 were used. Conversely, this significant decrease in spectral separation (indicated by TD values) described above is not present in Bands 3,4,7 and

Bands 4,5,7 of the Landsat-7 data. This may be due to Landsat-8 Band 6 narrower bandwidth (wavelength 1.57-1.65 µm) than Band 5 of Landsat-7 (wavelength 1.55-1.75µm).

Table 2.5: The importance of spectral bands of the Landsat-7 and Landsat-8 images computed by using RF based on the average normalized scores across the 10-repetition classification.

Importance order	Landsat-7		Landsat-8	
	Band	Average normalized score	Band	Average normalized score
1	4	4.145	5	3.455
2	7	3.155	7	3.101
3	5	2.224	4	2.171
4	1	1.804	6	1.759
5	3	1.573	3	1.655
6	2	1.418	1	1.5
7			2	1.256

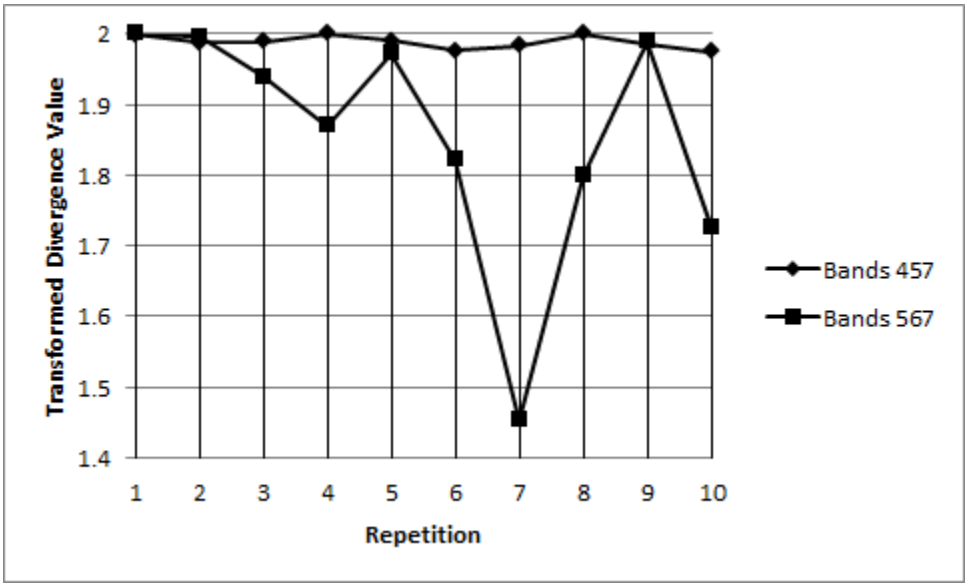

Figure 2.12: the Transformed Divergence values (TD) between dry vegetation and evaporite rocks derived from the classification using Bands 4, 5, 7 and Bands 5, 6, 7 of the Landsat-8 data.

2.4.6 Effect of radiometric resolution of Landsat-8 image on classification

To evaluate the effect of the 16-bit radiometric resolution of the Landsat-8 data on classification results, the original 16-bit Landsat-8 imagery used in the study was scaled to 8 bit by using a

linear transformation using PCI Geomatica image analysis software (http://www.pcigeomatics.com). The four classifiers were applied to the 8-bit Landsat-8 data. The volume of the 8-bit image decreased significantly to only ¼ of the volume of the original 16-bit image. The overall accuracy values derived from the 10 repetitions of the three advanced classifiers applied to the 8-bit Landsat-8 data only varied around 1% compared to the classification on the 16-bit Landsat-8 data. However, the overall accuracy values derived by applying MLC to the 8-bit Landsat-8 data were significantly increased for each repetition and the average overall accuracy value was increased from 79.28% (using the 16-bit Landsat-8 data) to 84.63% (using the 8-bit Landsat-8 data) (Figure 2.13). The majority classification map created by applying MLC to the 8-bit Landsat-8 data (Figure 2.14) presents similar patterns of the regional distribution of bedrock lithologies compared to the majority maps generated by using other classifications on either the Landsat-7 or Landsat-8 16-bit images (Figure 2.6). In addition, the overall accuracy value derived from the 8-bit Landsat-8 majority map against the independent test data was increased from 72.35% (when the original 16-bit Landsat-8 data was used) to 82.21%. Further investigation on why compressing the Landsat-8 data from 16 bit to 8 bit has an substantial impact on MLC classification results but little impact on NN, SVM, and RF revealed that the compressing process dramatically decreased the statistics (i.e., mean, variance, and covariance) of the training data that are important parameters in the MLC classification process (Richard and Jia 2006). Thus, re-scaling the Landsat-8 data from 16 bit to 8 bit may cause the training areas to approach a more normal distribution thus leading to the better performance of MLC.

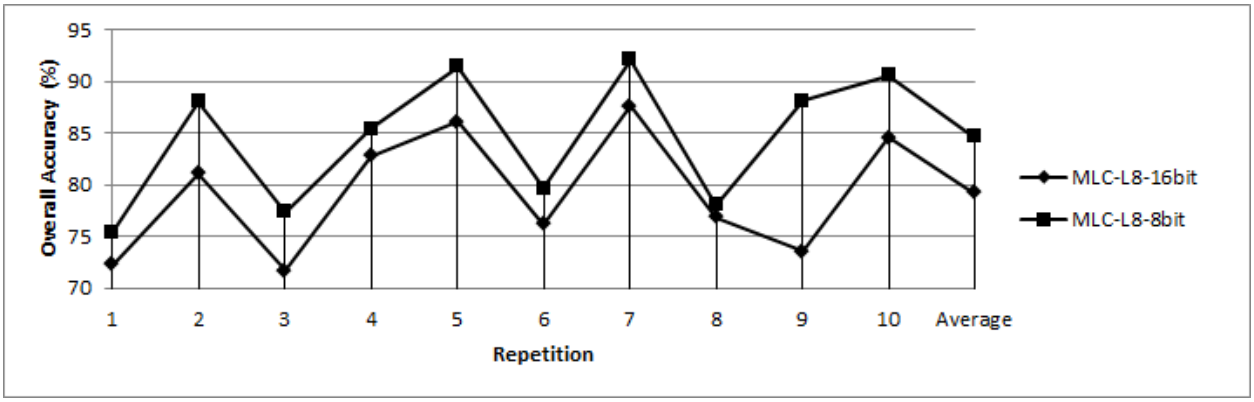

Figure 2.13: The increase of the overall accuracy values over 10 repetitions of MLC applied to the 8-bit Landsat-8 image (squares), compared to the results derived from applying MLC to the original 16-bit Landsat-8 image (diamonds).

2.5 Discussion

This study has indicated that advanced classification techniques can play a useful role in the generation of bedrock and surficial material maps of Canada's North given the appropriate geologic environment. The training data used in the study present relatively complex frequency distributions due to spatial variability in such factors as vegetation cover, moisture content, and complicating overlying weathered residual material. Overall, NN(Neural Network), SVM(Support-Vector Machine), and RF(Random Forest) performed better than MLC(Maximum Likelihood classifier) especially for the Landsat-8 data when training data were not normally distributed, demonstrating the advantages provided by these advanced algorithms mentioned in the *Introduction* section.

NN and SVM provided similar results in mapping lithologies and surficial materials from the perspective of visual comparison of the majority classification maps, a statistical comparison of overall accuracy values and cross-comparison agreement analyses. RF performed similarly to NN and SVM from the aspect of visual evaluation and cross-comparison agreement analyses on the majority classification maps. However, it was also observed in this study that RF provided

similar results to MLC when using the Landsat-7 data. This observation suggests that MLC is not an optimal classifier when training data are skewed but it can still generate acceptable (but not best) classification results despite the spectral variability and non-normal distributions of the training data.

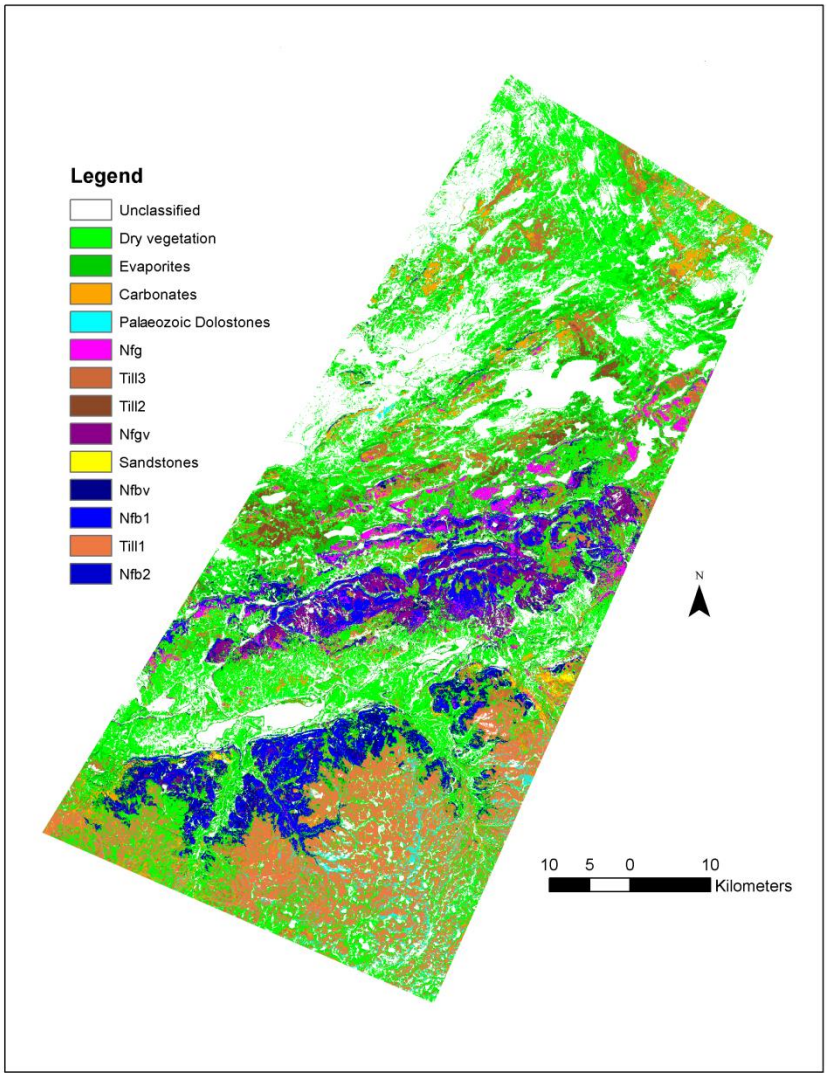

Figure 2.14: The majority classification map generated by using MLC applied on the 8-bit Landsat-8 data.

The study also observed that classification results derived from RF contained variation. For example, a RF classification was conducted using the Landsat-7 data 11 times using identical

parameters that include the same training and test dataset used in Repetition 1 (Table 2.2), 200 decision trees and 2 sampled input bands(these two parameters are the optimal setting in the case that was discussed in the section of 2.3). The overall accuracy values are shown in Figure 2.15. The test revealed that the classification results from RF varied to a certain degree (the highest is 87.08% and the lowest is 84.54%) though the *out-of-bag* accuracy is stable. In that case, the highest overall accuracy classification was obtained on the 10th repetition. However, it is not surprising to observe such variation because in RF, at each node, only a randomly sampled portion of the training data are used to train the classifier and the remainder are used to validate the result. In addition, at each node, only a randomly sampled subset of the input variables (i.e., bands) is used to build up the trees in the forest.

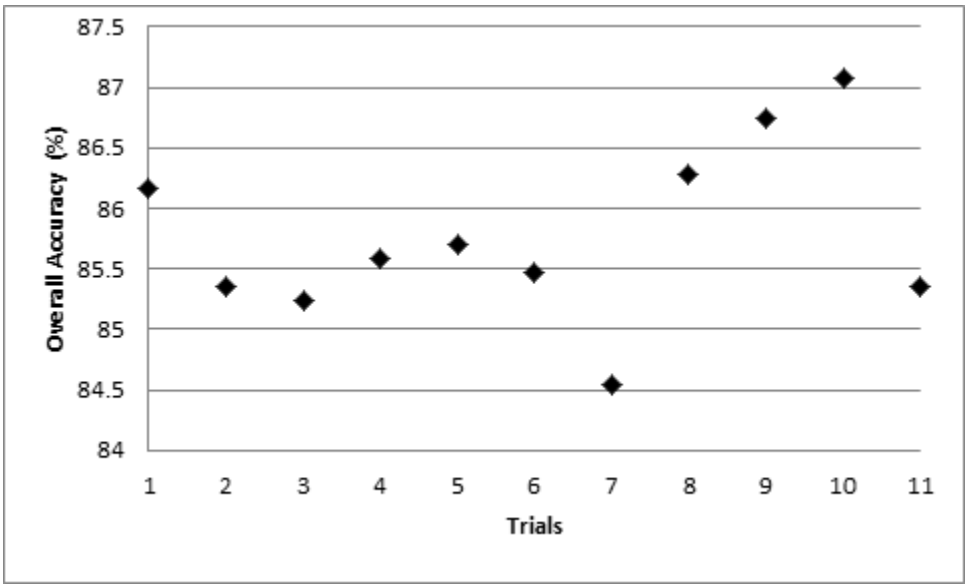

Figure 2.15: Variation in classification overall accuracy values (diamonds) with identical parameters in RF.

Landsat-8 offers some advantages such as an additional band in the blue portion of the electromagnetic spectrum, narrower bandwidth, and 16 bit radiometric resolution providing a greater dynamic range when compared to Landsat-7. The statistics of training data will be

changed when re-scaling (compressing) the radiometric resolution from 16 bit to 8 bit. This can compress the statistics of the training data into a semi-normal distribution affecting the performance of MLC in a positive fashion. The improved performance of the 8-bit Landsat-8 data over the 16-bit data was evidenced by higher mean overall accuracy values across the 10 repetitions of the MLC classification and a higher overall accuracy value for the majority map against the independent test data. Bands 4,5,6,7 of the Landsat-8 data are narrower than the corresponding Landsat-7 bands, and the narrower bands can offer more or less discrimination capability, depending on targets to be classified. The narrower bandwidth in the Landsat-8 data may capture less spectral reflectance, causing confusion among features to be classified. This may be a further reason for the Landsat-8 data slightly lower performance when compared to the Landsat-7 data.

This study has also demonstrated the importance of NIR and SWIR bands of both the Landsat-7 and Landsat-8 data for classifying lithology. Moreover, it revealed that the red band (Band 4) in Landsat-8 played a relatively more important role compared to Band 6 of the Landsat-8 data when they were used to discriminate dry vegetation, evaporite rocks, and till1, which is opposite to the band importance order for the Landsat-7 data (Table 2.5). Overall, the good performance of the Landsat-8 data (only slightly less than Landsat-7) indicates that Landsat-8 will provide acceptable data and continuous global coverage of the Earth's land surface for the next 10 years. This will be of great benefit for geological studies.

The selection of training data is critical for success of image classification techniques. The identification of representative training areas is especially difficult because the training areas for lithological classes present relatively complex frequency distributions. Even within a lithological class, spectral variability may be present due to variability of major element

mineralogy, weathering, moisture content levels, and dry vegetation cover. This variability in training data and the non-normal distributions have an effect on MLC and less so on NN, SVM, and RF classifiers. We thus recommend the use of advanced classifiers, presented in this study, for mapping bedrock and surficial materials in Canada's North.

2.6 Conclusions

This study has demonstrated that in mapping bedrock and surficial materials where training data present complex spectral signatures for which not a single frequency distribution model can be applied, advanced classification algorithms can play an important role in the geological mapping process. The advanced classification algorithms are recommended for producing geological maps of Arctic areas. The majority classification maps, together with classification uncertainty maps, can provide first order geological information in poorly mapped areas of Canada's North. In addition, these maps can provide preliminary information to support field based mapping studies.

Landsat-8 data, especially when used with advanced classification algorithms, can produce comparable classification results to the Landsat-7 images, suggesting that the Landsat-8 satellite can provide reliable and continuous image sources for future geological mapping.

The critical role of representative training data in remotely sensed image classification is paramount in classification exercises. The variation in classification results and overall accuracy values across the 10 repetitions of the four classification algorithms, compared in this study, demonstrate the effect that training area variability has on the classification process. This is why majority and uncertainty maps, presented in this study, are attractive as they average out both the statistical and spatial uncertainty in the classification process, focusing on pixels that display a more certain classification. Finally, the cross-validation classification method applied in the

study can assist in identifying poorly quality training areas that could be modified or even removed from the classification process.

2.7 Acknowledgements

This work was supported by the Geological Survey of Canada (GSC) under the Remote Predictive Mapping Project (RPM), part of the Geo-mapping for minerals and energy (GEMS) program.

Chapter 3 A Comparison of Different Remotely Sensed Data for Classifying Bedrock Types in Canada's Arctic: Application of the Robust Classification Method and Random Forests

Abstract: The Geological Survey of Canada, under the Remote Predictive Mapping project of the Geo-mapping for Energy and Minerals program, Natural Resources Canada, has the mandate to produce up-to-date geoscience maps of Canada's territory north of latitude 60°. Over the past three decades, the increased availability of space-borne sensors imaging the Earth's surface using increasingly higher spatial and spectral resolutions has allowed geologic remote sensing to evolve from being primarily a qualitative discipline to a quantitative discipline based on the computer analysis of digital images.

Classification of remotely sensed data is a well-known and common image processing application that has been used since the early 1970s, concomitant with the launch of the first Landsat (ERTS) earth observational satellite. In this study, supervised classification is employed using a new algorithm known as the Robust Classification Method (RCM), as well as a Random Forest (RF) classifier, to a variety of remotely sensed data including Landsat-7, Landsat-8, Spot-5, Aster and airborne magnetic imagery, producing predictions (classifications) of bedrock lithology and Quaternary cover in central Victoria Island, Northwest Territories. The different data types are compared and contrasted to evaluate how well they classify various litho-types and surficial materials; these evaluations are validated by confusion analysis (confusion matrices) as well as by comparing the generalized classifications with the newly produced geology map of the study area. In addition, three new Multiple Classification System (MCS) methods are proposed that leverage the best characteristics of all remotely sensed data used for classification.

Both RCM (using the maximum likelihood classification algorithm, or MLC) and RF provide good classification results; however, RF provides the highest classification accuracy because it uses all 43 of the raw and derived bands from all remotely sensed data. The MCS classifications, based on the generalized training dataset, show the best agreement with the new geology map for the study area.

Keywords: image classification, Random Forest (RF), Maximum Likelihood classification (MLC), Robust Classification Method (RCM), Multiple Classification System (MCS), geological map

3.1 Introduction

Over the past three decades, the increased availability of space-borne sensors imaging the Earth's surface using increasingly higher spatial and spectral resolutions has allowed geologic remote sensing to evolve from being primarily a qualitative discipline to a quantitative discipline based on the computer analysis of digital images. Many image processing and analytical techniques have been developed to aid the interpretation and extraction of useful geologic information from remotely sensed imagery. Image classification is perhaps the best known and the most widely used digital image analysis technique since the advent of the ERTS-1 (Landsat) series of satellites in the early 1970s.

The Geological Survey of Canada, under the Remote Predictive Mapping (RPM) project (Desnoyers and Harris 2003; Schetselaar et al. 2007; Harris 2008; Harris et al. 2008a), part of the Geo-mapping for Energy and Minerals (GEM) program of Natural Resources Canada, has the mandate to produce up-to-date geoscience maps of Canada's territory north of latitude 60°. The expansive territory to be mapped, in combination with the required resources and cost of setting up field mapping campaigns, requires new approaches to providing geological information. Remote Predictive Mapping (RPM), which emphasizes the use of various types of satellite and geophysical imagery to assist in producing geological maps, in part addresses this requirement (Harris 2008, and papers therein). Producing the required geological information can be undertaken by visual interpretation of enhanced imagery, by employing machine learning

algorithms to assist in extracting information, or by a combination of the two methods (Schetselaar et al. 2007; Harris et al. 2008a, 2012a). Predictive mapping is obviously not a replacement for field work; however, the intelligent use of remotely sensed, geophysical and geochemical data in concert with legacy geological data (maps, databases) can produce maps that focus field work on areas that have more complex signatures (Harris 2008). Furthermore, RPM also provides first-order geologic predictive maps in poorly mapped areas or areas where field work is not possible.

The use of machine learning algorithms, primarily classification, for mapping rocks has been applied mostly in arid environments (Rowan et al. 1987; Macias 1995; Glikson and Creasey 1995; El Rakaiby 1995; Van de Meer et al. 1995; Rowan et al. 2005; Kavak 2005; Peña and Abdelsalam 2006; Saadi and Watanabe 2009; Rajendran et al. 2012). A number of examples can also be found for Arctic environments (Leverington 2001, 2010; Lorenz 2004; Harris et al. 2005, 2008a, c, 2009, 2010; Wickert et al. 2008; Schetselaar and Ryan 2008; Leverington and Moon 2012; Behnia et al. 2012). Even though Arctic environments are not covered by trees, ubiquitous lichen and snow cover and a short season for acquisition of suitable imagery can make the application of machine learning algorithms for geological mapping a challenge. However, the studies listed above for northern environments do indicate that the classification of various remotely sensed data, including geophysical data, can make a useful contribution to mapping activities. Copious public domain remotely sensed data (www.geobase.ca) are freely available and can be applied to operational mapping programs such as RPM using both visual interpretation and machine learning methods. The application of machine learning algorithms offers the advantages of objectivity, speed and the ability to cover large areas efficiently and within short time frames. However, expert knowledge is paramount to the success of machine

learning algorithms, as training data are required for the geological features to be successfully classified.

Classification algorithms can be broadly divided into two categories – unsupervised and supervised. Unsupervised algorithms require very little user input and are designed to find like classes (or clusters) within an *n*-dimensional data (feature) space, whereas supervised classification requires much more user input to train and guide the classification process. Supervised classification is based on *a priori* knowledge of the features to be classified. The spectral signature of representative samples of the different surface cover types (information classes), referred to as *training areas*, are used to train the classifier and, by using a classification algorithm, to recognize spectrally similar areas for each class on an image dataset.

In this study, supervised classification using a new algorithm known as the Robust Classification Method (Harris et al. 2012b) is applied to a variety of remotely sensed data, including Landat-7, Landsat-8, Spot-5, Aster and airborne magnetic imagery, producing predictions (classifications) of bedrock lithology and Quaternary cover in central Victoria Island, Northwest Territories. A Random Forest classifier is also employed to (1) rank the predictive power of each input band, and (2) create a classified map. Both of these classifiers are 'ensemble classifiers', as they produce multiple predictions (classified maps) that are combined using a voting procedure whereby the predominant class, on a pixel basis, is selected for the final map. Using confusion analysis (confusion matrices) and comparing the generalized classifications with the newly produced geology map of the study area (Rainbird et al. 2013a, b, c and d), these different data types are compared and contrasted to evaluate how well they classify various litho-types and surficial materials. We also introduce three new multiple classification system (MCS) maps that compare favourably to the geology map of the study area. ENVI™ image processing

software in concert with ArcGIS™ and the EnMap Box software package (http://www.enmap.org/?q=enmapbox) were used to perform the classifications.

3.2 Study Area

The study area lies within the Minto Inlier on Victoria Island (Figure 3.1). The Minto Inlier, first recognized and mapped by Thorsteinsson and Tozer (1962), is a northeast-trending belt of early Neoproterozoic sedimentary rocks of the Shaler Supergroup that are intruded by gabbro–diorite sills and dykes and capped by coeval flood basalt of the Franklin intrusive (magmatic) event. The succession was gently folded, forming an open syncline and a smaller anticline, before deposition of an unconformably overlying shallow marine sandstone and carbonate succession of Cambro–Ordovician age. The Shaler Supergroup includes, in ascending stratigraphic order, the Rae Group, Reynolds Point Group, Minto Inlet Formation, Wynniatt Formation, Kilian Formation and Kuujjua Formation (Rainbird et al. 1994, 1996; Figure 3.1). With the aid of recent detailed field observations, the Wynniatt Formation was divided into four lithostratigraphic sub-units, namely black shale, lower carbonate, stromatolitic carbonate, and upper carbonate members (Thomson et al. 2014). Basaltic rocks of the Natkusiak Formation are exposed in the south, and related gabbro–diorite sills and dykes (Franklin Intrusions) were intruded into the sedimentary strata throughout the area. Recently, new geological maps of the study area have been produced using detailed field observations and visual interpretation of Spot-5 stereo imagery (Rainbird et al. 2013a, b, c and d). A portion of these maps is presented in Figure 3.2a, which shows the detailed geology, whereas Figure 3.2b is a generalized lithological map that will be used for comparisons with the generalized lithological and surficial classification maps produced in this study. Table 3.1 includes descriptions of the geology shown in Figure 3.2a (detailed lithology and surficial materials column), as well as the generalized lithology in Figure

3.2b (generalized lithology and surficial materials column).

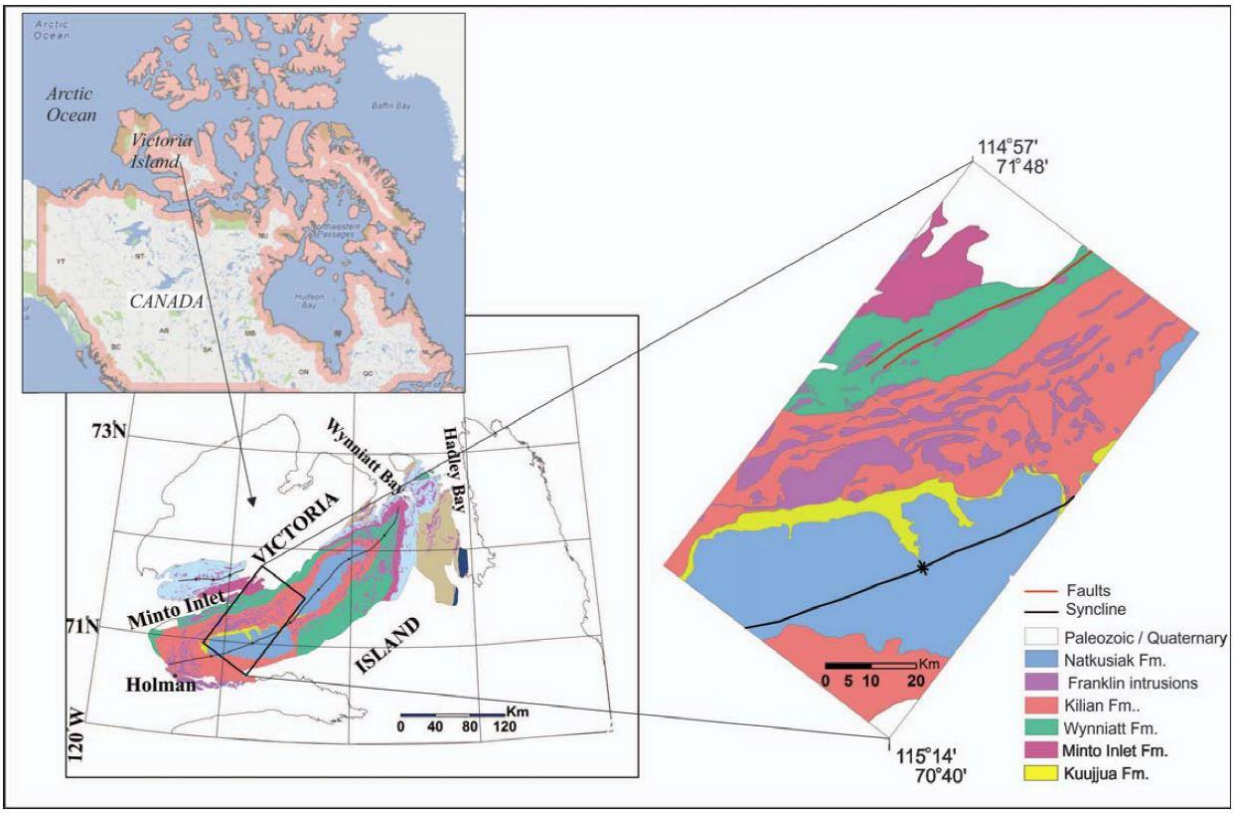

Figure 3.1: Location of study area in the Minto Inlier of Victoria Island, Northwest Territories, Canada, and legacy geology map showing main lithological units (geology from Hubert et al. (2005), modified from Thorsteinsson and Tozer(1962)).

3.3 Data and Processing

The types and descriptions of remotely sensed data used for classification purposes are presented in Table 3.2. Landsat-7 orthorectified images (both images acquired on 14 July, 2002) were downloaded from the United States Geological Survey (USGS) Global Visualization Viewer (http://glovis.usgs.gov), and two-level 1A Spot-5 scenes acquired on 22 August, 2009 were purchased from Blackbridge (http://www.blackbridge.com/). Both the Landsat-7 and Spot-5 data in 8-bit format were radiometrically balanced, mosaicked and then subset to the extent of the study area (Figure 3.1). The Landsat-8 data acquired on 4 July, 2013 were downloaded from

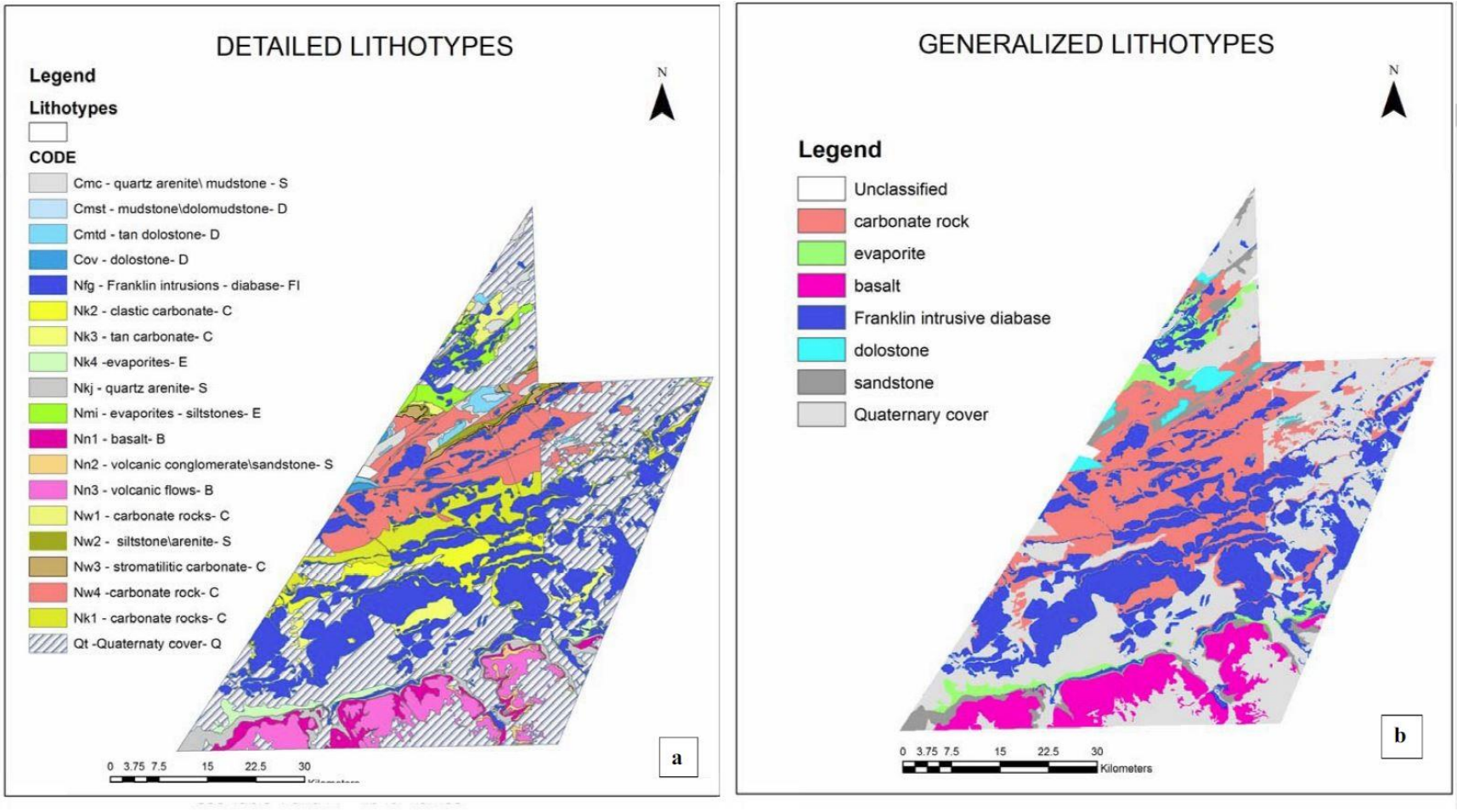

Figure 3.2: New geology map of the study area in the Minto Inlier of Victoria Island, Northwest Territories, Canada. (a) Detailed bedrock geology; see Table 3.1 for descriptions of lithological units. (b) Generalized lithology; see Table 3.1 for a description of how the generalized litho-types were derived from Figure 3.2a. Geology modified from Rainbird et al. (2013a, b, c, d).

Table 3.1: Description of bedrock types and surficial classes (training areas) used for classification of the Landsat, Spot and Aster data using Robust Classification Method–maximum likelihood classification (RCM–MLC) and Random Forest (RF). A total of thirteen lithological-surficial types were classified. These thirteen categories were generalized into seven major litho-types, based on the detailed lithology (Figure 3.2a) for comparison with the new generalized geology map (Figure 3.2b).

Detailed lithologies and surfical materials	**Description**	**Training area**	**Generalized lithologies and surficial materials**
(Figure 3.2a-see legend-symbols listed below)		(13 spectral units used for classification)	(Figure 3.2b- 7 generalized lithological-surficial classes)
Natkusiak basalt (1) – Nn1	Extends SE-NW in southern part. Low reflectance in VNIR. Moderate reflectance in SWIR.	Nfb1- (4 polygons – 184 pixels)	**Basalt**
Natkusiak basalt (2) – Nn2	Low reflectance in VNIR and SWIR. Distinguishable from basalt (1) in SWIR bands giving a layering appearance.	Nfb2 – (4 polygons – 179 pixels)	
Natkusiak basalt (3) – Nn3	Basalts covered by thin vegetation layer.	Nfbv – (5 polygons – 87 pixels)	
Franklin Intrusive Diabase (1) - Nfg	Extends in the central and northern parts of the area. Very similar to basalts in VNIR. Has slightly higher reflectance in SWIR-2 compared to basalts.	Nfg – (9 polygons – 420 pixels)	**Franklin intrusive rocks (diabase)**
Franlkin intrusives-diabase (2) – Nfg	Gabbro/diorite covered by thin vegetation	Nfgv – (4 polygons – 187 pixels)	
Wynniatt Fm. (carbonates) Nw1, Nw3, Nw4	Include all carbonate members of Wynniatt Fm. High reflectance in SWIR, moderate in VNIR.	Carbonates – (10 polygons - 293 pixels)	**Carbonates**
Cambro/Ordovician (carbonates) – Nk1, Nk2, Nk3, Nk4	Very similar spectral response with Wynniatt FM. Slightly different in SWIR bands.		
Paleozoic dolstones – Cmtd, Cov, Cmst	high spectral response with	Dolstones – 6 polygons – 82 pixels)	**Dolstones**
Kuujjua Fm. (quartz-arenites) – Njk Nw2, Cmc – clastic unit- mostly covered by	Occur in contact with the basalts. Low to moderate reflection in VNIR. Moderate reflectance in	Sandstones – (4 polygons – 178 pixels)	**Sandstones**

vegetation	SWIR-1, moderate to high in SWIR-2.		
Evaporite rocks – Nmi (Minto Inlet Fm), Nk4	Includes evaporates of Minto and Killian Fms. Low reflection in VNIR, slightly higher reflection in SWIR.	Evaporites – (8 polygons – 140 pixels)	**Evaporites**
Till1 – Q		Till1 – (6 polygons-191 pixels)	**Quaternary**
Till2 – Q		Till2 – (8 polygons – 156 pixels)	
Till3 - Q		Till3 – (8 polygons – 224 pixels)	
Dry Veg	Dry veg missed by biomass mask	Dry veg	Not classified

Table 3.2: Summary of remotely sensed data and associated acquisition characteristics.^

	Number of channels	**Spectral range**	**Spatial resolution**	**Dynamic range**	**Date**	**Sun Azimuth/Elevation (degrees)**
LANDSAT-7	3-V 1-NIR 2-SWIR 2-TIR 1-pan	0.450–0.515 µm (B) 0.525–0.605 µm (G) 0.630–0.690 µm (R) 0.750–0.900 µm (NIR) 1.55–1.75 µm (SWIR) 2.08–2.35 µm (SWIR) 10.40–12.50 µm (TIR) 0.520–0.900 µm (pan)	30 m (XS) 60 m (TIR) 15m (pan)	8 bits	07/14/2002	172.9 / 40.1
SPOT-5	2-V 1-NIR 1-SWIR 1-pan	0.500-0.590 µm (G) 0.610–0.680 µm (R) 0.780–0.890 µm (NIR) 1.58–1.75 µm (SWIR) 0.480–0.710 µm (pan)	10 m (XS) 20m (SWIR) 2.5m (pan)	8 bits	8/22/2009	174.3 / 30.4
LANDSAT-8	4-V 1-NIR 2-SWIR 1-TIR 1-pan	0.43-.45 (B) 0.45 -0.51 (B) 0.53 – 0.59 (G) 0.64 – 0.67 (R) 0.85 – 0.88 (NIR) 1.57 – 1.65 (SWIR) 2.11 – 2.29 (SWIR) 10.6-11.1 (TIR) 11.5 -12.51 (TIR) 0.5 – 0.68 (PAN)	15 m (pan) 30m (XS) 30m (TIR)	16 bit	7/4/2013	176. 1 / 41.9
ASTER (two	3 –VNIR	0.52-0.60 (G)	VNIR -	Bands 1	9/17/2005	South scene –

mosaicked scenes) – 1B data (radiometric and geometric coefficients applied – radiance at sensor)	6 –SWIR 6-TIR	0.63-0.69 (R) 0.76 – 0.86 (NIR) 1.6 – 1.7 (SWIR) 2.145 – 2.185 (SWIR) 2.185 – 2.225 (SWIR) 2.235 – 2.285 (SWIR) 2.295 – 2.365 (SWIR) 2.360 – 2.430 (SWIR) 8.125 – 8.475 (TIR) 8.475 – 8.825 (TIR) 8.925 – 9.275 (TIR) 10.25 – 10.95 (TIR) 10.95 – 11.65 (TIR)	15m SWIR – 30m TIR -90m	– 9 (VNIR and SWIR -8 bit) Bands 10 -14 (TIR -16 bit)	181.1 / 20.7 North Scene – 182.3 / 20.2
Airborne Magnetic data	1 channel	Total field (magnetic susceptibility		16 bit	1 km line spacing and 60m sample spacing; gridded using a minimum curvature function

^ XS = multispectral, TIR = thermal, NIR = near infrared, SWIR = short-wave infrared, V = visible, µm = microns (1* 106m).

the USGS in 16-bit format; these data covered the entire study area. Two Aster scenes (LEVEL1B) were purchased from the USGS; both were acquired on 17 September, 2005 at relatively low sun angles (Table 3.2). These were the best (cloud free) Aster images that could be sourced for the study area. However, the late season acquisition and consequent low sun angle may affect the classification results in a negative fashion (see Discussion section). The two images were radiometrically balanced and mosaicked.

Airborne total field magnetic data acquired by the Geological Survey of Canada (http://gdr.agg.nrcan.gc.ca/gdrdap/dap/search-eng.php) were also used to help classify the various rock types. These data were collected along north-south flight lines spaced 1 km apart. Sample spacing was approximately 60 m along-line and the data were gridded using a minimum curvature algorithm. The patterns of tones depicted on a total field magnetic map show the strength of the measured magnetic field and hence reflect how magnetic the rocks are. Trends in the anomalies therefore represent the distribution of magnetic minerals and reflect the geological make-up of the area.

All optical imagery was first corrected for atmospheric effects using the dark subtraction method (Gupta 1991) and enhanced by contrast stretching. In addition, the Aster radiance data were corrected for cross-talk, which is caused by signal leakage from band 4 into adjacent bands 5 and 9 (Kalinowski and Oliver 2004; Biggar et al. 2005). All images were visually and statistically assessed after all corrections were applied to ensure good quality (radiometrically balanced) data suitable for digital classification.

To obtain better discrimination between various lithological features, areas covered with water, ice and vegetation were masked using the Landsat-7 and Spot-5 imagery. The Landsat-7 image was used for masking vegetation as it was acquired in July when biomass would be most extensive. The well-known Normalized Difference Vegetation Index (NDVI) was used to characterize the vegetation distribution on the Landsat-7. NDVI, which has a dynamic range from −1 to +1 (Chuvieco and Huete 2010), is defined as:

$$NDVI = (NIR - RED) / (NIR + RED) \qquad (Eq.\ 3.1)$$

where NIR = near infrared band and RED = red band.

An upper threshold was defined to separate vegetation from non-vegetation and was used to mask all imagery. Applying the Landsat-7 vegetation mask to all data was undertaken to facilitate an unbiased comparison between the different classifications. The Spot-5 data appeared to show the most snow and ice, so these data were used to create a snow and ice mask that was, once again, applied to all imagery. This was accomplished by applying an upper threshold to the histograms for Spot-5 bands 2 and 3 (green and red), in which snow and ice is highly reflective.

Because topography can affect spectral analysis (Drury 1993; Jensen 2005), a 1:50,000 Digital Elevation Model (DEM) mosaic of the Canadian Digital Elevation Data (downloaded from Canadian Council on Geomatics website: www.geobase.ca) was created for the study area,

and shadows were modelled to produce hill shade images using the sun azimuth and elevation values for each sensor (see Table 3.2 for sun azimuth and elevation values). A mask was created from these images and overlaid on the final classification maps to screen out areas of topographic shadows and diffuse reflection from back slopes that could result in spectral misclassification.

3.3.1 Calculation of ratios

Calculation of band ratios is a common image processing technique used to highlight spectral reflectance differences between minerals as a function of wavelength as well as to reduce topographic illumination effects (Drury 1993; Jensen 2005). Table 3.3 presents well-documented ratios applied to the optical data in this study in order to highlight different minerals. These ratios were also used in concert with the raw data to classify different rock types.

Table 3.3: Summary of ratios applied to the remotely sensed data.^

Data	Ratio (band numbers)	Comment
LANDSAT 7	3/1	Ferric iron
	4/3	Biomass - vegetation
	5/7	Clay
	5/4	Ferrous iron
LANDSAT 8	6/5	Ferrous iron
	6/7	Clay
	5/4	Biomass - vegetation
	4/1b	Ferric iron
ASTER – VNIR & SWIR	6/7	Phyllic alteration
	5/6	Kalonite,muscovite,phengite
	5+7/6	Phyllic alteration
	6+9/8	Amphibole / MgOH
	6+9/7+8	Epidote, chlorite, amphibole
	7+9/8	Epidote, Chlorite, Carbonate
	4/3	Ferric oxides
	4/2	Gossan
	3/2	Biomass - vegetation

4/1	Gossan
2/1	Ferric iron
6/8	Amphibole

^References for these ratios can be found in Drury (1993), Abrams and Hook (1995), Lillesand and Keiffer (2004), and Kanlinowski and Oliver (2004).

3.3.2 Training areas

Once the images were masked, the spectral–lithological classes were defined (training area and description columns, Table 3.1). This was based on expert geologic knowledge of the study area as well as the spectral diversity observable in the Landsat, Spot and Aster images. Six main lithological classes (Table 3.1), including basalt of the Natkusiak Formation, Franklin intrusive rocks (gabbro–diorite sills and dykes), carbonate of the Wynniatt Formation, carbonate of the Cambro–Ordovician succession, evaporite rocks of the Minto Inlet and Killian Formations, and quartz arenite of the Kuujjua Formation (Figure 3.1) were defined based on the new geological map (Figure 3.2a) and field information.

Although vegetation, water and ice were excluded in all images by masking (discussed above), there were still some small areas of dry (low chlorophyll) vegetation present in the resulting images. To avoid the interference of these spectra with other rock spectra, an additional class was created to represent dry vegetation. Some of the classes that were selected on the basis of expert geologic knowledge showed variable spectral responses in different parts of the images. This could be attributed, in part, to the complicating spectral signatures of overlying glacial sediments and/or other overburden materials or vegetation cover. For example, in some parts of the study area, the Natkusiak basalt exhibits a lower reflectance that is visually distinguishable from neighbouring basaltic areas. This apparent spectral difference between basalts could be attributed to primary layering and compositional differences, grain size, and secondary processes such as weathering. To account for these apparent spectral differences, two separate basalt

classes were assigned, as well as a third class for those parts that were covered by thin, dry vegetation. Similarly, a separate class was assigned to gabbro–diorite covered by thin vegetation. Three surficial classes (glacial tills) were defined based on the spectral responses observable on the Landsat-7, Landsat-8, and Spot-5 imagery, making a total of thirteen classes used for the classification (see Table 3.1).

To facilitate a comparison with the new geology map, both the detailed lithological map (Figure 3.2a) and training areas were generalized into six basic rock types and one surficial type (Figure 3.2b) to evaluate which type of imagery provided the closest representation of the mapped geology (Table 3.1). The seven basic lithologic types that display different spectral reflection characteristics were basalt, Franklin intrusive rocks, carbonate rock, evaporite, sandstone, dolostone, and Quaternary cover (Figure 3.2b). The number of polygons and pixels for each training class is shown in Table 3.1.

The well-known maximum likelihood classification algorithm (MLC), a parametric classifier, was employed with the Robust Classification Method (RCM) to evaluate training area statistics for normal (Gaussian distributions) using first-order statistics and visual analysis of histograms (both cumulative and non-cumulative). This analysis confirmed that all distributions were normal or very close to normal, thus justifying the use of the MLC algorithm. Hereafter we refer to this algorithm as RCM-MLC.

3.3.3 Training area separability

For a successful classification, the training areas must represent each rock type, be spectrally pure, and statistically separable. The separability of a pair of probability distributions can be measured by means of divergence, which is defined in terms of the likelihood ratio:

$$Lij\ (x) = p(x|\omega i) / p(x|\omega j) \qquad (Eq.\ 3.2)$$

where $p(x|\omega i)$ and $p(x|\omega j)$ are the values of the ith and jth spectral class probability distributions, respectively, at position x (Richards and Jia 2006). The separability of the training dataset was examined using the transformed divergence (TD) statistic, which is considered to be a better measure than simple divergence (Richards and Jia 2006). TD values range between 0 and 2. Values greater than 1.9 indicate that the class pairs have good separability, whereas values less than 1 indicate that the class pairs are not spectrally separable and perhaps should be combined into one class. Values between 1.5 and 1.9 indicated moderate to good separability.

Figure 3.3a and 3.3b are plots of TD for each of the thirteen lithological –surficial classes and seven generalized lithological–surficial classes (see Table 3.1), respectively. In general, the training areas are statistically separable for all data except for the Aster thermal imagery (thermal infrared, or TIR). The Landsat-7 and Landsat-8 data offer the best separability, closely followed by Aster, whereas Spot-5 offers lower separability. The separability results indicate that classification can proceed, although the TIR results would be marginal.

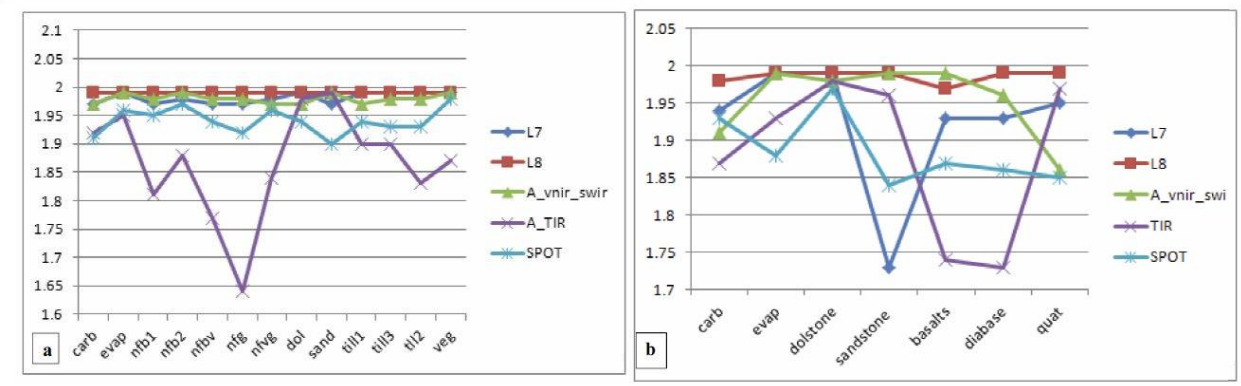

Figure 3.3:(a) Transformed Divergence (TD) values calculated from the training areas for the thirteen lithological–spectral classes for Landsat-7 (L7), Landsat-8 (L8), Aster (VNIR and SWIR; A_vnir_swir), Aster thermal (A_TIR), and Spot-5 (SPOT). See Table 3.1 (training data column) for the description of each lithological symbol along the x-axis. (b) Transformed Divergence (TD) values calculated from the training areas for the seven generalized lithologic

classes for Landsat-7 (L7), Landsat-8 (L8), Aster (VNIR and SWIR; A_vnir_swir), Aster thermal (TIR) and Spot-5 (SPOT). See Table 3.1 (generalized lithologies and surficial materials column) for descriptions of the generalized litho-types (along the x-axis) and how they were derived from the detailed geology (detailed lithologies and surficial materials column). For both TD plots, litho-type classes are shown along the x-axis and TD values along the y-axis.

3.4 Classification

Supervised classification is based on the premise that *a priori* knowledge of the features to be classified exist (i.e. training areas); this knowledge is then used to find similar areas based on the training statistics in an *n*-dimensional dataset.

A number of approaches that make use of training areas for not only producing a classification but also for validating the classification are typically employed. The same training dataset can be used for producing the classification and for validating the classified map using a confusion matrix. However, using the same training dataset for both classification and validation leads to a statistical bias that will result in an inflated value of the overall accuracy (Jones and Vaughan 2010). A more robust approach is to define one training dataset for classification and another independent set for validation, or else use a specified portion of the training dataset for classification and validation. These latter approaches provide a much more robust and statistical validation of the classified map, as an independent set of training areas are used to construct the confusion matrix.

Training area selection is often difficult because of the variability of spectral responses over the same terrain resulting from seasonal differences variation in moisture content and atmospheric effects. Furthermore, the area to be classified may be inaccessible (typical of many areas in the north) and therefore cannot always be verified by a field visit. Also, the features to be classified (bedrock and surficial materials) may present either a limited or wide range of spectral responses, making classification difficult. In some cases there are several people

involved in the selection of the training areas, also leading to variability because of different user knowledge and experience. Therefore, training area variability reflects inconsistencies in the properties inherent in the training area selection process. In this study, one of us (RR) guided the training area selection based on his knowledge of the geology and the spectral responses exhibited by different bedrock types in the remotely sensed imagery. Two ensemble classification algorithms were used to assist in quantifying the uncertainty within the training dataset.

3.4.1 Robust Classification Method (RCM) Classifier

The Robust Classification Method (RCM; Harris et al. 2012b) is an ensemble classifier that is based on a randomized and repeated sampling of a training dataset in concert with traditional cross-validation of the classification results. A series of predictions (classified maps) and associated uncertainty maps and statistics are produced. This method employs any supervised classification algorithm available in the ENVI™ suite of software; in this study we use the MLC. RCM is especially useful for assessing the effects of spectral and spatial variability in the classification process. Specifically, this method provides a majority classification and variability map, and 'confusion' statistics; these quantify the uncertainty in the classification process with respect to statistical (spectral) variability in the training dataset as well as identifying areas that show spatial variability in classification. RCM provides a more robust estimate of overall accuracy as it is based on producing a number of classifications (in this case, ten) that can be leveraged to not only provide an overall average classification accuracy but also supply an estimate of classification uncertainty (uncertainty map; see below). The RCM input parameters used herein included ten iterations, each of which sampled a random selection of 50% of the training data for classification and 50% for independent validation.

The majority classification map produced by RCM is based on the most 'popular' class on a pixel-to-pixel basis for the ten iterations, and the uncertainty maps are based on variability in the classification on a pixel-to-pixel basis. Average rule images that show the strength of membership of each pixel to a specific class, similar to a class probability image, are also generated. Harris et al. (2102b) provide a more detailed description of the RCM approach. RCM using the maximum likelihood classification algorithm (RCM–MLC) was applied to the twelve combinations of remotely sensed data shown in Table 3.4. Two sets of classifications were produced, based on the above data combinations: one for the thirteen spectral unit classes and one for the seven generalized (bedrock and surficial) classes (Table 3.1) to facilitate a comparison with the generalized geology map (Figure 3.2b).

Table 3.4: Data combinations used for classification.

Landsat-7	6 spectral bands – 3 visible, 1 NIR and 2 SWIR
Landsat -8	7 spectral bands – 4 visible (two blue bands), 1 NIR and 2 SWIR
Aster	9 spectral bands - 2 visible, 1 NIR and 6 SWIR
Aster	5 TIR bands
Spot-5	4 bands – 2 visible, 1 NIR and 1 SWIR
Landsat -7	6 spectral bands + 4 ratios (see Table 3.3)
Landsat- 8	7 spectral bands + 4 ratios (see Table 3.3)
Aster	9 spectral bands + 12 ratios (see Table 3.3)
Aster	7 – 6 spectral bands + total field magnetic data
Landsat- 8	7 spectral bands + total field magnetic data
Landsat –7	6 spectral bands and 4 ratios + total field magnetic data (for generalized classification only)
Landsat -8	7 spectral bands and 4 ratios + total field magnetic data (for generalized classification only)

3.4.2 Random Forests (RF) Classifier

Random Forests (RF) is an ensemble, and multiple decision-tree classifier that offers a number of advantages for classification:

(1) data can be binary, categorical or continuous;

(2) the classifier performs internal cross-validation through 'bootstrapping', which provides a robust estimate of classification accuracy using *out-of-bag* estimates;

(3) it is a non-parametric classifier and is relatively insensitive to outliers in the training data;

(4) it requires little user input (*m*, the number of decision trees, and *n*, the number of variables for each decision tree);

(5) it produces a classification map, but more importantly, probability maps (strength of membership in each lithological class); and

(6) it ranks the input variables with respect to their importance in the predictions.

Random Forests was originally developed by L. Breiman and A. Cutler at the University of California, Berkeley (Breiman et al. 1984; Briemen 2001). Training data are required for this approach, similar to other supervised classifiers. For each tree (the number of decision trees, *m*, is determined by the operator), a random selection of the input variables (i.e. remotely sensed image bands, *n*) is made. The number of variables selected for each tree is a fraction of the total number of variables; the square root of the number of variables is often used. Each tree employs a 'bagging' process (i.e. 'bootstrap' sample; Brieman 1996) whereby approximately two-thirds of the training areas (pixels) are used to create a prediction (referred to as *in-bag*) and one-third to validate the accuracy of the prediction (referred to as *out-of-bag*, or *oob*). This random sampling with replacement of the training dataset is undertaken for every tree. In-bag data are used to create multiple decision trees that are applied to produce independent classifications. At each node of the individual decision tree, the best split is chosen from a random sample of variables. Each tree is grown to the maximum extent with no pruning. We used the *Gini* index to determine the impurity at each node:

$$\text{Gini Index} = 1 - \sum_c (p^2(c|t)) \quad (Eq.\ 3.3)$$

where c = number of classes (e.g. litho-types), t = node of a tree, p = relative frequency of c (a given litho-type class).

The stop criteria for splitting each node is based on the minimum of samples in a node (we used 1) and the minimum impurity in a node (we used 0) allowing full growth of the decision tress (no pruning). Thus, an ensemble of trees (predictions) is created and a voting procedure is employed to assign the majority class to each pixel in the final prediction map. According to Brieman (2001), Gislason et al. (2006) and Menze et al. (2009), RF is not sensitive to noise or over-fitting and there is no need for cross-validation as it is estimated internally. However, as with any supervised classification method, an independent check of the training dataset of each litho-type is still required to calculate an unbiased and more robust estimate of classification accuracy. Additionally, the probability of membership in each class is also generated, which can be used to assess the uncertainty of the RF classification. Figure 3.4 summarizes the RF classification process.

Another very useful aspect of RF is that it calculates the importance (predictive power) of each variable in the classification process. This is accomplished by:

(1) For each tree the *oob* samples are permutated in the repsective variable and then put down the tree and the number of correct classifications are caclulated (*nP*);
(2) The *in-bag* training samples (original) are put down the tree and the number of correct classifications are calculated (*nC*); *nC–nP* calculated;
(3) Calculate *Nc–Np*.

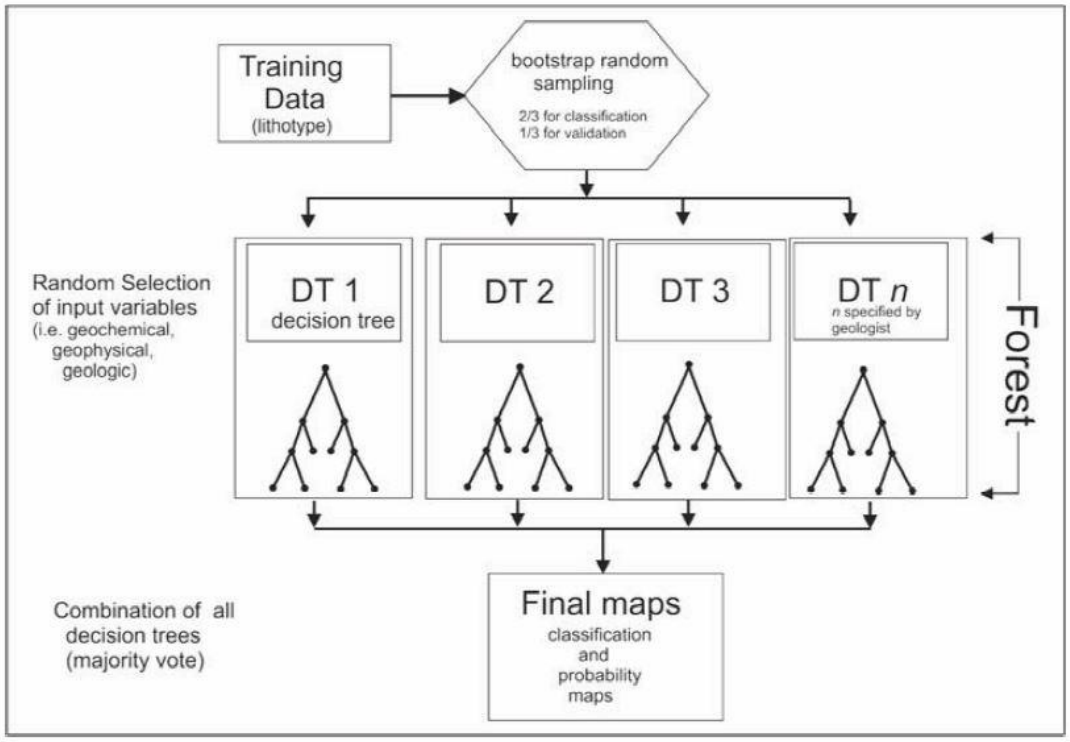

Figure 3.4: Summary of the Random Forest (RF) classification algorithm.

The average of the differences of the accuracies for all trees is the raw importance of each variable. However, to provide a more robust estimate the raw variable importance is divided by the respective standard deviation creating a normalized variable importance value. A high normalized value has a high importance for the entire RF and *vice versa* for a low number.

A RF classification map using all the available raw and ratio data from the Landsat-7 and Landsat-8, Aster (VNIR: visible near infrared; SWIR: shortwave infrared; and TIR) and magnetic data, totalling 43 input bands, was produced along with a probability-of-membership map (same as a rule map for RCM–MLC), which was used to reduce the uncertainty by only accepting pixels for each class that showed a strong membership (>95%). In addition, the 43 input bands were ranked for their predictive power using the method as discussed above. The RF parameters used include *n* (number of variables to create each tree), which was set to the square root of the total number of variables, and *m* (number of decision trees), which was set to 100

through experimentation. As pointed out by Rodriguez-Galiano et al. (2014), the number of trees is proportional to the classification accuracy until the *oob* error converges. Once the error converges at a given number of trees, the number of random variables selected for each tree (*n*) minimally alters the classification accuracy. In this study, the *oob* error stabilized at 100 decision trees (*m*) in all of our experiments.

3.4.3 Ensemble Classifications

Ensemble methods (Benediktsson and Swain 1992; Doan and Foody 2007; Benediktsson et al. 2007; Banfield et al. 2007) leverage multiple models (i.e. classifications) to obtain a better prediction than could be derived from individual predictions. The flexibility of ensembles can, in theory, enable them to over-fit the training data more than a single prediction would, but in practice, some ensemble techniques, especially bagging (Breiman 1996; Freund and Schpire 1999) tend to reduce problems related to over-fitting of the training data. In addition, they provide a more robust estimate of the overall classification accuracy. Both classification algorithms used in this study, RCM–MLC and RF, are ensemble classifiers as they produce multiple predictions (classifications), which are combined using a voting procedure that selects the most popular class on a pixel basis.

3.4.4 Multiple Classification Systems (MCS)

Multiple classification systems (MCS; Polikar 2006; Benediktsson et al. 2007; Waske and Braun 2009) are a fairly recent concept; they do not refer to a specific classification algorithm but describe a more general classification strategy. MCS can involve combining classifications based on a number of different algorithms, variants of the same algorithm, or classifications derived from different datasets. In this case, the latter approach was utilized by combining classifications

derived from Landsat, Spot-5, Aster and airborne magnetic data. Specifically, we use a *stacking* approach whereby the data combinations that produced the highest classification accuracies for the generalized lithological classes (seven in total) were combined in an ensemble fashion by (a) taking the majority class (most 'popular') for each pixel (*MCS majority map*), and (b) only classifying a pixel when the class was in agreement with all input classifications comprising the stack (*MCS agreement map*).

As part of this MCS process, and because multiple classifications are being combined, uncertainty measures such as the range and/or standard deviation of the classification on a stacked pixel basis may be calculated. In this study, a standard deviation calculation was used on the pixel stack (derived from the 100 decision trees for RF and 10 iterations for RCM–MLC) to estimate uncertainty in the MCS classification process. The standard deviation map was then thresholded and used to mask-out uncertain areas on the MCS classification maps.

A variation of the MCS theme was also added by selecting the data combination that provided the highest classification on a class-to-class basis and producing a classification map from these selections. Thus, this *MCS best bands majority classification map* contains the class from the data combination that provided the highest classification accuracy. To further decrease the uncertainty in this classification image, the average *rule or probability images* (strength-of-class membership, discussed above) generated by RCM–MLC were used to create the classification map. Each *rule image* for each data combination that provided the highest classification accuracy was thresholded at the 95% level, thereby presenting, for each class, the pixels having the strongest membership. Again, this produced a map where only the most certain pixels were classified.

3.5 Classification and Results

3.5.1 RCM–MLC classifications: thirteen classes

Table 3.5 presents the results of the RCM–MLC classification for all data combinations for all thirteen lithological–surficial classes (Table 3.1), and Figure 3.5 is a box and whisker plot of the average accuracies over the ten iterations of RCM–MLC for each data combination. Figure 3.6 is a plot of the average producer's and user's accuracy. Figure 3.7a shows the majority classification map for the thirteen classes derived from the data combination that produced the overall RCM–MLC highest classification accuracy (Landsat-7 raw and ratios data – 89.7%) and Figure 3.7b shows the associated RCM–MLC uncertainty map.

Table 3.5: Results of Robust Classification Method–maximum likelihood classification (RCM–MLC) showing average overall, producer's and user's accuracies and average class accuracy for each data combination for each of the thirteen lithological–spectral classes (Table 3.1). Bold numbers are the highest classification accuracies. ^

Class	**L7_r**	**L8_r**	**L7_rr**	**L8_rr**	**ASTER VNIR SWIR_r**	**ASTER -TIR**	**ASTER _rr**	**L7_rr _mag**	**L8_rr _mag**	**SPOT**
Carbonates	85.5	97.9	86.2	94.9	85.4	65	59.3	96	**98**	79
Evaporites	87.3	**98.2**	89.7	87.3	89.4	60.9	51.2	90.8	84.7	91.5
Basalts Nfb1	79.3	55.9	**87.9**	67.1	11.9	12.2	25.2	54.3	40.2	76.7
Basalts Nfb2	**92.6**	72.0	86.7	57.1	41.0	23.7	56.8	91.1	77.9	78.8
Basalts Nfbv	74.6	**88.2**	75.8	83.8	81.9	24.6	58.9	69.1	66.3	74.1
Franklin Diabase Nfg	91.2	92.3	92.6	**95.2**	81.9	36	83.7	91.2	87.8	87.6
Franklin Diabase Nfgv	73.7	92.2	89	**94.2**	21.2	7	41.0	47.7	72.8	82.5
Dolostones	83.2	**95.3**	80.1	**96.4**	56.7	15.1	65.2	65.2	74.6	74.8

Sandstone	61.7	78.0	84.6	74.6	62.3	**93.8**	32.8	50.7	58.7	47.1
Till1	90.5	79.1	**93.5**	75.3	81.3	70	79.5	76.5	64.3	62.1
Till2	92.6	78.9	**97.6**	91.8	60.9	20.2	53.2	73	75.9	57.8
Till3	89.4	88.0	92.8	**96.3**	79.3	32.7	75	95.4	86.7	71.8
Vegetation	**97.2**	70.6	95.8	91.1	45.3	16.4	13.4	82	80.6	90.5
Overall accuracy	86.3	84.2	**89.7**	86.2	65.1	42.9	57.3	77.1	77.5	76.1
Average Producer's accuracy	85.5	83.1	**88.4**	84.6	61.3	40	54.5	81.3	74.9	74.3
Average Users Accuracy	87.3	86.6	**90.1**	87.4	64.6	42.5	57.6	86.3	81.2	75.4

^ Key to abbreviations: L7_r = Landsat-7 (raw); L8_r = Landsat-8 (raw); L7_rr = Landsat-7 (raw and ratios); L8_rr = Landsat-8 (raw and ratios); ASTER VNIR_ SWIR_r = Aster very near infrared and shortwave infrared (raw); ASTER–TIR= Aster thermal bands; ASTER_rr = Aster VNIR and SWIR (raw and ratios); L7_rr_mag = Landsat-7 (raw, ratios and magnetics); L8_rr_mag = Landsat-8 (raw, ratios and magnetics); SPOT = Spot-5 (raw).

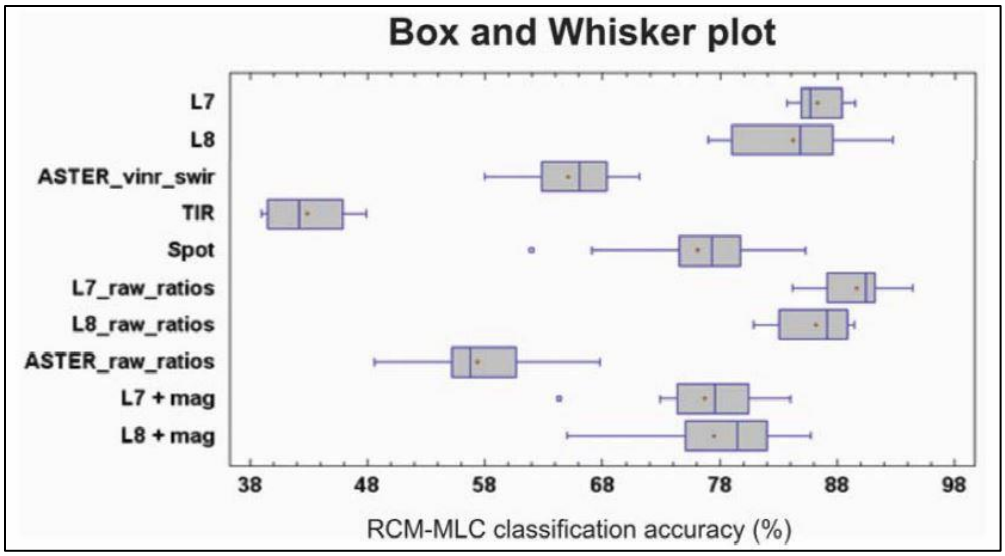

Figure 3.5: Box and whisker plot showing average (shown as a cross symbol) and median (shown as a vertical line) classification accuracies as well as the range in accuracies for ten iterations of Robust Classification Method–maximum likelihood classification (RCM–MLC) (using various data combinations) for thirteen lithological–spectral classes.

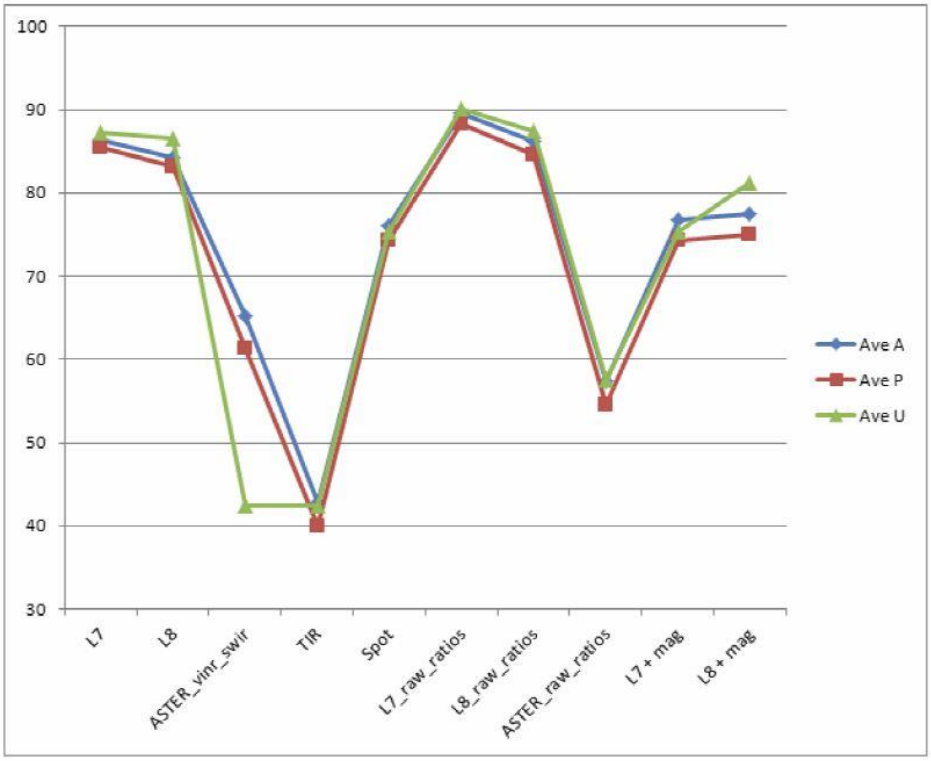

Figure 3.6: Plot of average overall accuracy (Ave A), average Producer's accuracy (Ave P) and average User's accuracy (Ave U) for ten iterations of the Robust Classification Method–maximum likelihood classification (RCM–MLC; using various data combinations) for thirteen lithological–spectral classes. The data combination is shown on the x-axis and classification accuracy (0–100%) along y-axis.

With respect to RCM–MLC, the highest classification accuracies were obtained from Landsat-7 (raw and ratios), Landsat-7 (raw), Landsat-8 (raw and ratios), and Landsat-8 (raw), all of which have an overall accuracy greater than 80%. The other combinations, including Aster (especially the TIR) and Spot-5, have lower classification accuracies. The results from Aster are generally lower and no doubt relate to the acquisition parameters (Table 3.2), primarily a lower sun angle, which would result in a lower signal-to-noise ratio.

With reference to Table 3.5, the carbonate rocks are classified most accurately on the Landsat-8 (raw and magnetics) data, as these rocks have very low magnetic susceptibility (compared to the basalts and Franklin diabase) and very high reflectance in the visible bands and

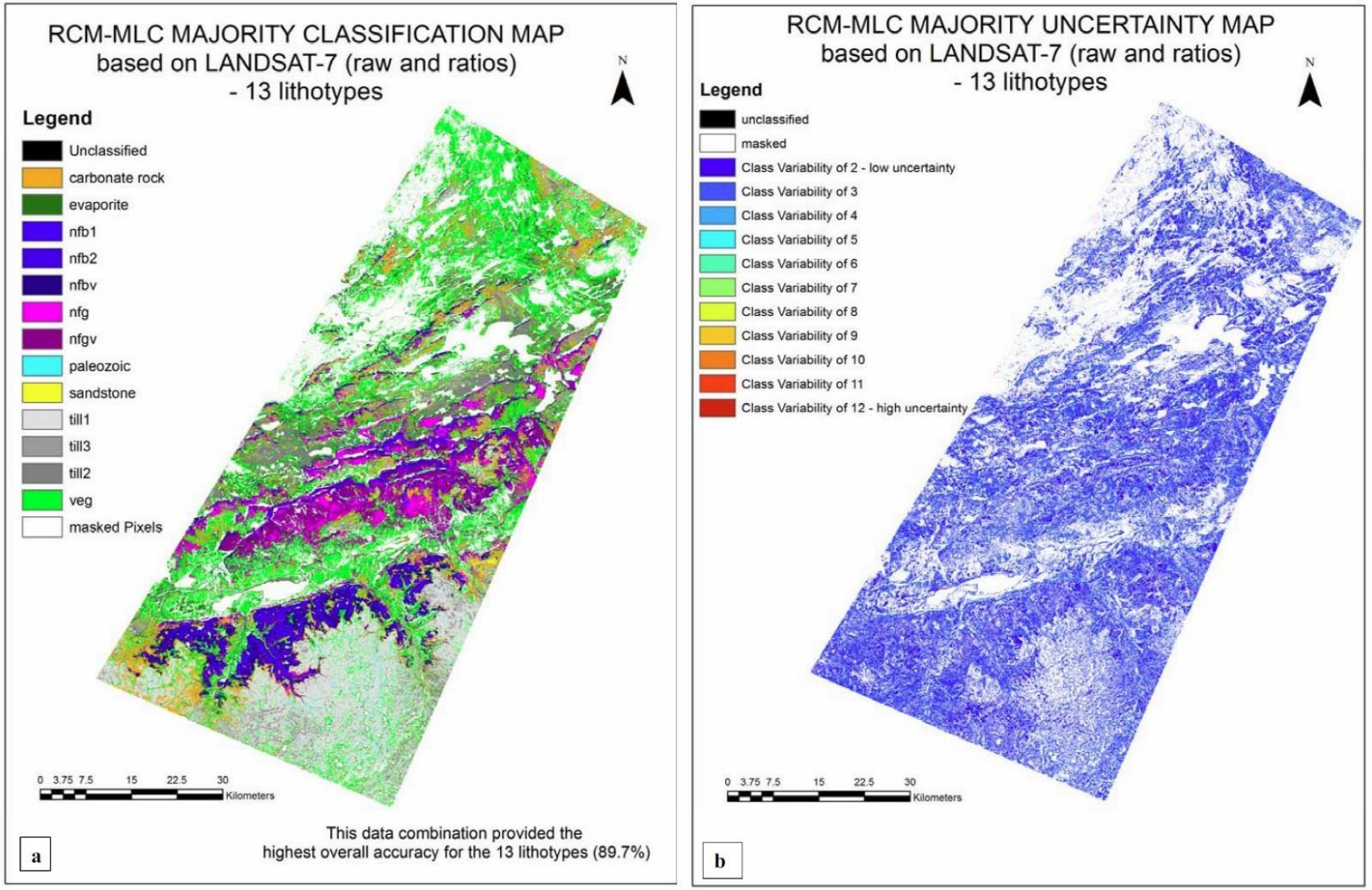

Figure 3.7:(a) Robust Classification Method–maximum likelihood classification (RCM–MLC) Majority Classification Map of thirteen classes, produced from the Landsat-7 (raw and ratios) data combination that produced the highest overall accuracy (89.7%) of all data combinations. See Table 3.1 for description of lithologic units. (b) RCM–MLC uncertainty map. Note that few pixels are uncertain; hence warmer colours are difficult to see on the uncertainty map.

band 6 (SWIR), but low reflectance in band 7 (SWIR). Note that the accuracies for carbonate are consistently high for all data combinations with the exception of Aster. Evaporites are also classified with high accuracy on all data combinations, with the exception of Aster (both VNIR–SWIR and TIR); Landsat-8 (raw data) provided the highest classification accuracy (98.2%).

The Natkusiak basalts, which comprise three distinct spectral subunits (see Table 3.1), are best classified on combinations of the Landsat-7 and Landsat-8 data. Landsat-7 (raw and ratios) provide the highest classification accuracy for Nfb1 (87.9%), Landsat-7 (raw) for Nfb2 (92.6%) and Landsat-8 (raw) for Nfbv (88.2%). It is interesting to note that addition of the magnetic data does not provide higher classification accuracies compared to the Landsat-7 and Landsat-8 optical data used alone. The Franklin intrusive rocks (diabase), which comprise two spectral sub-units (Table 3.1), are best classified on the Landsat-8 (raw and ratio) data combination (Nfg – 95.2%; Nfgv –94.2%). Inclusion of the Landsat-8 ratios improves the overall classification accuracy from 3 to 5% for the Nfg and Nfgv sub-units, respectively. Dolostones are again best classified on the Landsat-8 data, especially when the ratios are included (96.4%).

In general, the Aster TIR data offer little separation between the different rock types; however, the TIR data provide appreciably higher classification accuracy for sandstone (93.8%) than the other data combinations. The data combination that offers the second highest classification for sandstone (84.6%) is the Landsat-7 (raw and ratios). Visual examination of the thermal data indicates that the training areas used for classification of sandstone are thermally much cooler than other rock types, especially at the 8.125, 8.475 and 8.925 μm wavelengths. This is also confirmed by the separability values (TD) (Figure 3.3a), which reveal that sandstone has the highest TD value for all thirteen classes.

The Landsat-7 and Landsat-8 (raw and ratios) data provide the highest classification accuracies for the three spectrally distinct till classes. Again, inclusion of the ratio data increases the accuracies by an average of 4%. The Aster and Spot-5 data provide lower classification accuracies when compared to the Landsat data. It is interesting to note that inclusion of magnetics with the Landsat-7 (raw and ratio) data increases the classification accuracy from 92.8 to 95.4%, very close to the best accuracy provide by Landsat-8 (raw and ratio; 96.3%). Visual analysis of the magnetic data indicates that till unit 3 has a lower magnetic signature than the others, which may reflect lower magnetic susceptibility of the underlying bedrock (not known because of glacial cover), and/or a lower magnetite content in the overburden compared to other till units.

3.5.2 RCM–MLC generalized lithological classifications: seven classes

Figure 3.8 is a box and whisker plot of the average RCM–MLC accuracies for nine data combinations (Table 3.4) based on the seven generalized lithological–surficial classes (Table 3.1). Table 3.6 provides a summary of overall accuracy, producer's and user's accuracy, as well as classification accuracy by class for all the data combinations. Note that TIR was included with the Aster (raw and ratio) data combination mainly because of its ability to discriminate sandstone. Also note that for the generalized classifications two more data combinations were added – Landsat-7 and Landsat-8 (raw, ratios, and magnetic data) (Table 3.4). These were included because of the additional value the ratio and magnetic data provided for classifying the thirteen lithological–spectral classes, discussed above. The data combinations comprising the Landsat-7 and -8 (raw), Landsat-7 and -8 (raw and ratios), Landsat-7 and -8 (raw, ratios and magnetics) and the Aster (raw and ratios) data offer classification accuracies in excess of 85%; the Landsat-8 (raw and ratios) offers the highest overall, producer's, and user's accuracies of 88.4, 87.7 and

90.1%, respectively (Table 3.6). Figure 3.9a shows the majority classification map for the seven generalized lithological–surficial classes derived from the data combination that produced the highest overall classification accuracy (Landsat-8 raw and ratios – 88.4%), and Figure 3.9b shows the associated uncertainty map. The Aster (raw) and Spot-5 data combinations resulted in significantly lower accuracies. In general, the classification accuracies are higher for the generalized lithologies (seven classes) versus the more detailed lithologies and thirteen spectral sub-units for all data combinations. This is expected because of the reduction and generalization of the detailed lithological classes.

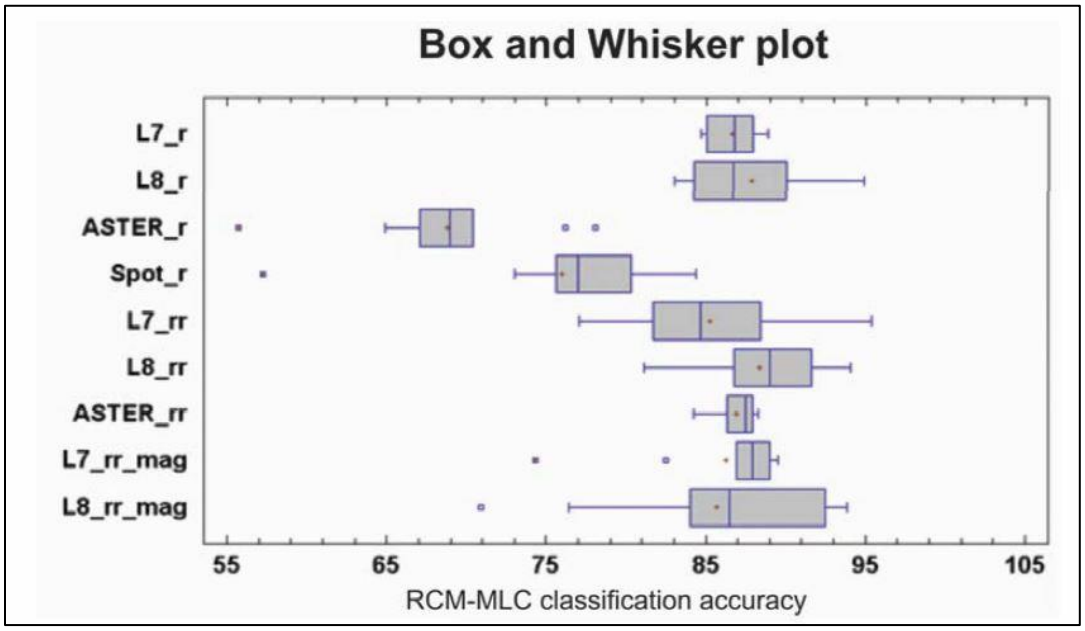

Figure 3.8: Box and whisker plot showing average (shown as a cross symbol) and median (shown as a vertical line) classification accuracies, as well as the range in accuracies for the ten iterations of the Robust Classification Method–maximum likelihood classification (RCM–MLC) for the seven generalized lithological–surficial classes (see Table 3.1).

The carbonate rocks, once again, are most accurately classified on the Landsat-8 (raw; 96.2%) and the Landsat-8 (raw, ratios, magnetics; 95.3%) data combinations (Table 3.6). Addition of the ratios and magnetic data in this case does not improve the overall accuracy.

Table 3.6: Results of Robust Classification Method–maximum likelihood classification (RCM–MLC) showing average overall, producer's and user's accuracy and average class accuracy for each data combination for each of the seven generalized lithologic–surficial classes. Bold numbers are the highest classification accuracies. ^

Class	**L7_r**	**L8_r**	**L7_ rr**	**L8_ rr**	**Aster VNIR SWIR_r**	**ASTE R_rr_ TIR**	**L7_r_mag (L7_rr_mag)**	**L8_r-mag (L8_rr_mag)**	**SPOT**
Carbonates	87	**96.2**	89	93.3	81.5	69.6	80.8 (92.2)	91.8 **(95.3)**	78.9
Evaporites	82.4	82.4	89.2	84.4	73.7	89.6	**95.5 (98.6)**	85.4 **(95.3)**	93.8
Basalts	**90.1**	87.8	87.5	81.4	63.4	**90.9**	**90.7** (86.4)	80.2 (76.1)	71.2
Franklin Diabase	85.1	87.4	79.9	**86.4**	68.5	**88.9**	80.4 (82.8)	82.8 (**90.5**)	72.9
Dolostones	91.0	89.7	87.9	**96.8**	68.5	73.2	72 (53.8)	88.9 **(91.2)**	83.4
Sandstone	62.9	70.6	62.7	73.1	30.1	**86.7**	42.2 (77.8)	72.7 (41.9)	66.4
Quatrernary – till cover	92.7	90.4	92.6	**94.9**	77.8	91.8	93.9 (91.5)	93.2 (92.4)	79.5
Overall accuracy	86.6	**87.9**	85.3	**88.4**	68.8	86.9	83.2 (86.3)	86.1 (85.7)	76
Average Producer's accuracy	83.8	85.8	84.5	**87.7**	66.7	84.3	79.8 (83,1)	85 (83.7)	78.1
Average Users Accuracy	86	89.6	87.3	**90.1**	65.8	86.2	81.7 (87.7)	88.1 (87.1)	73.3

^ Key to abbreviations: L7_r = Landsat-7 (raw); L8_r = Landsat-8 (raw); L7_rr = Landsat-7 (raw and ratios); L8_rr = Landsat-8 (raw and ratios); ASTER VNIR_ SWIR_r = Aster very near infrared and shortwave infrared (raw); ASTER_rr_TIR = Aster, all bands, ratios and thermal bands; L7_r_mag = Landsat-7 (raw, magnetics); L7_rr_mag = Landsat-7 (raw, ratios and magnetics); L8_r_mag = Landsat-8 (raw, magnetics); L8_rr_mag = Landsat-8 (raw, ratios and magnetics); SPOT = Spot-5 (raw).

Again, recall that carbonates are easy to discriminate using optical data. The evaporites are classified most accurately (98.6%) on the Landsat-7 (raw, ratios, magnetics) data combination, although they have a very low magnetic susceptibility. Addition of the ratio data again helps to increase the overall accuracy by approximately 6% and the magnetic data by approximately 5%. The basalts are best classified using, suprisingly, the Aster (raw and ratios), Landsat-7 (raw,

magnetics) and Landsat-7 (raw), all offering greater than 90% accuracy. In this case, the ratio data do not improve classification accuracies. The magnetic data are useful because the basalts have an enhanced magnetic response (higher magnetic susceptibility). The Landsat-8 (raw, ratio, magnetics) data combination results in the highest classification accuracy for the Franklin intrusive rocks. Addition of the ratios and the magnetic data, because of their high magnetic response, offers an increase in accuracy of approximately 4%. The dolostones are best classified by the Landsat-8 (raw and ratios) data combination, which is 5% higher than any other data combination. Inclusion of the ratios improved the accuracy by approximately 7% compared to the raw data only. The Aster data provided the highest classification accuracy for sandstones (VNIR and SWIR; raw and ratios), for reasons relating to the thermal data, discussed above. The accuracy is on the order of 15% higher than for any other data combination. Again, the Landsat-8 (raw and ratios) data provided the best classification of Quaternary deposits (94.9%).

3.5.3 Random Forest (RF) classification

Figure 3.10a shows the RF classification map and Figure 3.10b is a RF map in which uncertain pixels have been removed, as discussed in a previous section. Overall, the RF map produced the highest classification accuracy of 99%, which exceeds the highest overall RCM–MLC classification accuracy, that of the Landsat-8 (raw and ratios) data combination (88.4%), by approximately 10%. However, one has to keep in mind that all 43 bands (raw and ratios) from the Landsat-7, Landsat-8, and Aster (VNIR, SWIR, TIR) were used for classification. The RCM–MLC classifications used various data combinations, discussed above, but not the total number of bands.

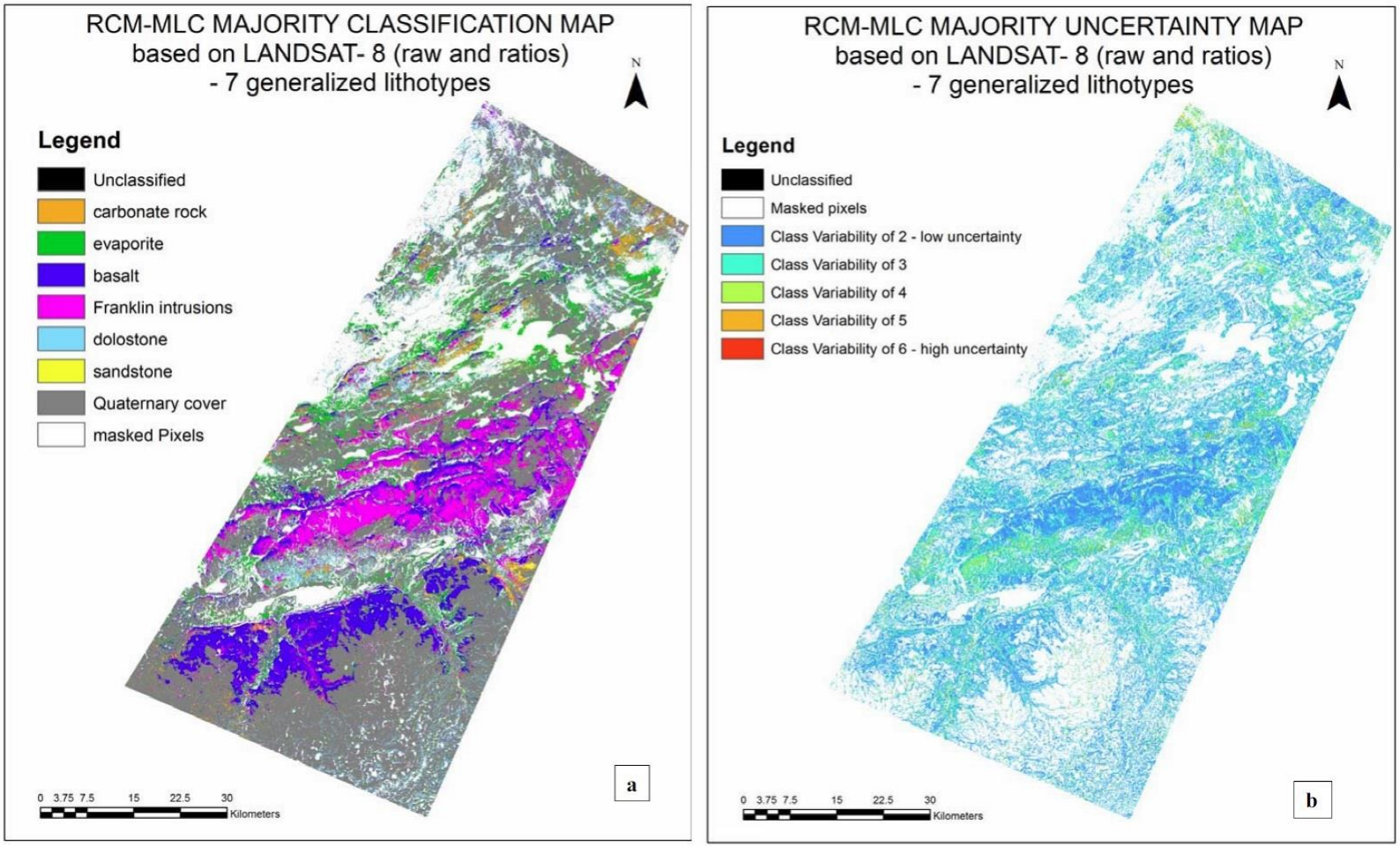

Figure 3.9: (a) Robust Classification Method–maximum likelihood classification (RCM–MLC) Majority Classification Map of seven generalized lithological–surficial classes, produced from the Landsat-8 (raw and ratios) data combination that produced the highest overall accuracy (88.4%) of all data combinations. (b) RCM-MLC uncertainty map.

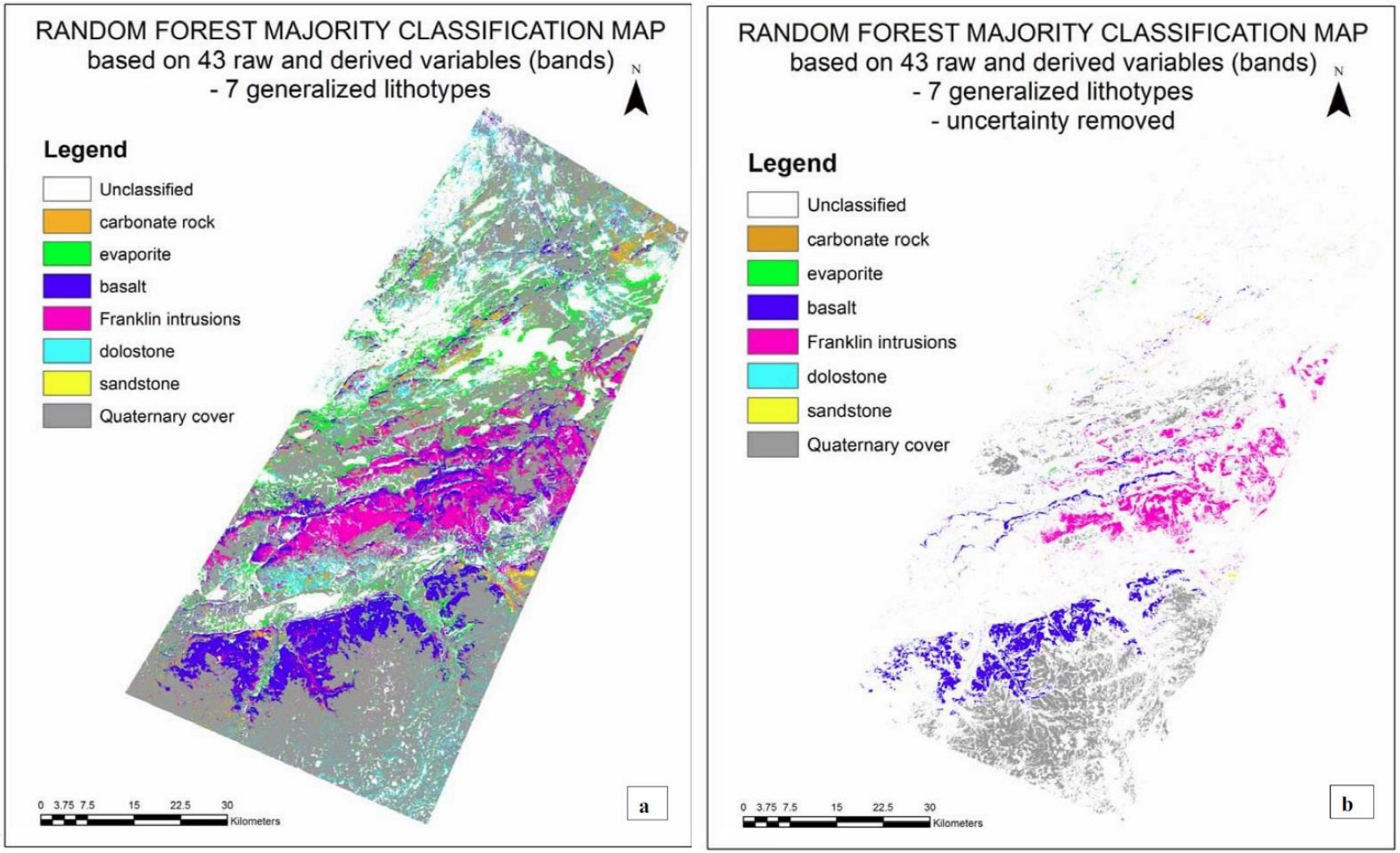

Figure 3.10: (a) Random Forest (RF) classification map of seven generalized lithological–surficial classes using all 43 input bands; an overall accuracy of 99% was achieved. (b) Associated RF classification map with uncertainty removed.

The other purpose of using the RF classifier was to rank the predictive power of each input band, the results which are shown on Figure 3.11 (*Gini* index) and Table 3.7. The groups of variables shown in Table 3.7 correspond to the breakpoints as indicated in the *Gini index* plot (Figure 3.11). With respect to the ranking of the bands that provide the most predictive power, the ferrous Landsat-7 ratio and Landsat-8 band 5 (NIR) (Group 1 – Table 3.7 and Figure 3.11) provide the most significant contributions. Group 2, also significant contributors, contain two Landsat-8 ratios, ferrous iron and biomass. Group 3 primarily comprises SWIR and NIR bands derived from Landsat-7 and Landsat-8. Of interest is the Aster NIR band, also included in this group, which provides moderate predictive power. Also of note are the total field magnetic data found in Group 6. The ferrous ratios and magnetic data are useful predictors, especially for the basalt and Franklin intrusive rocks, which are characterized by ferrous iron minerals and magnetite that result in high magnetic susceptibilities. However, the classes characterized by low magnetic susceptibility can also be separated using the magnetic data, as discussed above.

3.5.4 Multiple Classification System –seven classes

Figure 3.12 shows the *MCS majority classification map*, discussed above, in which each pixel was classified with the majority class, from the ensemble stack of data combinations that provided an overall classification accuracy in excess of 80% (Table 3.6). These included Landsat-7 (raw), Landsat-8 (raw), Landsat-7 (raw and ratios), Landsat-8 (raw and ratios), Landsat-7 (raw, ratios and magnetics), Landsat-8 (raw, ratios and magnetics) and Aster (raw, ratios and TIR).

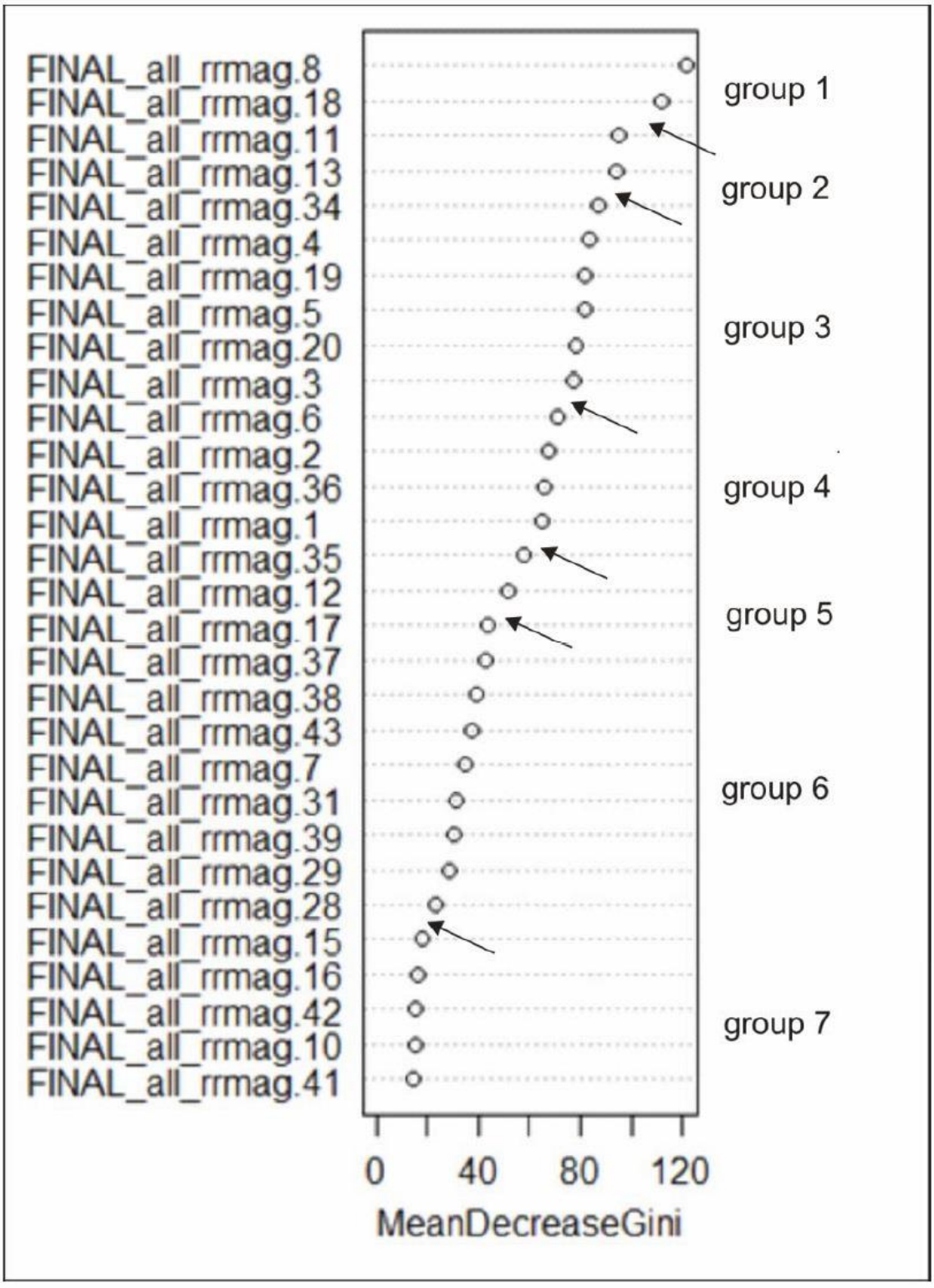

Figure 3.11: Gini index plot showing variable (band) importance. Arrows indicate breakpoints that divide the bands into groups. See Table 3.7 for description of bands composing each group.

Figure 3.13a is an ensemble map that shows the class agreement (MCS *classification agreement map*) for the best seven classifications of the generalized litho-types (>80% overall accuracy; Table 3.6), whereas Figure 3.13b shows the uncertainty associated with this map. Recall that the uncertainty map is based on the standard deviation of the majority vote for each pixel. Figure 3.13c is the *MCS agreement map* with the uncertainty, composed of the pixels in class 5 and 6 (most uncertain = less agreement; Figure 3.13b) removed. Note that not all the pixels in Figure 3.13c have been classified, because uncertainties in the classification process associated with snow, ice, high biomass, and pixels that do not show agreement, as well as low reflection from back slopes, have all been removed.

Table 3.7: Ranking of 30 of the 43 variables (image bands) constituting input to the Random Forest (RF) classification.

RANK (group—see Figure 3.11)	**Number In *gini index* plot (Figure 3. 11)**	**DATA**	**BAND**
Group 1			
1	8	L7	5/4 – ferrous iron
2	18	L8	Band 5 – NIR
Group 2			
3	11	L8	6/5 – ferrous iron
4	13	L8	5/4 biomass
Group 3			
5	34	ASTER	Band 4 (NIR)
6	4	L7	Band 4 (NIR)
7	19	L8	Band 6 (SWIR)
8	5	L7	Band 5 (SWIR)
9	20	L8	Band 7 (SWIR)
10	3	L7	Band 3 (red)
Group 4			
11	6	L7	Band7 (SWIR)
12	2	L7	Band 2 (green)
13	26	ASTER	Band 6 (SWIR)
14	1	L7	Band 1 (blue)
Group 5			
15	35	ASTER	Band 5 (SWIR)
16	12	L8	6/7 - clay
Group 6			
17	17	L8	Band 4 (red)

18	37	ASTER	Band 7 (SWIR)
19	38	ASTER	Band 8 (SWIR)
20	43	Total Field Magnetics	
21	7	L7	5/7 -clay
22	31	ASTER	4/1 -
23	39	ASTER	Band 9 (SWIR)
24	29	ASTER	4/2
Group 7			
25	28	ASTER	4/3
26	15	L8	Band 1b – (blue)
27	16	L8	Band 3 (green)
28	42	ASTER	BAND 3 (NIR)
29	10	L7	4/3
30	41	ASTER	Band 2 (red)

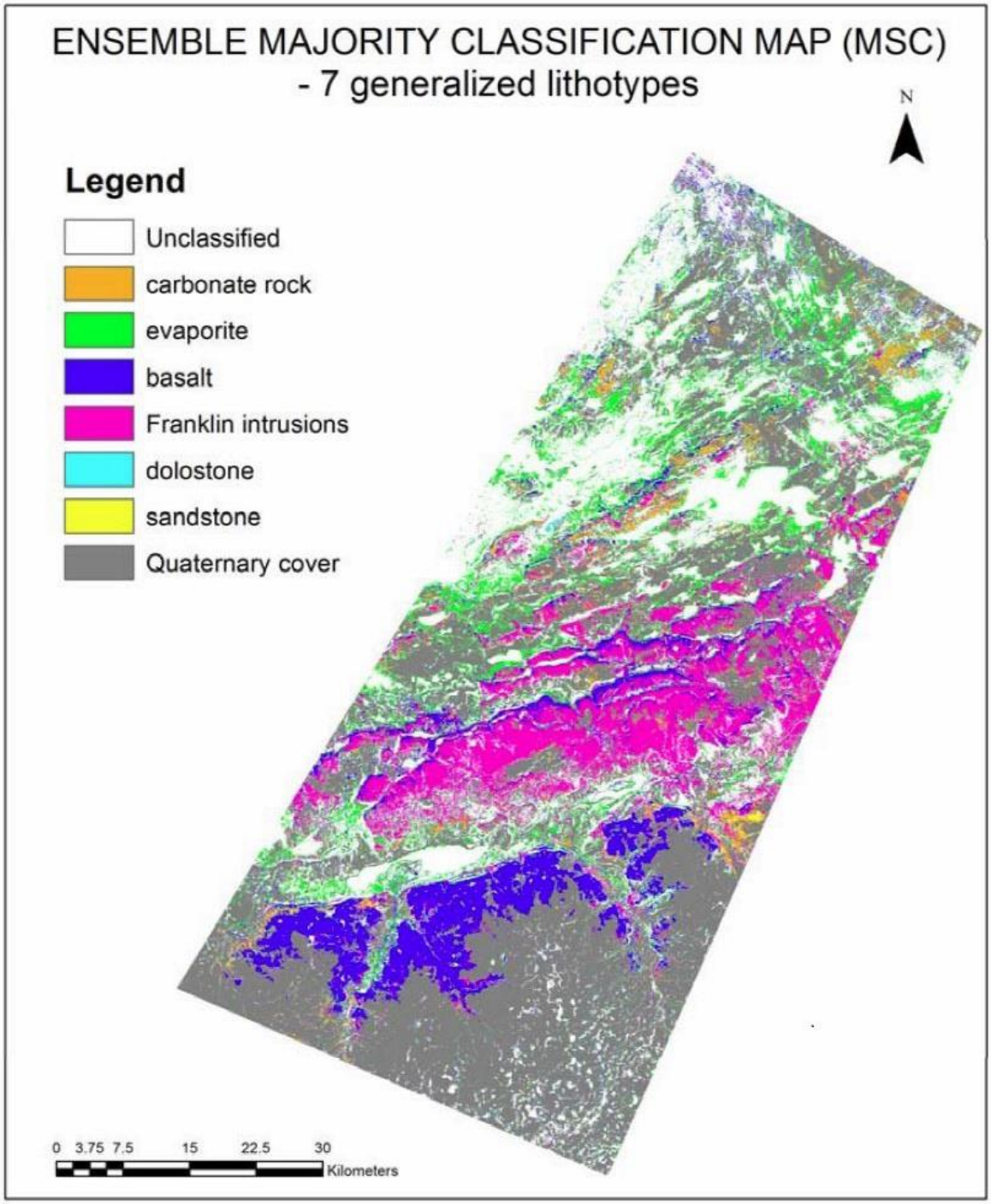

Figure 3.12: MLC Multiple Classification System (MCS) Majority Classification Map.

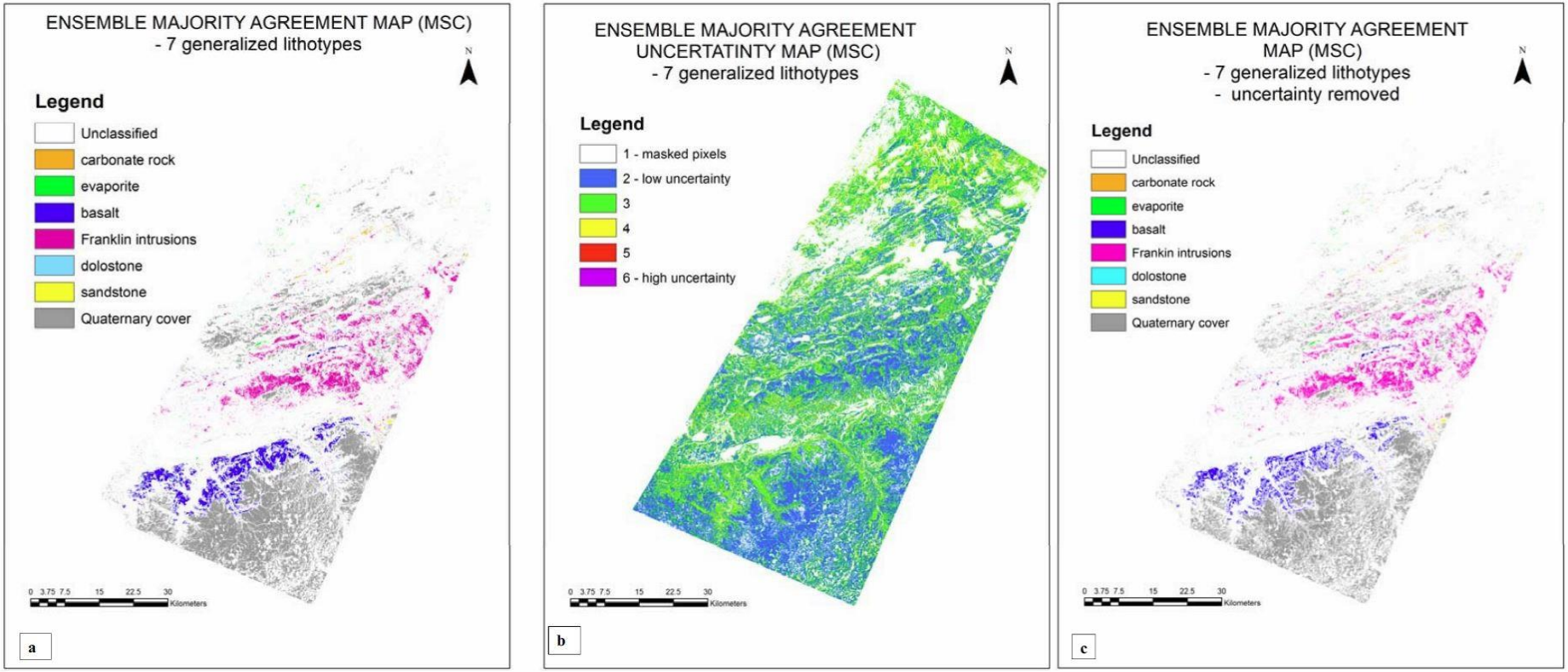

Figure 3.13: (a) Multiple Classification System (MCS) Classification Agreement Map. (b) MCS Agreement Uncertainty Map. (c) MCS classification Agreement Map with uncertainty (class 5 and 6) from the ensemble uncertainty map shown in b, removed.

The data combination that provided the highest classification accuracy for each of the seven generalized lithological classes (Table 3.8) was used to produce a MCS *best bands majority classification map* (Figure 3.14). In addition, the *average rule images* were used for each class to include only pixels that have a membership greater than 95% probability, as discussed above. For sandstone, which had the lowest classification accuracy, the best classification was achieved by the Aster (raw, ratios and TIR) data combination, whereas evaporites had the highest classification accuracy produced from the Landsat-7 (raw, ratios and magnetics) data combination.

Table 3.8: Best band data combination for each of the seven generalized lithological-surficial classes.

Class	**Data combination**	**Class accuracy (%)**
Carbonates	LANDSAT 8 –raw bands	96.2
Evaporites	LANDSAT 7 –raw, ratios, mag	98.6
Basalt	ASTER – raw ,ratios	90.9
	LANDSAT7, raw, ratios, mag	90.7
Diabase	LANDSAT 7 – raw ,ratios, mag	90.5
Dolostone	LANDSAT 8 – raw, ratios	96.8
Sandstone	ASTER- raw, ratios	86.7
Quaternary	LANDSAT 8 – raw, ratios	94.9

Figure 3.15 is a MCS map that combines the three MCS maps discussed above – the MCS *majority classification*, *agreement* and *best bands* maps. This was produced by first taking pixels that were in agreement on all three maps and then pixels that were in agreement on a combination of two of the three maps. As shown in the legend, only two classes, basalt and carbonate rocks, were in agreement on all three MCS maps; the others were in agreement on only two of the MCS maps. This indicates that the basalts and carbonate rocks were easiest to classify as they have unique spectral and magnetic signatures.

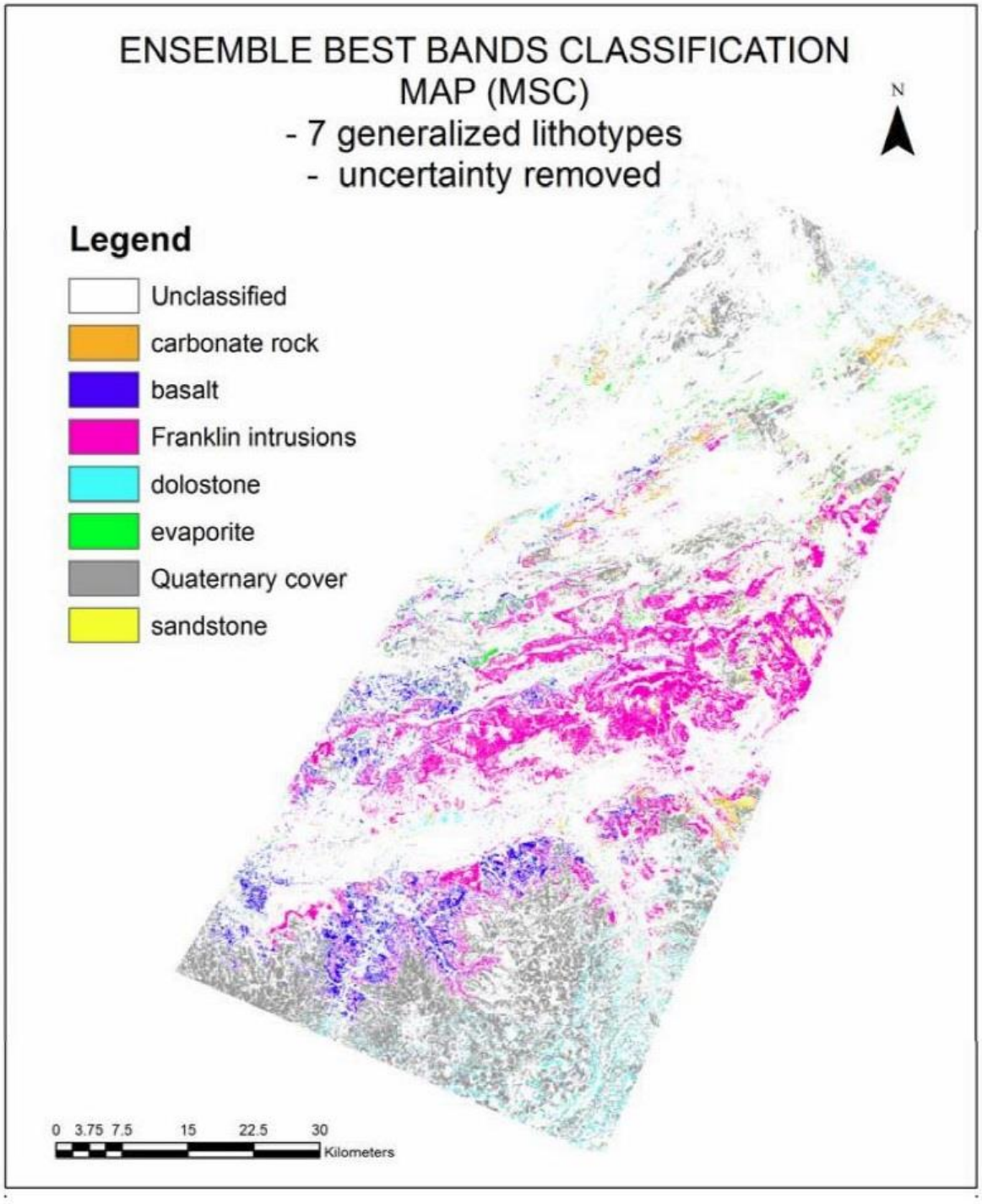

Figure 3.14: Multiple Classification System (MCS) Best Bands Classification Map using thresholded average rule images (>95%) to reduce uncertainty (see text).

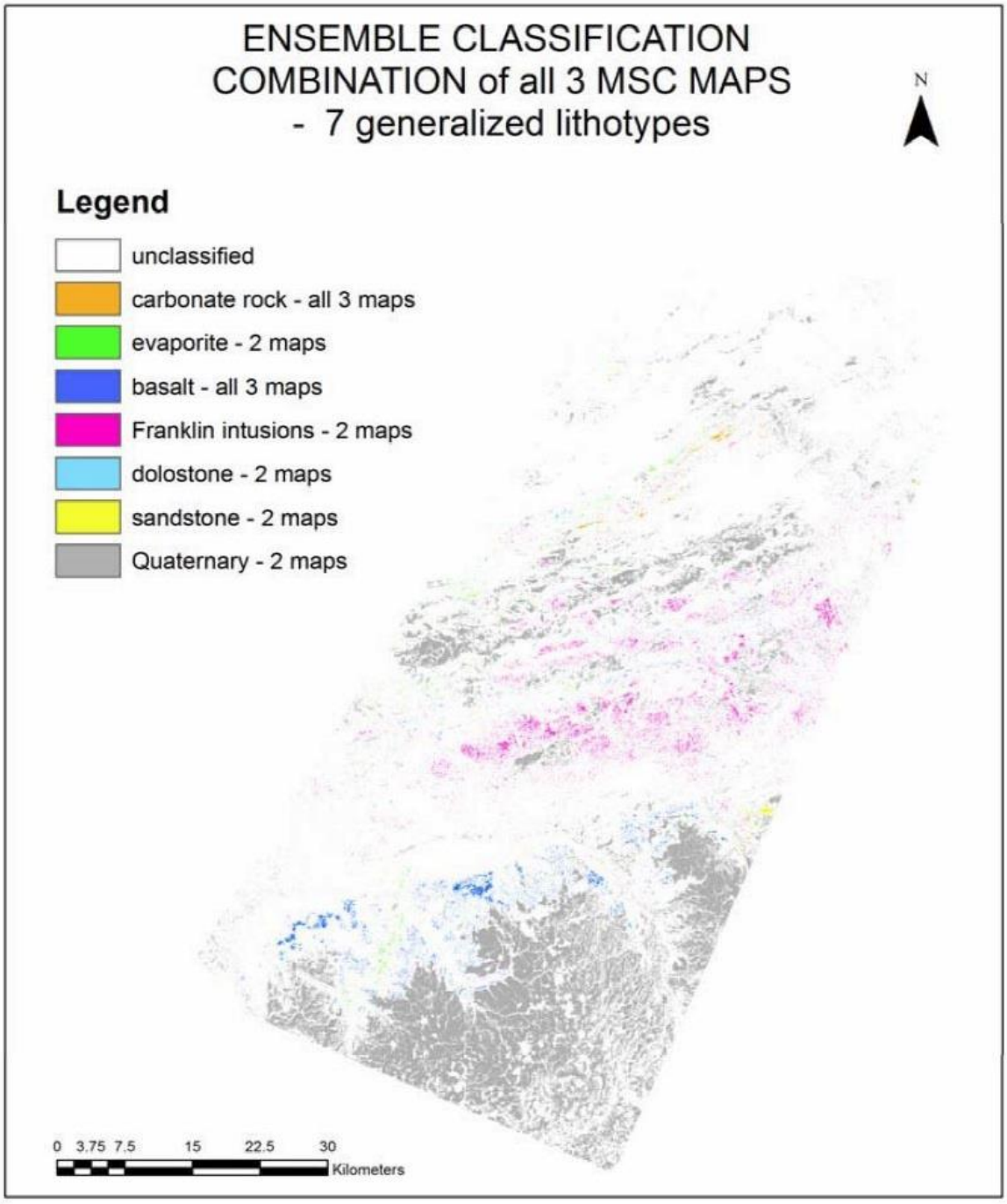

Figure 3.15: The Multiple Classification System (MCS) Combination Map combines MCS agreement (Figure 3.13a), best bands (Figure 3.14), and majority classification (Figure 3.12) maps.

3.5.5 Comparison to generalized geology map

A comparison of the generalized lithology (Figure 3.2b) with the generalized classification maps

(seven generalized lithological classes) generated by both RCM–MLC and RF was undertaken, as discussed in a previous section. The generalization was necessary, as spectral heterogeneity recorded on the more detailed classification maps (thirteen classes) occurred within each of the generalized lithological units shown on Figure 3.2b. Table 3.9 summarizes the classification agreement with the generalized lithology. The data combination that provided the best agreement when compared to the generalized geology map was achieved by Landsat-8 (raw, ratios and magnetics, 46.1%) followed by the Landsat-7 (raw, ratios and magnetics, 44.9%). The RF classification (Figure 3.10), using all 43 input bands, resulted in an agreement of 42.4% when compared to the generalized geology map even though it provided the highest classification accuracy of 99% using cross-validation (*oob*) of the training areas. It is clear that addition of the ratios and magnetics data has improved the overall accuracy when compared to the generalized lithology map.

Table 3.9: Summary of the overall classification accuracy for various data combinations compared to the generalized geology map (Figure 3.2b).

Data	**Accuracy**	**Kappa**
LANDSAT – raw bands	43.9	0.27
ASTER –raw bands and ratios	34.2	0.16
LANDSAT 7 - raw, ratios, magnetics	44.9 (1)	0.27
LANDSAT 7 - raw bands and ratios	43.9	0.27
LANDSAT 8 – raw bands	39.4	0.24
LANDSAT 8 - raw bands and ratios	42.4	0.25 (Figure 3.9a)
LANDSAT 8 - raw, ratios, mag	46.1 (1)	0.27
ENSEMBLE agreement map	69.8	0.557 (Figure 3.13a)
ENSEMBLE agreement map with uncertainty removed	71.6	0.5912 (Figure 3.13c)
ENSEMBLE RCM majority classification map	48.4	0.32 (Figure 3.12)
Ensemble best bands RCM classification	57.7	0.37 (Figure 3.14)
RF classification	42.4	0.25 (Figure 3.10a)
RF classification with uncertainty removed	71.8	0..6 (Figure 3.10b)

The *MCS best bands* classification (RCM–MLC; Figure 3.14) offers a higher accuracy (57.7%) than the RF and RCM–MLC classifications, indicating that from a MCS classification

perspective, a *best band* approach is a worthy undertaking. However, the *MCS agreement map* (Figure 3.13a) and agreement map with areas of uncertainty masked (Figure 3.13c) provided accuracies of 69.8% and 71.6%, respectively. The RF classification, when uncertainty was removed as discussed previously, also offered a high accuracy of 71.8%. However, the increasing accuracy is proportional to the degree of uncertainty. A low tolerance for uncertainty in the classification process results in fewer pixels being classified and, as one would expect, a higher overall classification accuracy.

3.5.6 Uncertainty analysis

The RCM–MLC uncertainty (variability) map (Figure 3.7b) produced for the data combination that provided the best classification (i.e. the highest overall accuracy) of the thirteen classes, which was the Landsat-7 (raw, ratios), was compared to the majority classification map (Figure 3.7a) to determine if there was a relationship between class and areas of high classification uncertainty. The classes with the highest uncertainty (class 7 on Figure 3.7b) included the sandstones and evaporites. The same comparison was made with the generalized lithological classes, in which Landsat-8 (raw, ratios) provided the highest accuracy (Figure 3.9a, b). The class that experienced the highest uncertainty was, once again, sandstone. This indicates that sandstone is characterized by the most spectral diversity and was not fully represented by the associated training areas. However, as mentioned above, the sandstones were separable using the Aster TIR data, because of their unique thermal response.

3.6 Discussion

This study has indicated that supervised classification techniques can play a useful role in the generation of bedrock and surficial maps of northern environments, providing that the rocks and

overburden demonstrate good spectral separation. These suitable environments tend to occur in Arctic Island and coastal regions where less lichen cover occurs because of favourable meso-climate conditions. The classified maps presented herein can be generated to accept more uncertainty but with large areas (i.e. more pixels) classified, producing what looks more like a traditional geologic map.

The advantage of semi-automatic classification techniques, in addition to their efficiency and speed, is that metrics that assess the uncertainty of the classification map can also be generated. However, as demonstrated here, a reduction in uncertainty results in fewer pixels being classified, albeit with greater overall confidence. These maps can be used as preliminary maps to aid in field mapping activities, as well as within a GIS environment, where they can guide more detailed visual interpretation to fill-in unclassified pixels.

This study shows that a new MCS classification approach using *majority classification*, *agreement* and *best band* maps can result in better agreement with a generalized lithological map (Figure 3.2b). It should be noted that the geology map used for comparison purposes was produced by using an integrated approach comprising the visual interpretation of stereo Spot-5 images and detailed field work.

Landsat-7 and Landsat-8 (raw and ratios) data combinations resulted in the highest classification accuracies using RCM–MLC. Landsat-8 offers an additional band in the blue portion of the electromagnetic spectrum and wavelengths similar to Landsat-7, except that the bandwidths are narrower. Both images were acquired in July, which provides optimum conditions because of higher sun elevation angles. The higher radiometric resolution of Landsat-8 (16-bit) did not offer any significant advantages with respect to classification accuracy compared to Landsat-7 (8-bit). It is comforting to confirm the good performance of Landsat-8,

which will provide uninterrupted world coverage once Landsat-7, the workhorse of remote sensing applications for many years, expires.

The poor performance of Aster is no doubt related to image acquisition in mid-September, when the sun angle was only 20°, resulting in a lower signal-to-noise ratio and an increased topographic effect such as shadowing and low reflectance from backslopes. The two Aster scenes were corrected and mosaicked, and although care was taken to radiometrically balance the scenes, these factors may also negatively affect the classification results. As such, the comparison with Landsat may not be made on a level playing field; i.e. better results may have been obtained from the Aster data if acquisition had occurred in July. The VNIR and SWIR Aster bands provided better results than the TIR bands, except for sandstone, the thermal properties of which were much different than the surrounding rocks. None of the TIR bands were selected by the RF classifier as being strongly predictive (Table 3.7). Thermal data are better collected during the night to suppress solar heating effects; however, northern environments, although dark for six months of the year, are completely snow-covered during these winter months, suppressing thermal signatures from the underlying rocks.

The Spot-5 data also provided lower classification results than the Landsat data. This may also be related to the lower sun angle on the date of acquisition (22 August), and also because of fewer spectral bands. Behnia et al. (2012), in another remote study of Victoria Island, note that Landsat-7 data using RCM–MLC provided a higher classification accuracy than Spot-5. They attributed this result to the higher spectral resolution of Landsat, which records reflectance in the blue and SWIR portions of the electromagnetic spectrum. They also concluded that the lower spatial resolution of Landsat actually acts as a natural filter by generalizing the surface and presenting less spectral variability than the higher spatial resolution of Spot-5 data.

The addition of ratios for all sensors (Landsat-7, Landsat-8, and Aster) improved overall classification accuracies for both the (thirteen) detailed and (seven) generalized lithological lithological classes. Ratios are advantageous to the classification procedure as they help to reduce illumination effects related to topography and, more importantly, emphasize spectral differences associated with individual minerals and groups of minerals that help to identify various litho-types. In fact, the best predictor, established through the RF classifier, was a Landsat-7 band 5/4 ratio (Table 3.7) designed to detect rocks with a high ferrous iron content.

Inclusion of magnetic data (total field or converted to magnetic susceptibility) in the classification process is also recommended, as it can help to identify rocks characterized by either high or low magnetic susceptibilities, as demonstrated in this study and by Martel et al. (2005), Harris et al. (2008b, c), and Schetselaar and Ryan (2008). Gamma ray spectrometry data, although not used in this study, should also be included in the classification process as it records information on the content of uranium, thorium and potassium in rocks at the earth's surface. These elements are present in most rock-forming minerals: potassium is a common constituent in feldspar, biotite, and muscovite, whereas uranium and thorium generally occur in low concentrations in a wide range of accessory minerals. Unfortunately, no gamma ray data were available for the study area; however, Harris et al. (2009) and Ford et al. (2008) have shown that these data are extremely useful in the classification of bedrock types.

The selection of representative training areas is obviously a key factor in the application of machine learning algorithms such as classification to the mapping of bedrock or surficial materials. Even within classes, spectral variability may be present. Ensemble classification algorithms such as RCM–MLC and RF are used to account for variability in training area signatures by employing *bagging* and cross-validation methods (see Harris et al. 2012b). This

can assist in identifying training areas that can be modified and/or deleted from the classification process. Furthermore, these classification algorithms provide a more robust estimate of classification accuracies and generate uncertainty metrics to accompany the classified map.

Comparing the RCM–MLC and RF algorithms is difficult because different datasets were used: for RF, all 43 bands were employed, whereas RCM–MLC used various subsets of the 43 bands. Use of MLC with a restricted number of training areas would result in the *Hughes effect* (Hughes 1968; Oommen et al. 2008; Alonso et al. 2011), which describes how classification accuracy depends on the number of training samples. In general, the separability of classes (as well as the number of statistical parameters defining the classes) increases with the number of dimensions (bands). In classification, a fixed number of training samples, generally proportional to the number of bands to be classified, are used, and at some point the classification accuracy can decrease if the number of training samples is insufficient. High dimensional data are generally more difficult to work with, as many bands can increase the noise factor and hence error factor, requiring many training samples to be used. Story and Congalton (1986) suggested that at least 50 pixels per map class should be used in the training process, whereas Jensen (2005) suggested 10*n* pixels for each class, where *n* equals the number of bands (channels) used in the classification. Other researchers suggest 10 to 30*n* pixels per class (Mather 1999; Piper 1999). Inclusion of the total number of input bands (43) for the RCM–MLC classification could lead to an overestimation of accuracy because of the limited number of training areas used in this study (see Table 3.1). However, in using RCM–MLC, the *Hughes effect* can be reduced because the iterative random sampling of the training dataset (*bagging*) produces a series of classifications as well as independent cross-validations (see above and Harris et al. 2012b). RF is also less affected by the *Hughes effect* because the *bagging* process is used to create, in this case, 100 trees

(classifications) and an equal number of validation estimates. Furthermore, RF has been shown, in a number of studies, to outperform other classifiers (Gislason et al. 2006; Crisci et al. 2012; Craknell and Reading 2014). In addition, RF provides a predictive rank for each input band and is a non-parametric classifier. In this study, the highest overall classification accuracy (99%) was achieved using RF, bearing in mind that all 43 bands (raw and ratios) derived from the Landsat, Aster and Spot data were used in the classification experiments.

In northern terranes where remotely sensed data are often acquired at lower sun angles, such as the Aster data used herein, it is recommended to either topographically correct remotely sensed data or, in the present case, to account for illumination effects associated with topography. This was accomplished simply by incorporating a 1:50,000 DEM (Canadian Digital Elevation Data) in which hillshade maps using the sun azimuth and elevation values were determined from the acquisition parameters of each sensor. The hillshade maps were then thresholded to mask areas in shadow and backslope areas that reflect less energy, both of which can add an element of confusion to the classification process.

Comparison of a classified map derived from even moderate-resolution remotely sensed data with a traditional geologic map is problematic as the two are constructed using entirely different processes. Although the geology map used here for comparison to the classified maps did incorporate a component of remote sensing (visual interpretation of stereo Spot-5 imagery), the map was also based on detailed field work. Furthermore, the cognitive processes involved in traditional map-making, which involves the incorporation of process information and generalization based on field experience, is more difficult to capture by machine-learning algorithms. Remote sensors collect data on a grid cell basis (30*m* for Landsat), in which much

spectral heterogeneity is recorded. A traditional geological map is often a generalized model of a geological environment in that it presents more homogeneous classes. The fact that remotely sensed data records more heterogeneous information does not render the classified map useless; in fact, just the opposite occurs, as indicated in this study. The more detailed classification (thirteen classes) produced a map with spectral sub-units within the main lithological groups. To create a level playing field for the comparison, we generalized both the geological map and the training areas to seven basic lithological–surficial classes. The majority of the classification maps showed some agreement (40 to 50%) with the generalized geology map. However, an ensemble MCS classification approach, here employing *majority agreement* and *best band* maps, resulted in higher classification accuracies in excess of 70% when compared to the geology map. The trade-off is that fewer pixels are classified, but with more confidence. Furthermore, inclusion of uncertainty metrics (i.e. RCM–MLC variability, RF probability and ensemble range maps) increases overall accuracies, again at the expense of the number of pixels classified. The geologist has the option of using classification maps in which all pixels are classified other than those that have been purposely masked, or else using a map that has had fewer pixels classified, but with less spectral uncertainty.

3.7 Conclusions

This study has demonstrated that classification of remotely sensed optical data can significantly contribute to the geologic mapping process in a geological environment in which the spectral expression of rocks is optimal. These environments tend to occur in Arctic islands and coastal environments in Canada's North. A map in which all non-masked pixels are classified or a map with fewer pixels classified, but with more spectrally certainty, can be produced. The maps can provide first order geologic information in poorly mapped or unmapped areas, as well as

preliminary maps to support and focus field-based studies. For example, a classified map produced from an ensemble classification in which very uncertain areas have been removed(e.g. Figure 3.13c) could form the basis of a more detailed visual analysis of various remotely sensed data (2D and 3D) in a GIS environment using a *heads-up* interpretation process and touch-screen technology (see Harris et al. 2012a).

Landsat-8 data produced classification results similar to those of Landsat-7, which will ensure the reliable and continuous use of the Landsat series of satellites for geological mapping into the future. The performance of Aster data was limited by the season of acquisition (low sun angle), indicating that for Arctic research, mid-June to mid-August data acquisition is best for optimal results. It is expected that Aster would have been more useful if a July acquisition date could have been found. Certainly, the Aster imagery did classify certain litho-types more accurately than did Landsat-7 or Landsat-8, although the overall performance was inferior. Aster TIR data may see only limited applications in northern environments because of snow cover and day-time solar heating biases. Spot-5 data provides higher spatial resolution, which is very useful for detailed visual interpretation, but has lower spectral resolution, lacking reflectance information in the blue and SWIR wavelengths.

For geological classification, both raw optical data and calculated ratios should be used to directly focus on spectral differences related to mineralogy, primarily in iron-, clay-, and carbonate-bearing rocks. Magnetic and gamma ray spectrometry data, depending on the geological environment, should also be incorporated.

Both RCM–MLC and RF produced high classification accuracies. The RF algorithm may be preferred, as other studies indicate that it marginally outperforms other classification algorithms by 2–10% with respect to classification accuracies. However, more research is

required to confirm this. Furthermore, RF does not require normal distribution of the training data to be classified, provides a ranked list of the best predictors, is less sensitive to the *Hughes effect*, and does not overfit the data.

A new MCS approach, using a *majority classification agreement* and *best bands* map derived from different remotely sensed data types, is recommended for other geologic classification studies in northern environments. This requires more data acquisition, but as shown in this study, results in higher classification accuracies when compared to the new geology map. This MCS classification approach could be useful for surficial mapping applications using multiple image acquisitions (time series) by the same or different sensors over the northern summer (mid-June to end of July). This time-series approach may prove useful for capturing biomass and soil moisture differences that may be helpful in the mapping of surficial materials.

3.8 Acknowledgements

The authors would like to thank the Geological Survey of Canada under the Remote Predictive Mapping Project (RPM), part of the Geo-mapping for Energy and Minerals (GEM) program, for supporting and funding this research. We would also like to thank two anonymous reviewers whose suggestions and comments greatly improved this manuscript. This is Geological Survey of Canada contribution number 20130439.

Chapter 4 A Systematic Investigation of Geostatistical Image Fusion for the Improvement of the Spectral Fidelity and Spatial Detail in Landsat MS Imagery

Abstract: This is the first systematic investigation into the assumptions of image fusion using regression kriging (RK) -a geostatistical method – illustrated with Landsat MS and SPOT panchromatic images. The efficiency of different linear regression and kriging methods in the fusion process is examined by visual and quantitative indicators. Results indicate a trade-off between spectral fidelity and spatial detail preservation for the GLS (generalized least squares regression) and OLS (ordinary least squares regression) methods in the RK process: OLS methods preserve more spatial detail, while GLS methods retain more spectral information from the MS images but at greater computational cost. Under either OLS or GLS, UK (universal kriging) fusion generates smoother (less noisy) results, but OK (ordinary kriging) preserves more spatial details. Under either OK or UK with either OLS or GLS, the spherical variogram improves spatial details from the panchromatic image, while the exponential variogram maintains more spectral information from the MS image. Overall, RK-based fusion methods outperform conventional fusion approaches from both the spectral and spatial point of view.

Keywords: geostatistical, regression kriging, data fusion

4.1 Introduction

Ideally, remotely sensed Imagery with high spectral and spatial resolution can provide good spectral discrimination and more resolved geometric information and thus improve land cover classification and mapping. However, there is often a trade-off between spectral and spatial resolution. For example, existing multispectral (MS) imagery (e.g. Landsat) has a number of electromagnetic bands but relatively coarse spatial resolution; on the other hand, high spatial resolution panchromatic (PAN) imagery tends to depict object geometry in more detail but contains less spectral information. Thus, data fusion techniques have been developed to merge

high-spatial-resolution PAN images and high-spectral-resolution MS images to create a single image that can improve feature identification, extraction and classification.

Image fusion has gained intensive attention in the remote sensing community since the launch of SPOT 1 in 1986 (Zhang 2004). Most research has focused on the fusion of a high-spatial-resolution PAN imagery with high-spectral-resolution MS images. This fusion process is called image pan-sharpening. A pan-sharpened MS image has a higher spatial resolution while ideally preserving the spectral fidelity of the MS inputs, thus aiding the discrimination between different land covers or features.

Various image fusion methods have been reviewed (Dong et al. 2009; Pohl and Van Genderen 1998; Zhang 2010). Pohl and Van Genderen (1998) placed image pan-sharpening approaches into two classes: (1) color related techniques (e.g. IHS (Intensity, Hue, Saturation) and HSV (Hue, Saturation, Value) techniques) and, (2) statistical/numerical methods (e.g. Brovey, Principal Components Analysis and wavelet-based approaches). In remote sensing applications, the most widely used approaches include IHS, Brovey, PCA, arithmetic combinations, and wavelet-based fusion techniques (Švab and Oštir 2006; Zhang 2008a). However, those conventional approaches suffer color distortion because the radiometry of the input MS image is modified in the fused outputs. To reduce color distortion, various adjustment approaches have been investigated to achieve satisfactory fusion results. Some successful examples include a new IHS fusion method proposed by Choi (2006), discrete wavelet transformation (e.g. Pei et al. 2010), and an improved high-pass filter additive technique (Gangkofner et al. 2007). However, with satellites launched since 1999, such as IKONOS, QuickBird, and Landsat 7, fusion induced radiometric distortion has been a significant problem when using traditional and adjusted fusion techniques (Zhang 2004; Nikolakopoulos 2008). The

significant color distortion is mainly caused by the spectral range extension within newer satellite panchromatic images (Zhang 2004).

Many studies have focused on investigating new techniques to reduce fusion induced color distortion. One such effort is the application of geostatistics within image fusion. For example, Petrie et al. (2002) used inverse kriging to fuse IKONOS MS bands with an IKONOS panchromatic band. Memarsadeghi et al. (2005) and Pardo- Igúzquiza et al. (2006) implemented image fusion based on cokriging. Meng et al. (2010) employed regression-kriging to fuse IKONOS and QuickBird images. Unlike the conventional fusion and adjustment techniques, geostatistical fusion methods integrate spatial correlation between high spatial resolution images and low spatial resolution MS images in the fusion process. Geostatistical methods can help overcome issues of conventional fusion methods, such as color distortion and limitations on the number of bands that can be fused. In addition, Geostatistics based image fusion is flexible and can incorporate ancillary data sets (e.g. topographic maps and thematic maps) into the fusion procedure.

In this study, the capability of geostatistical fusion techniques to merge medium-resolution Landsat MS images and SPOT-5 panchromatic images is investigated systematically. Unlike previous research on geostatistical fusion methods that focused on merging high spatial resolution images like IKONOS and QuickBird, this research examines the ability of geostatistical fusion to merge free medium-resolution Landsat MS images and SPOT-5 PAN images. The effect of using various regression and kriging approaches, such as ordinary kriging (OK) vs. universal kriging (UK) and ordinary linear squares regression (OLS) vs. generalized linear squares regression (GLS) is detailed.

4.2 Methods

Geostatistics examines not only the statistical distribution of a set of sampled data but also the spatial correlation among the sampled data. When kriging, the value of a random variable at an unobserved location is interpolated from nearby observations based on the spatial correlation structure present in the observed data. Regression kriging (RK), an example of combined interpolation, involves a linear regression with residuals being adjusted by kriging and finally the adjusted residuals are added to the linear estimate to form the final result (Hengl, 2009). This fusion method explicitly separates the regression process (i.e., trend estimate) from residual interpolation, allowing the use of other forms of regression (e.g. random forest) rather than the simple linear regression technique (Hengl et al. 2007). Meanwhile, the combined fusion process also allows various forms of kriging, like ordinary kriging and universal kriging, to be adopted in the interpolation of regression residuals. In addition, RK-based fusion methods are flexible because ancillary data like soil texture data or a DEM can be introduced into the process as independent variables (e.g. Odeh et al. 1995; Hengl et al. 2007).

4.2.1 Study area and data

The study area is located in Victor Mine, Ontario, Canada (centered at latitude 52° 49’ 15” and longitude -83° 53’ 00”), shown in Figure 4.1. The main land covers within the study area include built-up areas, water bodies, wetlands and vegetated terrain. Some vegetated areas are covered by coniferous or broadleaf trees, and some areas are covered by sparse woods. Landsat-5 Thematic Mapper (TM) MS and SPOT-5 PAN images were acquired in this region. The Landsat-5 TM imagery (30 m spatial resolution, path 23, row 23) was downloaded from www.usgs.gov and the SPOT-5 orthorectified panchromatic image (10 m spatial resolution) was

downloaded from Canadian Council on Geomatics website: www.geobase.ca. The SPOT panchromatic image was acquired on 11 September 2006 with a sun elevation angle of 41.296 degrees. The Landsat image was acquired on 1 September 2006 with a sun elevation angle of 42.877 degrees.

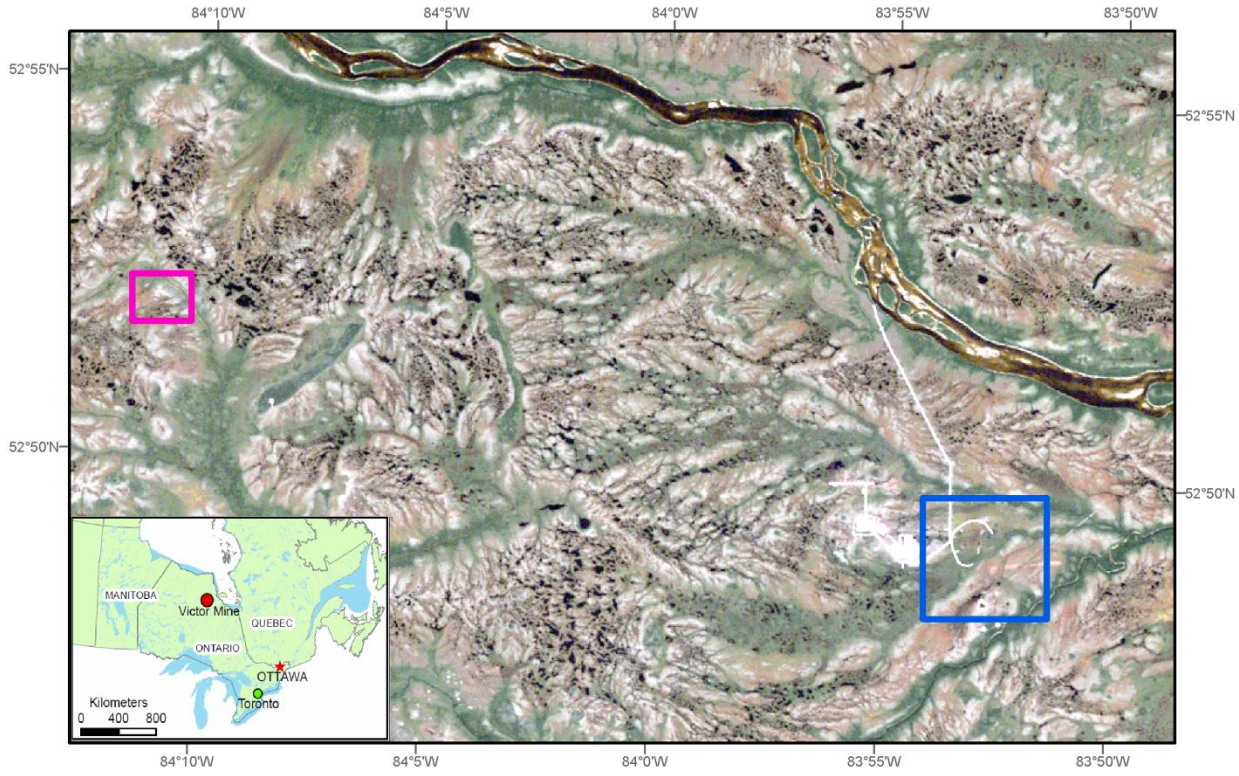

Figure 4.1: The true color composition of the Landsat-5 image shows the study area that is located in Victor Mine, Ontario (red dot in the lower-left map). The purple rectangle shows the location of the small clip within the study area used to investigate which fusion methods are more efficient, and the blue rectangle shows the location of the large clip within the study area where the pixel values are skewed.

Before implementing geostatistical image fusion, the Landsat image was resampled to 10 meters to match the SPOT image. The nearest neighbor method was used in the resampling process in order to retain the original pixel values. The Landsat MS image and the SPOT PAN image were then co-registered to the Universal Transverse Mercator coordinate system (zone

17N) with NAD83 datum and the co-registration error (RMSE: Root Mean Square Error) was less than half of the SPOT PAN image pixel size.

4.2.2 Fusion methods

Regression kriging (Odeh et al. 1994) is a combined prediction method involving various combinations of linear regression and kriging (Li and Heap 2008). That is, the fusion result is the addition of a fitted stochastic component computed by applying a linear regression method and then kriging the residuals. The most common method for fitting a regression line is the least squares approach, which can be further divided into OLS and GLS. OLS yields the maximum likelihood estimates for the regression coefficients if the residuals are normally distributed. GLS is a modified version of OLS that can be applied when the observations are correlated, whereas OLS assumes that the observations are independent. As such, when spatial correlation between observations exists, then GLS will outperform OLS (Cressie 1993; Webster and Oliver 2007; Hengl 2009). Theoretically, when GLS is applied in a RK procedure, the regression process starts with an OLS regression and follows several iterations of computing GLS coefficients and updating the residuals(Schabenberger and Gotway 2004; Hengl et al. 2007). However, Kitanidis (1993) and Hengl (2007) suggested that empirically, a single iteration of computing GLS coefficients based on OLS residuals is often acceptable because more iterations will not significantly improve the kriging interpolation results. Therefore, in this research, one iteration regression is applied: OLS is first used to compute the deterministic part of the regression, and then the regression model is updated using GLS whose coefficients are computed based on the OLS residuals. For the GLS regression, four different observation correlation structures including lag-1 autocorrelation (AR1 in Table 4.1), Gaussian, spherical, and exponential structures are tested in order to determine which correlation structure could provide the optimal

fusion result. In addition, even though the data in the study are spatially correlated, OLS is still tested in the fusion process where the spatial correlation of the observations is disregarded when conducting regression, due to the substantially smaller computational costs for OLS than GLS. Another intention of using OLS is to test if kriging of the residuals from OLS regression could generate comparable fusion results to those derived from the GLS based RK fusion methods.

Various kriging methods, such as OK and UK can be used to interpolate regression residuals. The difference between OK and UK is the assumptions on the trend component of the predictors. OK assumes the mean of the random field is an unknown parameter that is constant in the local neighborhood of each estimation point. Much like OK, UK also assumes the mean of the random field is unknown, but fits a linear or higher-order trend to estimate the spatially varying value with different coefficients for each point applied when fitting a trend. Both OK and UK are tested in this research in order to investigate which kriging method is more appropriate from the perspective of fused image quality and computational costs.

The definition of the variogram functional form is critical to kriging. The empirical variogram is first examined and then a theoretical variogram model is chosen to fit the empirical variogram. The fitted variogram model is then used as the distance weighting function within the interpolation process. There are three commonly used transitive variogram models which include spherical, exponential, and Gaussian equations. Both spherical and exponential models are characterized by linearity near the origin while the Gaussian model is parabolic (Isaaks and Srivastava 1989). In this research both the exponential and spherical variogram models are used but the Gaussian model is not applied because the sample semi-variogram does not present parabolic behavior near the origin.

Kriging can operate locally or globally. In local kriging, the pixels within a local window centered on the location to be estimated are used in the interpolation process; whereas, in global kriging all pixels within the study area are used in the estimate. Global kriging is more computationally expensive. In this research, global kriging is examined as well as local kriging with different sizes of local windows (e.g. the range size window and user-specified windows).

The performance of various combinations of linear regression methods and kriging methods (listed in Table 4.1) is first investigated on a small image chip (the purple rectangle in Figure 4.1, 47 columns and 40 rows of the MS bands) within the study area where the pixel values are normally distributed. The fusion results are evaluated both visually and quantitatively. In order to evaluate the degree to which the methods can be generalized, the methods performing best on the image chip are then used to fuse a larger scene (103 columns and 99 rows of MS bands shown by the blue rectangle in Figure 4.1) in which the MS and PAN pixels are non-normally distributed. In addition, for comparison with other conventional fusion methods, HSV (Chiesa and Tyler 1990), Brovey (Vrabel 1996), Gram-Schmidt (GS, Laben and Brower 2000), discrete wavelet (Mallat 1989; Ranchin and Wald 2000), and A Trous wavelet transforms (Nunes et al. 1999), are also applied in the study. The fusion results from the different RK methods listed in Table 4.1 are compared to those generated by these conventional fusion approaches.

Table 4.1: The list of RK fusion methods with various combinations of linear regression methods and kriging methods.

	Regression		**Kriging**		
Name	**Regression Method**	**Correlation Structure**	**Kriging**	**Variogram model**	**Searching Window**
UK_gls_AR1_rvgm_Exp_globe	GLS	AR1	UK	Exponential	Global
UK_gls_AR1_rvgm_Exp_range	GLS	AR1	UK	Exponential	Range

UK_gls_AR1_rvgm_Sph_globe	GLS	AR1	UK	Spherical	Global
UK_gls_AR1_rvgm_Sph_range	GLS	AR1	UK	Spherical	Range
UK_gls_Exp_rvgm_Exp_globe	GLS	Exponential	UK	Exponential	Global
UK_gls_Exp_rvgm_Exp_range	GLS	Exponential	UK	Exponential	Range
UK_gls_Exp_rvgm_Sph_globe	GLS	Exponential	UK	Spherical	Global
UK_gls_Exp_rvgm_Sph_range	GLS	Exponential	UK	Spherical	Range
UK_gls_Gaus_rvgm_Exp_globe	GLS	Gaussian	UK	Exponential	Global
UK_gls_Gaus_rvgm_Exp_range	GLS	Gaussian	UK	Exponential	Range
UK_gls_Gaus_rvgm_Sph_globe	GLS	Gaussian	UK	Spherical	Global
UK_gls_Gaus_rvgm_Sph_range	GLS	Gaussian	UK	Spherical	Range
UK_gls_Sph_rvgm_Exp_globe	GLS	Spherical	UK	Exponential	Global
UK_gls_Sph_rvgm_Exp_range	GLS	Spherical	UK	Exponential	Range
UK_gls_Sph_rvgm_Sph_globe	GLS	Spherical	UK	Spherical	Global
UK_gls_Sph_rvgm_Sph_range	GLS	Spherical	UK	Spherical	Range
UK_ols_rvgm_Exp_globe	OLS	None	UK	Exponential	global
UK_ols_rvgm_Exp_range	OLS	None	UK	Exponential	Range
UK_ols_rvgm_Sph_globe	OLS	None	UK	Spherical	Global
UK_ols_rvgm_Sph_range	OLS	None	UK	Spherical	Range
OK_ols_rvgm_Exp	OLS	None	OK	Exponential	Local
OK_ols_rvgm_Sph	OLS	None	OK	Spherical	Local
OK_gls_AR1_ rvgm _Exp	GLS	AR1	OK	Exponential	Local
OK_ gls_AR1_ rvgm _Sph	GLS	AR1	OK	Spherical	Local
OK_ gls _Gaus_ rvgm _Exp	GLS	Gaussian	OK	Exponential	Local
OK_ gls _Gaus_ rvgm _Sph	GLS	Gaussian	OK	Spherical	Local
OK_ gls _Exp_ rvgm _Exp	GLS	Exponential	OK	Exponential	Local
OK_ gls _Exp_ rvgm _Sph	GLS	Exponential	OK	Spherical	Local
OK_ gls _Sph_ rvgm _Exp	GLS	Spherical	OK	Exponential	Local
OK_ gls _Sph_ rvgm _Sph	GLS	Spherical	OK	Spherical	Local
Cokriging			Co-kriging		

4.3 Evaluation of the Fusion Results in the Small Scene.

Evaluation of fused results can be achieved qualitatively and quantitatively. Qualitative approaches include visual comparison of the color between original MS images and fused images, and examination of the spatial details between original PAN and fused images. Visual comparison provides an overall impression in terms of fused image quality when compared to the original MS and/or PAN image. However, visual comparison is inevitably subjective. Conversely, more objective quantitative approaches include the application of a set of indicators to measure spectral and spatial similarities between original MSS and PAN images and fused result.

4.3.1 Quality indicators applied in the research

Various indicators for spectral similarity and spatial resolution improvement can be drawn from existing literature (Zhang 2008b; Klonus and Ehlers 2009; Jagalingam and Hegde 2015). In this research, three commonly-used spectral similarity indicators are used to quantitatively evaluate the similarity between the original MS image and the fused imagery: the Correlation Coefficient (CC), Spectral Discrepancy (SPD), and ERGAS (Relative Non-dimensional Global Synthesis Error)**.** CC measures the correlation between the original MS and the fused results. Higher CC values indicate that the fused result more strongly co-varies with the original MS image and the ideal value of CC is 1 (Gonzalez-Audicana et al. 2004; Guo and Liu 2011). SPD measures the spectral difference between the original MS image and the fused image and, as such, quantitatively assesses the level of spectral fidelity (Guo and Liu 2011). The ideal value of SPD is zero, which means there is no loss/modification of spectral information in the fused result. ERGAS is a global dimensionless quality indicator of spectral distortion that is independent of

the number of spectral bands and the resolution of the input image (Wald 2002). The ideal value for ERGAS is 0 (Gonzalez-Audicana et al. 2004)

To evaluate the spatial resolution improvement of fused imagery, the high frequency data from the fused images is compared to that from the SPOT PAN image using a method proposed by Zhou et al (1998). A high-pass Laplacian filter is firstly applied to both the fused image and the SPOT PAN image to extract high frequency data. The correlation coefficients between the high-pass filtered fusion result and the high-pass filtered PAN image (HPCC) are then used as an index of the spatial similarity quality. A higher HPCC indicates that more spatial information from the PAN image has been injected into the fusion result. The ideal value of HPCC is 1.

To achieve an overall quantitative index to compare fusion results from both spectral and spatial aspects, the Structure Similarity (SSIM) index (Robertson et al. 2014) is applied in the study. The SSIM index consists of three independent components: Luminance (L), Contrast (C), and Structure (S). The L and C components are computed by comparing the fusion results to the Landsat-5 MS image and the S component is computed by comparing the fusion results to the SPOT PAN image. The final SSIM index for each fusion method is then derived by multiplying the L, C, and S components, giving an equal weight to each of them. The kernel size used to compute the SSIM index is 5 by 5 pixels.

4.3.2 Visual comparison

The evaluation is conducted on the fusion results created by employing all the methods listed in Table 4.1 on the Landsat-5 MS bands 2, 3, 4 (shown in Figure 4.2*al*) and SPOT-5 panchromatic image (shown in Figure 4.2*ak*). Three bands are chosen so that comparison could be made with HSV and Brovey pan-sharpening. Figure 4.2 shows the fusion results for the small image chip (for the order of the images please see the column called Index in Table 4.2). In general, RK-

based fusion preserves more spectral information from the original Landsat MS image when compared to conventional fusion methods. In contrast to OLS-based fusion, the GLS methods yield smoother results. Among the GLS-based fusion methods, spatial and spectral characteristics are similar, despite the different correlation structures, variogram models and kriging methods (either UK or OK) applied. In addition, there are negligible differences between global and local universal kriging when the same regression method is applied. Among the OLS-based fusion results, OK exhibits more color distortion than the UK kriging, but the former presents more resolved spatial information. In terms of the conventional fusion methods, HSV and Bovey exhibit the greatest amount of color distortion. With the exception of the discrete wavelet transform, all the conventional fusion methods retain spatial information.

4.3.3 Quantitative indicator computation and cluster analysis

Table 4.2 presents the quantitative evaluation indicators (CC, SPD, ERGAS, HPCC, and SSIM) derived from the fusion results shown in Figure 4.2. The average CC, SPD, HPCC, and SSIM for bands 2, 3, and 4 are listed in Table 4.2. To better understand the differences among the methods, a cluster analysis is conducted on the quantitative indicator values. The ward clustering criterion is implemented as the agglomeration method to create hierarchical clustering (Ward 1963; Caliński and Harabasz 1974). Figure 4.3 shows the dendrogram of the hierarchical clustering, where all the fusion methods are joined together in a hierarchical fashion from the closest (i.e. most similar, smaller height values) to the furthest apart (i.e. most dissimilar, larger height values) using the Euclidean distance among the quantitative indicator values of individual fusion methods.

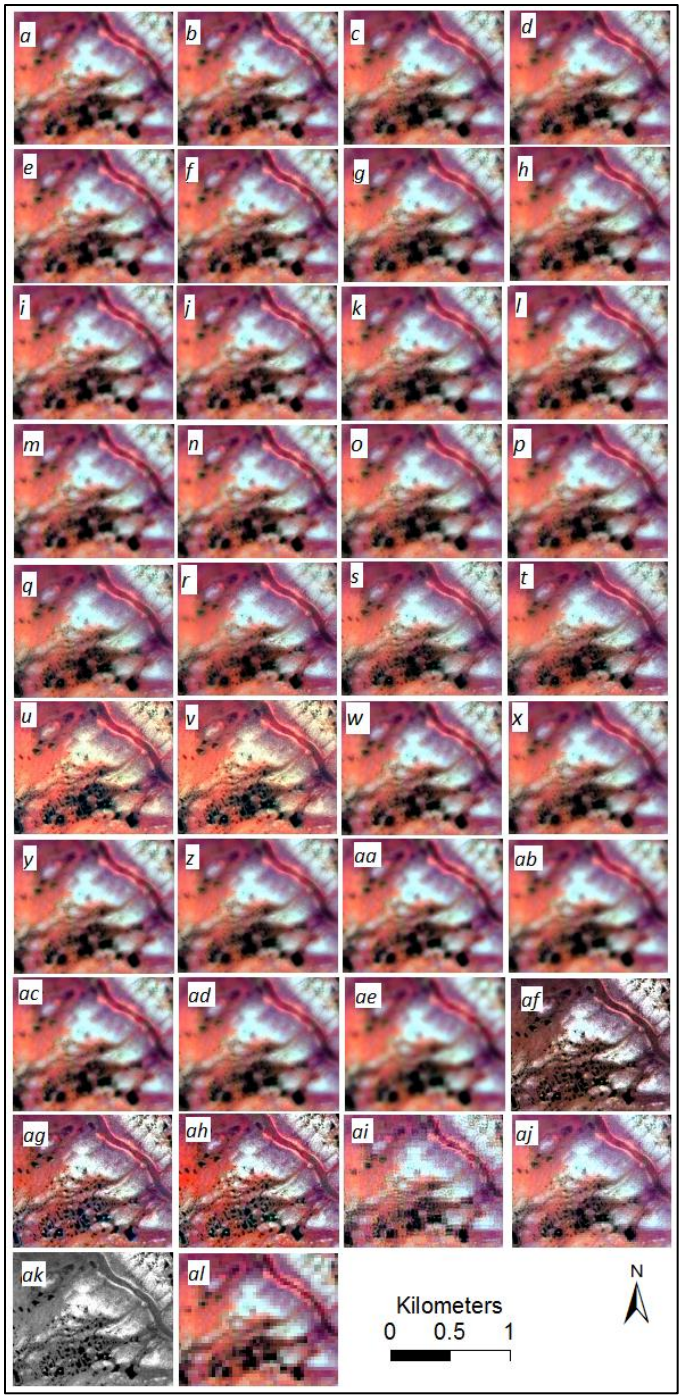

Figure 4.2: The fusion results generated by applying the RK-based fusion methods listed in Table 4.1 are presented by way of false color composition of the fused images (i.e. Band 4-Red, Band 3-Green, Band 2-Blue). The indexes of the figures are listed in the column called Index in Table 4.2. The original SPOT panchromatic and Landsat MS bands 2, 3, 4 (10 m resolution) are shown in *ak* and *al*, respectively.

Table 4.2: Quantitative evaluation indicator values for the fusion results shown in Figure 4.2. The mean values of the CC, SPD, HPCC, and SSIM for Bands 2, 3, 4 are listed.

	Index	CC	SPD	ERGAS	HPCC	SSIM
UK_gls_AR1_rvgm_Exp_globe	*a*	0.963	1.2843	1.0318	0.8707	0.7933
UK_gls_AR1_rvgm_Exp_range	*b*	0.9611	1.2935	1.0586	0.8133	0.7906
UK_gls_AR1_rvgm_Sph_globe	*c*	0.9519	1.5105	1.1686	0.9651	0.8319
UK_gls_AR1_rvgm_Sph_range	*d*	0.9526	1.489	1.1604	0.9388	0.8315
UK_gls_Exp_rvgm_Exp_globe	*e*	0.9634	1.2768	1.0279	0.8626	0.7907
UK_gls_Exp_rvgm_Exp_range	*f*	0.9615	1.2863	1.0545	0.8092	0.7883
UK_gls_Exp_rvgm_Sph_globe	*g*	0.9532	1.486	1.1537	0.9596	0.8283
UK_gls_Exp_rvgm_Sph_range	*h*	0.9538	1.4659	1.1461	0.9339	0.8279
UK_gls_Gaus_rvgm_Exp_globe	*i*	0.9634	1.276	1.0274	0.862	0.7904
UK_gls_Gaus_rvgm_Exp_range	*j*	0.9617	1.2839	1.0517	0.8111	0.7881
UK_gls_Gaus_rvgm_Sph_globe	*k*	0.9549	1.4541	1.1348	0.9507	0.8235
UK_gls_Gaus_rvgm_Sph_range	*l*	0.9554	1.4367	1.1283	0.9269	0.8227
UK_gls_Sph_rvgm_Exp_globe	*m*	0.9634	1.2767	1.0279	0.8625	0.7906
UK_gls_Sph_rvgm_Exp_range	*n*	0.9612	1.2935	1.0622	0.8047	0.7889
UK_gls_Sph_rvgm_Sph_globe	*o*	0.9533	1.4844	1.1527	0.9593	0.8281
UK_gls_Sph_rvgm_Sph_range	*p*	0.9539	1.4646	1.1452	0.9336	0.8276
UK_ols_rvgm_Exp_globe	*q*	0.9395	1.6913	1.3012	0.9704	0.8575
UK_ols_rvgm_Exp_range	*r*	0.9386	1.6725	1.3109	0.8803	0.8489
UK_ols_rvgm_Sph_globe	*s*	0.9252	1.9047	1.4461	0.9906	0.8693
UK_ols_rvgm_Sph_range	*t*	0.9273	1.8609	1.424	0.9595	0.8701
OK_ols_rvgm_Exp	*u*	0.903	2.156	1.6816	0.991	0.8756
OK_ols_rvgm_Sph	*v*	0.8944	2.259	1.7488	0.996	0.8818
OK_gls_AR1_rvgm_Exp	*w*	0.9604	1.3351	1.0961	0.7889	0.8074
OK_gls_AR1_rvgm_Sph	*x*	0.9531	1.4925	1.1920	0.7952	0.8278
OK_gls_Exp_rvgm_Exp	*y*	0.9633	1.2794	1.0581	0.7817	0.7913
OK_gls_Exp_rvgm_Sph	*z*	0.9565	1.4350	1.1545	0.7850	0.8060
OK_gls_Gaus_rvgm_Exp	*aa*	0.9676	1.1886	0.9958	0.7634	0.7413
OK_gls_Gaus_rvgm_Sph	*ab*	0.9611	1.3521	1.1016	0.7589	0.7368
OK_gls_Sph_rvgm_Exp	*ac*	0.9635	1.2757	1.0555	0.7812	0.7901
OK_gls_Sph_rvgm_Sph	*ad*	0.9567	1.4311	1.1519	0.7842	0.8043
Cokriging	*ae*	0.9578	1.4033	1.1138	0.1547	0.6538
HSV	*af*	0.7617	50.9333	37.2142	0.935	0.3066
Brovey	*ag*	0.8011	30.533	16.8257	0.9514	0.7208
Gram-Schmidt	*ah*	0.7967	3.2562	2.4557	0.9959	0.8601
Discrete Wavelet	*ai*	0.9448	1.6803	1.3274	0.5894	0.6498
A Trous Wavelet	*aj*	0.9093	2.1127	1.6378	0.8054	0.8622

The cluster analysis suggests that HSV and Brovey performed similarly but very differently from all the other methods due to their very large ERGAS and SPD values as

indicated by the large height value of the HSV and Brovey group. HSV fusion induces a significant degree of color distortion (high SPD and ERGAS values) but the spatial information was enriched (higher HPCC). The color distortion could be easily explained by the fact that in HSV fusion, the Value band is firstly derived from the false color composition of Landsat Bands 4, 3, 2 and then is replaced by the SPOT PAN image; but the histograms of the Value band and the SPOT PAN image are different from each other. The Brovey fusion result also exhibits significant color distortion for the same reason.

All the RK-based fusion methods perform similarly and are grouped at the same height. If the dendrogram tree is cut at the height of approximately 1, all the RK fusion methods can be agglomerated into two groups: GLS-based fusion methods and OLS-based fusion methods (the yellow polygons in Figure 4.3). Furthermore, if the tree is cut at the height of around 0.5, the GLS-based fusion methods can be further clustered into two groups: exponential-variogram-based and spherical-variogram-based fusion methods (the green polygons in Figure 4.3). If the tree is cut at the height of around 0.4, the OLS-based fusion methods can be clustered into two groups: OK-based and UK-based fusion methods (shown by the blue polygons in Figure 4.3). Thus the cluster analysis suggests that: (1) GLS is different from OLS-based fusion; (2) variogram models contribute to different performance of the fusion methods; and (3) OK and UK perform differently when applied to OLS regression residuals.

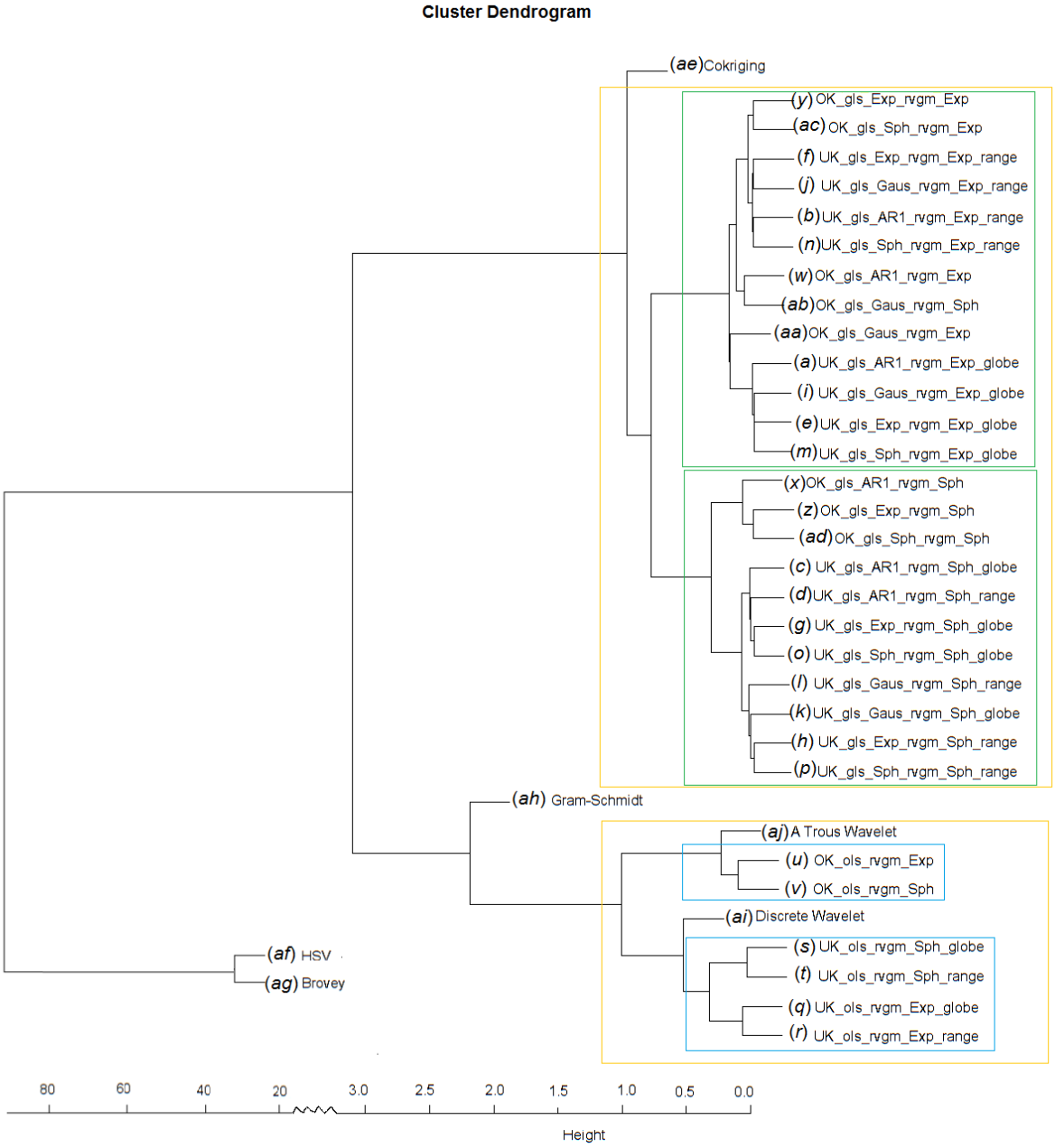

Figure 4.3: The dendrogram tree derived by cluster analysis shows similarity levels among the fusion methods. The dendrogram node heights (i.e. the x-axis values) are the cluster criterion values which are on a scale of distances (Murtagh and Legendre 2011). The heights of the vertical lines indicate the level of similarity between various clusters. The lower the height, the more similar the fusion methods.

In addition, the cluster analysis suggests that cokriging and GLS-based fusion preform similarly as is visually expressed by their preservation of spectral information in the MS image. However, cokriging and GLS yield visually blurred results, an effect most evident in cokriging which has the least improvement in spatial detail (lowest HPCC value). Discrete Wavelet and the A Trous wavelet fusion methods are similar to the OLS-based methods; however, the former is more similar to the OLS+UK and the latter is closer to OLS+OK. The A Trous wavelet fusion result presents considerately better spatial detail improvement and less noise when compared to the discrete wavelet fusion result. However, both of the wavelet based fusion results introduce more color distortion and depict less spatial detail when compared to the RK results. The Gram-Schmidt fusion method injects the highest level of spatial detail but at the cost of significant spectral distortion.

As discussed, the cluster analysis would place RK based fusion methods in two groups: GLS- and OLS-based groups. These two broad groups are also evident using boxplots (Figure 4.4). An ANOVA test shows that these two groups are significantly different ($p <= 0.05$). The GLS-based methods exhibited smoother images and less spatial resolution improvement compared to the OLS-based fusion methods whose fusion results contained more noise (lower CC, higher SPD and ERGAS values) but appreciable spatial resolution improvement (higher HPCC values). This difference could be explained by the fact that if an observation correlation structure is applied, the GLS regression process will use this correlation (i.e., coherrence) to derive the estimated values, which will decrease noise and thus generate smoother but less spatially detailed results. The OLS-based fusion methods, indicated by the SSIM values, outperform the GLS-based fusion methods. Moreover, the main difference between the OLS- and the GLS-based methods is the processing time. GLS methods take exponentially longer to

compute than OLS methods, particularly as the number of GLS iterations is increased in order to derive GLS coefficients. The dendrogram tree (Figure 4.3) also indicates that each of the GLS methods are more similar to each other, as indicated by smaller height values, than the group of the OLS methods. In terms of the contribution of different observation correlation structures in GLS (i.e., AR1, Exponential, Gaussian, and Spherical), both the ANOVA and the cluster analysis indicate that none of the four correlation structures outweigh the others.

Both UK and OK are employed in the RK process but without any visually noticeable differences in results. A one-way ANOVA test on the quantitative indicator values is used to compare the 20 UK with the 10 OK results and shows no significant difference with respect to spectral information distortion (Figure 4.5). However, the UK based fusion methods gain more spatial detail from the PAN image, demonstrated by both the HPCC boxplot in Figure 4.5 and one-way ANOVA test on the HPCC values. If only the OLS-based RK fusion methods are taken into account, there are subtle differences between OK and UK fusion methods as previously discussed.

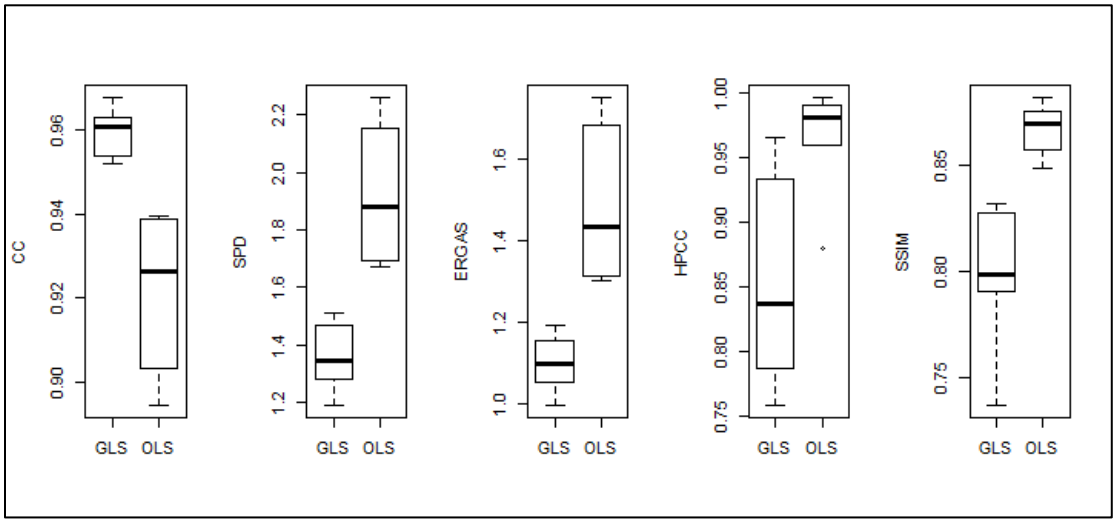

Figure 4.4: Boxplots of CC, SPD, ERGAS, HPCC, and SSIM of the OLS- and GLS-based clusters. The black horizontal lines within the boxplots indicate the positions of the mean values.

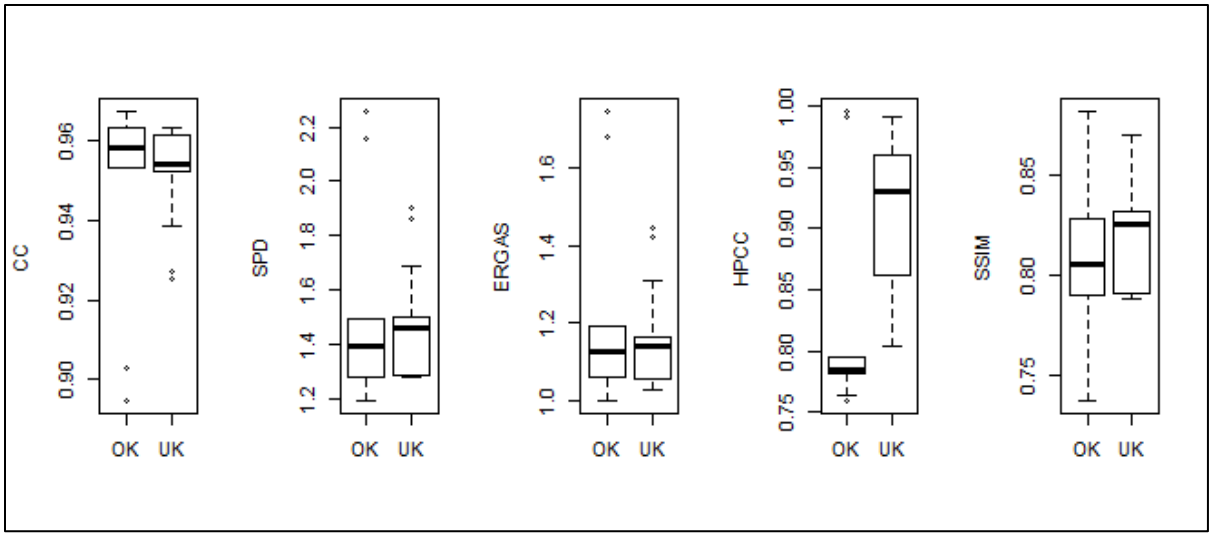

Figure 4.5: Boxplots of CC, SPD, ERGAS, HPCC, and SSIM of the UK- and OK-based fusion methods.

Two types of variogram structures (i.e., exponential and spherical) are employed in the process of interpolating regression residuals using either UK or OK. The cluster analysis reveals that all the GLS+RK fusion methods, with either UK or OK, form two different clusters based on the type of variogram used (shown by the green polygons in Figure 4.3). To investigate if these two groups perform differently from each other, all the RK-based fusion methods are paired up based on the same parameters (including regression methods, correlation structures, and kriging methods (OK or UK)) applied. For example, the UK_gls_Exp_rvgm_Sph_range (*h* in Table 4.2) and the UK_gls_Exp_rvgm_Exp_range (*f* in Table 4.2) form a pair. Figure 4.6 shows the CC, SPD, ERGAS, HPCC, and SSIM values of the 15 paired fusion methods. For each pair, the CC value of the exponential-based fusion method is slightly higher than that of the spherical-based method, and the SPD, ERGAS, HPCC, and SSIM values of the exponential-based method are lower than those of the spherical-based method. The paired T-test is then implemented on each of the CC, SPD, ERGAS, HPCC, and SSIM index variables. This T-test, together with Figure 4.6, suggests that using the exponential variogram model in a RK fusion process preserves

significantly more spectral information from the original MS bands; while using the spherical variogram model adds more spatial details in the fusion results from the PAN image. Overall, the spherical-based methods are more efficient than the exponential-based fusion methods as indicated by the SSIM. Thus, we can see that the variogram model plays an important role if kriging is applied to interpolate correlated regression residuals because each variogram yields different nuggets, ranges, and distance weights.

Three types of search windows are assessed in the kriging process (i.e., global, range, and a specified local window). A global search window uses all pixels within the study area in the interpolation process (called global kriging). A range search window (called range kriging) uses pixels within the distance defined by the range of the fitted variogram model. A specified-size search window (called local kriging) uses the pixels within one sixth of the extent of the study area from the unknown-value pixel. Both the cluster analysis and the one-way ANOVA show that there is no significant difference between the global kriging methods and range kriging fusion methods nor between the range kriging and local kriging fusion methods from the perspective of spectral fidelity and spatial sharpening. However, the local kriging method is statistically different from the global kriging fusion methods from the aspect of spatial detail improvement. The critical difference among the three types of search windows is the processing time. The more pixels involved, the more computation time needed. The global kriging fusion method is far more computationally burdensome compared to local methods (local kriging and range kriging fusion methods).

Based on visual examination, qualitative indicator comparison, cluster analysis, and computation time costs, the OLS regression followed by UK with spherical variogram (*t*: UK_ols_rvgm_Sph_range) and OLS regression followed by OK with spherical variogram (*v*:

OK_ols_rvgm_Sph), are considered as optimal RK-based fusion methods and therefore are applied to a larger study area for further investigation.

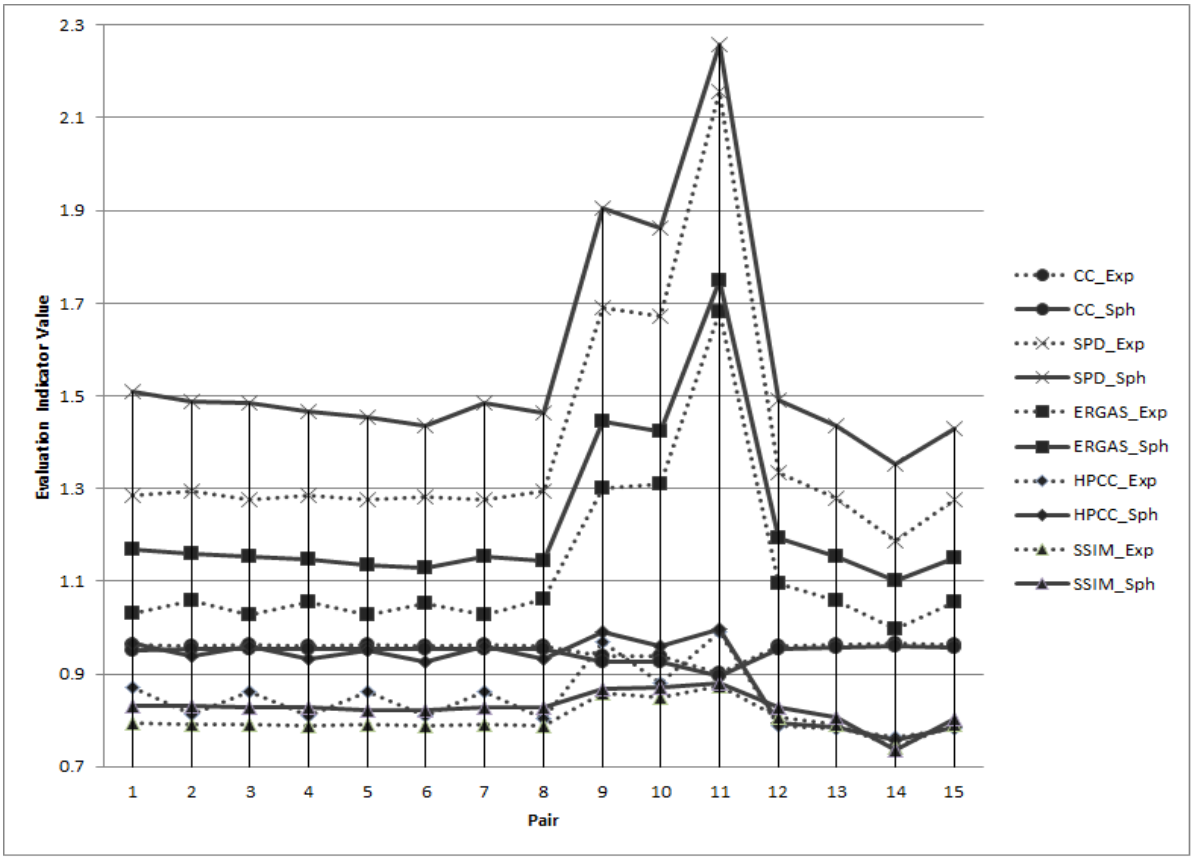

Figure 4.6: The CC, SPD, ERGAS, HPCC, and SSIM of the paired exponential-based (dashed lines) and spherical-based fusion methods (solid lines).

4.4 Applying the Selected Methods to A Larger Sub-area

In order to examine to what degree the study methods can be generalized, the selected two fusion methods are applied to a larger image chip within the study area where the pixels values are heavily skewed to the right rather than normally distributed. The fusion results generated from the selected RK fusion methods as well as HSV, Brovey, Gram-Schmidt(GS), and A Trous wavelet are evaluated.

The fused bands 1, 2, 3 are shown in Figure 4.7 for comparison with HSV and Brovey which only accept three bands. The quantitative evaluation indicator values are present in Table 4.3. The color of the RK images is closer to the original Landsat MS bands, though there is an obvious 'glow' surrounding an access road. It can be observed that the UK result is smoother than the OK result because UK uses a local trend to interpolate the regression residuals while the OK uses a constant local mean; conversely, OK exhibits richer geometry details than UK. The quantitative indicator values show that the UK method retains slightly more spectral information from the coarse MS image than the OK method, but the OK result presents more spatial details as indicated by higher HPCC values. Similar to the fusion results for the small image chip, HSV presents the greatest color distortion compared to the other methods, followed by Brovey and Gram-Schmidt. In terms of spatial detail sharpening, both Gram-Schmidt and Brovey present considerately better image sharpening than HSV and the A Trous wavelet. The A Trous wavelet preserves a similar amount of spectral information to the RK method as indicated by the CC, SPD and ERGAS values. However, the A Trous wavelet generates a lower level of spatial detail improvement compared to RK as indicated by a lower HPCC value. In addition, this wavelet fusion result is smoother than the other conventional fusion methods. Overall, the RK-based fusion methods provide better fused images that preserve richer spectral information from the original Landsat MS image and provide finer spatial details gained from the SPOT PAN image.

Table 4.3: Quantitative evaluation indicator values of the fusion results shown in Figure 4.7. The mean values of the CC, SPD, HPCC, and SSIM for Bands 1, 2, and 3 are listed.

Fusion Method	CC	SPD	ERGAS	HPCC	SSIM
UK_ols_rvgm_Sph_range	0.8887	1.5520	4.0502	0.9533	0.9626
OK_ols_rvgm_Sph	0.8740	1.7103	4.3716	0.9993	0.9568
HSV	0.8252	32.4769	49.5253	0.7333	0.3810
Brovey	0.8234	3.6805	7.4516	0.9392	0.9296
Gram-Schmidt	0.8217	1.9565	5.2812	0.9580	0.9272
A Trous Wavelet	0.8770	1.7081	4.2492	0.8014	0.9343

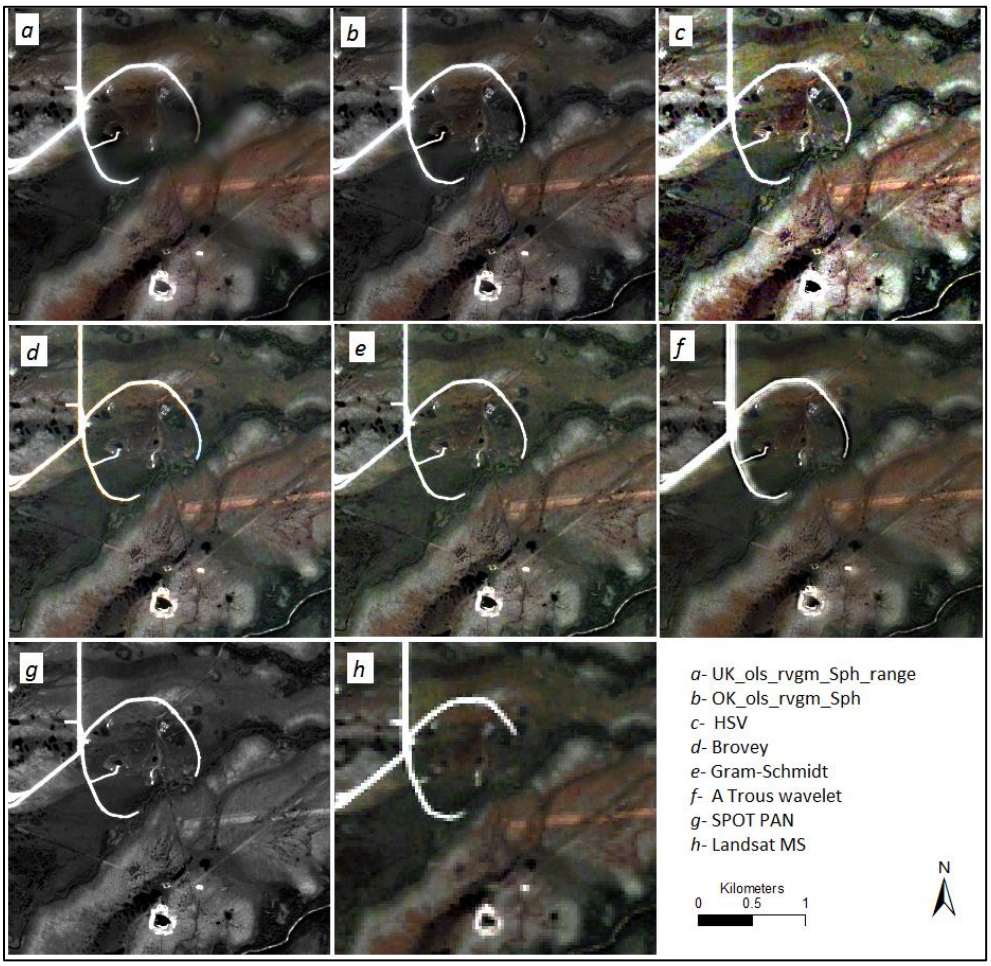

Figure 4.7: The fusion results for the larger clip within the study area are presented by true color composition of the fused images. The true color composition of the original Landsat MS images (resampled to10 m resolution) and the SPOT panchromatic image are also presented (*h* and *g*, respectively).

4.5 Discussion and Conclusions

In this study, regression kriging (RK) is applied to merge a Landsat MS image and a SPOT PAN image to form a single image having improved spatial resolution while retaining the spectral characters from the multispectral (MS) image. The RK process combines a regression between the Landsat MS image and the SPOT panchromatic (PAN) image with kriging of the regression residuals. The effect of two types of linear regression methods (i.e., Generalized Linear Squares

(GLS) and Ordinary Linear Squares (OLS)) and two kriging methods (i.e., Ordinary Kriging (OK), and Universal Kriging (UK)) is examined in the study.

The study has revealed the different roles of GLS and OLS in the fusion process. The GLS-based fusion results are smoother, while the OLS results present richer spatial information. Visual comparison and quantitative indicator values show that there is a trade-off between the preservation of spectral information and spatial detail improvement for the GLS- and OLS-based fusion methods. More importantly, GLS involves greater computational demands with little tangible benefits. Although GLS performs better than OLS when there is a spatial correlation present in the observations (Cressie 1993; Webster and Oliver 2007; Hengl 2009), the choice of which regression approach is optimal (GLS or OLS) is based on the application of the fusion results and affordability of related computation costs. In that respect, this research suggests that OLS residuals are satisfactory in the kriging process when compared with GLS. In addition, our study has revealed that different observation correlation structures do not impose a significant impact on the fusion results.

In general, regardless of which regression method (either GLS or OLS) is applied, the UK-based fusion methods do not exhibit significant differences when compared to OK, except for the longer time required to compute UK. However, if OK is applied to interpolate OLS regression residuals, the fusion results reveal more spatial details.

Cokriging is ineffective when fusing the Landsat MS and SPOT panchromatic images and the result shows the largest loss of spatial information. Visual examination of the cokriging fusion result shows that the features in the image are blurred together. Cokriging used in the fusion process requires that all variogram and cross-variogram models of the dependent variables (i.e., the MS MS image) and predictor variables (i.e. the SPOT PAN image) should have the

same shape and range (Tiho and Dagnogo 2015). When this condition is not met, cokriging will not produce satisfactory results. In the research on the small image chip, the condition is met for MS band 2 and the SPOT PAN band, but not for MS bands 3 and 4, which results in poor performance of cokriging in the study.

Some studies have demonstrated that the spherical variogram model is appropriate in a kriging process because it meets the spatial characteristics of many geographic phenomena (Meng et al. 2010; Wu et al. 2015). This study has revealed that, from the aspects of both the spectral and spatial characteristics, the RK-based fusion methods using a spherical variogram generate better fusion results than the methods using an exponential model. However, similar to the GLS-based fusion methods and the OLS-based methods, there is a trade-off between spectral fidelity and resolved spatial details for the spherical-based and the exponential-based fusion methods. According to the quantitative indicator values, the exponential-based fusion methods tend to preserve more spectral information from the MS image, while the spherical-based fusion methods gain more spatial details from the PAN image. Thus, the choice between the spherical or exponential model depends on the intended application of the fused image.

In addition, there is no substantial difference between the fusion results generated by using different types of search windows in the interpolation process of regression residuals, though the computation costs are different. However, more trials are needed to determine an optimal local search window. Moreover, we undertook RK using custom functions in the R language, which has considerable overhead as a high level object based language, and building a custom RK procedure in C or Fortran would likely reduce the computation times to very reasonable levels.

In terms of performance of the conventional fusion methods, HSV yields the most color distortion, followed by Brovey. However, Brovey depicts richer spatial information in the fused image. The Discrete and A Trous wavelet methods preserve spectral information sufficiently, but less spatial information from the panchromatic image is added into the fusion results (especially for the Discrete wavelet method) compared to the RK fusion results. The study has shown that all the RK-based fusion methods produce better quality fused images from both the spectral and spatial point of view.

Two heterogeneous sub-areas are tested in order to explore the generalizability of RK to other regions and datasets. The MS pixels and panchromatic pixel values are normally distributed for the smaller clip of the study area, while the distribution of the pixels for the larger clip has a long tail, even after applying value transformations. The test on these two heterogeneous areas shows that the RK-based fusion results are satisfactory and the fusion images derived from the RK-based methods are more spectrally and spatially accurate than existing fusion approaches.

The study has demonstrated that RK-based fusion methods can be applied to fuse Landsat MS images and SPOT panchromatic images based on the spatial correlation present in the observations and residuals. Moreover, the OLS-based fusion methods generate satisfactory results with good spectral fidelity and resolution improvement when compared to the GLS methods. The OK-OLS-based fusion method is the simplest and most computationally effective approach compared to the UK-OLS-based fusion method, though both are optimal fusion approaches.

4.6 Acknowledgements

This study was funded by a Canadian Space Agency Grant to Dr. M. Sawada. We appreciated the support provided by Dr. J. McLaughlin, Mr. B. Hamel of Ontario Ministry of Natural Resources for data preparation.

Chapter 5 Assessing Landsat Archive Data for Land Cover Change Detection in the Addington Highlands, Canada

Abstract: This paper investigates the application of two change detection techniques: a post-classification change detection (PCCD) method and a threshold-based change detection (TBCD) method, to quantify land cover changes in the Addington Highlands, Ontario, Canada. Within the PCCD method, a Support Vector Machine (SVM) algorithm is firstly used to classify two pairs of bi-season Landsat images. Subsequently, through GIS overlay, the generated thematic maps identify changes in deciduous and mixed forest, coniferous forest, non-forest vegetation, non-vegetated, and water over a 20 year period (1994 to 2013). Approximately 336 ha of non-vegetated areas were reforested and the total forest cover (mostly deciduous and mixed forest) increased by around 132 ha. Landscape metrics indicate little change in spatial structure since 1994, although the forest composition at different locations has changed. Within the TBCD method, Landsat multispectral and panchromatic images from different times are fused using a PCA (principal component analysis) technique. Next, the fused and the raw Landsat images are differenced. Finally, changes in vegetated areas are identified based on the generated differenced image. The TBCD method indicates approximately 47.6 ha non-vegetated areas have changed to vegetated and 107 ha areas changed conversely since 1994. Because the TBCD uses only one band of the Landsat image, whereas the PCCD method uses all available spectral information, results of the two methods differ. The PCCD method effectively generates detailed from-to land cover change maps, but error may be multiplicative from image classification. The TBCD method provides a simple and fast way to detect changes in vegetated areas with superior differentiation between vegetation types but with the drawback of resolution-induced spurious changes.

Keywords: Land cover, change detection, post-classification, threshold-based, PCA

5.1 Introduction

Accurate and timely information about land cover dynamics is important for resource management and ecosystem research. Remotely sensed imagery offers a wide selection of spatial, spectral and temporal resolution and a synoptic view (Chen et al. 2012, Hussain et al.

2013; Lu et al. 2014) and is, therefore, frequently employed to provide accurate and continuous land cover change information. Change detection techniques based on remotely sensed images have been intensively investigated for decades because they offer a cost-effective alternative to ground-based survey for the generation of timely change information (e.g. Hayes and Sader 2001; Forkuo and Frimpong 2012; Huang and Song 2012; Borrelli, Rondón, and Schutt 2013).

The various change detection techniques have been extensively reviewed elsewhere, e.g., Singh (1989), Coppin et al. (2004), Lu et al. (2004), Radke et al. (2005), Chen et al.(2012), Bhagat (2012), Alqurashi and Kumar (2013), Hussain et al. (2013), Hecheltjen et al. (2014), Lu et al (2014), and Tewkesbury et al (2015). Approaches to change detection can be placed into two primary groups: pre-classification and post-classification (Singh 1989; Peiman 2011; Hussain et al., 2013).

Pre-classification approaches generate binary "change" and "non-change" maps by directly comparing pixel values or derivatives of pixel values (e.g., NDVI value) from images taken at two different times, using various algorithms such as image differencing, image ratioing, or image regression. The prerequisites for such comparisons include rigid image co-registration, radiometric calibration and atmospheric correction, as well as radiometric normalization in order to reduce spurious results.

Post-classification change detection (PCCD) approaches generate "from-to" change trajectory maps by comparing classified thematic maps. These classified maps are produced separately by applying either pixel-based or object-based algorithms on images from different times. Therefore, the final change detection accuracy is entirely dependent on the accuracy of individual classifications; however, the aforementioned strict prerequisites for pre-classification methods are less significant for PCCD techniques (Coppin et al., 2004; Hussain et al., 2013;

Tewkesbury et al., 2015). Thus, PCCD techniques are the most established and widely used change detection approaches, especially when images from different sensors with various pixel sizes are required.

Image fusion is a technique that merges two or more images for the same area in different wavelengths and/or from different sensors in order to provide enriched spatial and/or spectral information (Pohl and Van Genderen 1998). Some studies employed image fusion techniques in change detection workflows to compensate for the insufficiency of single-source remotely sensed imagery (e.g. Lu et al. 2008; Zeng et al. 2010; Gong et al. 2012; Du et al. 2012; Rokni et al. 2015). For example, Lu et al. (2008) used a threshold-based change detection (TBCD) method to distinguish between changed and non-changed vegetated areas based on the fusion of a SPOT HRG panchromatic image (Pan) from 2003 with a Landsat TM multispectral image (MS) from 1998, using a principal component analysis (PCA) fusion algorithm. Rokni et al. (2015) detected lake surface area changes using a PCCD method based on the fusion of a Landsat TM 2010 MS image and Landsat ETM+ 2000 panchromatic image. Du et al. (2012) detected changes by fusing images created from various image differencing algorithms. Such studies have demonstrated that change detection based on fused images can generate improved results over the use of a single-source input.

Different change detection techniques have their own strengths and weaknesses and not a single technique is optimal for all cases (Lu et al. 2004; Coppin et al. 2004; Lu et al.2005; Alqurashi and Kumar 2013). Thus, in practice, selecting an optimal detection method is often based on the outcome of comparing the results from different detection methods.

The purpose of the study is two-fold. The first is to evaluate the performance of PCCD techniques based on multi-temporal images in quantifying land cover changes, and the second is

to assess the capability of TBCD approaches based on fused images in differentiating change/non-change vegetated areas. The change/non-change vegetated area maps generated from the TBCD method are then compared to those from the PCCD method.

5.2 Study area

The study area (Figure 5.1) is located in a remote rural region of the Addington Highlands, Ontario, Canada (77°16′20″W, 45°01′01″N). The map on the right-side of Figure 5.1 shows the current topographic information (downloaded from www.geogratis.gc.c) of the study area. The natural landscape in the study area consists of forest, lakes, rivers, streams, wetland, farms, and built-up areas. Forest dominates the study area and is a mixture of deciduous and coniferous trees, together with the presence of fire scars and bedrock outcrops. The economy in this region is based on forestry, fishing, and recreation (McLeman 2010). Due to the importance of forest cover and water bodies to the community economy, determining land cover change, especially forest cover changes, will eventually help facilitate an understanding of the impact that such land cover change may have on human activities and regional economics.

5.3 Methodology

In this study the post-classification change detection (PCCD) approach is firstly applied to detect change trajectories from 1994 (Time 1) to 2013 (Time 2), resulting in a detailed "from-to" land cover change thematic map, a land cover change matrix, and landscape metrics. Secondly, a threshold-based change detection (TBCD) approach based on fused images is employed to generate change and non-change vegetated areas. Within the TBCD method, the PCA-based (principal component analysis) fusion method is used to develop differenced images from which change/non-change vegetated area maps are created. Figure 5.2 summarizes the main steps

applied in the study including data collection and pre-processing, the generation of the detailed from-to land cover change maps by PCCD, and the creation of binary change/non-change maps by TBCD.

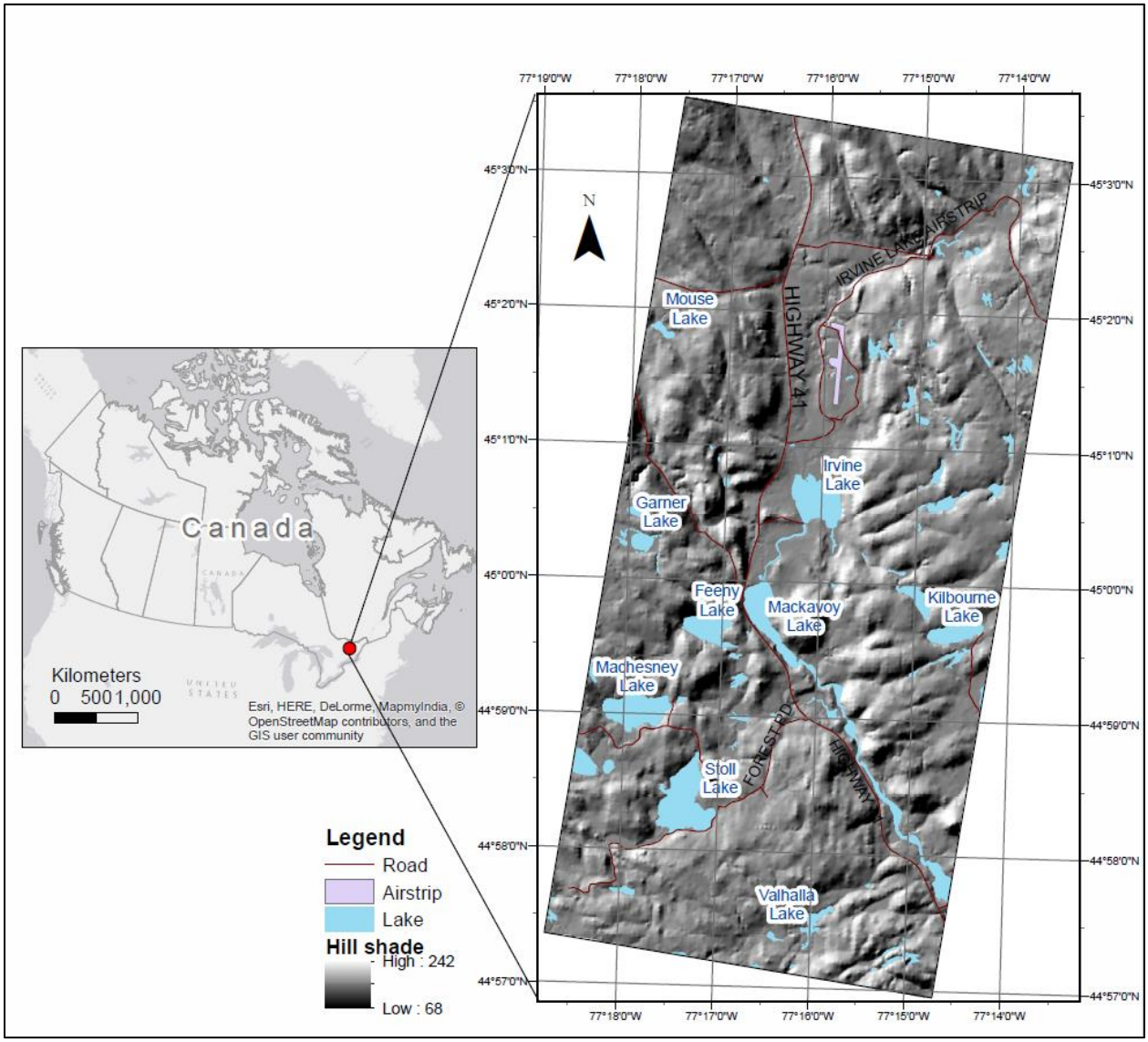

Figure 5.1: Location of the study area: Addington Highlights, Ontario, Canada. The right map shows the current topographic information of the study area. Data source:www.geogratis.gc.ca

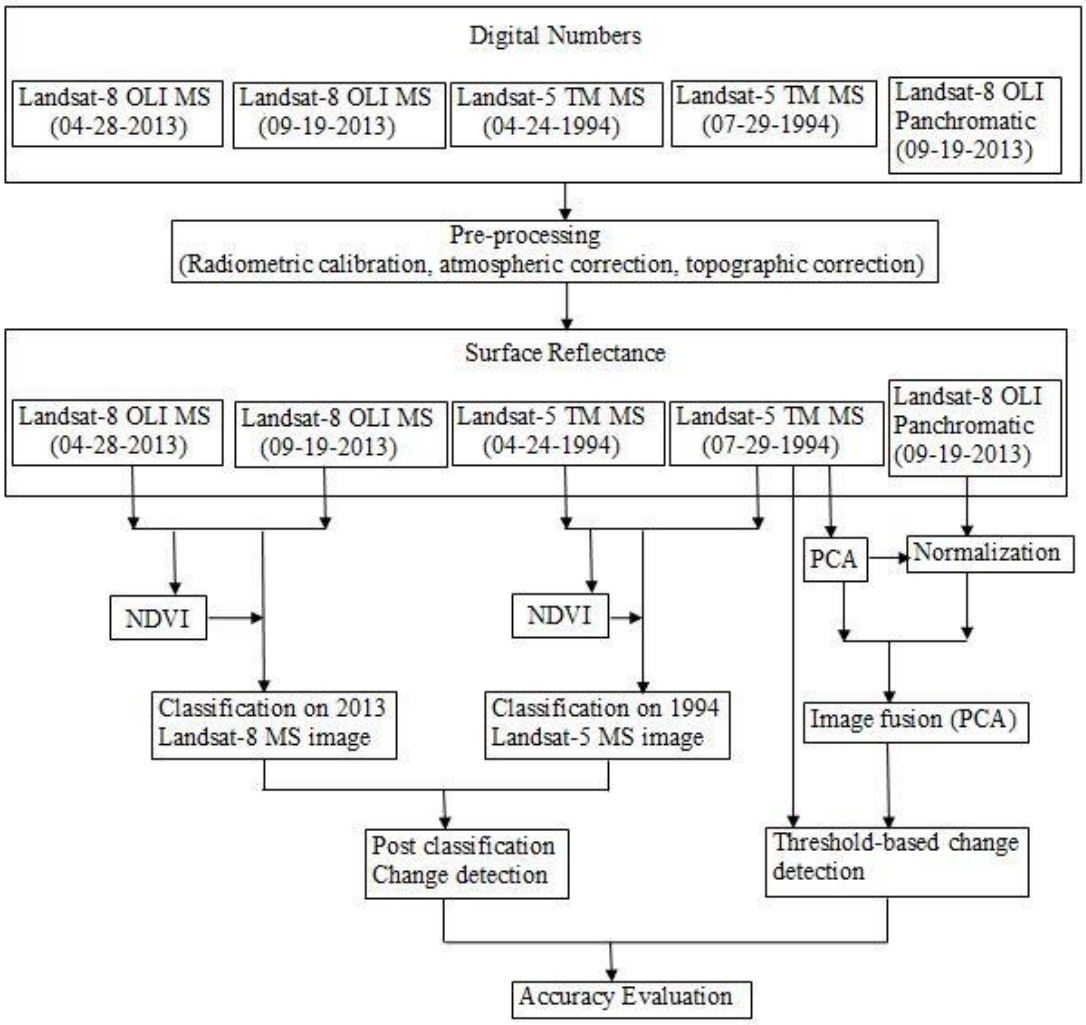

Figure 5.2: The main steps applied in the study from data acquisition to change detection accuracy assessment.

5.3.1 Data and data pre-processing

The majority (over 80%) of the land surface in the study area is covered by forest. Thus the land cover change study focuses on forest cover changes. The selection of appropriate image acquisition dates is crucial to forest cover change detection because of phenological changes caused by the natural variability of inter-annual precipitation and temperature (Coppin et al. 2004). Previous studies have suggested that for forest cover change detection, summer and winter are optimal times for image acquisition because of the phenological stability at those

times (Hame 1988). However, in this study, the land surface of the study area is covered by snow in winter and, moreover, the surface conditions in images taken in winter are often obscured by cloud. In addition, the study area contains mainly mixed deciduous and coniferous forest, therefore, the spring images highlight coniferous cover whereas the summer images show both deciduous and coniferous forest. Thus, to ensure snow-free and cloud-free (or less than 20%) imagery, images acquired in spring and summer (bi-season images) are chosen to sufficiently represent leaf-off and leaf-on conditions in the study region. Two pairs of bi-season Landsat images covering the study area over a 20 year span from 1994 to 2013 were selected. The first pair was acquired by sensor Landsat-5 TM on April 24 and July 29 of 1994 for leaf-off and leaf-on seasons, respectively. The second pair was taken by Landsat-8 OLI_TIRS on April 28 (leaf-off) and September 19 (leaf-on) of 2013. All the images (Level 1 Terrain Corrected (L1T) product) were downloaded from the US Geological Survey (www.usgs.org) with acquisition details in Table 5.1.

Accurate image registration and radiometric correction of multi-temporal images are important in order to achieve geometric integrity and radiometric consistency, especially when threshold differencing approaches are used for change detection (Coppin et al. 2004; Lu et al. 2005; Huang and Song 2012; Lu et al. 2014). All Landsat images were already co-registered accurately so that no further registration process was needed. The coordinate system of all the images is Universal Transverse Mercator of Zone 18 N with datum WGS-1984. The radiometric correction conducted here includes radiometric calibration, atmospheric correction, and topographic correction. The digital numbers recorded in the acquired images were calibrated to top-of-atmosphere (TOA) reflectance based on the ancillary information provided in the image metadata, and then calibrated to surface reflectance in order to remove atmospheric influence

using the Dark Object Subtraction (DOS) method (Gupta 1991) based on the presence of dark water bodies within the images. Topographic correction was conducted using the Cosine Correction method introduced by Civco (1989) to remove the impact of the relief variability on spectral reflectance (the study area is a highlands terrain). The 10-meter resolution Provincial DEM data (Ontario Ministry of Natural Resources, 2005-2006) were used for the topographic correction.

Table 5.1: Summary of the Landsat images used in the study.

Sensor	Number of Bands	Wavelength (micrometers)	Resolution (meters)	Radiometric Resolution	Acquisition date	Sun Azimuth/ Elevation angle (degrees)	Path/row
Landsat-5 TM	3 Visible 1 NIR 2 SWIR	0.45-0.52 (blue) 0.52-0.60(green) 0.63-0.69(red) 0.76-0.90(NIR) 1.55-1.75(SWIR 1) 2.08-2.35 (SWIR 2)	30	8-bit	1994-04-24 1994-07-29	133.7/ 50.1 125.0/ 53.4	16/ 29 16/ 29
Landsat-8 OLI	4 Visible 1 NIR 2 SWIR 1 pan	0.43 - 0.45 (Coastal aerosol) 0.45-0.51 (blue) 0.53-0.59 (green) 0.64-0.67(red) 0.85-0.88(NIR) 1.57-1.65(SWIR 1) 2.11-2.29 (SWIR 2) 0.50-0.68 (pan)	30 (MS) 15 (Pan)	16-bit	2013-04-28 2013-09-19	147.9/ 56.4 156.3/ 44.4	16/ 29 16/ 29

Note: NIR-Near Infrared, SWIR- Shortwave Infrared, Pan- Panchromatic

The Normalized Difference Vegetation Index (NDVI) was calculated for each Landsat image. Furthermore, the difference between the NDVI values from leaf-on and leaf-off season in the same year was also calculated. Finally, the preprocessed Landsat images acquired in the same year from both the leaf-on and leaf-off seasons, together with two NDVI index bands and one NDVI difference band, were stacked together to form one data set for image classification.

5.3.2 Post-classification change detection method (PCCD)

Two pairs of bi-season images are used to detect detailed from-to changes on the following five land cover types: deciduous and mixed forests (containing sparse coniferous trees), coniferous forests, non-forest vegetation, non-vegetated, and water. Water is easily distinguished from all the other land cover types because of its very low reflectance across the image wavelength spectrum. The non-vegetated areas include wetland, roads, residential areas, bare soil and bedrock outcrops and they are distinguishable from forests, grassland, and cropland because of the lesser chlorophyll content in both the spring and summer images. The non-forest vegetation areas include annual croplands and grasslands in the study. Grasslands and Croplands that are planted with annual crops will show or respond as bare soil in the spring images but present a pink color in the standard false color composite summer images. Thus they are separable from forested land cover. Coniferous trees show green in the true color composite and dark red in the standard false color composite of the image acquired in the spring and summer time because of foliage being on coniferous trees throughout a year. Unlike coniferous forests, deciduous and mixed forests respond in a fashion similar to bare soil in the true color composite of the spring images due to little or no foliage on the trees in early spring, but in the summer images deciduous and mixed forests exhibit green in the true color composite and red in the standard false color composite of the image because of the abundant foliage on the trees in the summer time. Such spectral reflectance variations with season enable the separation of deciduous and mixed forests from coniferous forests.

Training data are vital for a successful supervised classification and are required to be representative of the units to be classified. The training areas for the 2013 bi-season pair are firstly identified based on visual examination of true color composite from the 2013 Landsat image in different seasons and the Ontario 2008 DRAPE (Digital Raster Acquisition Project for

the East, 0.2 meter resolution) aerial photographs (DRAPE 2008). The selection of the training data for the 1994 bi-season pair is based on editing the training areas from the 2013 image as well as examination of a 1987 black/white aerial photo (collected from the National Air Photo Library of Canada), a CIRCA 2000 land cover image (CIRCA 2001), 2008 DRAPE aerial photos, and a true color composite of the 1994 image. The editing of the 2013 training data set to form the 1994 training data set includes discarding training areas if cover had clearly changed over the 20 years and/or adding new training areas by visually examining the aforementioned sources. Table 5.2 lists the number of polygons and pixels for the 1994 and 2013 training data. Non-forest vegetation cover contains fewer training pixels due to the very small percentage of croplands and grasslands present in the study area. Figure 5.3 plots the spectral signatures based on the 1994 summer Landsat-5 TM imagery for the five land cover types. The spectral separability of the 1994 and 2013 training are then evaluated using the Transformed Divergence (TD) values and all the TD values (>1.99) indicate that the training data are viable for successful supervised classification (Richard and Jia 1996).

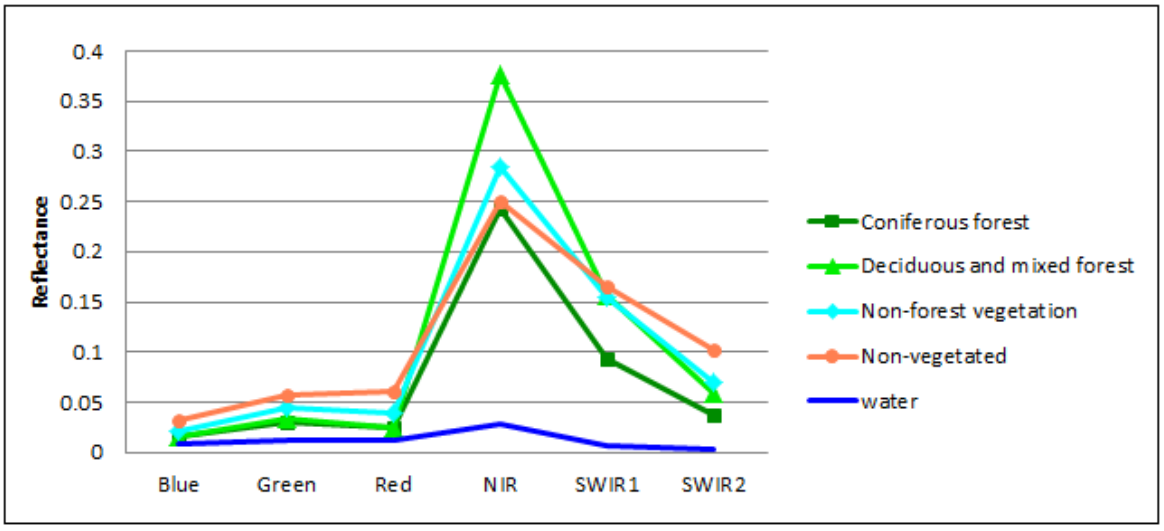

Figure 5.3: The signature plot of the land cover types investigated in the study based on the Landsat-5 image acquired on 29 July 1994.

Table 5.2: Training data used in the classification process.

	1994 TM images (# of polygon/ #of pixel)	2013 OLI images (# of polygon/ #of pixel)
Deciduous and mixed forest	21/254	27/212
Coniferous forest	31/167	19/156
Non-forest vegetation	3/15	6/28
Non-vegetated area	60/242	30/203
Water	11/152	12/156

In addition, compared to a single-date image, the TD value between deciduous and mixed forests and non-forest vegetation is increased by the combination of spring and summer images. Thus, bi-season imagery can be used to improve land cover classification accuracy, which has also been demonstrated in many studies (Yuan et al. 2005, Deng et al. 2008, and Sexton et al. 2013).

The Support Vector Machine (SVM) algorithm is adopted to label each pixel in the image to one of the five land cover types. SVM has proven to be superior to traditional classification methods like Maximum Likehood classification (Kuemmerle et al. 2009; Huang and Song 2012; Lu et al. 2014; He et al. 2015). In addition, Landsat images from 1993, 1995 and 2015 are used to detect non-real change phenomena that ensue from human activities and phenological variations in forest growth. For example, there are croplands located south of Irvine Lake (Figure 5.1) that seemed to have changed to bare soil in the 1994 bi-season images. However, a set of spring and summer Landsat-5 images acquired in 1993 and 1995 (downloaded from www.usgs.org) indicate that no changes occurred in these croplands. In this case the identification of these croplands were edited and re-labelled as a non-forest vegetation based on the 1993 and 1995 Landsat-5 images.

The classification accuracy evaluation is implemented by computing confusion matrices. The pixels used to compute confusion matrices are identified by randomly sampling pixels

across the study area. Theoretically, 30-50 pixels are required for each class in order to achieve reliable accuracy assessment (Congalton and Green 2008). In this study, 50 pixels are randomly sampled for each of the following land cover classes: coniferous forests, deciduous and mixed forests, and non-vegetated areas. However, only a few testing points are sampled for water and non-forest vegetation because water presents little variability and there is only a very small amount of non-forest vegetation in the study area. The overall classification accuracies for all are higher than 89%, and the Kappa coefficients are over 0.85, which meet the minimum 85% overall accuracy specified by the Anderson classification scheme (Anderson et al. 1976). Therefore, the classification results are viable for change detection.

The from-to change detection result is generated by applying overlay analysis, according to the PCCD, to the classified results from 1994 and 2013. The change detection accuracy is evaluated by applying a two-step approach introduced by Congalton and Green (2008). The first step of this approach is similar to a traditional single-date imagery classification accuracy assessment, but only focuses on the changed areas in the 2013 result over the 20 year period. A total of 150 points are randomly sampled from the changed areas in the 2013 image (i.e. a map class is different in 2013 when compared to 1994), and the identities of the 150 points are determined by visual examination of true and false color composites of the 2013 imagery and the 2008 DRAPE aerial photo. Among the 150 testing points, 59 points are coniferous forests, 52 deciduous and mixed forests, and 32 non-vegetated areas. Changes of water and non-forest vegetation classes do not occur in sufficient amounts and so only three testing points for water bodies and four for non-forest vegetated areas are selected for accuracy assessment. The accuracy of the 2013 classification within the changed areas is then assessed based on these 150 testing points. The overall accuracy is 88.7% and the kappa coefficient is 0.83.

The second step is a change/no-change validation. The from-to change detection map derived from PCCD is first converted to a binary change/no-change map. The sample size for the testing data is calculated based on Ginevan (1979)'s research. In this study, the map classification accuracy is greater than 0.85; if we run the risk of a 5% chance that we reject a map of acceptable accuracy (i.e. producer's risk) and a 5% chance that we accept a map of unacceptable accuracy (i.e. consumer's risk), then at least 93 samples are required. The change detection map will be accepted as accurate if eight or fewer sampled points are misclassified (Ginevan 1979). In this study 93 randomly sampled points are selected across the study area, among which 75 are detected correctly as unchanged, 12 correctly detected as changed, and six labelled as changed but unchanged in reality. Therefore, the change detection result is acceptable, and the computed overall change detection accuracy is 93.5%.

5.3.3 Threshold-based change detection (TBCD)

The threshold-based change detection (TBCD) method is used to develop a change/no-change vegetation map. The preprocessed Landsat-5 MS image acquired on 29 July 1994 and Landsat-8 Pan image acquired on 19 September 2013 are used because these images present the largest extent of leaf-on vegetation cover in the study area. In this method, the Landsat-8 Pan image is fused with the 1994 MS image using the PCA sharpening algorithm. In order to fuse the 15-meter Pan image and 30-meter MS image, the 30-meter 1994 MS image is firstly down-sampled to a 15-meter pixel size using the nearest neighbouring resampling method. Then, the Pan image is normalized to the first component (PC1) of the PCA result created from bands 1, 2, 3 of the resampled 1994 MS image using a pseudo-invariant feature (PIF) approach (Schott et al. 1988). Next the PC1 is replaced by the normalized Pan image. Finally an inverse PCA is applied to form the fused image (Lu et al. 2008). After image fusion, the spectral information of the

changed features in the 2013 image can be injected into the 1994 MS image. The differenced image between the fused 1994 image and the pre-processed 1994 MS image at the band-to-band basis is then developed using the formula $D_i=FTM_i - TM_i$, where FTM_i is band *i* of the fused image and TM_i is band *i* of the 1994 MS image. The change/non-change vegetated area map can be created by defining appropriate thresholds based on the differenced image.

Relative normalization of the 2013 Pan image according to the 1994 image is important in the TBCD method because it ensures spectral integrity of the unchanged features in the 1994 MS image and the 2013 Pan image. The PC1, generated from bands 1, 2, and 3 of the 1994 MS image, is used as a simulated image to normalize the 2013 Pan image because the wavelength of the 2013 Pan image falls within the spectral range of the three visible bands of the 1994 MS image (shown in Table 5.1). Various approaches can be used to achieve image normalization, such as histogram matching and the pseudo-invariant feature (PIF) approach (Yang and Lo 2000; Rahman et al. 2015). In this study, the normalization is accomplished using the PIF approach. The PIF approach assumes that the spectral reflectance characteristics of the PIF pixels remain constant in both the 1994 and 2013 images. The PIFs identified in the study include mature coniferous trees, bare soils, bedrock outcrops, roads, and deep water. A linear regression model is then applied to the pixel values of the PIFs in the PC1 and the 2013 Pan images. The coefficient of determination of the linear regression model derived in this study (r-squared) is 0.9188, and the residuals are normally distributed, all of which indicates a good model fit. Therefore, the normalization process can minimize the introduction of spectral distortion (caused by histogram mismatch between the 2013 Pan and the PC1 images) for the unchanged features in the fused image. Meanwhile, in the fused image, the pixels in changed areas will exhibit greater

color distortion and so change detection is measurable. Thus it is essential to choose an appropriate simulated image to normalize the high-resolution Pan image.

The identification of optimal bands that best reflect the features of interest is critical in a TBCD process (Lu et al. 2014). Each image band has unique characteristics and thus possesses different capabilities for distinguishing different land cover types. For example, some studies found that red band differencing is effective in identifying vegetation change (e.g. Chavez and Mackinnon 1994; Pilo et al. 1988). Lu et al. (2005) found that Landsat-5 TM band 5 was optimal in distinguishing vegetated areas in the Amazon. In this present study, only three visible bands are involved in the fusion process because the PCA-based fusion method only accepts three bands and the spectra of bands 1, 2, 3 of the Landsat-5 TM image and the Landsat-8 Pan image overlap. If the spectrum of the high-resolution Pan image covers the Near-Infrared (NIR) portion, for example, if a SPOT Pan image is used, then band 4 and band 5 of the Landsat-5 TM image could be used (e.g. Lu et al. 2008). In this study, among the three visible bands of the 1994 Landsat-5 TM image, band 3 (the red band) is identified as the optimal band to detect vegetation changes.

The identification of thresholds to generate binary change/non-change maps from the TBCD is required (Coppin et al. 2004; Alqurashi and Kumar 2013; Lu et al. 2014). Perfect image registration, radiometric calibration and normalization cannot be achieved in practice. Consequently differences in surface reflectance which are not caused by land cover changes may contaminate a threshold-based differenced image. Thus, the definition of suitable thresholds is critical in generating change detection results. The definition of thresholds is usually based on the statistics of differenced images. In most cases, unchanged areas account for most of the landscape, so that the pixels of unchanged areas are often normally distributed with a mean of

zero, and changed areas are located at the tails of the histogram of the differenced image. Cluster analysis, which roughly separates changed pixels from unchanged pixels, could help reveal groups of changed pixels. An interactive trial-and-error procedure is another often used method to define thresholds (Liu et al. 2004), whereby an analyst iteratively uses different thresholds and visually examines the results until satisfied. In this study, the thresholds are defined with the support of statistical cluster analysis on the differenced image and interactively adjusting thresholds and then examining the results on the screen. The changed/non-changed map generated herein, groups all the pixels into three classes: de-vegetated areas where pixel values are higher than a defined threshold t_1 (e.g. 0.007621 in this study), re-vegetated areas where pixel values are lower than a defined threshold t_2 (e.g. -0.018041 in this study), and non-changed areas where the pixel values are located within the range from t_2 to t_1.

The randomly sampled 93 points which are used in the second step of the accuracy assessment process in the PCCD process are applied to assess the binary change detection accuracy. Out of the 93 points, 84 unchanged points are correctly identified as unchanged and one changed point is identified as changed, but two unchanged points are labelled as changed and 6 changed points are classified as unchanged. Thus, the overall accuracy level is 91.4% for this TBCD method.

5.4 Results

5.4.1 Post-classification change detection (PCCD)

Figure 5.4 presents the SVM classification results in 1994 (Figure 5.4 (*a*)) and 2013 (Figure 5.4 (*b*)) with forest cover dominating the study area followed by water bodies. Non-vegetated areas

are sparsely distributed in the study area. The PCCD map is shown in Figure 5.5, and Table 5.3 lists the change matrix in the study area.

Table 5.3: Matrix of land covers and changes from 1994 to 2013 (unit: ha).

2013	1994					
	coniferous forest	Deciduous & mixed forest	Non-forest vegetation	Non-vegetated	water	total in 2013
Coniferous forest	2258.19	348.21	0	251.1	3.24	2860.74
Deciduous and mixed forest	485.1	1755.99	1.89	85.05	0.72	2328.75
Non-forest vegetation	0	0.36	30.06	3.24	0.18	33.84
Non-vegetated	108.18	81.45	6.48	502.83	20.7	719.64
water	14.94	4.59	0.9	36.63	232.02	289.08
total in 1994	2866.41	2190.6	39.33	878.85	256.86	

In general, around 336 ha of non-vegetated area was reforested over the 20 years, and the total forest cover increased by 132.48 ha (relative increase of 2.8%) from 1994 to 2013, among which, deciduous and mixed forests increased by 138.15 ha and coniferous forests showed little change (decreased 5.67 ha). In more detail, over the 20 years, around 485 ha of coniferous trees and 85 ha of non-vegetated areas have turned to deciduous and mixed forests; around 348 ha of deciduous and mixed forest and 251 ha of non-vegetated areas have changed to coniferous forests; around 108 ha of coniferous cover and 81 ha of deciduous cover changed to non-vegetated area. Though the total area of coniferous forests does not show many changes over the 20 years, the spatial configuration of coniferous forest cover has changed (Figure 5.5). Figure 5.6 illustrates one of the most extensively-changed areas where coniferous forests have changed to deciduous and mixed forests, which can also be visually observed in Figure 5.4 (the red rectangles). This area is located at west of the Stoll Lake and south of the Machesney Lake.

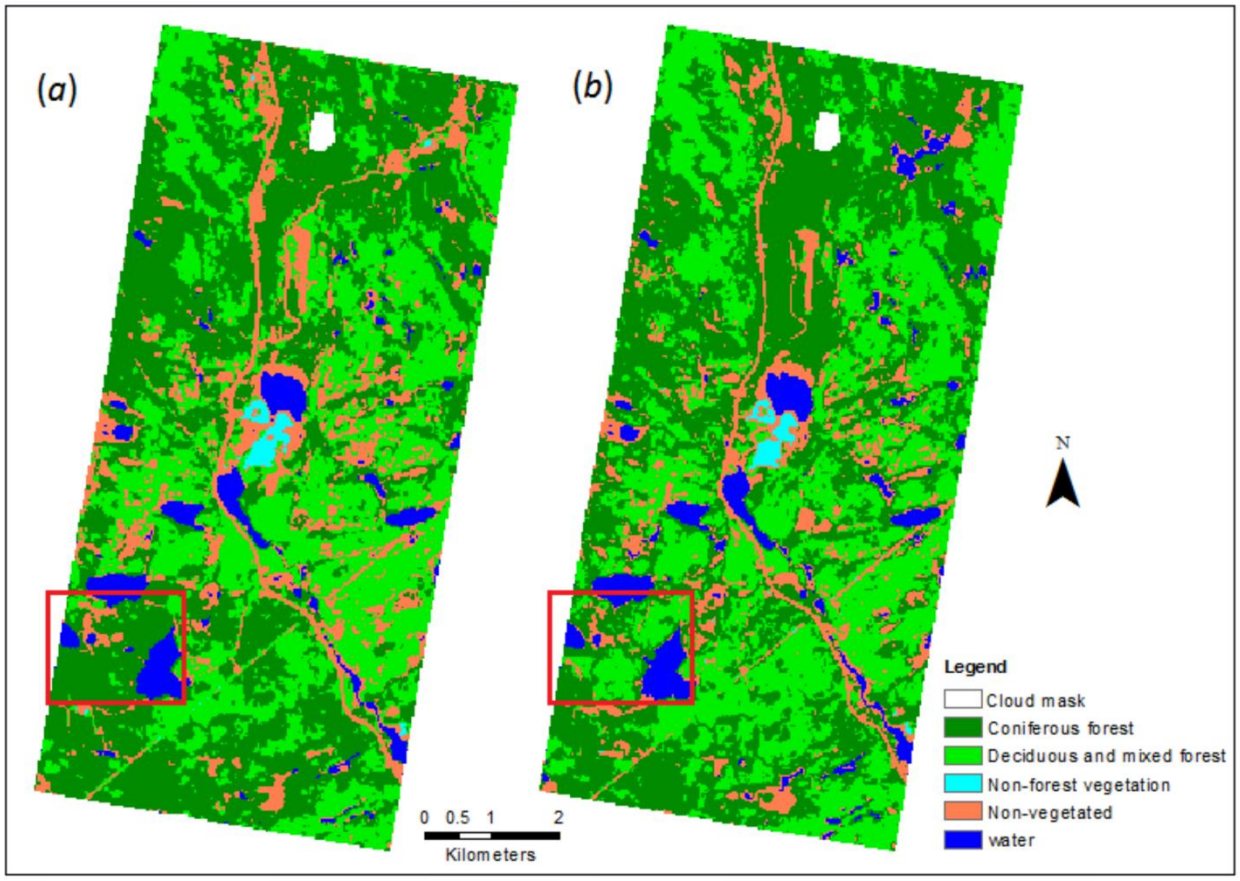

Figure 5.4: The SVM classification results: from the Landsat-5 TM image in 1994 (left) and from the Landsat-8 image in 2013(right). The red rectangles show one of the extensively-changed areas discussed in section 5.4.1 .

Figure 5.6 (*b*) and (*c*) are the standard false color composites of the spring and summer 1994 Landsat-5 images, respectively, and the standard false color composites of the spring and summer 2013 Landsat-8 image are shown in Figure 5.6 (e) and (*f*), respectively. The areas in 1994 with response in red in both the spring (leaf-off) and a summer image (leaf-on) are evergreen coniferous trees; while the areas in 2013 which show grey as bare soils in the spring image and then present a stronger red response in the summer image are deciduous and mixed forests - because very little foliage is on deciduous trees in the spring but a greater amount emerges by summer. Moreover, these changes from coniferous forests to deciduous and mixed forests over time can be confirmed by examining the historical aerial photos taken in July 1987

collected from the National Air Photo Library of Canada (Figure 5.6 (*a*)) and the 2008 DRAPE image taken in the leaf-off season (Figure 5.6 (*d*)).

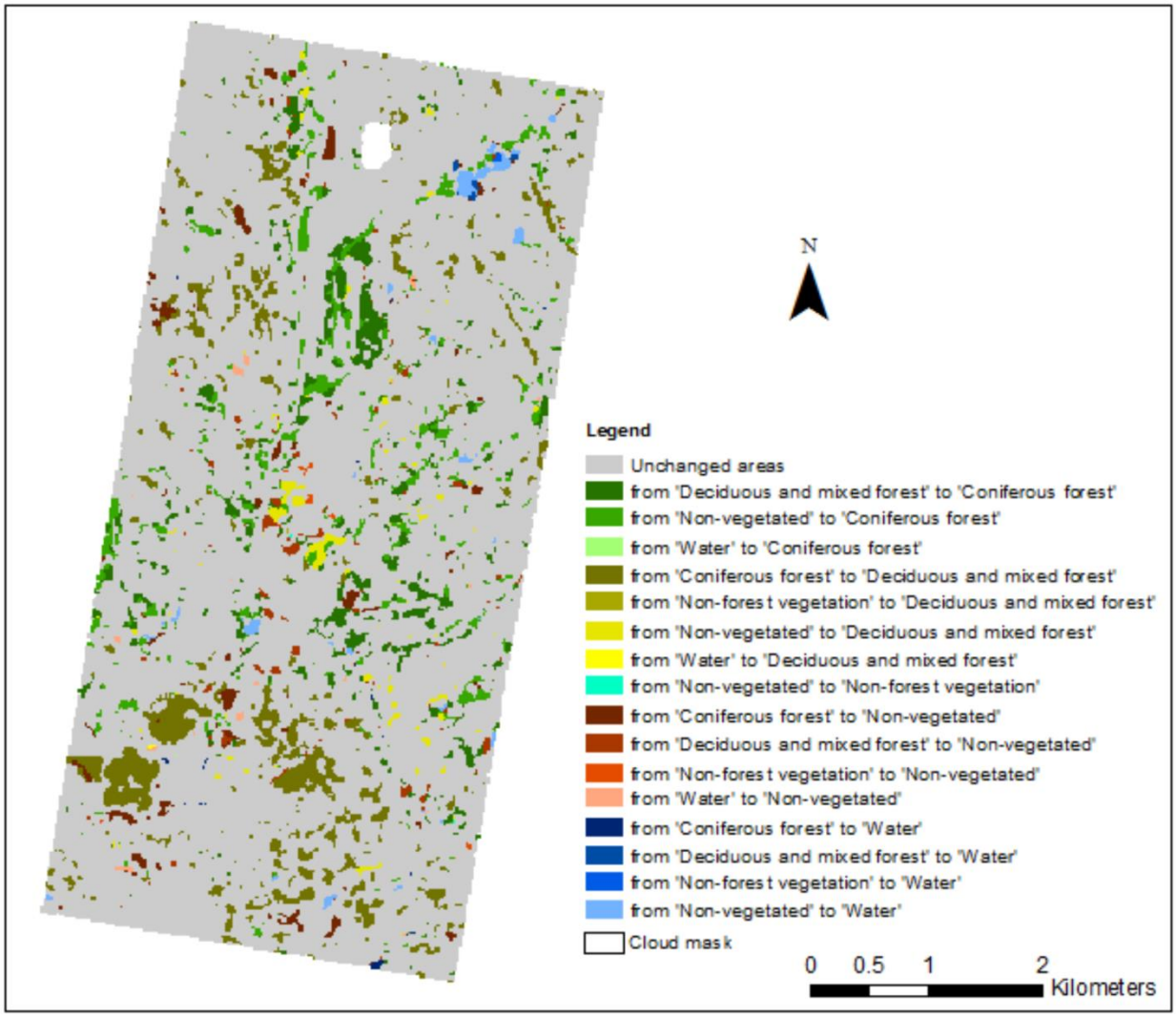

Figure 5.5: The post-classification change detection map over the 20 year period from 1994 to 2013.

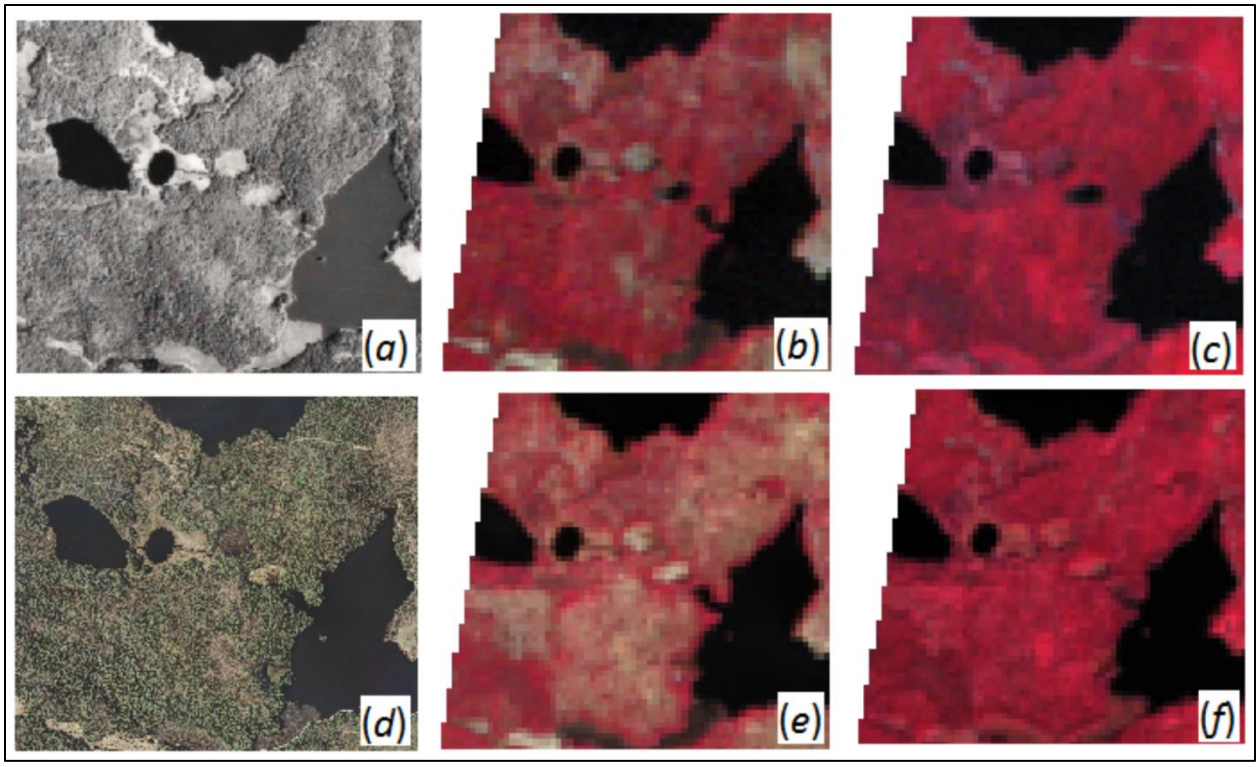

Figure 5.6: An example of the most extensively changed area. (a) the 1987 black/white aerial photo; (b) and (c) are the false color combination of the spring and summer of the 1994 image, respectively; (d) is the DRAPE aerial photo; (e) and (f) are the false color combination of the spring and summer of the 2013 image, respectively.

Various landscape metrics have been intensively applied to explore characteristics of land cover changes (e.g. Neel 2004; Liu and Weng 2013; Frate and Carranza 2013; Liu and Yang 2015; Del Castillo et al. 2015). In this study the metrics (Table 5.4) at the class level are used to portray the characteristics of the landscape: total class area (CA), percentage of landscape (PLAND), number of patches (NP), largest patch index (LPI), perimeter-area fractal dimension (PAFRAC), patch cohesion index (COHESION), and aggregation index (AI). All the metrics are computed based on the 1994 and 2013 classification results. The CA (unit: ha) records the total area of a landscape type, and the PLAND quantifies the proportional abundance of each patch type in the landscape. Both CA and PLAND measure landscape composition and can be used to compare among landscape types of varying sizes. The LPI equals the percentage of the

landscape comprised by the largest patch of the corresponding patch type. The NP and LPI have been mainly used as indicators for patch fragmentation (Frate and Carranza 2013). The decrease of LPI over time could indicate forest fragmentation, whereas the increment of LPI could mean the reverse process (Frate and Carranza 2013). The PAFRAC index has been used primarily as an index for patch shape complexity with a value ranging from 1 to 2 (Neel 2004). An increase of PAFRAC over time could indicate the generation of more complex land cover shapes, and thus may indicate extensive land cover change. The COHESION index measures the physical connectedness of a land cover type with values ranging from 1 to 100 and is commonly used for describing habitat connectivity (Frate and Carranza 2013). A decrease in that index could mean that the patch type becomes increasingly subdivided and less physically connected. The AI index measures the aggregation level of a patch type with values ranging from 0 to 100 and an increase of this index could imply more concentrated forest development (Liu and Weng 2013). The PLAND values in Table 5.4 show that deciduous and mixed forest increased around 2.2% and non-vegetated areas decreased around 2.6%. The NP value of coniferous forests decreased 15 while the LPI value increased by 6.6%, suggesting that over the 20 years, some small coniferous forest patches were aggregated into larger coniferous areas. Though the LPI of the deciduous and mixed forests decreased by 1.14%, the decrease of the NP value (18) and the increase of the CA and PLAND values suggest that the size of the largest deciduous and mixed forest patches shrank slightly, but some deciduous and mixed forest patches expanded their boundaries over the 20 years. The PAFRAC, COHESION, and AI values for all the land cover types in the study area do not show significant changes between 1994 and 2013, which implies that there is no extensive land cover change in the study area from 1994 to 2013. In short, the landscape metrics listed in Table 5.4, together with visual examination of the SVM classification

results (Figure 5.4), suggest that the landscape spatial structure in the study area has not undergone significant changes since 1994, although the composition of forest at different locations has changed (such as changes from coniferous to deciduous or vice versa).

Table 5.4: Landscape metrics for land cover types in 1994 and 2013.

	Coniferous forest		Deciduous and mixed forest		Non-forest vegetation		Non-vegetated		water	
	1994	2013	1994	2013	1994	2013	1994	2013	1994	2013
CA (ha)	2866.41	2860.74	2190.6	2328.75	39.33	33.84	878.85	719.64	256.86	289.08
PLAND (%)	46.0	45.9	35.2	37.4	0.6	0.5	14.1	11.5	4.1	4.6
NP	404	389	527	509	19	21	601	623	108	104
LPI (%)	7.92	14.52	5.19	4.05	0.29	0.26	2.74	2.25	0.51	0.53
PAFRAC	1.45	1.51	1.44	1.42	1.27	1.34	1.45	1.46	1.23	1.28
COHESION	98.20	99.09	97.48	97.21	90.16	89.79	94.10	93.17	88.64	89.76
AI	85.13	81.73	80.59	81.58	83.65	82.19	67.28	63.14	83.06	82.60

5.4.2 Change/non-change analysis based on threshold-based change detection (TBCD)

Figure 5.7(*a*) presents the band 3 differenced image between the pre-processed band 3 and the fused band 3 of the 1994 image. The 2013 Pan image (Figure 5.7(*b*)), the pre-processed band 3 of the 1994 image (Figure 5.7(*c*)), and the fused 1994 band 3 image (Figure 5.7(*d*)) are also shown in Figure 5.7 for comparison. Non-vegetated areas respond as bright color, vegetated areas exhibit dark color, and water is black in the 2013 panchromatic and the 1994 band 3 images. Figure 5.7 (*d*) indicates that the spectral information of the changed areas in the 2013 image has been injected to the fused 1994 image via replacing the PC1 with the 2013 Pan image. In the differenced image (Figure 5.7(*a*)), the re-vegetated areas are darker (close to black); de-vegetated areas become brighter (close to white); and the unchanged areas present as grey (e.g. pixels within water bodies and forests). Areas A and B shown in Figure 5.7 are two extensively

changed areas that are illustrated in more detail in Figure 5.8 (Area A: the images in the first row; Area B: the images in the second row). For example, the pixels at Locations 1, 3, 5, 6, and 8 in Figure 5.8, where the land cover changed from vegetated in 1994 to non-vegetated in 2013 respond as a change from dark (in 1994 band 3 image, Figure 5.8 (*b*) and (*f*)) to bright (in 2013 Pan image, Figure 5.8 (*a*) and (*e*)), and present as bright in the band 3 fused image (Figure 5.8 (c) and (g)) and differenced image (Figure 5.8 (*d*) and (*h*)). Conversely, at Locations 2, 4, and 7 in Figure 5.8, pixels which changed from non-vegetated (bright in the 1994 band 3 image, Figure 5.8 (*b*) and (*f*)) to vegetated (dark in the 2013 Pan image, Figure 5.8(*a*) and (*e*)) present very dark in the band 3 differenced image (Figure 5.8(*d*) and (*h*)). Although the pixels at Location 1 in the PCCD process are identified as non-changed based on the ancillary information, the example at Location 1, demonstrates that the PCA-based TBCD method can detect land cover change based on the spectral information acquired at two different times. In addition, Locations 9 and 10 are cloudy areas in the 1994 image while not in the 2013 image, and these two clouds are successfully detected using the PCA-based TBCD approach. Thus, the differenced image generated by the PCA fused images highlights vegetated area changes by adding the spectral information in the 2013 Pan image to the 1994 band 3 image. Thus the PCA-based TBCD method is capable of distinguishing vegetation changes.

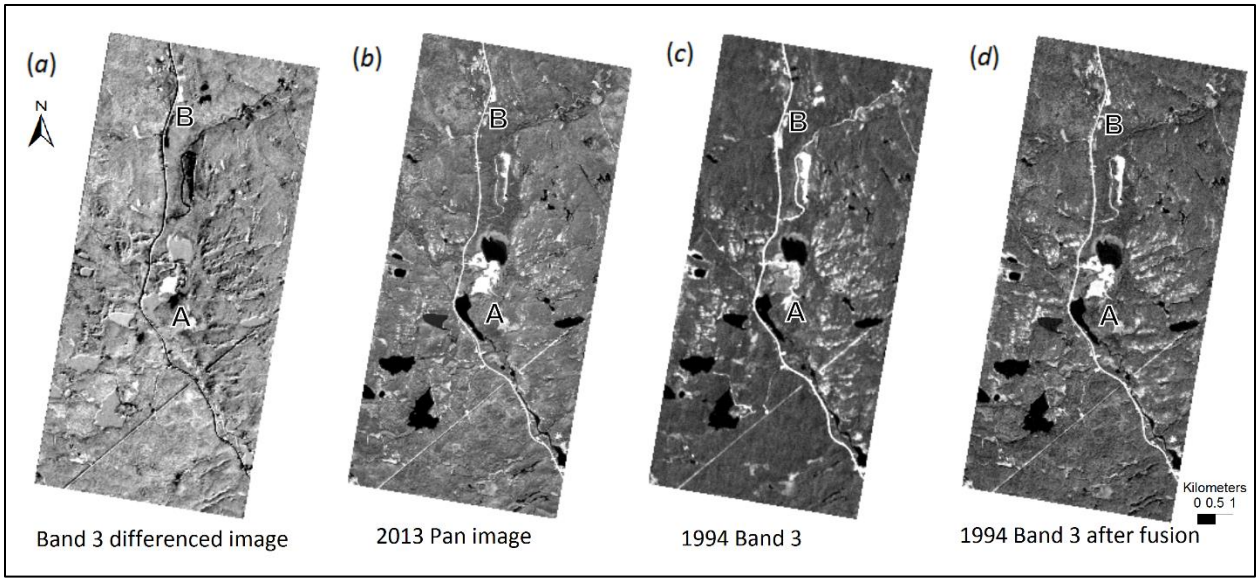

Figure 5.7: Images used in the TBCD process: (a) the Band 3 differenced image, (b) the 2013 Pan image, (c) the preprocessed 1994 Band 3, (d) the 1994 Band 3 after PCA image fusion. Areas A and B are two examples of the extensively changed areas and the more detail is shown in Figure 5.8.

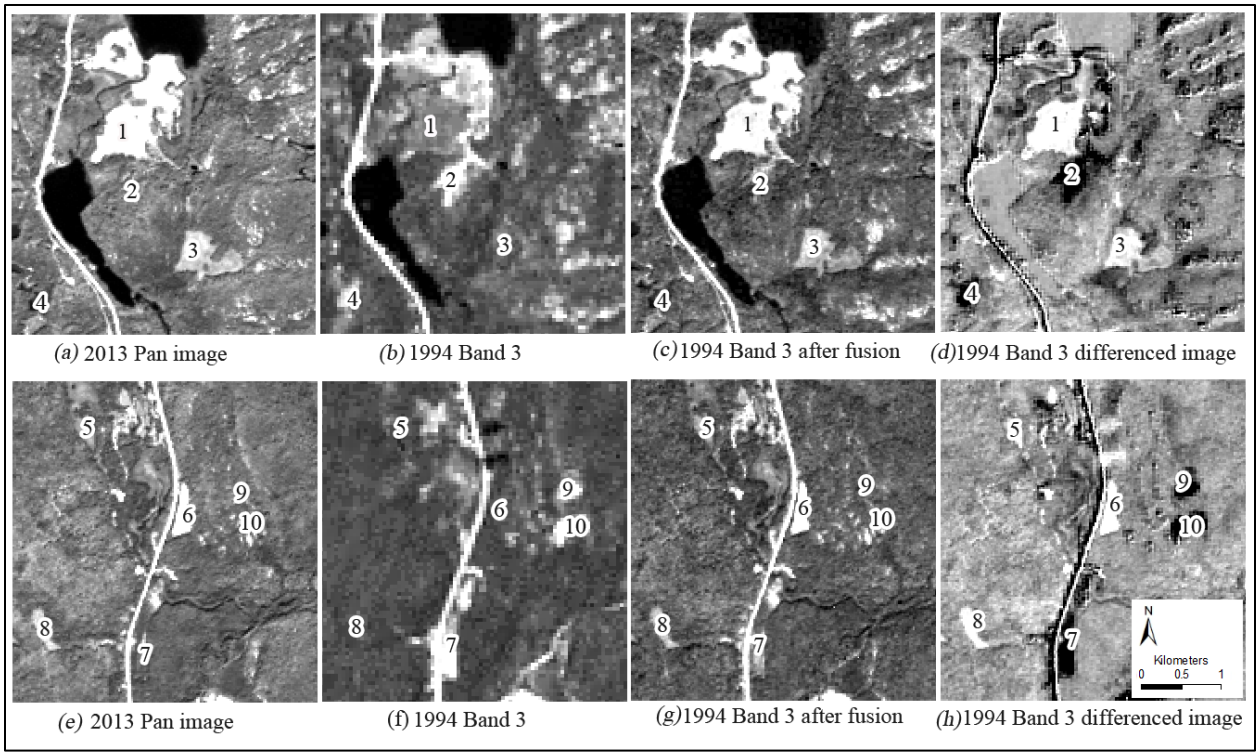

Figure 5.8: Examples of the TBCD results. Locations 1,3,5,6, and 8 are examples for de-vegetated areas and Locations 2, 4, and 7 are examples for re-vegetated areas. Locations 9 and 10 are clouds.

However the change detection map contains spurious changes for the pixels at edges of roads or feature boundaries such as water bodies (the explanation is left for the discussion below). Thus extra editing is needed to remove such spurious phenomena in the final change detection map. In addition, the abandoned Irvine Airstrip (Figure 5.1) is identified as re-vegetated in the TBCD result. This area is an abandoned Ministry of Natural Resources airstrip. The 1994 and 2013 NDVI images indicate that the NDVI values in this airstrip increased on average from 0.3 to 0.5 over the 20 years, which could imply this area was bare soil in 1994 and gradually re-vegetated. This slight change could result in the identification of this area as re-vegetated in the TBCD result, whereas in the PCCD, this area is classified as non-vegetated in both 1994 and 2013, and thus is identified as non-changed.

Figure 5.9 presents the vegetated area changes using the PCCD approach (*a*) and the PCA-based TBCD approach (*b*). The spurious changes along the roads and the changed pixels at the border of the study area in both of the change detection results are labelled as non-changed pixels. The TBCD map shows that over the 20 years, in total approximately 47.6 ha areas are detected as having changed to vegetated and 107 ha have changed to non-vegetated (if the croplands at the south of Irvine Lake are labelled as unchanged areas, as in the post-classification process, a total 92 ha would be labelled as a change to non-vegetated). The PCCD shows that in total 112 ha changed to vegetated area and 80.7 ha areas changed inversely. The vegetation change areas from the PCCD result are smaller than those listed in Table 5.3 because the spurious pixels along the road and the border of the study area are removed and edited to become non-changed pixels. In addition, many areas shown as re-vegetated or de-vegetated differ between the TBCD method and PCCD method. These differences could be caused by the unique

characteristics of these two different classification methods which are elaborated in the discussion.

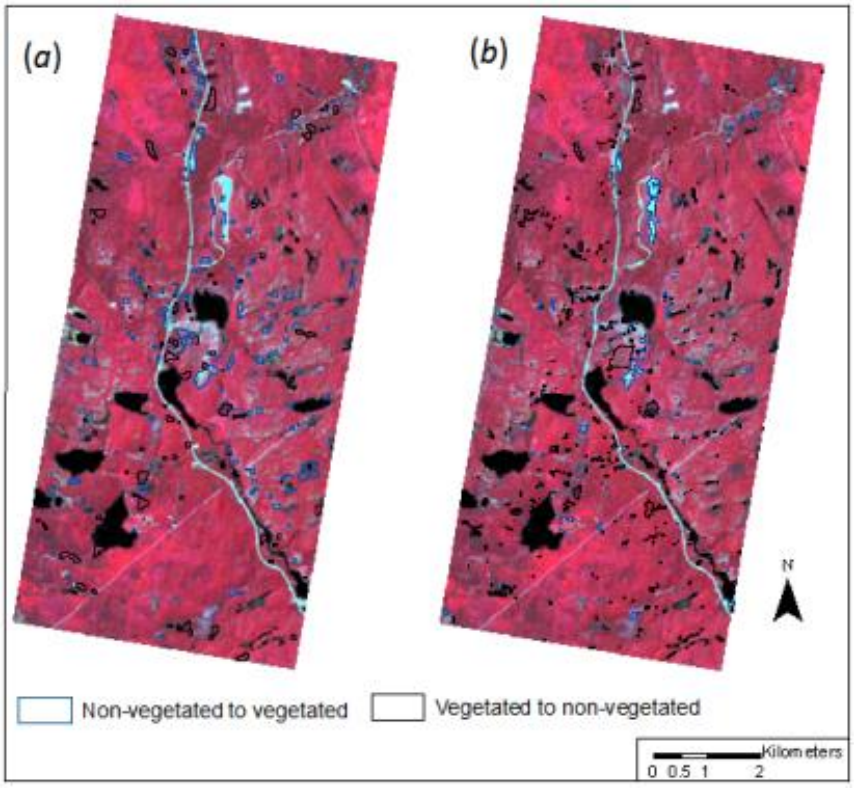

Figure 5.9: Changed and non-changed vegetated areas developed by the PCCD method (a) and the TBCD method (b). The false color composition of the 1994 summer image is used as the reference image.

5.5 Discussion

In the post-classification change detection (PCCD) approach, classification results often contain errors which contaminate change detection results. For example, all water bodies are well delineated by the SVM classifier in the 1994 and 2013 images; however, changes in precipitation and water levels in 1994 and 2013 will contribute to the differences in boundaries of lakes and wetlands. The shadows of trees (especially those at the edges of roads or boundaries of lakes and wetlands) will cause misclassification of these shadows as forest cover. For example, the Irvine Lake Airstrip is well distinguished in the 1994 classification result, but in the

2013 classification result this is labelled as forest. This difference could be caused by the effects of tree shadows or the different bandwidths between the Landsat-5 and Landsat-8 images. As such, in the change detection image, the pixels along this road are identified as changing from non-vegetated to forest.

Within the threshold-based change detection (TBCD), the PCA-fused image highlights vegetation change by infusing the information from the 2013 Pan image to the 1994 MS image and at the same time retaining the unchanged spectral information in the 1994 image. However, the TBCD method possesses several weaknesses. Firstly, in practice, the identification of thresholds is arbitrary and depends on the analyst's knowledge and experience. Secondly, due to the difference in spatial-resolution between the high-resolution Pan image and low-resolution MS image, pixels at edges of roads or feature boundaries (e.g. water bodies) could be spuriously labelled as changed features. For example, in the differenced results from the TBCD, edges of the roads are identified as changed only because of the difference in the spatial resolution between the Pan image and the MS image. The roads in the high resolution Pan image (Figure 5.8 (*a*) and (*e*)) are narrower and sharper than in the low resolution MS image (Figure 5.8 (*b*) and (*f*)). In addition, the pixels at the edges/boundaries are more spectrally mixed in the MS image. When fusing the Pan image with the MS image using PCA in the TBCD, the geometry of the roads is improved, which means the edges are more accurately outlined in the fused image (shown in Figure 5.8 (*c*) and (*g*)) and become more similar to those in the panchromatic image. This is one of the advantages of image fusion. However, based on the differenced image (Figure 5.8 (*d*) and (*h*)) which is derived by comparing the low resolution MS image to the fused MS image with improved spatial details, these unchanged edges are prone to misidentification as changed features (e.g. the dark pixels along the edge of the road shown in Figure 5.8 (*d*) and (*h*)

are identified as changed). These detected spurious changes need to be excluded from the changed pixel group and Figure 5.9(*b*) shows the TBCD result after this type of exclusion. However, further experiments in the future are required to evaluate the validity of the explanation. Thirdly, the TBCD only uses one band of the MS image or one derivative of the MS image (e.g. NDVI), thus, valuable spectral information in other bands (e.g. NIR and SWIR bands) is not utilized in the detection process, whereas the PCCD utilizes all available spectral information. This could limit the application of TBCD methods. For example in this study the TBCD method cannot identify changes between coniferous and deciduous forests. Finally, because we use the PCA method to operationalize the TBCD method, this restricts the choice of suitable bands for change detection to the three visible bands which are involved in the PCA analysis. The pool of possible suitable bands for change detection depends on the wavelength of the high-resolution Pan image for PCA fusion in the TBCD method. For example in this study, in order to avoid spectral distortion introduced by histogram mismatch between the Landsat-8 Pan image and the image to be replaced by the Pan image, the NIR and SWIR bands of the Landsat-5 image are not used in the PCA process due to their wavelength not overlapping with the Pan image. Thus they are not in the pool of potential suitable bands for the PCA fusion method in the TBCD.

There are various image fusion techniques available such as Brovey, IHS, Gram-Schmidt, and Wavelet-based fusion methods. Rokni et al. (2015) detected water surface change by applying a PCCD method on fused images and concluded that the Gram-Schmidt fusion method generated the best result when compared to the other three tested algorithms (i.e. the High Pass Filter, the Modified IHS, and the Wavelet-PC algorithms). Theoretically, the PCCD approach together with the four image fusion algorithms investigated in Rokni et al. (2015) could work in

this study to detect land cover changes; however, assessing those will be a future research topic. In this study, the RK-based fusion method proposed in Chapter 4 is not applied to the PCCD process because the 1994 Landsat-5 images did not contain a Pan band.

In this study, the PCCD methods take advantage of all available spectral information to generate thematic maps and then to create detailed change maps, thus the detected changes more likely occurred at the polygon level. Conversely, the TBCD result contains more small changed areas (even isolated pixels/artefacts). These differences are likely method related, in that, the TBCD method directly evaluates differences at the pixel level, whereas, the PCCD approach applies overlay analysis on thematic maps.

5.6 Conclusions

In this study, the land cover changes from 1994 to 2013 in the Addington Highlands area are quantified by applying a SVM-based post-classification change detection (PCCD) method and a PCA-based threshold-based change detection (TBCD) method using Landsat images. The PCCD results indicate that: 1) approximately 336 ha of non-vegetated areas were reforested and the total forest cover (mostly deciduous and mixed forests) increased by around 132 ha since 1994; and 2) the landscape metrics indicate little change in spatial structure since 1994, although the forest composition at different locations has changed.

Unlike the PCCD method that generates detailed land cover change information, the PCA-based TBCD method only detects changes in vegetated areas. The TBCD results indicate that approximately 47.6 ha non-vegetated areas converted to vegetated areas and 107 ha vegetated areas converted to non-vegetated, while the PCCD results show that 112 ha areas turned to vegetated areas and 80.7 ha converted to non-vegetated areas. These differences are likely caused by the different characteristics of the two change detection methods as discussed

above. However, more research is needed to determine the sensitivity of each result to the various assumptions.

In practice, the selection of an appropriate method for detecting changes using remotely sensed imagery depends on an array of factors which include the characteristics of the study area, the spatial resolution of the available remotely sensed imagery, the characteristics of the objects under investigation, and the objectives of the application. In addition, each change detection method has its own advantages. The TBCD method used in this study is simple, straightforward, and easy to implement and interpret, but it cannot provide detailed from-to change information. In addition, the TBCD approach has some weaknesses, including strict requirements on radiometric calibration and normalization, limited selection of a suitable image band, and selection of appropriate thresholds. The TBCD method is usually applied to a single band and the selection of the appropriate single band is limited by the spectrum of the high-resolution image which is used to pansharpen the MS image. Alternatively, PCCD minimizes the impacts of radiometric and geometric differences between multi-temporal images and takes full advantages of all available information to generate complete matrixes of from-to change information.

This study shows that the PCCD method can provide sufficient matrices of land cover change information with sufficient accuracy in a relatively remote forested area in Ontario, Canada. This study also demonstrates that the TBCD method is a simple and fast way to generate vegetated area change maps using the red band (band 3) of the Landsat-5 TM image and a Landsat-8 panchromatic image. Both of the PCCD and the TBCD methods work well in the study and the choice between them depends on the objective of the application.

Chapter 6 Summary and conclusions

The thesis investigates how remote sensing methods can be employed to assess and improve the use of free Landsat images in mapping land surface cover and change detection in Canadian remote regions. This investigation is accomplished from the following four perspectives: 1) determining the continuity of Landsat-8 images in mapping surficial information in Canadian Arctic environments, 2) selecting appropriate classification algorithms in the context of spectrally mixed pixels, minimal field sampling and ground truthing when mapping land cover, 3) applying geostatistically based image fusion techniques to improve Landsat image geometric details while retaining spectral integrity, and 4) examining the capabilities of post-classification and threshold-based methods in land cover change detection. Accordingly, four specific research objectives were formulated and listed in Chapter 1. This thesis research has successfully addressed these research objectives. Brief descriptions of the research work related to each research objective are provided below:

(1) The capability of Landsat-8 images in mapping Canada Arctic geological information was explored by comparing Landsat-8 to Landsat-7 images and by comparing Landsat-8 to SPOT-5, Landsat-7, and ASTER images in the western Minto Inlier on Victoria Island, Northwest Territories, Canada. The benefits of Landsat-8 (including 16-bit radiometric resolution, narrower bandwidths, and two new spectral bands) were examined. In addition, important bands from Landsat-7 and Landsat-8 images in mapping geological information were explored by using RF classification. This work demonstrates that Landsat-8 images can provide continuity with Landsat-7, and therefore, can going forward, serve as the major data source in mapping geological information in Canada's Arctic environments when Landsat-7 is decommissioned.

(2) The capabilities of various conventional and advanced image classification algorithms were investigated by applying a 10-iteration classification scheme on various satellite images in order to map geological information in Canada's North. Classification performance was compared visually and statistically using confusion matrices, uncertainty maps, and cross-comparison agreement. The effect of non-normally distributed training data on different classifiers was examined. The results indicate that when mapping geological information using Landsat images, the advanced classification algorithms (especially Neural Networks (NNs) and Support Vector Machine (SVM)) outperform Maximum Likelihood Classifier (MLC). In addition to well-established classification algorithms, a new Multiple Classification System (MCS) classification method was proposed to take advantage of the best characteristics of various data sources and thereby further improve classification accuracy.

(3) The systematic investigation of the RK (regression kriging)-based fusion method was conducted in two steps. The performance of the RK-based fusion methods using different combinations of linear regression methods and kriging approaches was firstly examined and compared to well-established conventional fusion methods such as Brovey and Gram-Schmidt in a small study area where pixel spectral values were normally distributed. Two optimal RK-fusion methods were then selected to fuse another Landsat image that contained non-normally distributed pixel values. The capability of various fusion methods in spatial improvement and spectral integrity was evaluated by visual examination, image quality index comparisons, and cluster analysis. The results indicated that the RK-based fusion methods are capable of fusing medium resolution

Landsat images and maintaining maximum spectral fidelity while providing sufficient resolution enhancement.

(4) The effectiveness of Landsat archive images in land cover change detection in the remote areas was assessed by applying both a post-classification (PCCD) method and a threshold-based change detection (TBCD) method. The PCCD method is implemented by firstly applying the SVM algorithm to classify two pairs of bi-season Landsat images with the resulting thematic maps overlaid in a GIS (Geographic Information System) to identify changes in deciduous, mixed, coniferous forests, non-forest vegetation, non-vegetated, and water from 1994 to 2013. In addition, the landscape metrics and the land cover change matrix were calculated to quantify land cover changes. The TBCD change detection method was implemented based on comparing fused multi-temporal Landsat images to the raw Landsat multispectral (MS) images. The PCA image fusion technique was used to merge Landsat MS images and panchromatic images acquired at different times. The thematic maps identifying vegetation change were developed by applying user-defined thresholds to the differenced images. This research suggests that a PCA-based image fusion method can be utilized to effectively synthesize Landsat images acquired at different times to detect cover changes.

In conclusion, the thesis research demonstrates the continuity of Landsat-8 images to extend the continuity of Landsat 5 and 7 and serve as a major data source for mapping surface cover in Canadian remote areas. The research also shows how the capabilities of two remote sensing methods, advanced classification algorithms and data fusion techniques, can improve the effective use of Landsat images for mapping land cover and detecting changes in the remote areas.

6.1 Research Contributions

This thesis concludes that Landsat 8 can be a major data source for mapping land cover in the future and that advanced classification algorithms and data fusion techniques can improve the effectiveness of Landsat images for mapping land cover and detecting changes in Canadian remote areas. The 16-bit radiometric resolution, narrower bandwidths and two new spectral bands in Landsat 8 do not present significant advantages to the mapping of geological information in Canadian Arctic environments. This research also reveals the importance of the near infrared (NIR) and short wavelength infrared (SWIR) bands of both the Landsat-7 and Landsat-8 images for classifying surficial materials in Canadian Arctic regions.

Furthermore, advanced classification algorithms, especially SVM, are effective in classifying geological information when training data are spectrally mixed and non-normally distributed. The well-designed cross-validation classification approach provided herein, presents an effective experimental scheme for the evaluation and comparison of various classifiers using different images. The findings also demonstrate that MLC is not an optimal classifier when non-normally distributed training data are employed, even though such an approach can generate acceptable classification results. In addition, a new MCS method is suggested that is capable of mapping surficial geological information. This method generates the majority classification, agreement and best band maps with satisfactory accuracies by leveraging the unique characteristics of various remotely sensed data used in the classification.

A further contribution to the Landsat community is demonstrating that the use of geostatistically based fusion methods outperform conventional image fusion approaches for maintaining spectral fidelity. The research systematically identifies the various trade-offs between spectral fidelity preservation and spatial detail improvement for the GLS (Generalized Linear Squares) and OLS (Ordinary Linear Squares) regression methods, for the UK (Universal

Kriging) and OK (Ordinary Kriging) kriging methods, and for the exponential and spherical variogram models in the RK fusion process. The RK fusion using the OLS regression and the OK kriging technique is recommended as optimal, from the perspectives of spatial resolution improvement, spectral fidelity preservation, and computational cost.

In addition, the thesis research demonstrates the effectiveness of the Landsat archive to detect land cover changes in Canadian remote areas. It also demonstrates that the SVM classifier can generate satisfactory classification maps which are the basis of successful post-classification change detection (PCCD) in a forested remote area. This research explores a new PCA fusion scheme to merge Landsat MS images and panchromatic images acquired at different times to identify changes.

Three typical Canadian landscapes were chosen in the research. However, the methodologies present in the paper can be applied in different areas. The 10-repetition cross validation approach can be implemented in image classification in either rural or urban areas, given the presence of spectrally mixed training data from medium resolution Landsat images. The PCCD and TBCD methods can also be applied to change detection practices in other regions. However, if the available training data in a PCCD process are normally distributed, the MLC method is preferred since it runs faster compared to the SVM method.

6.2 Suggestions for future research

The methodological research in this thesis contains a number of contributions to the remote sensing literature and extends methods that assess and improve the use of Landsat images in mapping the earth's surface and changes in remote areas. However, some specific research issues require further investigation:

(1) This research compares advanced classification algorithms to conventional classification algorithms when mapping Canadian remote areas. All the algorithms investigated in the research are pixel-based classifiers. The research on the application of object-based or hybrid classification methods (using both objects and pixels) in mapping land surface cover in remote areas will add useful knowledge for the use of Landsat images in Canada's north.

(2) When fusing images using the RK-based techniques, to what degree can surface topographic information such as a DEM add relevant information to improve image fusion results? Can other geostatistical based fusion methods, such as cokriging and area-to-point regression kriging (Wang et al. 2015), be used to successfully synthesize free Landsat images? These questions are open for further investigation.

(3) The thesis shows the PCA-fused images can be used in a TBCD process to detect vegetated area changes. Future research should include the investigation on the application of other image fusion methods such as Brovey, Gram-Schmidt (GS), and Wavelet-based to fuse multi-temporal Landsat image in a TBCD practice.

(4) A preliminary investigation on the use of RK-fused multi-temporal Landsat images in a TBCD process has been conducted in the course of the thesis research. The preliminary result shows that the RK-based fusion method proposed in Chapter 4 is not capable of fusing multi-temporal Landsat images to highlight changed features in a TBCD process. In future research, the current RK-based fusion method may need to add a function to adjust the weight or power of the panchromatic image in the fusion process in order to better inject the changed spectral information into the fused result. In addition, the RK-based fusion method is not applied in a PCCD process in this study due to the lack of

Landsat-5 panchromatic images. The future research on the RK-based fusion method should also focus on applying the RK-based fusion method in a PCCD practice.

Bibliography

Abedi, M., G. Norouzi, and A. Bahroudi. 2012. "Support Vector Machine for Multi-Classification of Mineral Prospectivity Areas." *Computers & Geosciences* 46: 272–283. doi:10.1016/j.cageo.2011.12.014.

Abrams, M., and S. Hook. 1995. *Aster Users Handbook*. Jet Propulsion Laboratory, Pasadena, CA, 133 p. Available from: http://gds.aster.ersdac.jspacesystems.or.jp/gds_www2000/service_e/u.tools_e/cross/GUIDEED.PDF.

Akar, Ö., and O. Güngör. 2012. "Classification of Multispectral Images Using Random Forest Algorithm." *Journal of Geodesy and Geoinformation* 1(2): 105–112. doi:10.9733/jgg.241212.1.

Alonso, M.C., J.A. Malpica, and A.Martinez de Agirre. 2011. "Consequences of the Hughes Phenomenon on some classification techniques." American Society for Photogrammetry and Remote Sensing 2011 Annual Conference, Milwaukee, WI, May 1-5th.

Alqurashi, A.F. and L. Kumar. 2013. "Investigating the use of remote sensing and GIS techniques to detect land use and land cover change: a review." *Advances in Remote Sensing 2(2):*193-204. doi:10.4236/ars.2013.22022

Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. 1976. "A land use and land cover classification system for use with remote sensor data." *USGS professional paper* 964. Sioux Falls, S.D., USA

Arora, M. K., K. C. Tiwari, and B. Mohanty. 2000. "Effect of Neural Network Variables on Image Classification." *Asian Pacific Remote Sensing and GIS Journal* 13: 1–11.

Banfield, R.E., L.O. Hall, K.W. Bowyer, and W.P. Kegelmeyer. 2007. "A comparison of decision tree ensemble creation techniques." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29: 173–180. dx.doi.org/10.1109/TPAMI.2007.250609.

Behnia, P., J. R. Harris, R. H. Rainbird, M. C. Williamson, and M. Sheshpari. 2012. "Remote Predictive Mapping of Bedrock Geology Using Image Classification of Landsat and SPOT Data, Western Minto Inlier, Victoria Island, Northwest Territories, Canada." *International Journal of Remote Sensing* 33 (21): 6876–6903. doi:10.1080/01431161.2012.693219.

Benediktsson, J.A., and P.H. Swain. 1992. “Consensus theoretic classification methods.” *IEEE transactions on Systems, Man, and Cybernetics* 22: 688–704, dx.doi.org/10.1109/21.156582.

Benediktsson, J.A., J. Chanussot, and M. Fauvel. 2007. “Multiple classifier systems in remote sensing: from basic to recent developments, *in* Kittler, H.J., and Roli, F., *eds.*, Multiple Classifier Systems.” *Proceedings of the 7th International Workshop*, Prague, Lecture notes in Computer Science 4472. Berlin: Springer.

Bhagat, V.S. 2012. “Use of remote sensing techniques for robust digital change detection of land: a review.” *Recent Patents on Space Technology* 2(2):123–144. doi: 10.2174/1877611611202020123

Biggar, S.F., K. Thome, J.T. McCorkel, and J.M. D’Amico. 2005. “Vicarious calibration of the ASTER sensor including crosstalk correction, *in* Butler, J.J., *ed.*, Earth Observing Systems X Earth Observing Systems X.” *Proceedings of the SPIE* 5882:408–415.

Borrelli, Pasquale, Luis Alfonso Sandia Rondón, and Brigitta Schütt. 2013. “The use of Landsat imagery to assess large-scale forest cover changes in space and time, minimizing false-positive changes.” *Applied Geography* 41: 147-157. doi:10.1016/j.apgeog.2013.03.010

Breiman, L. 1996. “Bagging predictors” *Machine Learning* 24: 123–140. dx.doi.org/10.1007/BF00058655.

Breiman, L. 2001. “Random Forests.” *Machine Learning 45* (1): 5–32. doi:10.1023/A:1010933404324.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. Monterey, CA: Wadsworth.

Caliński,T. and J. Harabasz. 1974. “A dendrite method for cluster analysis.” *Communications in Statistics – Theory and Methods* 3(1): 1–27. doi:10.1080/03610927408827101

Campbell, J E; Harris, J R; Huntley, D H; McMartin, I; Wityk, U; Dredge, L A; Eagles, S. 2013 “Remote predictive mapping of surficial earth materials: Wager Bay North area, Nunavut - NTS 46-E (N), 46-K (SW), 46-L, 46-M (SW), 56-H (N), 56-I and 56-J (S) Geological Survey of Canada.” Open File 7118, 42 pages. doi:10.4095/293158

Chavez, P. S., D. J. MacKinnon. 1994. “Automatic Detection of Vegetation Changes in the Southwestern United States Using Remotely Sensed Images.” *Photogrammetric Engineering and Remote Sensing* 60 (5): 571-583.

Chen, D. 2008. “A Standardized Probability Comparison Approach for Evaluating and Combining Pixel-Based Classification Procedures.” *Photogrammetric Engineering & Remote Sensing* 74 (5): 601–609. doi:10.14358/PERS.74.5.601.

Chen, X., T. A. Warner, and D. J. Campagna. 2007. “Integrating Visible, Near-Infrared and Short-Wave Infrared Hyperspectral and Multispectral Thermal Imagery for Geological Mapping at Cuprite, Nevada.” *Remote Sensing of Environment* 110 (3): 344–356. doi:10.1016/j.rse.2007.03.015.

Chen, G.,G.J. Hay, L.M.T. Carvalho, M.A. Wulder. 2012. “Object-based change detection.” *International Journal of Remote Sensing* 33(14): 4434–4457. doi: 10.1080/01431161.2011.648285

Chiesa, C. C., and W. A. Tyler. 1990. “Data fusion of off-nadir SPOT panchromatic imageswith other digital data sources.” *Technical Papers 1990, ACSM-ASPRS Annual Convention, Image Processing and Remote Sensing* 4: 86-98.

CIRCA. 2001. Land cover for agricultural regions of Canada, circa 2000. Agriculture and Agri-Food Canada.

Civco, D. L. 1989. “Topographic Normalization of Landsat Thematic Mapper Digital Imagery.” *Photogrammetric Engineering and Remote Sensing* 55(9):1303-1309

Choi, M. 2006. “A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter.” *IEEE Transactions on Geoscience and Remote Sensing* 44 (6): 1672-1682. doi: 10.1109/TGRS.2006.869923

Congalton R.G. and K. Green. 2008. *Assessing the accuracy of remotely sensed data: principles and practices*. FL: CRC Press

Coppin, P., I.Jonckheere, K. Nackaerts, B. Muys, E.Lambin. 2004. “Digital change detection methods in ecosystem monitoring: a review.” *International journal of remote sensing* 25(9): 1565-1596.

Chuvieco, E., and Huete, A. 2010. *Fundamentals of Satellite Remote Sensing*. Boca Raton, FL: CRC Press.

Cracknell, M. J., and A. M. Reading. 2014. “Geological Mapping Using Remote Sensing Data: A Comparison of Five Machine Learning Algorithms, Their Response to Variations in the Spatial Distribution of Training Data and the Use of Explicit Spatial Information.” *Computers & Geosciences* 63: 22–33. doi:10.1016/j.cageo.2013.10.008.

Cressie, N. 1993. *Statistics for Spatial Data, revised edition*. New York: John Wiley & Sons.

Crisci, C., B. Ghattas, and G.Perera. 2012. "A review of supervised machine learning algorithms and their applications to ecological data." *Ecological Modelling* 240: 113–122, dx.doi.org/10.1016/j.ecolmodel. 2012.03.001.

Del Castillo, E. M., García-Martin, A., Aladrén, L. A. L. 2015. "Evaluation of forest cover change using remote sensing techniques and landscape metrics in Moncayo Natural Park (Spain)." *Applied Geography* 62: 247-255. doi:10.1016/j.apgeog.2015.05.002

Demšar, J. 2006. "Statistical Comparisons of Classifiers over Multiple Data Sets." *Journal of Machine Learning Research* 7: 1–30.

Deng, J. S., K. Wang, Y.H. Deng, G.J. Qi. 2008. "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data." *International Journal of Remote Sensing* 29(16): 4823-4838. doi: 10.1080/01431160801950162

Desnoyers, D.W., and J.R. Harris. 2003. "Remote Predictive Mapping - A new tool for geoscience mapping (abstract)." *31st Yellowknife Geoscience Forum*. 19–21 November

Doan, H.T.X., and G.M. Foody. 2007. "Increasing soft classification accuracy through the use of an ensemble of classifiers." *International Journal of Remote Sensing* 28: 4609–4623, doi:10.1080/014311607 01244872.

Dong, J., D. Zhuang, Y. Huang, J. Fu. 2009. "Advances in multi-sensor data fusion: Algorithms and applications." *Sensors* 9(10): 7771-7784. doi:10.3390/s91007771

DRAPE : Orthoimagery. 2008. [computer file]. Ontario Ministry of Natural Resources.

Drury, S. 1993. *Image Interpretation in Geology, 2nd Edition*. London, UK: Chapman and Hall.

Du, P. S.Liu, P.Gamba, K.Tan, J. Xia. 2012. "Fusion of difference images for change detection over urban areas." *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal* 5(4): 1076-1086.

Dube, T., O. Mutanga. 2015. "Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas." *ISPRS Journal of Photogrammetry and Remote Sensing* 108: 12-32. doi: 10.1016/j.isprsjprs.2015.06.002

El Janati, M., A. Soulaimani, H. Admou, N. Youbi, A. Hafid, and K. P. Hefferan. 2014. "Application of ASTER Remote Sensing Data to Geological Mapping of Basement Domains in Arid Regions: A Case Study from the Central Anti-Atlas, Iguerda Inlier,

Morocco." *Arabian Journal of Geosciences* 7 (6): 2407–2422. doi:10.1007/s12517-013-0945-y.

El Rakaiby, M.L. 1995. "The use of enhanced Landsat-TM image in the characterization of uraniferous granitic rocks in the Central Eastern Desert of Egypt." *International Journal of Remote Sensing* 16:1063–1074. Doi:10.1080/01431169508954463.

Estoque, R.C., Y. Murayama. 2015. "Classification and change detection of built-up lands from Landsat-7 ETM+ and Landsat-8 OLI/TIRS imageries: A comparative assessment of various spectral indices." *Ecological Indicators* 56: 205-217. doi: 10.1016/j.ecolind.2015.03.037

Fauvel, M., J. Chanussot, and J. A. Benediktsson. 2006. "Evaluation of Kernels for Multiclass Classification of Hyperspectral Remote Sensing Data." *In ICASSP 2006, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, May 14–19, 813–816. New York: IEEE.

Foody, G. M., and A. Mathur. 2006. "The Use of Small Training Sets Containing Mixed Pixels for Accurate Hard Image Classification: Training on Mixed Spectral Responses for Classification by a SVM." *Remote Sensing of Environment* 103: 179–189. doi:10.1016/j.rse.2006.04.001.

Ford K., J.R. Harris, R. Shives, J. Carson, and J. Buckle. 2008. "Gamma Ray Spectrometry, *in* Harris, J.R., *ed.*, Remote Predictive Mapping: an aid for northern mapping: Geological Survey of Canada." Open File 5643, p. 53–97.

Forkuo, E. K. and A. Frimpong. 2012. "Analysis of forest cover change detection." *International Journal of Remote Sensing Applicaions* 2(4): 82-92.

Frate, L and M. L. Carranza. 2013. "Quantifying landscape-scale patterns of temperate forests over time by means of neutral simulation models." *ISPRS International Journal of Geo-Information* 2(1): 94-109. doi: 10.3390/ijgi2010094

Freund, Y., and R.E. Schapire. 1999. "A short introduction to boosting." *Journal of Japanese Society for Artificial Intelligence* 14: 771–780.

Gangkofner, U.G., P. S. Pradhan, and D. W. Holcomb. 2007. "Optimizing the high-pass filter addition technique for image fusion." *Photogrammetric Engineering and Remote Sensing* 73(9):1107-1118. doi:10.14358/pers.73(9):1107-1118

Ginevan,M. 1979. "Testing land-use map accuracy: another look." *Photogrammetric Engineering and Remote Sensing*, 45: 1371-1377

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson. 2006. "Random Forests for Land Cover Classification." *Pattern Recognition Letters* 27: 294–300. doi:10.1016/j.patrec.2005.08.011.

Glikson, A.Y., and J.W. Creasey. 1995. "Application of Landsat-5 TM imagery to mapping of the Giles Complex and associated granulites, Tomkinson Range, western Musgrave Block, central Australia." *AGSO Journal of Australian Geology and Geophysics* 16:173–193.

Gong, M., Z. Zhou, and J. Ma. 2012. "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering." *Image Processing, IEEE Transactions* 21(4): 2141-2151.

Gonzalez-Audicana,M., J. L. Saleta, R. G. Catalan, and R. Garcia. 2004. "Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition." *IEEE Transactions on Geoscience and Remote Sensing* 42(6):1291 - 1299. doi: 10.1109/TGRS.2004.825593

Grunsky, E., Harris, J.R., and McMartin, I. 2009. "Predictive Mapping of Surficial Materials, Schultz Lake Area (NTS 66A), Nunavut, Canada." *Society of Economic Geologists* 16: 177-198

Guo, Q. and S. Liu. 2011. "Performance analysis of multi-spectral and panchromatic image fusion techniques based on two wavelet discrete approaches." *Optik – International Journal for Light and Electron Optics* 122 (9): 811-819. doi: 10.1016/j.ijleo.2010.06.008

Gupta, R.P.1991. *Remote Sensing Geology*. Berlin: Springer-Verlag. doi:10.1007/978-3-662-12914-2.

Haack, Barry, Nevin Bryant, and Steven Adams. 1987. "An assessment of Landsat MSS and TM data for urban and near-urban land-cover digital classification." *Remote Sensing of Environment* 21(2): 201-213.

Hadigheh, S. M. H., and H. Ranjbar. 2013. "Lithological Mapping in the Eastern Part of the Central Iranian Volcanic Belt Using Combined ASTER and IRS data." *Journal of the Indian Society of Remote Sensing* 41 (4): 921–931. doi:10.1007/s12524-013-0284-1.

Hame, T. H. 1988. “Interpretation of forest changes from satellite scanner imagery. In Satellite imageries for forest inventory and monitoring; experiences, methods, perspectives, Research Notes No. 21.” *Department of Forest Mensuration and Management, University of Helsinki,* Helsinki, Finland, pp. 31–42.

Harris, J.R., *ed.* 2008. “Remote Predictive Mapping: an aid for northern mapping. Geological Survey of Canada,” Open File 5643, 306 p., 1 DVD. Doi:10.4095/225997.

Harris, J.R., D. Rogge, R. Hitchcock, O. Ijewliw, and D. Wright. 2005. “Mapping lithology in Canada’s Arctic: application of hyperspectral data using the minimum noise fraction transformation and matched filtering.” *Canadian Journal of Earth Sciences* 42: 2173–2193, http://dx.doi.org/10.1139/e05-064.

Harris, J.R., E.M. Schetselaar, T. Lynds, and E.A. de Kemp. 2008a. “Remote predictive mapping: A strategy for geological mapping of Canada‘s North, Chapter 2, *in* Harris, J.R., *ed.,* Remote Predictive Mapping: an aid for northern mapping: Geological Survey of Canada.” Open File 5643, p. 5–27. doi:10.4095/225997.

Harris J.R., E. Schestlaar, , E.A. de Kempand, , M. St-Onge. 2008b. “LANDSAT, Magnetic and Topographic Data for Regional Lithological Mapping –Southeast Baffin Island, Case study, *in* Harris, J.R., *ed.,* Remote Predictive Mapping: an aid for northern mapping: Geological Survey of Canada.” Open File 5643, p. 173–182, doi:10.4095/225997.

Harris, J.R., E. Martel, M. Currie, K. Pierce, M. Pilkington, and P.Keating. 2008c. “Snowbird Lake –NWT: Application of RPM methods, *in* Harris, J.R., *ed.*, Remote Predictive Mapping: an aid for northern mapping: Geological Survey of Canada.” Open File 5643, p. 161–171. doi:10.4095/225997.

Harris, J.R., K.L. Ford, and B.W. Charbonneau. 2009. “Application of gamma ray spectrometer data for lithological mapping in a cordilleran environment, Sekwi region, NWT.” *Special Issue of the Canadian Journal of Remote Sensing* 35: S12–S30. doi:10.5589/m09-022.

Harris, J.R., R. McGregor, and P. Budkewitsch. 2010. “Geological analysis of hyperspectral data over southwest Baffin Island: methods for producing spectral maps that relate to variations in surface lithologies.” *Canadian Journal of Remote Sensing* 36: 412–435. doi:10.5589/m10-072.

Harris, J.R., E. Schetselaar, and P. Behnia. 2012a. “Remote Predictive Mapping: An Approach for the Geological Mapping of Canada’s Arctic.” *Earth Sciences: InTech*: 495–524. Available from: http://cdn.intechopen.com/pdfswm/27604.pdf.

Harris, J. R., E. Grunsky, J. He, D. Gorodetzky, and N. Brown. 2012b. “A Robust, Cross-Validation Classification Method (RCM) for Improved Mapping Accuracy and Confidence Metrics.” *Canadian Journal of Remote Sensing* 38 (1): 69–90. doi:10.5589/m12-013.

Harris, J.R. W. Parkinson, A.Dyke, D. Kerr, H. Hussell, S.Eagles, M.Richardson, E.Grunsky. 2012c. “Predictive surficial geological mapping of Hall Peninsula and Foxe Basin Plateau, Baffin Island using LANDSAT and DEM data.” Open file 7038, Natural Resources Canada.

Harris, J. R., J. He, R. Rainbird, and P. Behnia. 2014. “A Comparison of Different Remotely Sensed Data for Classification Lithology in Canada’s Arctic: Application of the Robust Classification Method and Random Forests.” *Geoscience Canada* 41 (4): 557–584.

Harris, J. R., L. Wickert, T. Lynds, P. Behnia, R. Rainbird, E. Grunsky, R. McGregor, and E. Schetselaar. 2011. “Remote Predictive Mapping 3. Optical Remote Sensing—A Review for Remote Predictive Geological Mapping in Northern.” *Geoscience Canada* 38 (2): 49–83.

Hayes, D. J., and S. A. Sader. 2001. “Comparison of change-detection techniques for monitoring tropical forest clearing and vegetation regrowth in a time series.” *Photogrammetric engineering and remote sensing* 67(9): 1067-1075.

He, J., J.R. Harris, M. Sawada, P. Behnia. 2015. “A comparison of classification algorithms using Landsat-7 and Landsat-8 data for mapping lithology in Canada’s Arctic.” *International Journal of Remote Sensing* 36(8): 2252-2276.

Hecheltjen, A., F. Thonfeld, and G. Menz. 2014. “Recent Advances in Remote Sensing Change Detection–A Review.” *Land Use and Land Cover Mapping in Europe*. Netherlands:Springer. 145-178. doi: 10.1007/978-94-007-7969-3_10

Hengl,T., G. M. Heuvelink, D. G. Rossiter. 2007. “About regression-kriging: From equations to case studies.” *Computers & Geosciences* 33(10): 1301–1315. doi:10.1016/j.cageo.2007.05.001

Hengl, T. 2009. *A practical guide to geostatistical mapping*. Amsterdam: University of Amsterdam.

Huang, C., L. S. Davis, and J. R. G. Townshend. 2002. "An Assessment of Support Vector Machines for Land Cover Classification." *International Journal of Remote Sensing* 23 (4): 725–749. doi:10.1080/01431160110040323.

Huang, C. and K. Song. 2012. "Forest-Cover Change Detection Using Support Vector Machines." *Remote Sensing of Land Use and Land Cover*. P:191–206. doi: 10.1201/b11964-16

Hughes, G.F. 1968. "On the mean accuracy of statistical pattern recognizers." *IEEE Transactions on Information Theory* 14: 55–63, doi:10.1109/TIT.1968.1054102.

Hulbert, L.J., R.H. Rainbird, C.W. Jefferson, and P. Friske. 2005. "Map of mafic and ultramafic bodies related to the Franklin magmatic event, Minto Inlier, Victoria Island: Geological Survey of Canada." Open File No. 4928. Doi:10.4095/220616.

Hussain, M. D. Chen, A. Cheng, H. Wei, D.Stanley. 2013. "Change detection from remotely sensed images: From pixel-based to object-based approaches." *ISPRS Journal of Photogrammetry and Remote Sensing* 80: 91-106. doi:10.1016/j.isprsjprs.2013.03.006

Issaks, E.H. and R. M. Srivastava. 1989. *Applied Geostatistics*. New York: Oxford University Press.

Jagalingam, P., and A. V. Hegde. 2015. "A Review of Quality Metrics for Fuse Image." *Aquatic Procedia* 4: 133-142. doi: 10.1016/j.aqpro.2015.02.019

Jensen, J.R. 2005. *Introductory digital image processing: a remote sensing perspective.* Upper Saddle River, NJ: Pearson Prentice Hall, 3rd ed:.

Jones, H.G., and R.A. Vaughan. 2010. *Remote sensing of vegetation: Principles, Techniques, and Applications.* Oxford University Press 384 p.

Kalinowski, A., and S. Oliver. *compilers.* 2004. ASTER mineral index processing manual: Remote Sensing Applications, Geoscience Australia, internal report 36, 37 p., http://www.ga.gov.au/image_cache/GA7833.pdf.

Kavak, K.S. 2005. "Determination of palaeotectonic and neotectonic features around the Menderes Massif and the Gediz Graben (western Turkey) using Landsat TM image." *International Journal of Remote Sensing* 26:59–78, doi:10.1080/01431160410001709994.

Kavzoglu, T., and I. Colkesen. 2009. “A Kernel Functions Analysis for Support Vector Machines for Land Cover Classification.” *International Journal of Applied Earth Observation and Geoinformation* 11: 352–359. doi:10.1016/j.jag.2009.06.002.

Klonus,S. and M. Ehlers. 2009. “Performance of evaluation methods in image fusion.” *Information Fusion, 2009. FUSION'09. 12th International Conference on. IEEE*, 2009.

Kitanidis, P. 1993. “Generalized covariance functions in estimation.” *Mathematical Geology* 25(5): 525–540.

Kuemmerle, T., O. Chaskovskyy, J. Knorn, V.C. Radeloff, I. Kruhlov, W.S. Keeton,nd P. Hostert. 2009. “Forest cover change and illegal logging in the Ukrainian Carpathians in the transition period from 1988 to 2007.” *Remote Sensing of Environment* 113:1194–1207.

Laben, CA and B. V. Brower. 2000. “Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening.” *Technical Report, US Patent No. 6011875* Eastman Kodak Company.

LaRocque, A., B. Leblon, J. Harris, C. Jefferson, V. Tschirhart, and Y. Shelat. 2012. “Surficial Materials Mapping in Nunavut, Canada with Multibeam RADARSAT-2 Dual-Polarization C-HH and C-HV, LANDSAT-7 ETM+, and DEM Data.” *Canadian Journal of Remote Sensing* 38 (3): 281–305. doi:10.5589/m12-020.

Leverington, D.W. 2001. Discriminating lithology in Arctic environments from Earth orbit: an evaluation of satellite imagery and classification algorithms: Unpublished PhD thesis, Department of Geological Sciences, University of Manitoba, MB.

Leverington, D. W. 2010. “Discrimination of Sedimentary Lithologies Using Hyperion and Landsat Thematic Mapper Data: A Case Study at Melville Island, Canadian High Arctic.” *International Journal of Remote Sensing* 31 (1): 233–260. doi:10.1080/01431160902882637.

Leverington, D. W., and W. M. Moon. 2012. “Landsat-TM-Based Discrimination of Lithological Units Associated with the Purtuniq Ophiolite, Quebec, Canada.” *Remote Sensing* 4 (5): 1208–1231. doi:10.3390/rs4051208.

Li, J. and A. D. Heap. 2008. *A Review of Spatial Interpolation Methods for Environmental Scientists*. Canberra: *Geoscience Australia*.

Li, N., M. Frei, and W. Altermann. 2011. “Textural and Knowledge-Based Lithological

Classification of Remote Sensing Data in Southwestern Prieska Sub-Basin, Transvaal Supergroup, South Africa." *Journal of African Earth Sciences* 60 (4): 237–246. doi:10.1016/j.jafrearsci.2011.03.002.

Lillesand, M.L., R.W. Kiefer, and J.W. Chipman. 2004. *Introduction to Remote Sensing.* John Wiley and Sons: Canada, 763 p.

Liu, T., and X. Yang. 2015. "Monitoring land changes in an urban area using satellite imagery, GIS and landscape metrics." *Applied Geography* 56: 42-54. doi:10.1016/j.apgeog.2014.10.002

Liu, H, and Q. Weng. 2013. "Landscape metrics for analysing urbanization-induced land use and land cover changes." *Geocarto International* 28(7): 582-593. DOI: 10.1080/10106049.2012.752530

Liu, Y. S. Nishiyama, and T. Yano. 2004. "Analysis of four change detection algorithms in bi-temporal space with a case study." *International Journal of Remote Sensing* 25(11): 2121-2139.

Lorenz, H. 2004. "Integration of Corona and Landsat Thematic Mapper data for bedrock geological studies in the high Arctic." *International Journal of Remote Sensing* 25: 5143–5162.doi:10.1080/01431160410001705097.

Loizzo, R., L. G. Sylos, M. Pappalepore, P. Pieri, G. Pasquariello, and M. Antoninetti. 1995. "Multitemporal and Multisensor Signatures Evaluation for Lithologic Classification." *International Geoscience and Remote Sensing Symposium (IGARSS)* 3: 2209–2211.

Lu, D., P. Mausel, E. Bronaizio, E. Moran. 2004. "Change detection techniques." *International journal of remote sensing* 25(12): 2365-2401. doi: 10.1080/0143116031000139863

Lu, D., P. Mausel, M. Batistella, E. Moran. 2005. "Land-cover binary change detection methods for use in the moist tropical region of the Amazon: a comparative study." *International Journal of Remote Sensing* 26(1): 101-114.doi: 10.1080/01431160410001720748

Lu, D., M. Batistella, and E. Moran. 2008. "Integration of Landsat TM and SPOT HRG images for vegetation change detection in the Brazilian Amazon." *Photogrammetric engineering and remote sensing* 74(4): 421-430.

Lu, D., G. Li, and E. Moran. 2014. "Current situation and needs of change detection techniques." *International Journal of Image and Data Fusion* 5(1):13-38. doi: 10.1080/19479832.2013.868372

Macias, L.F. 1995. “Remote sensing of mafic-ultramafic rocks: Examples from Australian Precambrian terranes.” *AGSO Journal Australian Geology and Geophysics* 16: 163–171.

Mallat, S. G. 1989. “A theory for multiresolution signal decomposition: the wavelet representation.” *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 11(7): 674-693.

McLeman R. 2010. “Impacts of population change on vulnerability and the capacity to adapt to climate change and variability: a typology based on lessons from a hard country.” *Population and Environment* 31(5): 286-316.

Memarsadeghi, N., L. M. Moigne, D. M. Mount, and J. Morisette. 2005. “A new approach to image fusion based on cokriging.” *The Eighth International Conference on Information Fusion, July, 25–29, PA, USA: Philadelphia*. doi: 10.1109/ICIF.2005.1591912

Margono, Belinda Arunarwati, S. Turubanova, I. Zhuravleva,P. Potapov,A. Tyukavina,A. Baccini, S. Goetz, and M.C. Hansen. 2012. “Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010.” *Environmental Research Letters* 7(3): 034010.

Martel, E., J.R. Harris, M. Currie, and K. Pierce. 2005. “Snowbird Lake (NTS 65D) Remote Predictive Mapping and Geoscience Data compilation: Northwest Territories.” Open File 2005-08, 1CD-ROM.

Mas, J. F., and J. J. Flores. 2008. “The Application of Artificial Neural Networks to the Analysis of Remotely Sensed Data.” *International Journal of Remote Sensing* 29 (3): 617–663. doi:10.1080/01431160701352154.

Mather, P.M. 1999. *Computer Processing of Remotely Sensed Images, 2nd Edition*. Chichester : John Wiley and Sons.

Mei, A., Manzo, C., Fontinovo, G., Bassani, C., Allegrini, A., Petracchini, F. 2015. “Assessment of land cover changes in Lampedusa Island (Italy) using Landsat TM and OLI data.” *Journal of African Earth Sciences*. DOI: 10.1016/j.jafrearsci.2015.05.014

Meng, Q., B. Borders, and M. Madden. 2010. “High-resolution satellite image fusion using regression kriging.” *International Journal of Remote Sensing* 31(7): 1857-1876. doi: 10.1080/01431160902927937

Menze, B.H., B.M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F.A. Hamprecht. 2009. “A comparison of random forest and its Gini importance with standard

chemometric methods for the feature selection and classification of spectral data." *BMC Bioinformatics* 10(1):213. Doi:10.1186/1471-2105-10-213.

Mondal, A., S. Kundu, S. K. Chandniha, R. Shukla, and P. K. Mishra. 2012. "Comparison of Support Vector Machine and Maximum Likelihood Classification Technique Using Satellite Imagery." *International Journal of Remote Sensing and GIS* 1(2): 116–123.

Mountrakis, G., J. Im, and C. Ogole. 2011. "Support Vector Machines in Remote Sensing: A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 66: 247–259. doi:10.1016/j.isprsjprs.2010.11.001.

Murtagh, F. and P. Legendre. 2011. "Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm." *arXiv preprint arXiv*:1111.6285.

Mwaniki, M. W., M. S. Moeller, and G. Schellmann. 2015. "A comparison of Landsat 8 (OLI) and Landsat 7 (ETM+) in mapping geology and visualising lineaments: A case study of central region Kenya." *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Neel, M. C., K. M., and S. A. Cushman. 2004. "Behavior of class-level landscape metrics across gradients of class aggregation and area." *Landscape ecology* 19(4): 435-455.doi: 10.1023/B:LAND.0000030521.19856.cb

Nikolakopoulos,K. G.2008. "Comparison of nine fusion techniques for very high resolution data." *Photogrammetric Engineering and Remote Sensing* 74(5): 647-659. doi: 10.14358/PERS.74.5.647

Nitze, I., U. Schulthess, and H. Asche. 2012. "Comparison of Machine Learning Algorithms Random Forest, Artificial Neural Network and Support Vector Machine to Maximum Likelihood for Supervised Crop Type Classification." *In Proceeding of the 4th International Conference on Geographic Object Based Image Analysis(GEOBIA),* Rio de Janeiro, May 7–9.

Nunez, J., X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol. 1999. "Multiresolution-based image fusion with additive wavelet decomposition." *Geoscience and Remote Sensing,* IEEE Transactions on 37(3): 1204-1211

Odeh,I., Mcbratney,A., and D. Chittleborough.1994. "Spatial prediction of soil properties from landform attributes derived from a digital elevation model." *Geoderma* 63 (3):197-214.

Odeh,I., A. Mcbratney, D. Chittleborough. 1995. “Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging.” *Geoderma* 67(3): 215-226. doi: 10.1016/0016-7061(95)00007-B

Ontario Ministry of Natural Resources. Provincial Digital Elevation Model - Tiled Dataset, version 2.0.0 [computer file]. Peterborough : Ontario Ministry of Natural Resources, 2005-2006.

Oommen, T., D. Misra, N.K.C. Twarakavi, A. Prakash, B. Sahoo, and S. Bandopadhyay. 2008. “An Objective Analysis of Support Vector Machine Based Classification for Remote Sensing.” *Mathematical Geoscience* 40:409–424. doi:10.1007/s11004-008-9156-6.

Otukei, J. R., and T. Blaschke. 2010. “Land Cover Change Assessment Using Decision Trees, Support Vector Machines and Maximum Likelihood Classification Algorithms.” *International Journal of Applied Earth Observation and Geoinformation* 12: S27–S31. doi:10.1016/j.jag.2009.11.002.

Pal, M. 2005. “Random Forest Classifier for Remote Sensing Classification.” *International Journal of Remote Sensing* 26 (1): 217–222. doi:10.1080/01431160412331269698.

Pal, M., and P. M. Mather. 2005. “Support Vector Machines for Classification in Remote Sensing.” *International Journal of Remote Sensing* 26 (5): 1007–1011. doi:10.1080/01431160512331314083.

Paola, J. D., and R. A. Schowengerdt. 1995. “A Review and Analysis of Backpropagation Neural Networks for Classification of Remotely-Sensed Multi-Spectral Imagery.” *International Journal of Remote Sensing* 16 (16): 3033–3058. doi:10.1080/01431169508954607.

Pardo- Igúzquiza,E., M Chica-Olmo,and P M Atkinson. 2006. “Downscaling cokriging for image sharpening.” *Remote Sensing of Environment* 102: 86-98. doi: 10.1016/j.rse.2006.02.014

Patel, N., and K. K. Rampal. 1993. “Lithological Discrimination of Sun-Facing and Shadowed Portions of Aravalli Region.” *Advances in Space Research* 13 (11): 91–94. doi:10.1016/0273-1177(93)90207-R.

Peiman, Reihaneh. 2011. “Pre-classification and post-classification change-detection techniques to monitor land cover and land-use change using multi-temporal Landsat imagery: a case study on Pisa Province in Italy.” *International journal of remote sensing* 32(15): 4365-4381.doi: 10.1080/01431161.2010.486806

Peña, S.A., and M.G. Abdelsalam. 2006. “Orbital remote sensing for geological mapping in southern Tunisia: Implication for oil and gas exploration.” *Journal of African Earth Science* 44: 203–219. http://dx.doi.org/10.1016/j.jafrearsci.2005.10.011.

Petrie,G.M., P. G. Heasler, E. M. Perry, S. E. Thompson, and D. S. Daly. 2002. “Inverse kriging to enhance spatial resolution of imagery.” *In Algorithms and Systems for Optical Information Processing, Proceedings-SPIE*, 4789: 55-63. SPIE , Bellingham.

Pei, Y., J. Yu, H. Zhou, G. Cai. 2010. “The improved wavelet transform based image fusion algorithm and the quality assessment.” *Image and Signal Processing (CISP),* 2010 3rd International Congress on. Vol. 1. IEEE, 2010.

Pilon,P. P.J. Howarth, R.A. Bullock, P.O. Adeniyi. 1988. “An Enhanced Classification Approach to Change Detection in Semi-Arid Environments.” *Photo- grammetric Engineering and Remote Sensing* 54(12): 1709-1716.

Piper, J. 1999. “The effect of zero feature correlation assumption on maximum likelihood based classification of chromosomes.” *Signal processing* 12(1):49–57. doi:10.1016/0165-1684(87)90081-8

Pohl,C. and J.L. Van Genderen. 1998. “Multisensor image fusion in remote sensing: concepts, methods and applications.” *International Journal of Remote Sensing* 19(5): 823-854. doi: 10.1080/014311698215748

Polikar, R. 2006. “Ensemble based systems in decision making.” *IEEE Circuits and Systems Magazine* 6:21–45, http://dx.doi.org/10.1109/MCAS.2006.1688199.

Quinlan, J. R. 1993. C4.5: *Programs for Machine Learning*. San Matero, CA: Morgan Kaufmann.

Rabe, A., B. Jakimow, M. Held, S. van der Linden, and P. Hostert. 2014. EnMAP-Box, Version 2.0, software available at www.enmap.org

Radke, R.J., S. Andra , O. Al-Kofahi and B. Roysam. 2005. “Image change detection algorithms: A systematic survey.” *IEEE Transactions on Image Processing* 14(3): 294 -307. doi: 10.1109/TIP.2004.838698

Rahman, M. M., G.J.Hay, I. Couloigner, B. Hemachandran, J.Bailin. 2015. “A comparison of four relative radiometric normalization (RRN) techniques for mosaicing H-res multi-temporal thermal infrared (TIR) flight-lines of a complex urban scene.” *ISPRS Journal of Photogrammetry and Remote Sensing* 106: 82-94. doi:10.1016/j.isprsjprs.2015.05.002

Ranchin T and L. Wald. 2000. “Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation.” *Photogrammetric Engineering and Remote Sensing* 66(1): 49-61.

Rainbird, R.H., C.W. Jefferson, R.S. Hildebrand, and J.K. Worth. 1994. “The Shaler Supergroup and revision of Neoproterozoic stratigraphy in Amundsen, Northwest Territories.” *Geological Survey of Canada*. Current Research no. 1994-C: 61–70.

Rainbird, R.H., C.W. Jefferson, and G.M. Young. 1996. “The early Neoproterozoic sedimentary Succession B of northwestern Laurentia: Correlations and paleogeographic significance.” *Geological Society of America Bulletin* 108: 454–470. doi:10.1130/0016-7606(1996)108<0454:TENSSB>2.3.CO;2.

Rainbird, R. H., J. C. Harrison, E. M. Hillary, A. Ford, L. J. Hulbert, R. L. Christie, and F. H. A. Campbell. 2013a. Geology, Tectonic Assemblage Map of Hadley Bay, Victoria and Prince of Wales Islands, Nunavut - Northwest Territories; Geological Survey of Canada, Canadian Geoscience Map 75, (ed. prelim.); 1 sheet.

Rainbird, R. H., J. H. Bédard, K. Dewing, T. Hadlari, and D. Thomson. 2013b. Geology, Qiqittiivik, Victoria Island, Northwest Territories; Geological Survey of Canada, Canadian Geoscience Map 59, (ed. prelim.); 1 sheet.

Rainbird, R. H., R. L. Christie, J. C. Harrison, and A. Ford. 2013c. Geology, Tectonic Assemblage Map of the Ulukhaktok Area, Southwestern Victoria Island, Nunavut - Northwest Territories; Geological Survey of Canada, Canadian Geoscience Map 77, (ed. prelim.); 1 sheet.

Rainbird, R. H., J. H. Bédard, and N. Williamson. 2013d. Geology, Takiyuaqattak, Victoria Island, Northwest Territories; Geological Survey of Canada, Canadian Geoscience Map 104, (ed. prelim.); 1 sheet.

Rajendran, S., S. al-Khirbash, B. Pracejus, S.Nasir, A.H. Al-Abri, T.M. Kusky, and A. Ghulam. 2012. “ASTER detection of chromite bearing mineralized zones in Semail Ophiolite Massifs of the northern Oman Mountains.” *Exploration strategy: Ore Geology Reviews* 44: 121–135. doi:10.1016/j.oregeorev.2011.09.010.

Richard, J. A., and X. Jia. 2006. *Remote Sensing Digital Image Analysis: An Introduction*. Berlin: Heidelberg: Springer-Verlag.

Robertson, C., J. A. Long, F.S. Nathoo, T. A. Nelson, and C.C. Plouffe. 2014. "Assessing quality of spatial models using the structural similarity index and posterior predictive checks." *Geographical Analysis* 46(1): 53-74. doi: 10.1111/gean.12028

Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012. "An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 93–104. doi:10.1016/j.isprsjprs.2011.11.002.

Rodriguez-Galiano, Victor F., and Mario Chica-Rivas. 2014. "Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models." *International Journal of Digital Earth* 7(6): 492-509.

Rodriguez-Galiano, V.F., M. Chica-Olmo, and M.Chica-Rivas. 2014. "Predictive modelling of gold potential with the integration of multisource information based on random forest: a case study on the Rodalquilar area, Southern Spain." *International Journal of Geographical Information Science* 28: 1336–1354. doi:10.1080/13658816.2014.885527.

Rokni, K. A. Ahmad, K.Solaimani, S. Hazini. 2015. "A new approach for surface water change detection: integration of pixel level image fusion and image classification techniques." *International Journal of Applied Earth Observation and Geoinformation* 34: 226-234.doi: 10.1016/j.jag.2014.08.014

Rowan, L.C., C. Anton-Pachesco, D.W. Brickey, M.J. Kingston, A. Payas, N.Vergo, and J.K. Crowley. 1987. "Digital classification of contact metamorphic rocks in Extremadura, Spain, using Landsat thematic mapper data." *Geophysics* 52: 885–897, doi:10.1190/1.1442359.

Rowan, L.C., J.C. Mars, and C.J. Simpson. 2005. "Lithologic mapping of the Mordor, NT, Australia ultramafic complex by using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)." *Remote Sensing of Environment*. 99: 105–126, doi:10.1016/j.rse.2004.11.021.

Saadi, N.M., and K.Watanabe. 2009. "Assessing image processing techniques for geological mapping: a case study in Eljufra, Libya." *Geocarto International* 24: 241–253, doi:10.1080/10106040802556199.

Schabenberger, O., C.A. Gotway. 2004. *Statistical methods for spatial data analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Schetselaar, E.M., and J. Ryan. 2008. "Remote Predictive Mapping Case study of the Boothia Mainland area, Nunavut, Canada, *in* Harris, J.R., *ed.*, Remote Predictive Mapping: an aid for northern mapping: Geological Survey of Canada." Open File 5643, p. 261–280.

Schetselaar, E.M., J.R. Harris, T. Lynds, and E.A. de Kemp. 2007. "Remote Predictive Mapping (RPM): A Strategy for Geological Mapping of Canada's North." *Geoscience Canada* 34: 93–111.

Schott, J. R., C. Salvaggio, and W. J. Volchok. 1988. "Radiometric scene normalization using pseudoinvariant features." *Remote Sensing of Environment*26.1: 1-16. doi:10.1016/0034-4257(88)90116-2

Sexton, J.O., D.L. Urban, M.J.Donohue, C. Song. 2013. "Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record." *Remote Sensing of Environment* 128: 246-258. doi:10.1016/j.rse.2012.10.010

Singh, A. 1989. "Review article digital change detection techniques using remotely-sensed data." *International journal of remote sensing* 10 (6): 989-1003.

Stephens, M. A. 1974. "EDF Statistics for Goodness of Fit and Some Comparisons." *Journal of the American Statistical Association* 69: 730–737. doi:10.1080/01621459.1974.10480196.

Stevens, C W; Kerr, D E; Wolfe, S A; Eagles, S. 2013. "Predictive surficial materials and surficial geology derived from LANDSAT 7, Hearne Lake, NTS 85I, Northwest Territories." Geological Survey of Canada, Open File 7233, 2013; 22 pages, doi:10.4095/292394

Story, M., and R. Congalton. 1986. "Accuracy Assessment: a user's perspective." *Photogrammetric Engineering and Remote Sensing* 52: 397–399.

Švab, A., and K. Oštir. 2006. "High-resolution image fusion: methods to preserve spectral and spatial resolution." *Photogrammetric Engineering & Remote Sensin*. 72(5): 565-572. doi:10.14358/pers.72.5.565

Tewkesbury, A.P., A.J. Comber, N.J. Tate, A. Lamb, P.F. Fisher. 2015. "A critical synthesis of remotely sensed optical image change detection techniques." *Remote Sensing of Environment* 160:1-14.doi: 10.1016/j.rse.2015.01.006

Thomson, D., R.H. Rainbird, and G. Dix. 2014. “Architecture of a Neoproterozoic intracratonic carbonate ramp succession: Wynniatt Formation, Amundsen Basin, Arctic Canada.” *Sedimentary Geology* 299: 119–138. doi:10.1016/j.sedgeo.2013.11.005.

Thorsteinsson, R., and E. T. Tozer. 1962. Banks, Victoria, and Stefansson islands, Arctic Archipelago. Ottawa, ON: Geological Survey of Canada. Memoir 330.

Tiho, Seydou, and M. Dagnogo. 2015. “Aggregation and Spatial Coexistence of Earthworm Community in Grassy Savanna of Lamto (Côte d’Ivoire).” *Journal of Biology and Life Science* 6(2): 37-49. doi: 10.5296/jbls.v6i2.6975

Tottrup, C. 2004. “Improving tropical forest mapping using multi-date Landsat TM data and pre-classification image smoothing.” *International Journal of remote sensing* 25(4): 717-730. doi: 10.1080/01431160310001598926

Van der Meer, F., P.M. Van Dijk, and A.B. Westerhof. 1995. “Digital classification of the contact metamorphic aureole along the Los Pedroches batholith, south-central Spain, using Landsat Thematic Mapper data.” *International Journal of Remote Sensing*16: 1043–1062. doi:10.1080/01431169508954462.

Vittek, M, A Brink, F. Donnay,D. Simonetti, B. Desclee. 2014. “Land cover change monitoring using Landsat MSS/TM satellite image data over West Africa between 1975 and 1990.” *Remote Sensing* 6(1): 658-676.

Vrabel, J. 1996. “Multispectral Imagery Band Sharpening Study.” *Photogrammetric Engineering & Remote Sensing* 62(9): 1075-1084.

Wang, X., R. Niu, and K. Wu. 2011. “Lithology Intelligent Identification Using Support Vector Machine and Adaptive Cellular Automata in Multispectral Remote Sensing Image.” *Optical Engineering* 50 (7): 076201–076213. doi:10.1117/1.3598315.

Wang Q, W. Shi, P.M. Atkinson, Y. Zhao. 2015. “Downscaling MODIS images with area-to-point regression kriging.” *Remote Sensing of Environment* 166: 191-204.doi: 10.1016/j.rse.2015.06.003

Wald, L. 2002. *Data fusion: Definitions and Architectures - Fusion of Images of Diferent Spatial Resolutions*. Paris, France: Ecole des Mines de Paris. ISBN 2-911762-38-X

Ward, J.H., 1963. “Hierarchical grouping to optimize an objective function.” *Journal of the American Statistical Association* 58(301): 236–244.doi: 10.2307/2282967

Waske, B., J. A. Benediktsson, K. Árnason, and J. R. Sveinsson. 2009. "Mapping of Hyperspectral AVIRIS Data Using Machine-Learning Algorithms." *Canadian Journal of Remote Sensing 35 (Suppl. 1)*: SS106-116. doi:10.5589/m09-018.

Waske, B., S. van der Linden, J. A. Benediktsson, A. Rabe, and P. Hostert. 2010. "Sensitivity of Support Vector Machines to Random Feature Selection in classification of Hyperspectral Data." *IEEE Transactions on Geoscience and Remote Sensing* 48 (7): 2880–2889. doi:10.1109/TGRS.2010.2041784.

Waske, B., and M. Braun. 2009. "Classifier ensembles for land cover mapping." *ISPRS Journal of Photogrammetry and Remote Sensing* 64:450–457, http://dx.doi.org/10.1016/j.isprsjprs.2009.01.003.

Webster, R. and M. A. Oliver. 2007. *Geostatistics for Environmental Scientists*. Chichester: John Wiley & Sons (second edition).

Wickert, L.M., J.B. Percival, W.A. Morris, and J.R. Harris. 2008. "XRD and Infrared Spectroscopic Validation of Weathering Surfaces from Ultramafic and Mafic Lithologies Examined Using Hyperspectral Imagery, Cross Lake Area, Cape Smith Belt, Northern Quebec, Canada." *in* Proceedings of 2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 6–11 July 3. p. 362–365.

Wu, W., A. Xu, H. Liu. 2015. "High-resolution spatial databases of monthly climate variables (1961-2010) over a complex terrain region in southwestern China." *Theoretical and Applied Climatology* 119.1-2:353-362. doi: 10.1007/s00704-014-1123-1

Yang, X. and C. P. Lo. 2000. "Relative radiometric normalization performance for change detection from multi-date satellite images." *Photogrammetric Engineering and Remote Sensing* 66(8): 967-980.

Yazdi, M.M.Taheri, P. Navi, N. Sadati. 2013. "Landsat ETM+ imaging for mineral potential mapping: application to Avaj area, Qazvin, Iran." *International journal of remote sensing* 34(16): 5778-5795. doi: 10.1080/01431161.2013.797127

Yu, L., A. Porwal, E. Holden, and M. Dentith. 2012. "Towards Automatic Lithological Classification from Remote Sensing Data Using Support Vector Machines." *Computers & Geosciences* 45: 229–239. doi:10.1016/j.cageo.2011.11.019.

Yuan, Fei, K.E.Sawaya, B.C. Loeffelholz, E.B. Marvin. 2005. "Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal

Landsat remote sensing." *Remote sensing of Environment* 98(2-3): 317-328. doi: 10.1016/j.rse.2005.08.006

Zeng, Y. J. Zhang, J.L. Van Genderen, Y. Zhang. 2010. "Image fusion for land cover change detection." *International Journal of Image and Data Fusion* 1(2): 193-215. doi: 10.1080/19479831003802832

Zhang,J. 2010. "Mutlti-source remote sensing data fusion: status and trends." *International Journal of Image and Data Fusion* 1(1):5-24. doi: 10.1080/19479830903561035

Zhang, Y. 2004. "Understanding image fusion." *Photogrammetric Engineering and Remote Sensing* 70(6):657-661.

Zhang, Y. 2008a. "Panchromatic-sharpening for improved information extraction." In: Li, Z, Chen,J., Baltsavias, E (Eds.), Advances in Photogrammetry, remote sensing and spatial information sciences: 2008 ISPRS Congress Book. CRC Press, London,UK,185-204

Zhang,Y. 2008b. "Methods for image fusion quality assessment-a review, comparison and analysis." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37:1101-1109.

Zhou,J., D. L. Civco, and J. A. Silander. 1998. "A wavelet transform method to merge Landsat TM and SPOT panchromatic data." *International Journal of Remote Sensin*. 19(4): 743-757. doi: 10.1080/014311698215973

Zhu, Zhe, Curtis E. Woodcock, and Pontus Olofsson. 2012. "Continuous monitoring of forest disturbance using all available Landsat imagery." *Remote Sensing of Environment* 122: 75-91.