AUTOMATIC TEMPLATE-GUIDED CLASSIFICATION OF REMNANT TREES

PETER KENNEDY
Bachelor of Science, University of Lethbridge, 2010

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Geography
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

AUTOMATIC TEMPLATE-GUIDED CLASSIFICATION OF REMNANT TREES

PETER KENNEDY

Date of Defence: December 19, 2014

Dr. Karl Staenz Supervisor	Professor	Ph.D.
Dr. Jinkai Zhang Co-supervisors	Adjunct Professor	Ph.D.
Dr. Craig Coburn Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Howard Cheng Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Wei Xu Chair, Thesis Examination Committee	Professor	Ph.D.

To the Measure of all things.

Abstract

Spectral features within satellite images change so frequently and unpredictably that spectral definitions of land cover are often only accurate for a single image. Consequently, land-cover maps are expensive, because the superior pattern recognition skills of human analysts are required to manually tune spectral definitions of land cover to individual images. To reduce mapping costs, this study developed the Template-Guided Classification (TGC) algorithm, which classifies land cover automatically by reusing class information embedded in freely available large-area land-cover maps. TGC was applied to map remnant forest within six 10-m resolution SPOT images of the Vermilion River watershed in Alberta, Canada. Although the accuracy of the resulting forest maps was low (58 % forest user's accuracy and 67 % forest producer's accuracy), there were 25 % and 8 % fewer errors of omission and commission than the original maps, respectively. This improvement would be very useful if it could be obtained automatically over large-areas.

Acknowledgments

I am grateful for the opportunity to explore the ideas presented in this thesis and for the people that have supported me. I especially thank Dr. Karl Staenz for his patient support, Dr. Jinkai Zhang for sharing his remote sensing experience and for his friendship, Dr. Craig Coburn for his advice and practical help in the face of looming deadlines, and Dr. Howard Cheng for his dutiful oversight. I appreciate their generosity.

I also thank Dr. Anne Smith, Gary Larson, and Ingrid Oseen for welcoming me into their lab, Trevor Armstrong for his very helpful advocacy and assistance, and David Rolfson for patiently guiding me through the details of the Vermilion project.

I acknowledge fellow students James Banting for his mapping expertise and help, Kevin Riddell for Arduino demonstrations, Jurien van der Sluijs for discussions about spectral mixture analysis and Dutch lessons, Ilya Parshakov for helping me understand his research, and Gordon Logie for discussions about imaging and georegistration.

I thank the University of Lethbridge, the Alberta Terrestrial Imaging Centre, Alberta Agriculture and Rural Development, the Agriculture and Food Council of Alberta, and the NSERC CREATE Advanced Methods, Education and Training in Hyperspectral Science and Technology program whose funding supported me well throughout my research.

Contents

List of Tables

List of Figures

Lists of Symbols and Abbreviations

Symbol	Description
$\%\text{LAND}_\text{k}$	Percentage of land cover classified as class k.
a_{ij}	Spatial correspondence between target cluster K_i and template class C_j.
A	Overlapping area matrix containing elements a_{ij} where $1 \leq i \leq m$ and $1 \leq j \leq n$.
b	Number of bands in an image.
C_j	Template class j.
d	Number of cluster makers produced by TGC.
d	Distance between means of forest correspondence distributions.
f	Forest class.
f_{ij}	Number of pixels classified as class i that are actually reference class j.
f_{i+}	Number of pixels that were classified as class i.
f_{+j}	Number of pixels that are actually reference class j.
ϕ_f	TGC forest correspondence threshold.
I	Target image. A target image used to cluster an ROI in TGC.
K_i	Cluster i of a clustering X.
k_{min}	Minimum number of target pixels allowed in a TGC cluster.
λ	Cluster maker. Produces clusters from a target image.
Λ	Set of cluster makers.$\Lambda = \{\lambda_1, \ldots, \lambda_d\}$.
m	Number of clusters in the clustering of an ROI.
MAREA_f	Classification produced from a forest correspondence image with the minimum absolute REA_f.
NSD	Normalized spectral distance.
n	Number of classes in the template image.
N	Number of sample pixels in error matrix.
nf	Non-forest class.
OA	Overall classification accuracy.
PA_j	Producer's accuracy of class j.
ψ_i	Cluster purity of cluster K_i with respect to classes $C_1, \ldots, C_n$.
REA_j	Relative error of area for class i.
q	Number of target images input to TGC.
Q	Quality of clustering.
TI	Template Image. The land-cover map used to guide TGC.
UA_j	User's accuracy of class j.
w_j	Weighting factor for class C_j.
$\vec{w}$	Weight vector $(w_1, w_2, \cdots, w_n)$ for classes $C_1, \ldots, C_n$.
$\vec{x}$	Feature vector $\vec{x} = \{x_1, x_2, \ldots, x_b\}$.
X	Clustering $X = \{K_1, \ldots, K_m\}$.
y_i	Class label for cluster K_i.

Abbreviation	Description
AGCC	Alberta Ground Cover Characterization. A 30-m land-cover map of Alberta from year 2000 data.
ATIC	Alberta Terrestrial Imaging Centre.
ATS	Alberta Township Survey. A system for dividing Alberta into a grid of 6 x 6 mile townships.
CHM	Canopy Height Model. A surface indicating the height of the canopy above the ground.
DALA	Data-assisted Labeling Approach. An automated cluster and label technique.
EOSD LC 2000	Earth Observation for the Sustainable Development of Forests Land Cover map. A 25-m land-cover map of forested Canada from year 2000 data.
ETM+	Enhanced Thematic Mapper Plus. A Landsat 7 instrument with 30-m resolution.
FROM-GLC	Finer Resolution and Monitoring of Global Land Cover. The first 30-m global land-cover map.
GFC	Global Forest Cover. A set of 30-m global forest cover and forest change products from 2000-2012.
GLC-2000	Global Land Cover 2000. A 1000-m global land-cover map.
GLOBCOVER	Global Cover. A 300-m land-cover map.
GRS	Grid Reference System. A system of coordinates used to locate SPOT images.
GSD	Ground Sample Distance. The ground distance between pixel centres.
HRG	High Resolution Geometrical. A SPOT 5 instrument with multispectral and panchromatic sensors.
ISODATA	Iterative Self-Organizing Data Analysis Technique, A. A clustering algorithm.
IGSCR	Iterative Guided Spectral Class Rejection. An automated cluster busting algorithm.
LCC2000	Land Cover Circa 2000. A land-cover map of Canada.
LTCCF	Landsat Tree Cover Continuous Fields. A global 30-m map of tree cover.
LiDAR	A system that produces 3-d point clouds of land cover using lasers.
MLCT	MODIS Land Cover Type. A 500-m global land-cover map.
MERIS	Medium Resolution Imaging Spectrometer. An instrument on the Envisat satellite (now defunct).
MODIS	Moderate-resolution Imaging Spectroradiometer. An instrument on the Terra and Aqua satellites.
MS	Multispectral data.
MS+P	Multispectral and Panchromatic data.
MSS	Landsat Multispectral Scanner. An instrument on Landsats 1-5.

NLCD	National Land Cover Database. A 30-m land-cover map (and other information) for the conterminous United States.
NTS	National Topographic System.
Pan	Panchromatic data.
ROI	Region Of Interest.
RMSE	Root Mean Square Error.
SPOT	Satellite Pour l'Observation de la Terre.
SRTM	Shuttle Radar Topography Mission. Digital elevation map at 3 arc-second resolution.
SVM	Support Vector Machine. An algorithm for classification.
TGC	Template-Guided Classification. The algorithm tested in this thesis.
TM	Thematic Mapper. An instrument on Landsats 4 and 5.
VCF	Vegetation Continuous Fields. 250-m global tree-cover products from MODIS data.

Chapter 1

Introduction

Remote sensing is the only way to obtain comprehensive and current information about the Earth's land cover. However, although remote sensing data is comparatively cheap (or even free), the labour cost of land-cover classification is an important consideration when producing land-cover maps (DeFries and Chan, 2000). Ideally, land-cover classification would be completely automated with no human intervention. With reduced labour costs, land-cover information can be updated more frequently and be better customized to applications. However, Landgrebe (1999) advised that fully automatic systems are likely to be more limited and less robust (and possibly less economical) than human-assisted digital techniques. Semi-automatic classifiers (classifiers assisted by human interpretation skills) may be the best option (Landgrebe, 1999) because automatic systems cannot satisfy the subtleties of user needs (Landgrebe, 2002). To reduce human involvement, a classifier should be generally applicable, requiring minimal labour when applied to new data.

Accuracy without generality is useless (and vice-versa). It is easy to make a perfectly accurate classifier that is only applicable to a single image and which saves no labour. For example, it is possible to classify solely on the basis of a pixel's location in an image by transforming pixel coordinates into the correct class using a hard-coded look-up table. Theoretically, this classifier could be made 100 % accurate for a single image, but it has no ability to generalize to other images in a different location or with a different pixel grid. There is no labour savings, because there is no generalization.

This thesis focuses mainly on minimizing labour costs while maintaining acceptable accuracy through development of a classification algorithm that is robust to noise.

1.1 A Fundamental Problem of Land-cover Classification

A fundamental problem of land-cover classification is that classifiers are disposable; they only work for very limited (or unique) datasets. A classifier may classify a particular image accurately, but when applied to another image that was acquired under slightly different conditions (e.g., in a different year, atmospheric conditions, or location), the classifier often produces unacceptably inaccurate results. Conditions change so frequently that a classifier is normally only used for a single image (Schott, 2007, p. 405).

This problem occurs because the measured features of land-cover classes are highly variable. Feature variability causes different land-cover classes to have very similar features. When different classes have virtually the same features, classification is an ill-posed problem. The classes overlap in feature space and cannot be distinguished from each other, and therefore, classification accuracy is reduced.

As variability increases, class overlap increases. To produce acceptable accuracies, classifiers are overfitted to very narrow conditions that minimize variability. They are made for such a narrow set of conditions that they fail when applied generally and therefore, they are disposable (Figure 1.1).

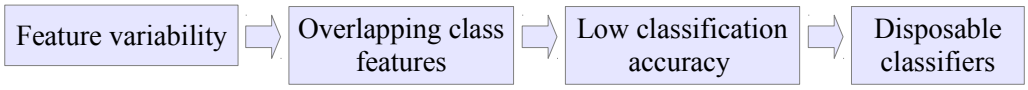

Figure 1.1: Feature variability causes the features of different classes to overlap, resulting in low classification accuracy and disposable classifiers.

For example, Figure 1.2a shows a scene of forest patches and fields in the Vermilion River watershed, Alberta, where trees appear as dark patches. However, Figure 1.2b shows a scene that is further north, near Chipewyan Lake, where many trees are very reflective (perhaps because they have already senesced). A classifier that assumes that all trees have low reflectance might perform well on the image of Figure 1.2a, but would fail to classify

the highly reflective trees in Figure 1.2b. As the example illustrates, feature variability can

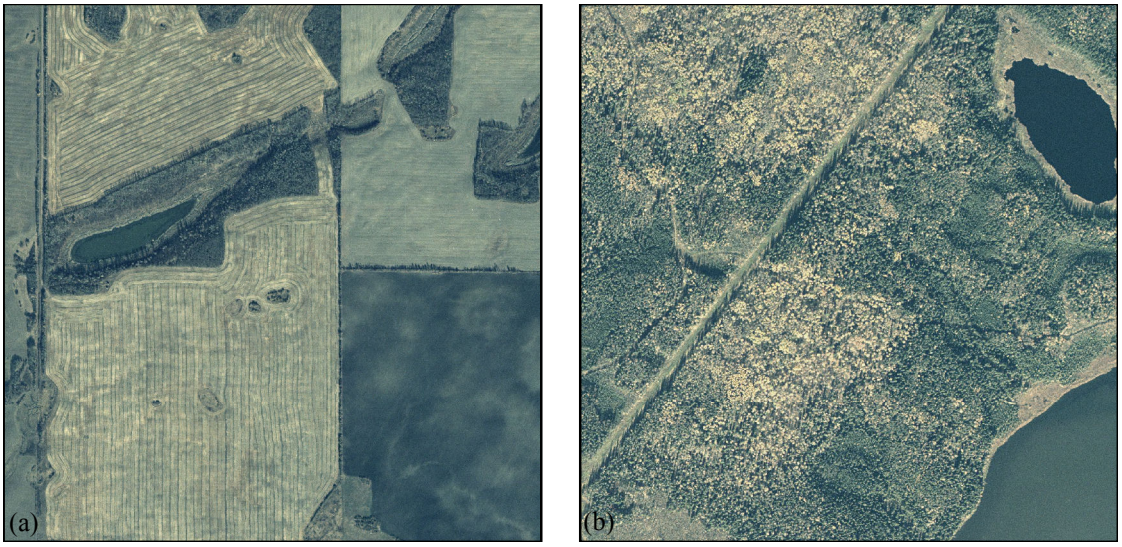

Figure 1.2: Land-cover classes usually do not have unique or invariant spectral signatures. Feature variability makes it difficult to define land-cover classes in terms of image features. (a) dark tree patches and some bright fields in the Vermilion River watershed, Alberta, and (b) dark trees and bright trees near Chipewyan Lake, Alberta.

cause a classifier that is accurate in one context to be inaccurate in another. For this reason, spectral signatures used to classify Landsat images are typically only applied to a single image or a group of radiometrically normalized images (Olthof et al., 2005).

Remote sensing research has intensely pursued technologies to generalize and reduce the labour of land-cover classification. Cheap, accurate, automatic land-cover classification is the highest goal (Landgrebe, 2002; Loveland et al., 2002). However, for operational applications, there is often an unpleasant conundrum: one may classify land cover manually and pay the price in time and money or classify land cover automatically and pay the price with lower classification accuracy.

1.2 Local Class Information in Land-cover Maps

One of the problems with automating land-cover classification is that class features vary spatially. This means that classifiers need to be tuned to local conditions to produce ac-

curate results. For example, because trees in northern Alberta have different features than trees in central Alberta, classifiers need to have rules that are different for each region. However, determining how class features vary spatially is expensive, because it requires human intervention to establish local definitions of each class.

A possible solution to the high cost of defining classes is found in existing land-cover maps. They contain spatially varying information that is validated and overseen by experts. To make a map for a particular area, the local definitions of classes must be known. These definitions are implicitly embedded in land-cover maps by the spatial distribution of classes. If this local class information could be reused for new classifications, there could be important cost reductions.

For example, the Earth Observation for the Sustainable Development of Forests (EOSD) land-cover map is a map of the forested regions of Canada containing 23 classes derived from 30-m Landsat ETM+ images (Wulder et al., 2008). Each image was automatically partitioned into hundreds of clusters of spectrally similar pixels and manually labelled through image interpretation that was assisted by maps, local knowledge, and other resources (Wulder et al., 2002). This manually interpreted information was embedded in each class label assigned to the map. If this information could be used to define classes accurately, it might be possible to produce an up-to-date land-cover map of the entire forested regions of Canada for little cost.

1.3 Automating Land-cover Classification using Existing Land-cover Maps

The main idea of this thesis was inspired by a single image. Figure 1.3 shows how moderately accurate forest pixels can be used to determine the correct class labels for each part of a forest/non-forest map. In this figure, green EOSD forest pixels are overlaid on top of a forest/non-forest ground-reference map. But which part of the ground-reference map is forest? White pixels or black pixels? The strong spatial correspondence of the green EOSD

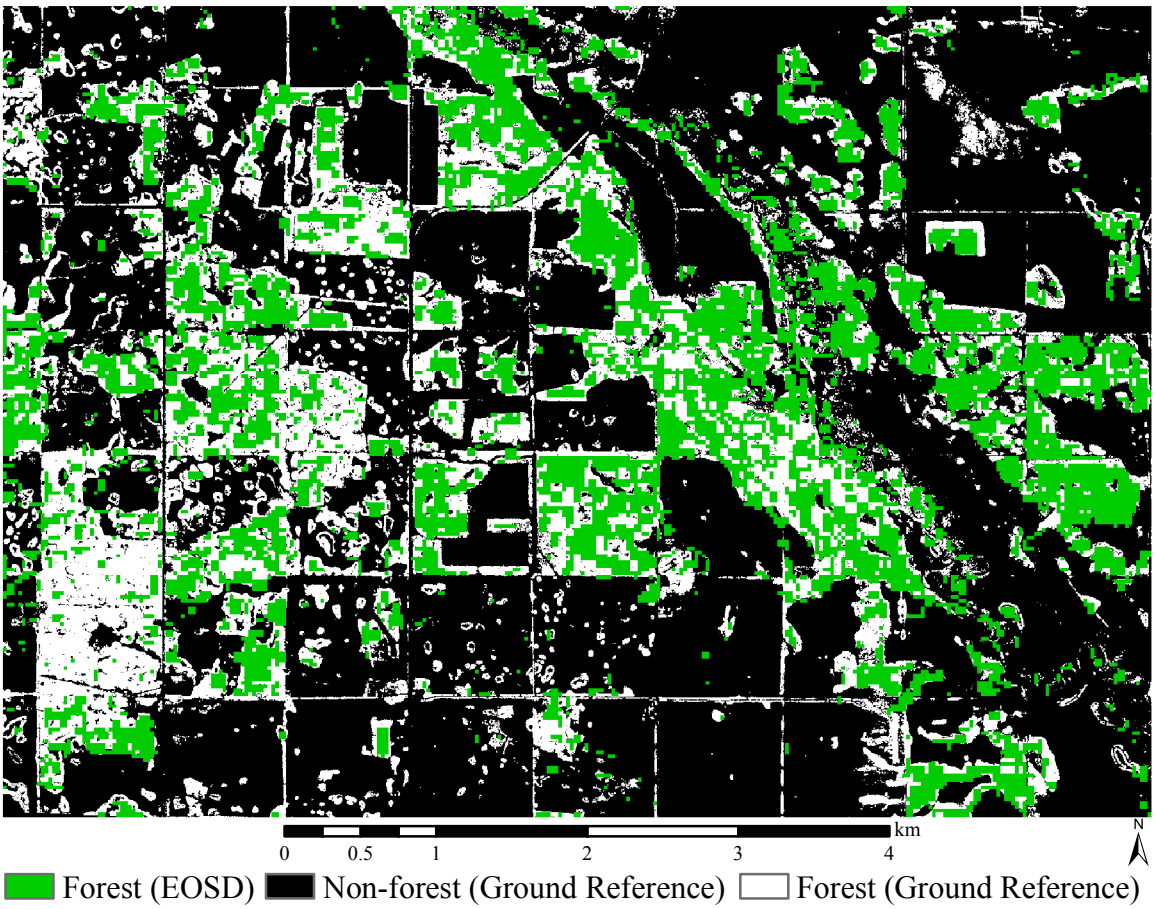

Figure 1.3: EOSD forest pixels (green) overlaying a forest/non-forest ground-reference map. Even if the legend did not specify which colour corresponded to which class, the appropriate label for each part of the ground-reference map could be determined by its spatial correspondence to the EOSD labels (white = forest, black = non-forest).

forest pixels to the white ground-reference pixels clearly indicates that the white pixels are forest pixels (and black pixels are non-forest pixels).

The hypothesis of this thesis is that land-cover classification could be automated by:

1. Partitioning an image into separate clusters (as with the EOSD), and

2. Labelling each part with the class of an existing land-cover map (e.g., EOSD) that it best corresponds to spatially.

An implementation of this idea, named Template-Guided Classification (TGC), is presented in this thesis with the goal of reducing the cost of land-cover classification. Numerous

land-cover maps have been produced at continental and global scales that could be used as templates for updating land-cover maps. Theoretically, automatic updates for global land-cover maps could be achieved with negligible labour costs. TGC might also be useful for increasing the spatial resolution of land-cover classifications. For example, the 25-m resolution EOSD might be used to label 10-m data.[1]

This thesis tests TGC by applying this approach to remnant tree classification of the Vermilion River watershed, Alberta, Canada.

1.4 Remnant Trees in the Vermilion River Watershed: A Case Study for Land-cover Classification

Remnant vegetation is native vegetation that remains after a region has been predominantly cleared (Millsom and Earl, 2001). Remnant trees, an important subcategory of remnant vegetation, are found with varying degrees of fragmentations, from small, scattered patches surrounded by agriculture (Figure 1.4) to large, undisturbed forests (e.g., in national parks or in other protected areas) (Gibbons and Boak, 2002; McIntyre and Hobbs, 1999). In particular, small (less than 1 ha), scattered remnant tree patches are exceptionally important for the environment.

1.4.1 The Importance of Small Remnant Tree Patches

Scattered tree patches are "keystone structures" (Manning et al., 2006); their ecological contributions are disproportionate to their small area (Fischer et al., 2010a). Within the relative homogeneity of agricultural landscapes, remnants function as islands of biodiversity, providing food, increased connectivity, and shelter for wildlife (Harvey and Haber, 1998). Remnant trees also influence the abiotic environment by producing cooler microclimates (Vetaas, 1992), increasing infiltration (Eldridge and Freudenberger, 2005), reducing

[1]Note that labels of the high-resolution ground-reference data in Figure 1.3 are clearly indicated by the lower resolution EOSD data.

Figure 1.4: Examples of small remnant tree patches within the Vermilion River watershed in central Alberta. These scattered patches are disproportionally valuable to the ecology of the agricultural landscape.

flood risk (Carroll et al., 2004), stabilizing soil, and reducing wind speeds (Millsom and Earl, 2001). Economically, they are sources of timber, native seed, tourism and recreation opportunities (Driver and Higgins, 2001). Small patches of remnant trees are especially valuable because they are anomalous; they provide additional functions to the surrounding ecosystem because they are different from the dominant land cover (Manning et al., 2006).

1.4.2 Remnant Tree Vulnerability

There is an urgency to the study of scattered remnant trees, because they are vulnerable and are (in many cases) no longer self-sustaining. Numerous studies describe the chronic decline and dieback of scattered trees at sites on every continent (Fischer et al., 2009; Gibbons et al., 2008; Maron and Fitzsimons, 2007; Manning et al., 2006; de Barros Ferraz et al., 2005; Jurskis, 2005; Ozolins et al., 2001). In many agricultural settings, remnant trees are growing old and dying without replacement. The rate of regeneration is low to non-existent, because of damage to seedlings from grazing (Fischer, et al., 2009), changes to soil (Granger et al., 1994), and increased competition from other plants which are more suited to the agricultural regime (Close et al., 2008). The model proposed by Landsberg and Wylie (1988) describes pathways of tree death and stress in rural Australia, which are

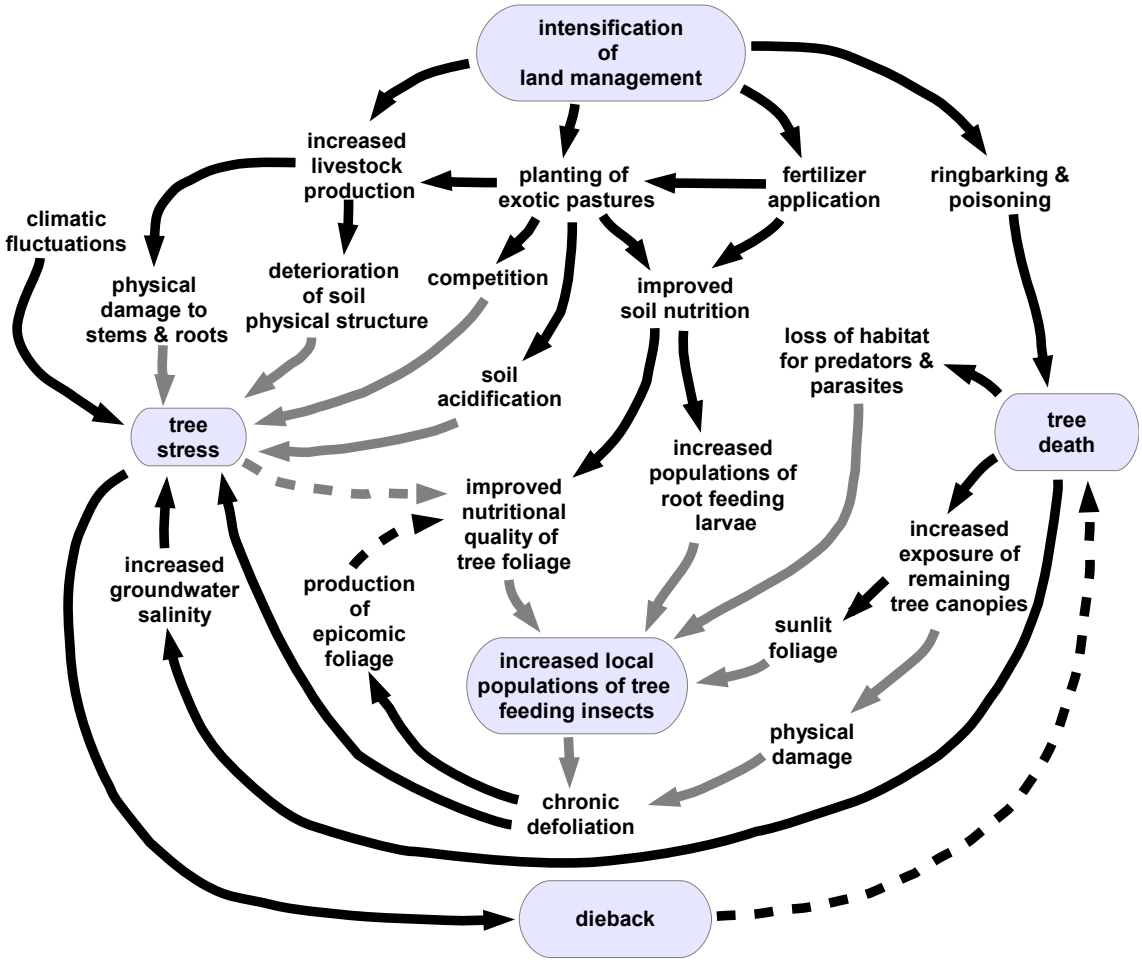

Figure 1.5: A model of rural dieback of remnant trees in Australia (Landsberg and Wylie, 1988). Pathways represented by black lines are supported by research results; grey lines are more speculative. Dashed lines indicate feedback pathways.

at least partially applicable to agricultural landscapes worldwide (Figure 1.5). Gibbons et al. (2008) predicted that without changes in agricultural management practices, mature trees will be lost from case study sites in Spain, United States, Costa Rica and Australia in 90-180 years. In Australia, the problem is so severe that two-thirds of trees on an average farm will be gone by 2100 (Fischer et al., 2010b). This will result in a regime shift, wherein the absence of trees causes cascading changes to the ecosystem so that it functions using very different processes and feedbacks (Fischer et al., 2010a).

1.4.3 Monitoring Remnant Trees in Alberta

Despite the importance and vulnerability of remnant trees, there is no comprehensive system for monitoring the extent and distribution of remnant trees in Alberta. The lack of monitoring prevents the decline of remnant trees from being quantified. There needs to be an accurate and cost-effective monitoring system to provide a foundation for policy decisions to protect remnant trees.

1.5 Objective

The objective of this research is to develop and evaluate an automatic land-cover classifier that extracts class information from existing large-area land-cover maps. To evaluate the classifier, remnant tree/non-remnant tree classifications were produced for the Vermilion River watershed in Alberta, Canada. The main criterion for evaluation is the operational utility of the classifier. In this context, the following questions were asked:

- Was the classifier accurate?
- Did it save labour?
- What were the problems?
- How can they be overcome?

1.6 Thesis Organization

This thesis is divided into seven chapters. Chapter 2 is a literature review of different classification strategies, with particular attention to stratification, automation, map reuse, and iterative partitioning algorithms. Chapter 3 introduces TGC, the algorithm developed by this research that automatically classifies remnant trees using existing land-cover maps. Chapter 4 presents the study area, data, and experiments used to evaluate TGC. Chapter 5

presents the experimental results, while Chapter 6 evaluates the operational utility of TGC by answering the four research questions. Chapter 7 presents conclusions and describes future work.

Chapter 2

Background

2.1 Large-area Land-cover Maps

Large-area land-cover maps are very interesting examples of the trade off between accuracy and automation. With large-area land-cover classification, there is a very strong financial incentive to minimize human intervention despite the need for locally adapted class definitions to handle high feature variability.[1] This tension of antagonistic goals has resulted in an interesting variety of classification systems (Table 2.1).

From the perspective of reuse, the main benefit of large-area maps is that they are a source of free, spatially extensive land-cover information. The main problem is that large-area maps are relatively inaccurate, because they trade accuracy for low cost per unit area. The intensive human oversight required for higher accuracies is too expensive.

The overall accuracies reported in Table 2.1 are all greater than 50 %. However, these numbers should be viewed with caution because accuracy assessment is so expensive that incorrect or vague assessments are common (Foody, 2002). Accuracies are sometimes estimated by inadequate samples or by comparison to other maps that also have errors (e.g., Olthof et al. (2009)), because it is too expensive to collect representative validation data.

Accuracies vary spatially according to the type and heterogeneity of land cover (Herold et al., 2008; Smith et al., 2002). Consequently, regional accuracies can be very different from reported values. For example, large, homogeneous areas of snow cover (e.g., Greenland) and desert regions (e.g., North Africa) are likely to have higher classification

[1] The high variability of large areas is a consequence of the first law of geography: "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). For example, the entire globe has more species of trees and more variability than a single hectare of the northern boreal forest.

accuracies than the Vermilion River watershed, which is in a transition zone between the boreal forest and grasslands. As summarized by Gong et al. (2013), overall accuracies of global land-cover maps (as calculated by third-parties) ranged from 11 % to 50 %. Homer et al. (2007) specifically warned against using the National Land Cover Database (NLCD) 2001 for local applications (e.g., county level use) due to the variability of accuracies on smaller scales.

Table 2.1: Selected large-area land-cover maps.

Name	Region	Reference	No. of Classes	Overall Accuracy	GSD (m)	Data Source	Stratification	Algorithm
NLCD 1992	USA	Vogelmann et al. (2001)	7, 21	81 % for 7 classes, 60 % for 21 classes	30	TM	31 regions of political and image boundaries, 100 clusters, logical and threshold models of ancillary data, decision tree leaves	Cluster regional mosaics, manually label
GLC 2000-NCA	North and Central America	Latifovic et al. (2004)	28	80 % agreement with other maps	1000	SPOT VEGETATION	Canada & Alaska, USA & Central America, 150 initial clusters from ISODATA	Cluster multi-temporal and composite data, manually label
GLC 2000	Global	Bartholome and Belward (2005); Mayaux et al. (2006)	22	69 %	1000	SPOT VEGETATION	18 geographic regions worldwide	Cluster multi-temporal and composite data, manually label
EOSD LC 2000	Forested Canada	Wulder et al. (2008)	23	Unknown	30	ETM+	NDVI strata (water; non, low, high vegetation), 241 clusters	Cluster individual images with texture channel, manually label
GLOBCOVER 2005	Global	Bicheron et al. (2008)	22	67 %	300	MERIS	22 regions worldwide delineated by natural discontinuities, poorly represented classes (e.g., urban and wetland), clusters	Cluster multi-temporal data in subregions, label automatically using rules-based procedure
LCC 2000	Northern Canada	Olthof et al. (2009)	15	82 % (76 plots)	30	ETM+	Bioclimatic subzones, soil categories, mountainous regions, 200 clusters reduced to 80 clusters	Cluster normalized mosaic, label manually using existing land-cover maps
MLCT (MCD12Q1)	Global	Friedl et al. (2010)	8, 10, 12, 14, 17	75 %	500	MODIS	18 ecoregions, urban areas, decision tree leaves	Apply boosted decision tree with manual training data to multi-temporal spectra and metrics
FROM-GLC2013	Global	Gong et al. (2013)	10	65 %	30	TM, ETM+	Training data selection by ecoregion	Apply SVM with manual training data
NLCD 2011	USA	Jin et al. (2013)	16	Unknown	30	Landsat	‘stable’ and ‘dynamic’ land-cover classes, decision tree leaves	Detect areas of change aided by the NLCD 2006, apply decision tree of multi-temporal data
GFC	Global	Hansen et al. (2013)	Forest cover (%)	Only change statistics reported	30	Landsat, MODIS, SRTM	Decision tree leaves	Apply bagged classification tree with manual training data to normalized multi-temporal image metrics
LTCCF	Global	Sexton et al. (2013)	Forest cover (%)	R^2 = 0.81, RMSE = 17 % when compared to test sites	30	Landsat, MODIS VCF	Agricultural areas, low tree cover areas, decision tree leaves	Use regression tree to relate Landsat spectra to 250-m MODIS Tree Cover VCF

EOSD = Earth Observation for the Sustainable Development of Forests, FROM-GLC = Finer Resolution and Monitoring Global Land Cover, GFC = Global Forest Cover, GLC = Global Land Cover, LC = Land Cover, LCC 2000 = Land Cover Circa 2000, LTCCF = Landsat Tree Cover Continuous Fields, MLCT = MODIS Land Cover Type, NCA = North and Central America, NLCD = National Land Cover Database, VCF = Vegetation Continuous Fields, TM = Thematic Mapper, ETM+ = Enhanced Thematic Mapper Plus, SPOT = Satellite Pour l’Observation de la Terre, MERIS = Medium Resolution Imaging Spectrometer, MODIS = Moderate-resolution Imaging Spectroradiometer, SRTM = Shuttle Radar Topography Mission.

2.1.1 Stratification

Stratification is one of the primary techniques for handling the variability of large areas (Hansen and Loveland, 2012). It breaks a large problem with highly variable features into several smaller problems with less variability. Its goal is to separate pixels that would otherwise be confused with each other into separate classification problems. For example, Olthof et al. (2009) used bioclimatic subzones, soil type, elevation data, and treeline data to divide Northern Canada into eight strata with similar land-cover types. Water and topographic shadows were found to have similar spectral features. Class confusion was reduced by labelling ambiguous features as water in a stratum of elevations below 300 m, and as topographic shadow for features above 300-m (because water is not usually found in mountains). Stratification can be thought of as adding a dimension to each object's feature vector to help distinguish between different land-cover types.

Instead of stratifying by elevation, Wulder et al. (2008) manually selected values of the Normalized Difference Vegetation Index (NDVI; Rouse et al., 1974) corresponding to water, non-vegetated, low-reflectance vegetation, and high-reflectance vegetation. By breaking images into strata with similar NDVI, the discrimination between land-cover classes was improved (Wulder et al., 2008). Stratifying by ecozones within images reduced class confusion between Douglas-fir (*Pseudotsuga menziesii*), Jeffrey pine (*Pinus jeffreyi*), and Ponderosa pine (*Pinus ponderosa*) (Jiang et al., 2004).

All large-area classification systems stratify implicitly or explicitly. Many systems classify each image independently, although this practice may decrease in the future with the increasing availability of standardized atmospherically corrected images, such as those available through the Web-Enabled Landsat Dataset (WELD) for the continental United States (Hansen et al., 2014). Other systems stratify by larger areas such as ecozones, which may reduce the problems of seam-line and model inconsistencies between strata (Hansen and Loveland, 2012). When classifying multi-image strata, it is important to atmospheri-

cally correct images and reduce differences in class signatures between images as much as possible. These image adjustments are problematic when differences are unpredictable (Landgrebe, 2002).

Stratification does not always improve classification accuracy. For example, suppose a set of observations is stratified by random assignment to different strata. Each stratum is then essentially the same classification problem as the original — there is no advantage to the stratification. The key to stratification is finding a feature that reduces the amount of overlap between difficult land-cover types.

Sometimes stratification causes error because there are exceptions to rules imposed by stratification. For example, Olthof et al. (2009) recognized that hard boundaries of bioclimatic strata are problematic because bioclimate is a continuum. Therefore, for objects near boundaries, strata were ignored (Olthof et al., 2009).

2.1.2 Cluster and Label

One of the early techniques for creating large-area land-cover maps was to cluster data from satellite images and to label the clusters by visually interpreting the images. The NLCD 1992, GLC 2000, EOSD 2000, LCC 2000,[2] and many other large-area maps were produced using this 'cluster and label' approach (Table 2.1).

Clustering is a process that partitions a set of objects into clusters (subsets of objects with similar features). Cluster and label's basic premise is that pixels in a cluster represent the same land cover and that effort is saved by manually labelling a few hundred clusters instead of labelling millions of pixels. Clustering itself requires no manual effort, because it can be automated by hard-coding suitable parameters. For example, the EOSD clustering workflow specified automated application of the *k*-means algorithm (MacQueen, 1967) with 241 output clusters, 12 iterations and a 1-% movement threshold for each stra-

[2]NLCD = National Land Cover Database, GLC = Global Land Cover, EOSD = Earth Observation for the Sustainable Development of Forests, LCC 2000 = Land Cover Circa 2000.

tum (Wulder et al., 2008).

From the perspective of land-cover classification, the goal of clustering is to partition different land-cover classes into separate clusters (Lark, 1995). Unfortunately, after clustering, clusters frequently contain objects from different land-cover classes (Lark, 1995). These mixed clusters greatly reduce clustering's usefulness for land-cover classification, because mixed clusters cannot be labelled without error. For example, a cluster that is 50 % forest pixels and 50 % non-forest pixels can only be labelled with 50 % accuracy. To improve cluster purities for the NLCD 1992, analysts manually inspected mixed clusters and developed local threshold-based rules that split mixed clusters in a process that was described by Vogelmann et al. (2001) as "time consuming" and "difficult."[3]

Cihlar (2000) reported that the mismatch between spectral and thematic classes can be reduced by hyperclustering, a process that creates many spectral clusters for each thematic class. The idea that smaller clusters are purer clusters is demonstrated by the large-area maps in Table 2.1, where the number of clusters per stratum range from 100 (NLCD 1992) to 241 (EOSD 2000). Kelly and White (1993) clustered Landsat TM data with up to 4096 clusters and found that as the number of clusters increased, the differences between pixel values and their cluster means decreased. This suggests that as the number of clusters increases, the percentage of pixels in mixed clusters will decrease, if it is assumed that classes do not have identical features. At its theoretical limit, hyperclustering can produce a cluster for each unique feature vector. Although this extreme hyperclustering would produce the highest possible cluster purities, manual labelling of such large numbers of clusters is not practical.

[3]Cluster purity is the percentage of pixels within a cluster that belong to the dominant land-cover class. Cluster purity is also called "Separability Index" (Lark, 1995).

2.1.3 Decision Tree Classification

Decision trees are an automated way to stratify data. The automation is guided by training data, which provide representative examples of the various classes. The key operation is the use of training data to estimate node purity, where a node is a subspace of feature space (akin to a stratum or cluster) and purity is a measure of the label homogeneity of training data within a node.[4] For example, a node that contains 100 forest pixels and 100 non-forest pixels would be assigned a lower purity score than a node with 199 forest pixels and one non-forest pixel.

The basic idea of decision tree classification is to repeatedly split mixed nodes into smaller child nodes until they are pure. To split a node, all possible partitions of the training data for each feature dimension are examined and their purity values quantified. The best partition is chosen (where the best partition has the purest nodes) to split the node into child nodes. This process continues recursively for each node until the purity gained from splitting is outweighed by the estimated loss of generality. The final result is a set of smaller, purer nodes (called leaves) that are a partition of the original root node.

Because decision trees evaluate splits by purity, they are able to select the "best" data by selecting the purest split. Theoretically, a decision tree can be supplied with various data of unknown quality and it will select the best data for each split. However, in practice, decision trees should not be given poor quality data, because random variation sometimes produces purer clusters than truly informative data (Witten and Frank, 2011).

Decision trees are an enormous improvement over 'cluster and label' techniques. Homer et al. (2004) estimated that the decision tree technique used for the NLCD 2001 required about 50 % of the labour of the 'cluster and label' technique used for the NLCD 1992. Consequently, many of the most recent large-area land-cover maps are based on decision

[4]Node purity is essentially the same idea as cluster purity. It is a measure of how mixed the classes are within a node. It can be measured in many ways (e.g., Gini impurity (Breiman et al., 1984), information (Quinlan, 1986)).

trees (e.g., MLCT, NLCD 2001, NLCD 2006, NLCD 2011, GFC, LTCCF).[5]

Decision trees have several properties that are very attractive for large-area land-cover classification (Homer et al., 2004). They are non-parametric classifiers; they do not require a specific data distribution (Hansen et al., 1996). This is very useful, because class features over large areas tend to have multi-modal distributions (McIver and Friedl, 2002). They are also generally insensitive to noise, which is a useful property for handling the variety of data qualities found in large-areas (Friedl and Brodley, 1997). Their speed makes them useful for processing large quantities of data (Homer et al., 2004; Friedl and Brodley, 1997) and permits advanced iterative and ensemble classification techniques, such as random forest, boosting, and bagging (Witten and Frank, 2011). In addition, many types of ancillary data can be integrated into decision trees, because they can handle both categorical and ratio data (Friedl and Brodley, 1997). Decision trees also provide insight into the classification problem because each criterion that partitions feature space is often simple and easily understood (Friedl and Brodley, 1997).

One of the main challenges of using decision trees is the collection of representative training samples (Friedl et al., 2010; Homer et al., 2004). When decision trees are used over large areas, it is a challenge to obtain adequate, well-balanced training data (Homer et al., 2004). For example, Friedl et al. (2010) collected 40,131 training samples from 1,860 sites for the 500-m MLCT but concluded that it was unrealistic to expect that these points captured all the variability of global features. Besides describing feature variability, training data samples should have class proportions that mirror those of the target data. McIver and Friedl (2002) showed how over-representation of a class in a decision tree's training data causes over-representation of the class in the output classification, a result confirmed by Song (2010).

[5]MLCT = MODIS Land Cover Type, NLCD = National Land Cover Database, GFC = Global Forest Cover, LTCCF = Landsat Tree Cover Continuous Fields.

2.1.4 Support Vector Machines

The Support Vector Machine (SVM) algorithm is a classification algorithm that uses training data to calculate hyperplane class boundaries while maximizing the margins between different classes (Cortes and Vapnik, 1995), and which minimize the distances from the boundaries of misclassified training data. It has several advantages for large-area classification. Like the decision tree, the SVM is non-parametric so it can be used on the multi-modal distributions found over large areas. The SVM is also robust to training data errors (Bruzzone and Persello, 2009; Song, 2010) and is frequently more accurate than other classifiers in remote sensing applications (Mountrakis et al., 2011). For example, the Finer Resolution and Monitoring of Global Land Cover (FROM-GLC) project produced the first global land-cover map with 30-m spatial resolution (Gong et al., 2013). It compared the results of maximum likelihood, decision tree, random forest and SVM classifiers and found that the SVM's overall classification accuracy was the highest (64.9 %).[6]

Kernels are frequently used to allow non-linear decision boundaries (Mountrakis et al., 2011). The SVM has a variety of parameters that can be difficult to optimize (Mountrakis et al., 2011). However, Gong et al. (2013) simply used the same parameters for each classification world-wide[7] so that adaptation to local conditions was governed solely by local training data.

Townshend et al. (2012) also compared classification algorithms and chose the SVM to produce a global map of forest cover change. Although the authors had previously used decision trees to produce global maps, they chose the SVM because of its demonstrated value in other studies and because of its resistance to training data error (Townshend et al., 2012; Song, 2010). Remarkably, Townshend et al. (2012) planned to train the SVM by

[6]Other overall accuracies: random forest (59.8 %), decision tree (57.9 %) and maximum likelihood (53.9 %).

[7]Radial basis function (RBF) kernel, cost of 100, and gamma of 0.1.

automatically selecting training samples (Huang et al., 2008).[8]

Selection of training data is the most costly aspect of SVM land-cover classification (Gong et al., 2013). The FROM-GLC project manually collected 91,433 training samples and 38,664 test samples using 27 image interpreters (Gong et al., 2013). This astounding number of samples was still considered to be less than ideal.

When compared to other classification algorithms, the SVM is more accurate in cases where there are relatively low number of training data. Song (2010) found that the SVM did not lose much accuracy when the training-data-to-test-data ratio was less than $\frac{1}{1000}$. Although the SVM ignores training data that are not near the decision boundary (i.e., data that are not support vectors) and which are not misclassified, over-representation in training data causes over-representation in the classification output (Song, 2010). Consequently, training data must be carefully selected to obtain adequate quality, distribution and number (Gong et al., 2013; Song, 2010).

2.1.5 Multi-temporal Data

Most of the maps in Table 2.1 used multi-temporal data to improve classification results. One of the most simple examples is the NLCD 1992, where analysts chose either leaf-on (summer) or leaf-off (spring) imagery for classification (Vogelmann et al., 2001).

One of the increasingly important uses of multi-temporal data is for production of multi-temporal composite images (Hansen and Loveland, 2012). A composite image consists of the best (i.e, least noisy) pixels acquired within a specific time period (e.g., 7-day, 10-day, 16-day, month, season). For example, the MODerate-resolution Imaging Spectroradiometer (MODIS) MOD13A1 product (Huete et al., 1999) contains an NDVI layer where each pixel is the highest quality pixel (at its location) of all images acquired in a sixteen-day pe-

[8]Forest pixels were automatically identified by the characteristic 'forest peak' in histograms of local windows within Landsat red-band images. Forests are typically darker than surrounding land cover and when there are sufficient forest pixels (and no water), they form a characteristic peak in the 'dark' region of the histogram (Huang et al., 2008).

riod (as determined by a model using maximum NDVI, cloud masking, and sensor angle).

The goal of multi-temporal image compositing is high-quality, cloud-free, standardized data. However, to maintain consistency between pixels acquired at different times and conditions, each image requires atmospheric correction. Despite theoretical models for atmospheric scattering and absorption, correcting atmospheric effects is difficult, because the state of the atmosphere is rarely known with high accuracy. The best operational atmospheric correction techniques have 2 to 5 % error that may be unsuitable for some classification applications (Landgrebe, 1999) and result in serious classification errors (Schott, 2007, p. 405).

Although composite image pixels are selected to be less noisy than the pixels of individual images, they still depend on complicated interactions of factors that are difficult to quantify. Instead of having a single, easily-identified spectral signature, classes contain a range of feature vectors that are distributed within a region of feature space (Price, 1994; Cochrane, 2000). For example, Figure 2.1 demonstrates how the red reflectance of two 500-m forest pixels (located within 2.5 km of each other in the Vermilion River watershed) varies between seasons (snow has higher reflectance than green vegetation), varies within seasons (varying amounts of snow and green vegetation), and has frequent spikes (clouds). The difference in reflectance between the two pixels also varies (bottom of Figure 2.1), demonstrating that the factors affecting reflectance vary spatially within the 2.5 km separating the pixels. Differences in snow, cloud, water vapour, and vegetation are some possible factors that vary on this scale. This example demonstrates the difficulty in defining a spectral signature for a class. Another important source of spectral variability is georegistration error. Unsystematic changes in imaging geometry may cause spatial errors in image pixels, resulting in nonplanimetric images (Jensen, 1996, p. 124).

Some recent large-area classification systems derive additional metrics from composite products (e.g., GFC, MLCT). For example, the MLCT is a global land-cover map that is

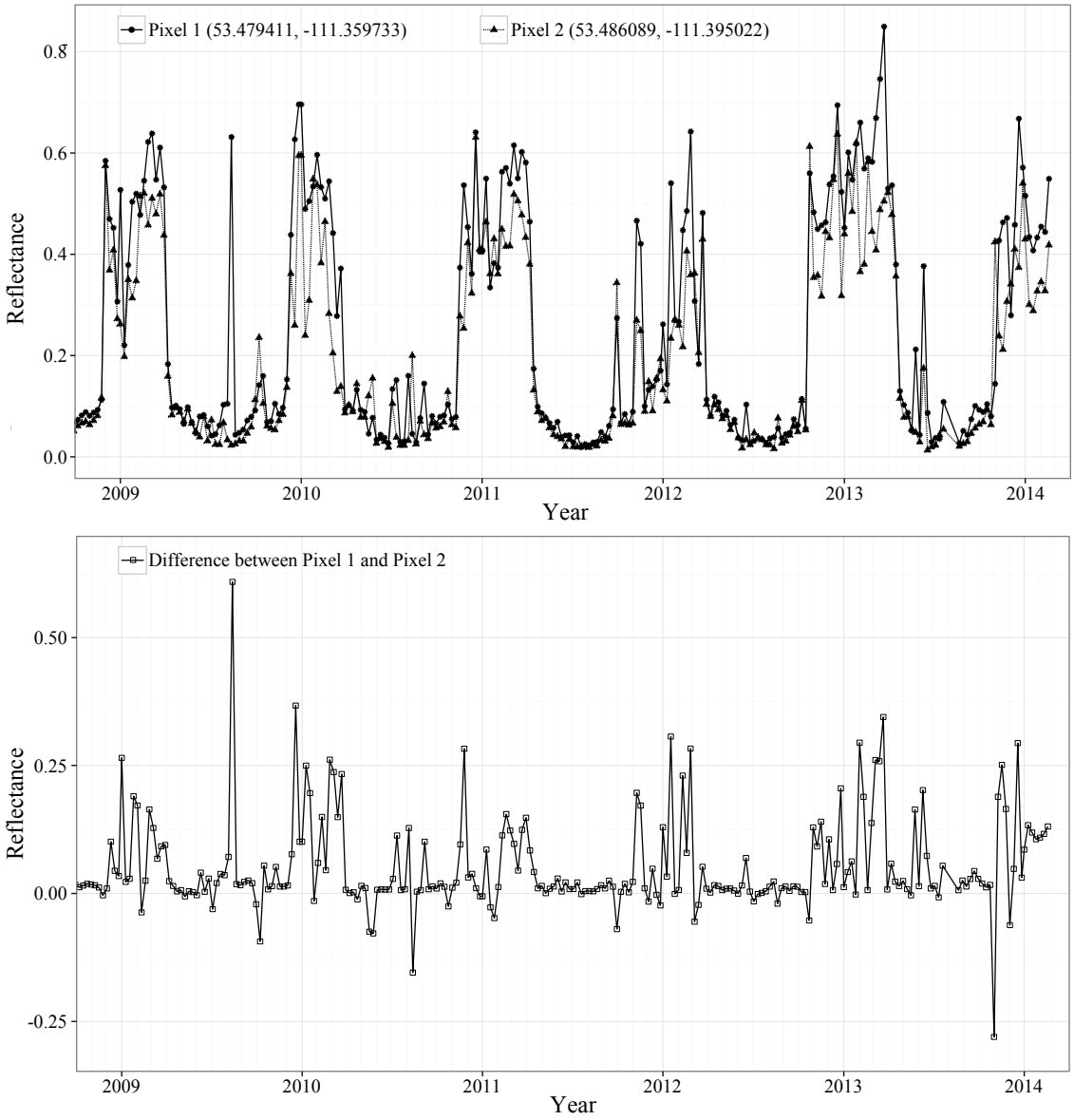

Figure 2.1: Time sequences of 16-day composite red reflectance for two forest pixels within the Vermilion River watershed; (top) snow causes high reflectance in the winter, green vegetation causes low reflectance in the summer, (bottom) the difference in reflectance between the two pixels is variable and sometimes large.

produced annually from MODIS data. It calculates the 32-day means (for twelve 32-day periods), the 365-day mean, the 365-day minimum, and the 365-day maximum for:

1. 16-day composites of MODIS bands 1-7 (the 'best' data for each pixel in a 16-day interval),

2. Enhanced Vegetation Index (EVI; Huete et al., 2002), and

3. MODIS Land Surface Temperature (LST; Wan et al., 2002)

for a total of 135 different features (Friedl et al., 2010).

Multi-temporal metrics improve classification accuracies by distinguishing land cover by its phenology and temporal variability (Friedl et al., 2010). For example, the original overall accuracy of the FROM-GLC was 64.9 %, but this was improved to 67.1 % using multi-temporal and ancillary data (Yu et al., 2013).

Simple phenological metrics can be very good for distinguishing between land-cover classes. For example, De Fries et al. (1998) suggested that maximum annual NDVI and mean annual NDVI were sufficient for classifying almost the entire globe into a scheme of 13 land-cover classes.

One of the primary difficulties with using multi-temporal imagery has been the cost of data. However, Hansen and Loveland (2012) predicted that automated pre-processing and open data policies will increase the use of large volumes of data. Currently, low resolution products from the MODIS Terra and Aqua instruments are the dominant data for multi-temporal analysis, because of their standardized products and daily revisit times. Although medium resolution instruments (e.g., 30-m Landsat OLI) are more limited by their revisit times, new land-cover characterization techniques that use all available datasets to reduce noise (Zhu et al., 2012) and the increasingly available cloud-free reflectance composites (Hansen et al., 2014) are very promising directions of research for improving classification accuracies and automation. New satellites are being launched, such as

the SENTINEL-2 constellation[9], that will provide opportunities for using denser time series at medium spatial resolution. Unfortunately, Very High Resolution (VHR) products (e.g., 1-m spatial resolution) are still very costly and VHR time series analysis is rare.

2.2 Reusing Existing Land-cover Maps for Classification

When a land-cover map is reused, the output land-cover map should be an improvement over the existing one. Land-cover map reuse has resulted in classifications with improved area coverage (Knorn et al., 2009; Olthof et al., 2009; Lang et al., 2008), thematic consistency (Olthof et al., 2009), ground sample distance (Sexton et al., 2013), and classification accuracy (Jin et al., 2013; Demir et al., 2013; Friedl et al., 2010; Xian et al., 2009). Table 2.2 is a summary of projects that have reused land-cover maps.

[9]Free access to Copernicus Sentinel Data. Accessed: 2014-03-19. http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Free_access_to_Copernicus_Sentinel_satellite_data

Table 2.2: Projects that reused existing land-cover maps.

Name	Reference	Overall Accuracy	Existing Map	Target Data	Algorithm
Data-assisted Labeling Approach	Lang et al. (2008)	86 %	1-m classified orthophotos resampled to 15 m	15-m fusion of TM and ETM+	– Hypercluster target data and overlay with 8,000 random samples from the classified images – Label each cluster with the most numerous class within the cluster
LCC 2000 (above treeline)	Olthof et al. (2009)	82 %	4 protected area land-cover maps; other maps were used to stratify clusters	30-m ETM+	– Hypercluster radiometrically normalized mosaics – Stratify clusters using climate and soil maps – Label clusters using dominant class from existing land-cover maps
Chain Classification	Knorn et al. (2009)	96 %	30-m Forest/Non-forest (derived from training data)	30-m ETM+	– Perform an initial classification using an SVM with manually-acquired training data – Use the classification as training data for an SVM classification of an overlapping image – Repeat
NLCD 2006	Xian et al. (2009)	78 % (16 classes), 84 % (8 classes)	30-m NLCD 2001	30-m TM, ETM+	– Get difference image between two co-registered Landsat images – Identify large differences as changed pixels – Classify changed pixels with a decision tree trained by sampling unchanged pixels that are labelled by the existing land-cover map
MLCT (MCD12Q1)	Friedl et al. (2010)	75 %	1000-m Collection 4 MLCT	500-m MODIS	– Calculate class-conditional probabilities using a decision tree – Adjust probabilities using prior-probabilities derived from MODIS Collection 4 land cover – Label each pixel with the most likely class
NLCD 2011	Jin et al. (2013)	Unknown	30-m NLCD 2006	30-m Landsat	– Identify changed pixels between two pairs of co-registered images using multiple indices – Classify changed pixels with a decision tree trained by sampling unchanged pixels that are labelled by the NLCD 2006
LTCCF	Sexton et al. (2013)	R^2 = 0.81, RMSE = 17 % when compared to test sites	250-m MODIS Tree Cover VCF, 250-m Cropland Probability, 30-m TDA-SVM classifications	30-m TM, ETM+	– Overlay 250-m MODIS VCF with Landsat reflectance and temperature data – Supplement low tree-cover areas with other datasets – Apply regression relating spectra to tree cover

EOSD = Earth Observation for the Sustainable Development of Forests, LCC 2000 = Land Cover Circa 2000, LTCCF = Landsat Tree Cover Continuous Fields, MLCT =-mODIS Land Cover Type, NLCD = National Land Cover Database, TDA = Training Data Automation, VCF = Vegetation Continuous Fields

One of the main problems with reusing land-cover maps is low label accuracy. Large-area maps in particular tend to have lower accuracies than smaller maps (Section 2.1). Consequently, one of the most important preconditions for map reuse is the ability to handle label errors.

2.2.1 Data-Assisted Labeling Approach

Data-assisted Labeling Approach (DALA) is a technique that reuses land-cover maps and has excellent potential for error correction (Lang et al., 2008). The DALA algorithm is essentially hyperclustering with automated labelling using a majority rule. DALA automatically partitions a Landsat scene into clusters of similar pixels. Labels sampled from a coincident land-cover map are then used to automatically label each cluster. Each cluster is assigned the class from the map that dominates the cluster (i.e., the class whose intersection with the cluster has the greatest area).

The type of clustering applied to the target data strongly affects DALA's classification accuracy (Lang et al., 2008). Depending on the number of clusters and convergence threshold parameters used to cluster, overall accuracies for DALA ranged from 67 % to 86 %.[10] Accuracies increased as the number of clusters increased (presumably the effects of mixed clusters decreased).

In theory, DALA has great potential for error correction. For a cluster to be labelled correctly, the number of correct land-cover class labels within the cluster must be greater than for any other class. This requirement can correct very high percentages of label error. For example, class labels from the existing map that overlie a pure cluster only need to be >50 % accurate for the cluster to be labelled properly. For three classes, label accuracies can be as low as 33 %; for four classes, 25 %. These accuracies are achievable by large-area maps (Table 2.1).

[10] 86 % was achieved with ISODATA configured to produce 512 clusters using a 99-% convergence threshold.

DALA has some interesting algorithmic properties, because it does not use training data in the typical way. It uses labels (and their locations) to label structures within target data. As long as sufficient structural information about class boundaries can be extracted from the target data, any type of target data (e.g., from radar, LiDAR, or optical sensors) can be classified. Consequently, DALA does not require target data to be atmospherically corrected or calibrated in any way. This could be useful for automation.

The DALA algorithm is an example of semi-supervised classification, because it uses both labelled and unlabelled data (Zhu, 2008). Unlike supervised classifiers (which only use training data) or unsupervised classifiers (which only use target data), DALA extracts information from both training labels and target images.

2.2.2 Map Reuse using Cluster and Label

Olthof et al. (2009) produced a large-area map of Canada's north using a cluster and label technique in which existing land-cover products were used to both stratify and label clusters. Multiple Landsat images were radiometrically normalized and mosaicked together. The mosaics were clustered and labelled using existing small-area land-cover maps. Although, in principle, the map could have been generated automatically using a technique similar to DALA, Olthof et al. (2009) used image interpreters to split mixed clusters and to label clusters.

2.2.3 Chain Classification

Labels from an existing map were paired with coinciding data of an overlapping target image and used to train an SVM that classified the target image (Knorn et al., 2009). This process was repeated in chain-fashion so that eight overlapping Landsat scenes were classified into forest/non-forest with 96 % overall accuracy. Georeferencing errors between the source labels and target images were reduced by using only the interiors of label patches as

training data.

Chain classification is conceptually similar to DALA, because it transfers labels between images (without transferring features) and partitions the feature space using target features. Consequently, target images do not have to be calibrated and do not have to be from the same imaging modality. For example, labels from a radar image could be used to classify a Landsat image.

2.2.4 MODIS Land Cover Types

The MODIS Collection 5 MLCT is a global 500-m land-cover map that reuses class labels from the MODIS Collection 4 MLCT (Friedl et al., 2010). It is produced by the following algorithm:

- calculate class-conditional probabilities using a boosted ensemble decision tree with 40,131 training samples,
- adjust class-conditional probabilities using prior probabilities estimated from the MODIS Collection 4 land-cover map (and other data), and
- label each pixel with the most likely class.

Each pixel's prior probability for each class is estimated from the MODIS Collection 4 land cover by calculating the percentage of the class label within a 151-pixel x 151-pixel window. The large window (151 km x 151 km) means that the prior probabilities of each pixel are affected by regional land cover.

Each prior probability is weighted by a confidence parameter that determines the influence of the prior probabilities relative to the class-conditional probabilities. This parameter was optimized through trial and error to weight class-conditional probabilities to have four times the effect of prior probabilities. Class-conditional probabilities are calculated for each class and pixel by a boosted ensemble decision tree (McIver and Friedl, 2002). They

are combined with the previously calculated prior probabilities using Bayes' rule (Gelman et al., 2013). Finally, each pixel is labelled with the most likely class.

Within the MLCT classifier, the influence of prior probabilities is relatively weak so that map labels are mostly determined by decision tree results. The probabilities derived from reused MODIS Collection 4 MLCT labels are less accurate than the probabilities determined by the decision tree.

The MLCT is very interesting, because it explicitly weights the influence of spectral data features relative to land-cover map labels. Confidence for the output land-cover label at each pixel is also reported, which could be very useful for other products that might reuse the MLCT.

2.2.5 National Land Cover Databases

The NLCD 2006 used the NLCD 2001 as a base map (Fry et al., 2011). In essence, the NLCD 2006 was developed by reclassifying areas that had changed since 2001. Areas of change were identified by comparing co-registered, radiometrically-normalized Landsat images from 2001 and 2006 (Jin et al., 2013; Xian et al., 2009). Change vectors between the images and other derived metrics (such as NDVI) were calculated for each pixel and combined using a series of complex rules to produce a measure of how much the spectra of each pixel had changed. Changed areas were classified using a decision tree with training data supplied by a sample of unchanged labels from the NLCD 2001 and their corresponding spectra from the 2006 image (Xian et al., 2009; Fry et al., 2011).

Development of the NLCD 2011 was conceptually similar, but used two pairs of Landsat images and a specialized method for calculating changes in forests (Jin et al., 2013). In addition, untrustworthy labels in the NLCD 2006 were identified by calculating the Normalized Spectral Distance (NSD), which determines whether a pixel is a spectral outlier

compared to the mean of its class. The NSD of a pixel is:

$$\text{NSD} = \sum_{x_i \in \vec{x}} \left[\frac{x_i - \mu_{ic}}{\sigma_{ic}} \right]^2, \tag{2.1}$$

where i is the Landsat band number, c is the land-cover class of the pixel, $\vec{x} = \{x_1, x_2, \ldots, x_b\}$ is the feature vector of the pixel, x_i is the pixel value for band i, μ_{ic} is the mean of band i for class c, and σ_{ic} is the standard deviation for band i and class c (Jin et al., 2013). The Integrated Forest Index (IFI; Huang et al., 2008) and Z-Score Distance (ZSD; Parshakov et al., 2014) are similar metrics.

In essence, the NLCD 2011 calculated a confidence value for each label of the NLCD 2006 using a statistical metric derived from the NLCD 2006 labels and two Landsat images from 2006. The NLCD 2011 also used each pixel's label from the NLCD 2006 to predict whether the pixel was "dynamic" (likely to change land cover) or "stable" (less likely to change) and adjusted the sensitivity of the change detection algorithm accordingly (Jin et al., 2013). All of this information was combined to produce a map of the potential for spectral change.

The NLCD 2011 reused the NLCD 2006 in a very clever and appealing way. Presumably, the influence of the spectral data relative to that of the NLCD 2006 labels has been optimized by factors similar to the confidence parameter of the MLCT. However, at its foundation, the NLCD relies on comparison of spectral features that vary from one date to another in ways that may be unpredictable and difficult to correct.

2.2.6 Regression with an Existing Product

Sexton et al. (2013) used a regression tree to produce a global 30-m map of percent-tree cover using the 250-m MODIS VCF tree-cover layer (DiMiceli et al., 2011). The median of the MODIS VCF tree-cover from 2000 to 2005 was calculated for each pixel and regressed to Landsat-derived surface reflectance and temperature data. Data from the MODIS Crop-

land Probability Layer (Pittman et al., 2010) and Training-Data Automation and Support Vector Machines (TDA-SVM) algorithm (Huang et al., 2008) were used to improve accuracy in areas of low tree cover. The new 30-m product showed good agreement with high-accuracy LiDAR-derived tree-cover data (Coefficient of determination (R^2) = 0.81, Root Mean Square Error (RMSE) = 17 %).

2.3 Cluster Busting

Cluster busting is an iterative technique for purifying clusters. Like the decision tree algorithm, cluster busting evaluates the purity of each cluster and iteratively partitions clusters that are not sufficiently pure. The main differences are as follows:

- Cluster busting applies multi-variate clustering to the target image; the standard decision tree algorithms partition training data band by band (to produce a model that is used to partition the target image), and
- Cluster busting merges all mixed clusters into a single cluster before splitting.

A selection of cluster busting projects are summarized in Table 2.3.

Cluster busting has been performed manually (Jiang et al., 2004; Jensen et al., 1987), semi-automatically (Wayman et al., 2001), and automatically (Phillips et al., 2012; Musy et al., 2006; Song et al., 2005; Scrivani et al., 2001). Manual cluster busting is essentially a cluster and label technique with advanced clustering to produce pure clusters. The semi-automatic and automatic cluster busting techniques require training data (and therefore human oversight) to guide classification.

Cluster busting has produced excellent results. Jiang et al. (2004) performed manual cluster busting over the entire Pacific Northwest of the United States and obtained 91% overall accuracy when classifying mature and old conifer forest.

Table 2.3: Cluster busting projects.

Region	**Area (km^2)**	**Data Source**	**Reference**	**Classes**	**Overall Accuracy (%)**	**Initial Clusters**
South Carolina	0.34	5.6-m aircraft MSS	Jensen et al. (1987)	Water, forests, and swamps	Unknown	50
Virginia	18,888	30-m TM	Wayman et al. (2001)	Forest, Non-forest	80-89	500
Virginia	23,000	30-m ETM+	Scrivani et al. (2001)	Forest, Non-forest	89	500
Pacific northwest	253,666	30-m ETM+	Jiang et al. (2004)	Mature conifer, Old conifer, Other	91	60-80
Paraguay	Unknown	30-m ETM+	Song et al. (2005)	Forest transition classes	60	256
Minnesota, Virginia	124,987	30-m ETM+	Musy et al. (2006)	Forest, Non-forest	82-95	100
Virginia	3 image mosaic	30-m ETM+	Phillips et al. (2007)	Forest, Non-forest	90	70
Virginia	Partial image	30-m ETM+	Phillips et al. (2012)	Forest, Non-forest	80-89	10-25

Wayman et al. (2001) developed a semi-automated cluster busting technique named Iterative Guided Spectral Class Rejection (IGSCR) that was later automated (Musy et al., 2006) and implemented in parallel using 64 processors to improve execution times (Phillips et al., 2007). Wayman et al. (2001) initially hyperclustered a Landsat TM image into 500 clusters using ISODATA (Ball and Hall, 1965). Cluster purity was evaluated using training data (like with decision trees). Mixed clusters were then merged and reclustered until they were sufficiently pure or contained less than 10 training samples.

Wayman et al. (2001) reported that after cluster busting, some clusters were still impure (and could not be labelled accurately). Accordingly, the entire scene was reclassified using a maximum likelihood classifier that was trained with the labels and mean features of pure clusters. Presumably, by using a hybrid technique in which only pure clusters were used to train a supervised classifier, the pixels in the mixed clusters were classified more accurately

than they would have been by their dominant class labels.[11]

Song et al. (2005) automated cluster busting and labelling, but only achieved an overall accuracy of approximately 60 % when classifying forest changes. They speculated that more training data were needed to achieve higher accuracy. In contrast, the SVM approach achieved a much higher overall accuracy of 92 %.

Cluster busting classification accuracies are sensitive to a variety of factors. Wayman et al. (2001) stressed the importance of cloud-free scenes and high quality training data. Musy et al. (2006) found that IGSCR's overall classification accuracy decreased linearly with the amount of fragmentation in an image. Automated cluster busting requires sufficient training data. Phillips et al. (2012) used 29,000 training data points to classify a single Landsat scene.

Not surprisingly, IGSCR's clustering parameters also affected classification accuracy. In a grid search, Phillips et al. (2007) varied parameters for the homogeneity (i.e., cluster purity) threshold and the initial number of clusters. Overall classification accuracies ranged from 84.4 % (60 clusters, 90 % homogeneity) to 88.9 % (70 clusters, 70 % homogeneity). They concluded IGSCR parameters should be adjusted when training data or images change.

The IGSCR classifier and its variants have been applied over large areas with overall accuracies that ranged from 80 to 95 %. Although IGSCR is an automated form of cluster busting, it still has a high labour cost, because of the need for extensive, high-quality training data. Presumably, IGSCR has never been applied to data having spatial resolutions higher than 30 m.

[11] See Phillips et al. (2012) for more IGSCR labelling techniques.

2.4 Clustering and Mixed Clusters

Cluster purity is very important, because it determines the maximum accuracy with which a cluster can be labelled. In practice, mixed clusters are the primary source of error in unsupervised classification. However, the problem of maximizing cluster purity is difficult, because clusters and land-cover are defined in different domains. For example, the k-means algorithm produces clusters that minimize within-cluster variance (MacQueen, 1967).[12] That is, for a set of feature vectors $\vec{x} = \{x_1, x_2, \ldots, x_b\}$, k-means produces a partition $X = \{K_1, K_2, \ldots, K_m\}$, where the mean of K_i is $\vec{\mu}_i$, such that:

$$X = \underset{X}{\operatorname{argmin}} \sum_{K_i \in X} \sum_{\vec{x} \in K_i} \|\vec{x} - \vec{\mu}_i\|^2. \tag{2.2}$$

Suppose a forest is defined as an area containing woody vegetation that is at least 3-m high and that an image was partitioned into 100 % pure forest clusters and 100 % pure non-forest using Equation 2.2. The same clustering can be considered impure simply by changing the conceptual definition of forest to exclude what were previously defined as forest pixels. For example, using the same clustering, some forest clusters would be impure if the forest height requirement was changed from 3-m to 5-m, assuming there are instances of 3-m woody vegetation in the image.

When land-cover classes occupy distinct, widely-separated regions of feature space, cluster purity can be high. For example, Li et al. (2014) tested 15 different classification algorithms and found that Iterative Self-Organizing Data Analysis Technique (ISODATA) clustering (Ball and Hall, 1965) was almost as accurate as the best supervised algorithm for classifying urban land cover. However, when different land-cover classes exist as a continuum of spectral features, the boundaries of clusters cannot be expected to accurately delineate an arbitrary boundary between conceptual classes. In contrast, supervised clas-

[12] k-means is the most popular clustering algorithm in remote sensing (Tso and Mather, 2009, p. 54).

sification has the advantage that arbitrary boundaries in feature space can be specified by example using training data.

Cluster purity is further complicated by the variety of clustering algorithms and definitions of clusters. No single clustering algorithm is able to recognize all types of structure (Jain et al., 1999). Therefore, there is a need to experiment to determine the application-specific accuracy and generality of clustering algorithms. Duda and Canty (2002) clustered and manually labelled Landsat TM data with five different clustering algorithms using farmland, forest, water and built-up classes. They found that overall classification accuracies ranged from 73.1 % (extended k-means) to 92.8 % (fuzzy k-means), but emphasized that these accuracies may have been influenced by essentially subjective choices of the number of clusters and land-cover classes.

2.5 Spatial Resolution and Mixed Pixels

The spatial resolution of a target image data can have a very strong effect on classification accuracy. As spatial resolution increases, feature variability tends to increase (Aplin, 2006; Townshend, 1980). For example, trees can be imaged at such high-spatial resolution that bark, branches, and leaves can be clearly distinguished. However, trees can also be imaged at low resolution so that they appear as featureless green blobs. Despite the additional information, it is often more difficult to classify spatially high-resolution data because of the high variability within classes and the resulting feature overlap (Hsieh et al., 2001; Townshend, 1980).

However, in some cases, high-resolution data can decrease feature overlap by reducing the area of mixed pixels (Hsieh et al., 2001). Mixed pixels are pixels that are a blending of spectrally distinct objects. For example, a pixel that straddles the boundary between forest and road is likely to have a feature vector that differs from both.

The effect of mixed pixels on classification accuracy is surprisingly strong. In simu-

lations, Lechner et al. (2009) classified small, isolated patches of land cover. The patches were classified at all possible orientations and positions with respect to the pixel grid. Using a perfect classifier (that always labelled a pixel with its dominant class), the average classification accuracy of 3.3-pixel x 3.3-pixel patches was 75 %. This means that 25 % of patch area was misclassified because of mixed pixels dominated by the matrix. Consequently, for large numbers of tree patches of this size and shape, 75 % is the best accuracy that can be expected! Therefore, for accurate classification, the pixel size needs to be many times smaller than the patch size (Lechner et al., 2009).

In landscapes with many small or convoluted patches of different classes (where mixed pixels are common), land-cover classification is especially problematic (Smith et al., 2002). In a comparison of four global land-cover maps, Herold et al. (2008) found the most disagreement in transition zones between ecosystems and attributed this to mixing of land cover.

2.6 Classification Accuracy Assessment

Congalton and Green (2008, p. 70) attested that single pixel sampling units are very poor for assessing accuracy because of registration errors, mixed pixels, and because minimum mapping units are typically larger than a single pixel. They suggested that clusters of pixels (e.g., a 3 x 3 pixel sampling unit) or polygons be used to minimize errors from misregistration. In the case of the Vermilion River watershed, larger sampling units are not suitable for resolving small remnant tree patches. Within large tree patches, a single pixel sampling unit is just as effective as larger units.

2.7 Considerations for Template-Guided Classification

Clearly, there has been a lot of expertise invested in producing land-cover maps of the world. Experts have spent thousands of hours interpreting images. Hundreds of thousands

of training data points have been collected worldwide, and a wealth of operational decision rules have been developed. Template-guided classification may provide a way to extract this expert information.

The most serious problem with reusing existing land-cover maps seems to be the high rates of classification error. Gong et al. (2013) summarized overall accuracies of land-cover maps which ranged from 11 % up to 50 %. Without additional information, these accuracies may be too low to be useful for map reuse. However, an advantage of using large-area maps is that there is a lot of data. Random error might be greatly reduced by averaging large numbers of samples.

The key to map reuse is the ability to correct high rates of classification errors. The SVM has been reported as having the best land-cover classification results with high percentages of mislabelled training data (Song, 2010; Bruzzone and Persello, 2009). For example, the SVM was found to classify forest change with over 90 % accuracy when there is up to 30 % training data error (Song, 2010). Knorn et al. (2009) demonstrated that the SVM was sufficiently robust to reuse forest maps to accurately classify new images (Knorn et al., 2009).

However, the theoretical error correction potential of pure clusters is even more attractive (Lang et al., 2008). If a cluster is pure, classification is 100 % accurate for up to 50 % label error. However, this accuracy depends on cluster purity that may be difficult to obtain. Nonetheless, good classification results have been obtained with 128 clusters (Lang et al., 2008).

Despite the good results of DALA, hyperclustering in large-area maps has been problematic because of mixed clusters (Vogelmann et al., 2001). With mixed clusters, the error correction potential of a cluster and label algorithm cannot be realized. Either cluster busting or decision tree methods can be used to purify clusters. Although cluster busting has been reported as being sensitive to training data quality (Wayman et al., 2001), robustness

to training data error would seem to be its strength.

Whether cluster busting or decision trees produce purer clusters is unknown. However, the hierarchical tree structure has been analysed extensively in the decision tree literature, and decision trees have many established techniques for improving classification accuracy. Therefore, it seems that producing pure clusters using a decision tree has more immediate potential.

Wayman et al. (2001) used pure clusters as training data for a maximum likelihood classifier, while Huang et al. (2008) used high probability forest pixels and high probability non-forest pixels as training data for an SVM. TGC could profit from using a similar approach to labelling impure clusters.

Land-cover classes that overlap in feature space are also a concern. If possible, overlapping classes should be reduced through automatic stratification with ancillary data (e.g., soil maps, digital elevation maps). Water masks have commonly been used to reduce confusion between water and other classes. In addition, data transforms such as vegetation indices, texture filters, and context filters may be helpful to increase separability.

Friedl et al. (2010) suggested that high-quality global land-cover classification will never be fully automated. Landgrebe (1999) argued that fully automated classification systems are likely to be more limited, and perhaps less cost-effective than systems using the "human assisted machine approach". Both of these conjectures are likely true. However, it can be argued that the proposed TGC algorithm is human assisted, because it uses human expertise embedded in land-cover maps. Whether sufficient information can be extracted is the subject of this thesis.

Chapter 3

Template-guided Classification

3.1 Overview

TGC is a new algorithm that uses an existing land-cover map as a template to automatically classify new images. It is conceptually similar to Sexton et al. (2013), because it uses an existing land-cover product as a template to guide a decision tree, but it is also closely related to semi-supervised 'cluster-busting' techniques (Wayman et al., 2001; Phillips et al., 2012). TGC finds relationships between target image features and large numbers of labels. The fundamental assumption is that while individual labels within a template land-cover map are often wrong, the central tendency of labels is correct.

The beauty of TGC is in its reuse of the expertise embedded in large-area land-cover maps. Maps like the EOSD and NLCD require expert knowledge and skill to produce. TGC uses class labels, which have already been adapted by experts to local conditions, to produce locally-adapted definitions for each class. Because there are many large-area land-cover maps that are freely available, TGC could (in principle) update, downscale, and improve the accuracy of land-cover data around the world automatically.

TGC is based on the two following observations:

1. Successful large-area maps have been made by hyperclustering images and labelling the clusters manually (e.g., Wulder et al. (2008)); and

2. The correct class labels for high spatial resolution clusters can be identified visually by overlaying the clusters with moderately-accurate low-resolution labels.

TGC automates classification by clustering a Region of Interest (ROI) and labelling each cluster with the class to which it best spatially corresponds (Figure 3.1). Template map

labels are an approximation of the true spatial distribution of class labels and are used to infer true class boundaries and labels.

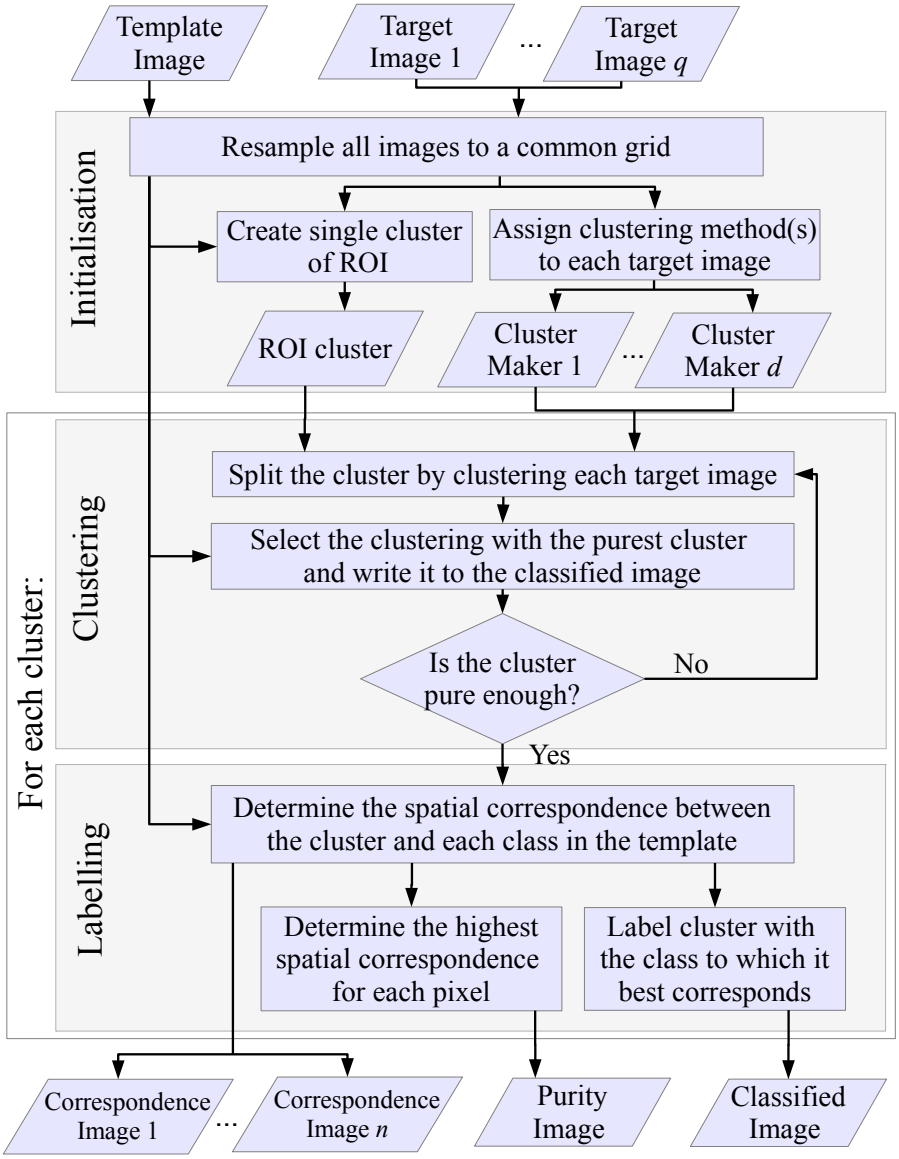

Figure 3.1: Overview of the TGC algorithm. To increase cluster purity, the ROI cluster is repeatedly split by the best clustering produced by the cluster makers. Once the clusters are sufficiently pure, the classified image is generated by labelling each cluster with its best corresponding class. A cluster purity image is produced to estimate the reliability of the TGC classification at each pixel. Correspondence images indicate the strength of the spatial correspondence of each pixel to each class.

3.1.1 Input

The minimum input for TGC is two overlapping images: a target image and a template image (Table 3.1). However, in its most general application, TGC uses multiple target images from various sensors with different resolutions. Target images are used to produce clusters for the ROI being classified; multiple target images may result in purer clusters, because more variety provides more features by which classes may be distinguished.

The template image is used to guide the clustering process to produce clusters that are sufficiently pure and large. They may have lower spatial resolution than the target images. All input images must be georeferenced (or registered to each other) so that spatial relationships between images can be calculated.

Table 3.1: TGC input and output.

Input	Target images	– One or more data images of any type or resolution – Used to produce clusters that distinguish between classes
	Template image	– Covers ROI to be classified – Possibly of lower spatial resolution than the target images – Must be 'moderately' accurate
Output	Classified image	– Classified ROI using the template image's classification scheme – May be of higher resolution than the original template image
	Spatial correspondence images (optional)	– Spatial correspondence of each cluster to each template class – Discriminants for labelling
	Cluster purity image (optional)	– Cluster purity of each cluster with respect to template image labels – Indicates the confidence of correct labelling

3.1.2 Output

TGC outputs a classified image with the same classification scheme as the template image. Optionally, TGC can output class correspondence images that map the spatial correspondence (Section 3.2.1) of each cluster to each template class, or a cluster purity image that

indicates the confidence for each label in the classified image. The output spatial resolutions are equal to the highest resolution of the input images.

3.2 Spatial Correspondence

The foundation of the TGC algorithm is the spatial relationship between template classes and target clusters. Spatial correspondence is the basis for selecting relevant input data, deciding the location of decision boundaries, and labelling pixels. Spatial correspondence is calculated between target clusters (i.e., sets of similar features) and template classes (Figure 3.2). The spatial correspondence a_{ij} between target cluster K_i and template class C_j is as follows:

$$a_{ij} = \left| K_i \cap C_j \right|, \tag{3.1}$$

where the spatial correspondences between all m clusters in clustering $X = \{K_1, \ldots, K_m\}$ and n classes in the template image $TI = \{C_1, \ldots, C_n\}$ are described by the overlapping area matrix A as follows (Baraldi et al., 2005; Beauchemin and Thomson, 1997):

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \tag{3.2}$$

3.2.1 Cluster Purity

Cluster purity is a measure of the strongest relationship of a cluster to any single class and is equal to the percentage of the cluster that intersects its dominant class. More formally,

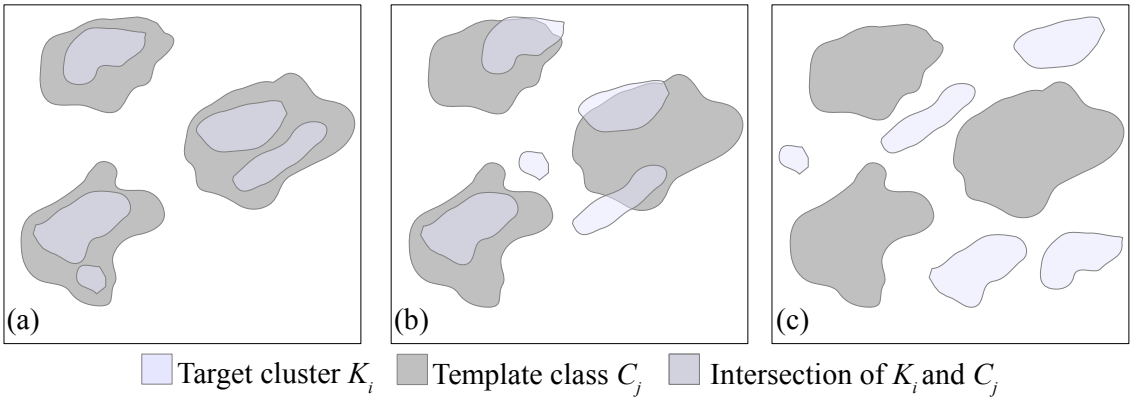

Figure 3.2: Examples of spatial correspondence between a target cluster K_i and a template class C_j; (a) perfect correspondence: $a_{ij} = |K_i|$, (b) partial correspondence: $a_{ij} < |K_i|$, and (c) no correspondence: $a_{ij} = 0$.

the purity ψ_i of cluster K_i with respect to classes $C_1, \ldots, C_n$ is:

$$\psi_i = \frac{\max\limits_{j \in \{1,\ldots,n\}} a_{ij}}{\sum_{j=1}^{n} a_{ij}}. \tag{3.3}$$

Within TGC, cluster purity is used to measure how well a target cluster fits the template image. The assumption is that cluster purity with respect to the template image classes is a good approximation of cluster purity with respect to the actual ground-reference classes.

3.3 Algorithm

3.3.1 Initialisation

To make processing faster and less complicated, the template image TI and target images $I_1, \ldots, I_q$ are resampled to the spatial grid of the input image having the highest spatial resolution. After the images have been regridded, there is no further need for projection and registration information, because corresponding locations within images have the same position in their respective data arrays. The intersection of all valid data defines the ROI for classification.

Clustering algorithms and parameters are defined by initialising cluster makers. A clus-

ter maker is a function object that inputs an ROI and outputs a clustering of the ROI. By default, TGC is configured to create a cluster maker for each target image, which outputs two clusters produced by the k-means algorithm with 12 iterations. However, any number of cluster makers could be produced automatically by combining target images, predefined clustering algorithms, and clustering parameters.

After initialisation, the ROI (defined as a single cluster K), the set of cluster makers ($\Lambda = \{\lambda_1, \ldots, \lambda_d\}$), and the regridded template image (TI) are passed to the template-guided clustering algorithm (Figure 3.1).

3.3.2 Template-guided Clustering

The goal of template-guided clustering is to produce large, pure clusters for the ROI. Clustering is vital to the success of TGC, because mixed clusters limit labelling accuracy and small clusters limit error correction (Figure 3.3). However, in practice it can be very difficult to produce large, pure clusters.

Template-guided clustering bridges the gap between features and classes by using a template image that describes the approximate locations of thematic classes. From the approximate class locations, cluster purity can be estimated by assuming that cluster purity with respect to the template is similar to cluster purity with respect to the true class labels, because the template labels are moderately accurate. Using this cluster purity estimate, TGC automatically selects the clustering with the best (i.e., purest) cluster from candidates produced by the cluster makers. This is important, because it allows TGC to try any number of data and algorithm combinations and to only use the best clustering results.

TGC evaluates the quality of clustering using a simple metric. The quality Q of clustering $X = \{K_1, \ldots, K_m\}$ is equal to the maximum purity of the clustering as follows:

$$Q = \mathit{GetPurityOfPurestCluster(X,TI)} = \max_{i \in \{1,\ldots,m\}} \psi_i, \tag{3.4}$$

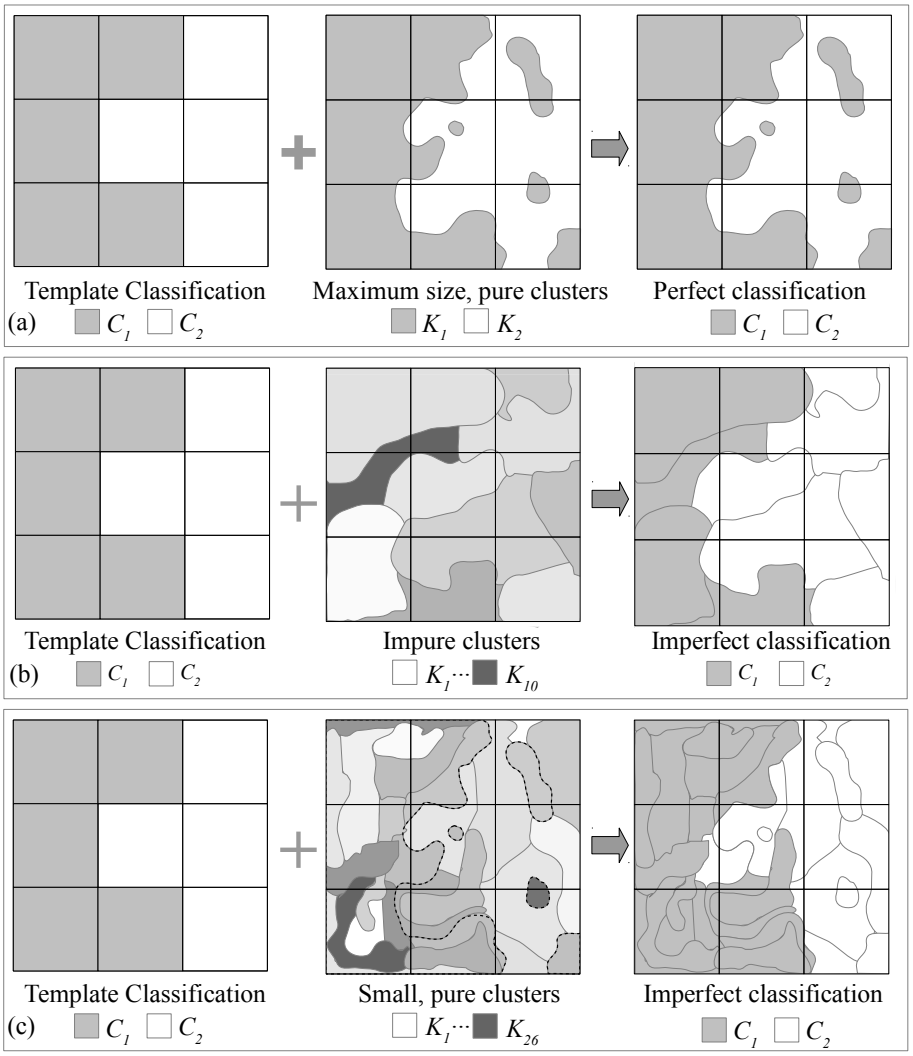

Figure 3.3: Cluster size and purity strongly affect classification accuracy; (a) large, pure clusters are ideal; clustering in which all clusters are pure and the number of clusters is equal to the number of classes will produce a 100 % accurate classification if the producer's accuracy of the template for each class is more than 50 %, (b) mixed clusters cause labelling errors because they contain more than one class, (c) small clusters are undesirable because as cluster size decreases, the ability of the clusters to reduce template label error also decreases. In theory, a clustering where each cluster consists of a single target pixel would produce output identical to the template image (errors included).

where each purity ψ_i is calculated using the overlapping area matrix for X and TI (see Equations 3.2 and 3.3, and Algorithm 1).

The procedure for selecting data for determining decision boundaries is like that com-

```
Input  : X = {K_1, ..., K_m} // Clustering with m clusters
Input  : TI // Template image of class labels
Output: Q // Purity of the purest cluster of X
Function GetPurityOfPurestCluster:
    // Calculate OAM for clustering
    A = CalculateOverlappingAreaMatrix (X,TI) ;      // Equation 3.2
    Q = 0;
    // Calculate purity of each cluster in clustering
    for i =0; i < m do
        // Get spatial correspondences of current cluster
        clusterCorrespondences = A [i,*];
        // Calculate cluster purity (Equation 3.3)
        ψ = max(clusterCorrespondences)/sum(clusterCorrespondences);
        if ψ > Q then
            Q = ψ;
        end
    end
    return Q
end
```

Algorithm 1: Returns the purity of the highest purity cluster of X.

monly used by decision trees but with a different objective function as follows:

$$X_b = \arg\max_X Q(X, TI), \tag{3.5}$$

where $X \in \{X_1, \ldots, X_d | |X| > k_{min}\}$. The best clustering X_b is selected from candidate clusterings $X_1, \ldots, X_d$ by choosing the clustering with the purest cluster that has more than k_{min} pixels (Algorithm 2).

Cluster purity estimates also provide a means for purifying mixed clusters. Much like a decision tree, TGC evaluates the quality of its clusters and iteratively splits mixed clusters to purify them. This assumes that the clustering method is better than random selection at separating pixels of different classes from each other. A cluster K is split if its purity is less than ψ_{min} (where ψ_{min} is a constant; typically $\psi_{min} \approx 0.95$).

TGC's basic algorithm (Algorithm 3) is identical to the following algorithm for con-

```
   Input  : K // Cluster to split
   Input  : TI // Template image of class labels
   Input  : Λ = {λ1,...,λd} // Set of d cluster makers
   Input  : kmin // Minimum number of target pixels allowed in a
            cluster
   Output: Xb // Clustering having the purest cluster
 1 Function GetBestClustering:
 2     ψmax = 0;
 3     Xb = NULL;
 4     // Cluster ROI using each cluster maker and find the best
 5     foreach λ in Λ do
 6         // Split cluster using cluster maker's image and method
 7         X = λ.clusterRoi(K);
 8         ψ = GetPurityOfPurestCluster (X,TI) ;          // Algorithm 1
 9         minSize = GetMinimumClusterSize (X);
10         if ψ ≥ ψmax and minSize ≥ kmin then
11             ψmax = ψ;
12             Xb = X;
13         end
14     end
15     return Xb
16 end
```

Algorithm 2: Splits a cluster using the best clustering out of several candidate clusterings. This function runs each cluster maker (target image and cluster method pair) and selects the clustering with the purest cluster with respect to the template image.

structing a decision tree:

1. Find the best clustering of the data,

2. Evaluate whether the clusters are acceptable, and

3. Split the clusters until they are acceptable or cannot be split further.

3.3.3 Template-guided Labelling

Template-guided labelling automatically labels clusters according to their spatial correspondence with the template classes. Each cluster is labelled with the class with which it

```
Input  : TI // Template image of class labels
Input  : Λ = {λ_1, ..., λ_d} // Set of d cluster makers
Input  : ROI // Region of interest
Input  : k_min // Minimum number of target pixels allowed in a
         cluster
Input  : ψ_min // Minimum cluster purity
Output: X // Clustering of ROI
1  Function templateGuidedClustering:
2      // Initialise cluster queue with roi
3      X = ROI;
4      queue = {ROI };
5      repeat
6          // Get cluster to split from queue
7          K = queue.pop();
8          // Split cluster with each and return the best clustering
9          X_b = GetBestClustering(K, TI, Λ, k_min) ;          // Algorithm 2
10         // Write split cluster to output image clustering
11         X = ReplaceClusterWithClusters(X, X_b);
12         // Add impure clusters to queue
13         foreach K_split in X_b do
14             // Split cluster?  (Equation 3.3)
15             if GetPurity(K_split, TI) ≥ ψ_min and |K_split| > k_min then
16                 queue = queue.push(K_split);
17             end
18         end
19     until queue.size() == 0;
20     return X
21 end
```

Algorithm 3: Template-guided clustering is designed to output a pure clustering of an ROI. It splits the input ROI by choosing the best clustering with respect to the template image (see Algorithm 2) and puts impure clusters in a queue for further splitting.

has the highest spatial correspondence. For cluster K_i, the label y_i is:

$$y_i = \arg\max_j A_{ij}, \tag{3.6}$$

where $j \in \{1, \ldots, n\}$, and A_{ij} is the overlapping area matrix defined by Equation 3.2. Both the label and the purity of each cluster are written to output images (Algorithm 4). The labelled image is TGC's classification of the input ROI. The purity image indicates the purity of each cluster and can be used to estimate the reliability of TGC at each pixel.

```
   Input  : X // Clustering of ROI
   Input  : TI // Template image of class labels
   Output: CI // Classification of ROI
   Output: PI // Cluster purity image for ROI
   Output: SI // Spatial correspondence images
 1 Function TemplateGuidedLabelling:
 2     // Copy ROI for output images
 3     CI = GetROI(X);
 4     PI = GetROI(X);
 5     for j =0; j < n do
 6         SI [*,j] = GetROI(X);
 7     end
 8     // Calculate OAM for clustering
 9     A = CalculateOverlappingAreaMatrix(X,TI);       // Equation 3.2
10     // Write label and purity for each cluster to output images
11     for i =0; i < m do
12         K = X [i];
13         clusterCorrespondences = A [i,*];
14         y = argmax(clusterCorrespondences);              // Equation 3.6
15         CI = WriteLabelToImage(y,K,CI);
16         p = max(clusterCorrespondences)/sum(clusterCorrespondences);
           // Equation 3.3
17         PI = WriteValueToImage(p,K,PI);
18         // Write cluster to each class's spatial correspondence
              image
19         for j =0; j < n do
20             SI [i,j] = WriteValueToImage(A[i,j],K,SI)
21         end
22     end
23     return CI,PI,SI
24 end
```

Algorithm 4: Template-guided labelling. A clustering and template label image are input; a labelled image, a purity image, and a spatial correspondence image for each class are output.

3.4 Correcting Bias in Template Images

Label error in template images is sometimes biased towards a particular class or classes. For example, a map may have a tendency to incorrectly label forest pixels as non-forest pixels. This error can be mitigated by weighting factors that multiply the actual number of pixels of each class. The goal is for the resulting products to better reflect the actual number of pixels on the ground for each class. Therefore, class C_j with k_j pixels on the ground should have the weighting factor w_j such that:

$$w_j = \frac{k_j}{|C_j|}. \tag{3.7}$$

For weighting vector $\vec{w} = (w_1, w_2, \cdots, w_n)$, the weighted overlapping area matrix A_w is:

$$A_w = A \cdot \text{diag}(w_1, w_2, \cdots, w_n) = \begin{pmatrix} w_1 a_{11} & w_2 a_{12} & \cdots & w_n a_{1n} \\ w_1 a_{21} & w_2 a_{22} & \cdots & w_n a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_1 a_{m1} & w_2 a_{m2} & \cdots & w_n a_{mn} \end{pmatrix}. \tag{3.8}$$

In practice, a template's $\vec{w}$ is estimated because k_j is unknown.

3.4.1 The Forest Correspondence Threshold

In the special case of a forest and non-forest classification scheme, template bias can be corrected by introducing the forest correspondence threshold, a value that specifies the minimum spatial correspondence to forest required for a pixel to be classified as forest. The forest correspondence threshold, represented by ϕ_f, is applied to a forest spatial correspondence image to produce a classified image. Every pixel with forest correspondence above ϕ_f is labelled forest, and every other pixel is labelled non-forest. For example, when $\phi_f = 0$, every pixel is labelled forest; when $\phi_f = 1$, every pixel is labelled non-forest.

The forest correspondence threshold is a simple parameter for correcting bias that was shown in this thesis to have a very strong effect on classification accuracy.

Chapter 4

Experimental Set-up

The TGC algorithm was tested on a real problem: the mapping of remnant trees in the Vermilion River watershed. This chapter describes the strategy for mapping remnant trees, classification experiments with TGC, and methods for assessing the results.

4.1 A Strategy for Simplifying the Mapping

In many cases, remnant trees can be identified by their age and species. However, determining tree age and species from multispectral data is a difficult problem. For example, van der Sluijs et al. (2014) obtained overall accuracies of 50 % or lower when using single date Landsat TM imagery to classify leading tree species in the northern boreal forest. Another complication is that age and species do not unambiguously identify remnant trees (Figure 4.1).

This research simplifies the remnant tree classification problem by mapping forest and non-forest classes without distinguishing remnant and non-remnant classes. With this strategy, all remnant trees are mapped as part of the forest class. Remnant studies may use the resulting forest/non-forest maps as masks for simplifying the production of true remnant/non-remnant maps.

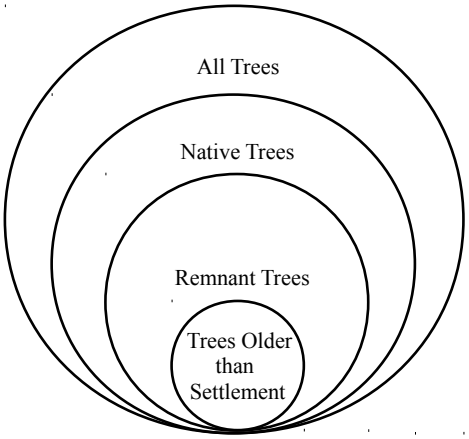

Figure 4.1: Venn diagram showing relationships between all trees, native trees, remnant trees and old trees. Remnants are not necessarily old because natural descendants of remnants are remnants. Artificially planted trees (e.g., windbreaks, plantations) are not considered remnants even if their species is native.

4.2 Study Area

The Vermilion River watershed is a relatively small (7,860 km^2) sub-watershed of the North Saskatchewan River basin, located approximately 150 km east of Edmonton, Alberta (Figures 4.2 and 4.3). It is an interesting study area for mapping remnant trees, because it has been heavily altered by human activities (NSWA, 2005). Many of its aspen groves and grasslands have been replaced by cultivated land and only 0.2 % of the watershed is covered by trees (NSWA, 2005). Thousands of small patches of trees ($<$1 ha) remain throughout the watershed, scattered within agricultural fields (often surrounding ponds, dugouts or swamps), and along roads or field boundaries. Identifying these small patches is a challenging problem. Larger tree patches are found within riparian, lakeshore, and other areas unsuited to agriculture. The watershed is a transition zone: to the north is boreal forest; to the south, grasslands.

Another challenge is caused by the thousands of small swamps and other waterbodies that are scattered throughout the landscape. These wetlands and shallow lakes, which comprise 5.6 % of the watershed's area (NSWA, 2005), have highly variable spectral signatures

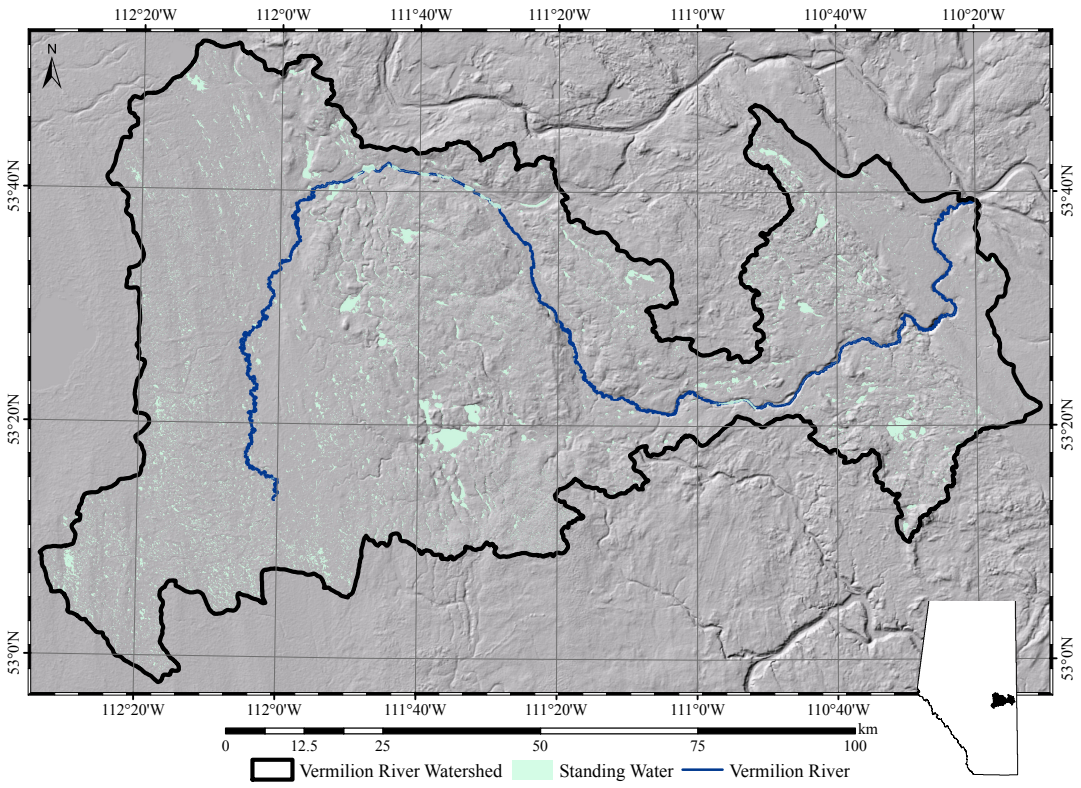

Figure 4.2: The Vermilion River watershed is dominated by agriculture and contains large numbers of small tree patches, wetlands, and shallow lakes. The landscape is fairly flat.

and are sometimes confused with trees (Ozesmi and Bauer, 2002). Larger swamps are also prevalent in the watershed, forming wide borders around shallow lakes and within river widenings.

The landscape is predominantly flat and is recharged by large volumes of snow melt. Large numbers of sloughs and wetlands provide internal drainage, such that the effective drainage of the Vermilion River is only about 30 % of the watershed's total area (NSWA, 2009). The Vermilion River itself can discharge up to 10 m^3/s in the spring and less than 0.02 m^3/s in mid-August,[1] resulting in distinctive vegetation communities within flood-prone areas. Peak discharge is thought to have increased, because of wetland drainage

[1]Vermilion River at Range Road 105 river data: 2013-04-01 to 2013-10-01. Accessed: 2014-01-09, http://www.environment.alberta.ca/apps/basins/DisplayData.aspx?Type=Figure&BasinID=4&DataType=1&StationID=RVERM105.

Figure 4.3: Scenes from the Vermilion River watershed (credits: Alberta Terrestrial Imaging Centre).

(NSWA, 2009).

The watershed is networked with roads (mainly gravel) that form a comprehensive grid of square mile sections based on the Alberta Township Survey (ATS) system.[2] Accordingly, there are many intersections that are suitable ground control points for georeferencing images. Oil and gas pipelines, cutlines, power lines, and the Trans-Canada Highway are other significant linear disturbances. There are two main towns, Vegreville and Vermilion, with populations of approximately 5,000 people.[3] Urban areas cover 0.5 % of the watershed (NSWA, 2005).

4.3 Template Images

Two large-area land-cover maps were used as templates for guiding classifications (Table 4.1 and Figure 4.4).

Table 4.1: Land-cover maps used as templates to guide classification experiments.

Product Name	**GSD (m)**	**Acquisition Date**	**Minimum Mapping Unit**	**Processing**
EOSD	25	1999-2001	30 x 30 m^2	– Mosaic EOSD NTS tiles 83A, 83H, and 73E – Merge all tree classes
GFC	30	2000-2012	Continuous field	– Apply 20-% tree-cover threshold to year 2000 tree cover (from 50°N 120°W to 60°N 110°W)

EOSD = Earth Observation for the Sustainable Development of Forests land-cover map, GFC = Global Forest Cover map.

4.3.1 EOSD Template

The EOSD LC 2000 is a 25-m land-cover map of the forested regions of Canada for the year 2000 (Wulder et al., 2008). Within Alberta, it was derived from the 30-m Alberta

[2]Legal land descriptions. Accessed: 2014-01-17, http://esrd.alberta.ca/forms-maps-services/air-photos/legal-land-descriptions.aspx.

[3]2013 Municipal Affairs Population List. Accessed: 2014-01-08, http://www.municipalaffairs.alberta.ca/documents/msb/2013_Municipal_Affairs_Population_List.pdf.

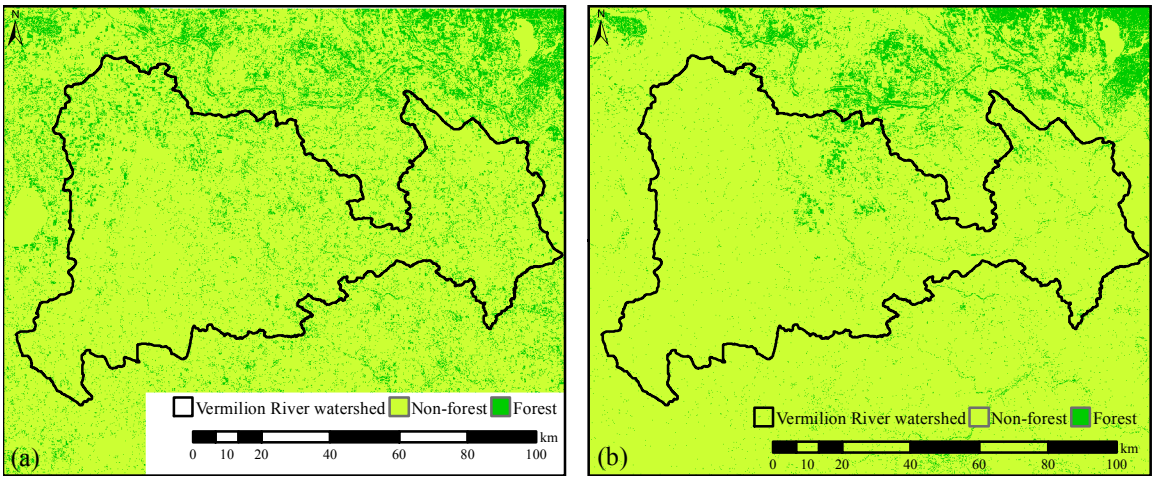

Figure 4.4: Template image area coverage; (a) 25-m EOSD template, and (b) 30-m GFC template.

Ground Cover Characterization (AGCC) map (Wulder et al., 2008). The AGCC itself was produced using an iterative clustering technique with manual labelling (Sanchez-Azofeifa et al., 2004). 30-m Landsat data were initially partitioned into 45 clusters and impure clusters were manually busted. Sanchez-Azofeifa et al. (2004) reported that labelling was the most difficult and time-consuming task in the classification process. Any pixel having tree cover greater than 6 % was considered treed.

The EOSD repackaged the AGCC as 1:250,000 scale map sheets defined by the National Topographic System (NTS) that were resampled to 25-m.[4] Map sheets 73E, 83A and 83H were downloaded[5] and mosaicked within the ENVI (Environment for Visualizing Images; Exelis Vis, Boulder, CO) image analysis system. All tree classes were merged into a single forest class and all other classes merged into a non-forest class. Map sheet 73D (an area south of the watershed) is not published, because it is not forested. Inspection of the images showed that the EOSD-AGCC systematically omitted forest pixels within the experimental ROIs.

[4]The National Topographic System of Canada. Accessed: 2014-01-17, `http://www.nrcan.gc.ca/earth-sciences/geography/topographic-information/maps/9763`.

[5]System of Agents for Forest Observation Research with Automation Hierarchies (SAFORAH). Accessed 2014-11-22, `http://www4.saforah.org/eosdlcp/ab.tar`.

4.3.2 GFC Template

The Global Forest Cover (GFC) map is a 30-m map of forest canopy cover for the year 2000 (Hansen et al., 2013).[6] It is part of a larger set of products that describe forest-cover change from the year 2000 to 2012. The canopy-cover product reports the canopy cover for all vegetation taller than 5-m, expressed as a percentage from 0 - 100 for each 30-m pixel. The product is an amazing achievement that used 654,178 Landsat 7 ETM+ images to characterize forest canopy cover and forest changes worldwide. Data selection, cloud-free pixel compositing, and normalization were automated. Relationships of multi-temporal metrics to training data were modelled with a bagged classification tree (Hansen et al., 2013; Potapov et al., 2012). Although the GFC change products were extremely accurate (> 99 % overall), no forest cover accuracies were reported (Hansen et al., 2013).

To produce a forest/non-forest map, pixels having forest cover values above 20 % were labelled as forest, and all others were labelled as non-forest. Even using this relatively low canopy-cover threshold to define forests, the resulting map had high rates of forest omission.

4.3.3 Water Masking

New templates were produced by masking water in the EOSD and GFC template images. The mask was derived from waterbody shapefiles from the National Hydro Network (NHN)[7] as shown in Figure 4.5. They were merged to provide complete coverage of the target images and rasterized to 10-m pixels by ArcMap's *Feature To Raster* tool. Masked EOSD ($EOSD_m$) and masked GFC (GFC_m) templates were produced using ENVI's *Apply Mask* function.

[6] Global Forest Change 2000-2012 Data Download. Accessed: 2014-09-23. http://earthenginepartners.appspot.com/science-2013-global-forest/download.html

[7] National Hydro Network. Accessed 2014-09-24, http://geobase.ca/geobase/en/data/nhn/index.html

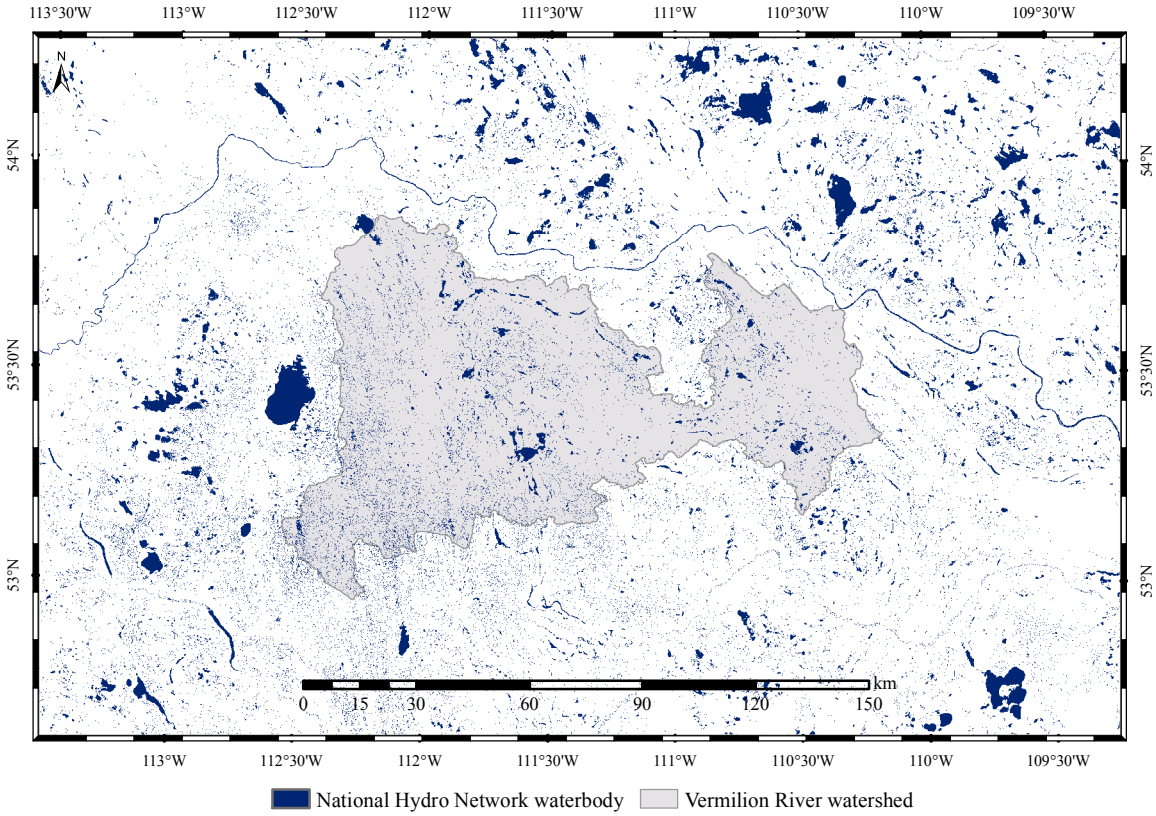

Figure 4.5: A water mask derived from National Hydro Network waterbody shapefiles.

4.4 Target Images

4.4.1 SPOT 5 Images and Experimental ROIs

TGC used six 10-m SPOT 5 HRG MS images[8] and six 2.5-m SPOT 5 HRG Pan images as target data for the Vermilion River watershed (Table 4.2). The images were obtained at level 1A (radiometric correction only) and then automatically georefersenced, and orthorectified by Geomatica. Each multispectral image was atmospherically corrected using aerosol and visibility parameters derived from a database of MODIS products (Zhang et al., 2009). The spectral bands and spatial resolutions of the HRG instrument are shown in Table 4.3.[9]

[8] SPOT = Satellite Pour l'Observation de la Terre; HRG = High Resolution Geometrical; MS = Multispectral; Pan = Panchromatic.

[9] SPOT Resolution and Spectral Bands. Accessed on 2014-03-25, `http://www.astrium-geo.com/en/194-resolution-and-spectral-bands`.

Table 4.2: SPOT 5 HRG target images used in experiments.

Target Image ID	Image Type	GSD (m)	Image Dates	GRS ID
553824108080718372 32J	MS	10	2008-08-07	538-241
553824208080718373 22J	MS	10	2008-08-07	538-242
554124111063018172 22J	MS	10	2011-06-30	541-241
554124210072218145 81J	MS	10	2010-07-22	541-242
554224108062318032 42J1	MS	10	2011-07-19	542-241J1
554224111071918515 32J6	MS	10	2011-06-30	542-241J6
553824106060318523 11B	Pan	2.5	2006-06-03	538-241
553824208081818254 01B	Pan	2.5	2008-08-18	538-242
554124111063018171 92B3	Pan	2.5	2011-06-30	541-241
554124211073018402 11B	Pan	2.5	2011-07-30	541-242
554224108061184159 1B	Pan	2.5	2008-06-11	542-241
554224111071918515 12B9	Pan	2.5	2011-07-19	542-241B9

SPOT = Satellite Pour l'Observation de la Terre, HRG = High Resolution Geometrical, MS = Multispectral, Pan = Panchromatic.

Table 4.3: SPOT 5 HRG bands and spatial resolutions as used by TGC.

Band Name	Band Range (nm)	GSD (m)
Panchromatic	480 - 710	2.5 (superresolution)
B1: Green	500 - 590	10
B2: Red	610 - 680	10
B3: Near Infrared	780 - 890	10
B4: Mid Infrared	1,580 - 1,750	20

Each SPOT MS image defined an experimental ROI (Figure 4.6a). Additional ROIs were defined by the intersection of pairs of SPOT MS and SPOT Pan images having the same Grid Reference System (GRS) coordinates (Figure 4.6b). For convenience, these two types of ROIs are referred to as MS (defined by the extent of a multispectral image), and MS+P (defined by the intersection of corresponding multispectral and panchromatic images).

Each multispectral image was atmospherically corrected using aerosol and visibility parameters derived from a database of MODIS products (Zhang et al., 2009). The mid-infrared band was supersampled from 20-m to 10-m spatial resolution. Panchromatic im-

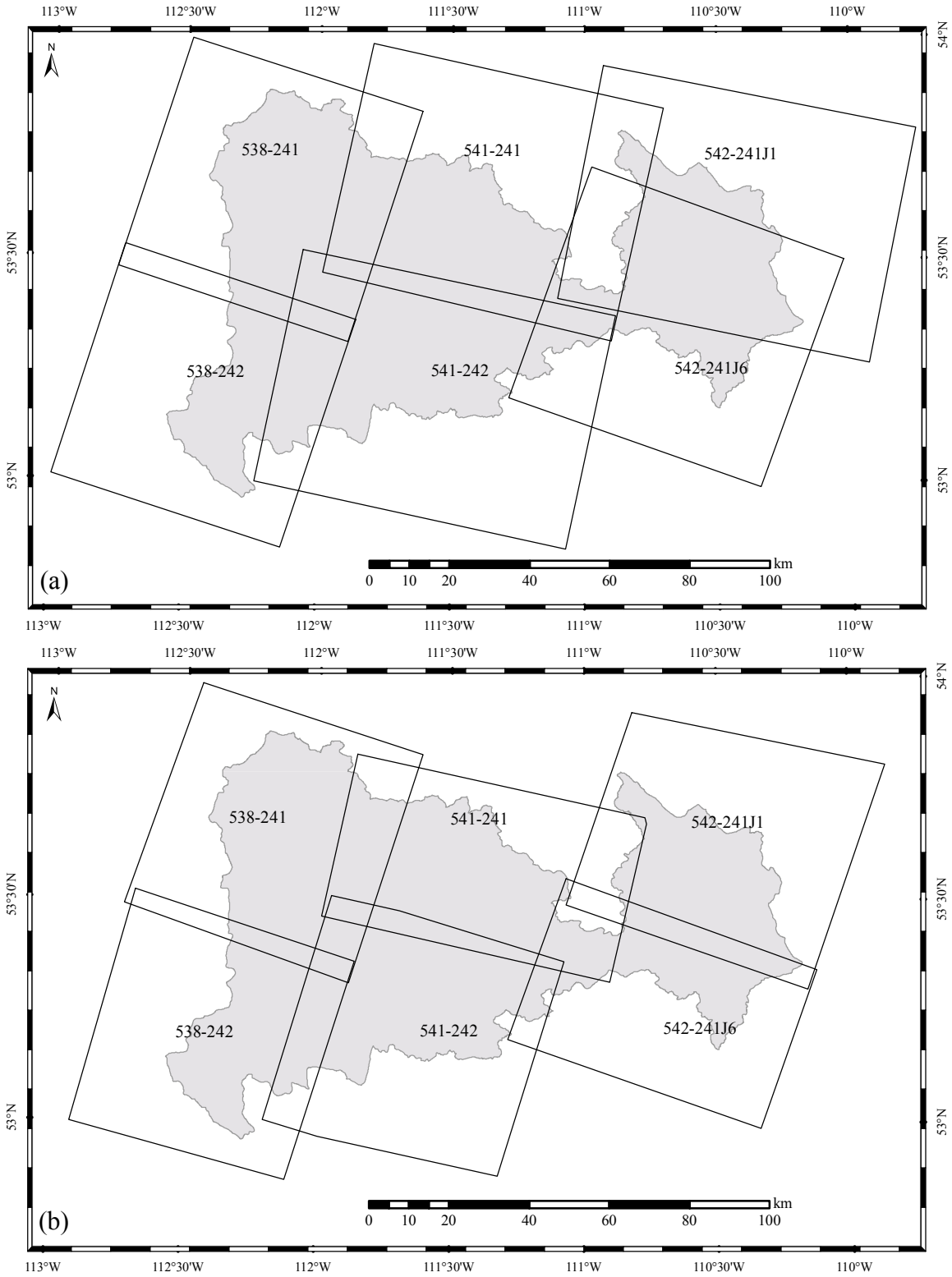

Figure 4.6: ROIs for (a) multispectral experiments, where each ROI was defined by the extent of a SPOT MS image, and (b) multispectral and panchromatic experiments, where each ROI was defined by the intersection of a pair of SPOT MS and Pan images.

ages were resampled to 10-m using aggregation-based resampling in ENVI.

4.4.2 1-m Orthophotos

Two 1-m colour orthophotos were classified by TGC (Table 4.4). Each orthophoto corresponded to a township defined under the ATS system[10] where each township is 6 x 6 square miles (9.656 x 9.656 km^2) and overlaps adjacent townships by approximately 700-m on each side. Both orthophotos were acquired in the summer of 1999 by the Prairie Farm Rehabilitation Administration (PFRA) and used by TGC without any additional processing.

Table 4.4: 1-m orthophoto target images.

Image Township ID	Image Type	GSD (m)	Image Date
T52-R10-W4	Colour orthophoto	1	Summer 1999
T52-R12-W4	Colour orthophoto	1	Summer 1999

4.5 Template-guided Classification Implementation

TGC was implemented as a script in IDL using ENVI libraries on a 64-bit Windows 7 computer with an Intel Core i5-2500 CPU and 16 GB of RAM. Parameters for template images, target images, purity threshold, forest correspondence threshold, and minimum cluster size varied according to the experiment. For example, Figure 4.7 shows an example of applying TGC to SPOT HRG MS data using the EOSD template. All clusters were split using the *k*-means algorithm.

[10] Alberta township survey system. Accessed: 2014-01-09, http://esrd.alberta.ca/recreation-public-use/recreation-on-agricultural-public-land/alberta-township-survey-system.aspx.

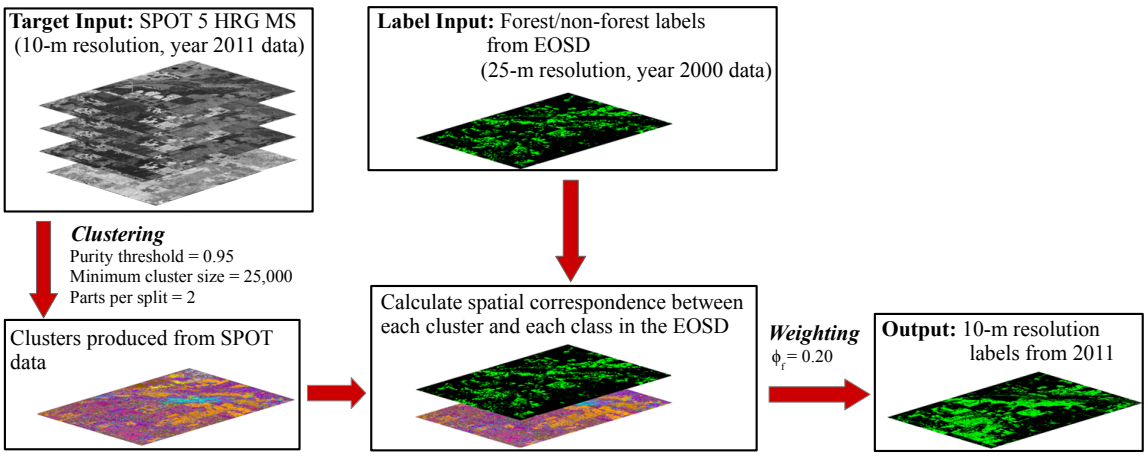

Figure 4.7: An example of TGC applied to a SPOT 5 MS image using the EOSD template.

4.6 Experiments

4.6.1 TGC Watershed Classifications

Experimental classifications were run using the template image, target image, and parameter configurations shown in Table 4.5. Forty-eight forest correspondence images were produced by combining two templates (EOSD, GFC), two masking states (unmasked, masked), and two datasets (MS, MS+P) for six different ROIs.

To evaluate the ability of TGC to achieve good results without locally adapted parameters, forest/non-forest classifications were produced by thresholding forest correspondence images with the same ϕ_f value for each image. For each template image, a constant ϕ_f value was used to classify all images in the watershed as follows:

1. For EOSD-derived images (masked and unmasked), $\phi_f = 0.2$, and

2. For GFC-derived images, $\phi_f = 0.06$.

The constants were selected to produce classifications of 538-241 (i.e., the ROI with GRS ID 538-241) with relatively equal errors of commission and omission (with the respective template). In other words, TGC parameters were optimized for a single ROI and applied to all ROIs.

In addition, each forest correspondence image was classified by thresholding with 34 different ϕ_f values ranging from 0 to 0.5. These classifications were used to evaluate how ϕ_f affected classification accuracies. The accuracies of each classification were calculated using ground-reference data (Section 4.7).

For convenience, TGC experiments were referred to by their ROI and template names. For example, the experiment using multispectral data and the EOSD template was named "MS EOSD."

Table 4.5: Experimental parameters for 10-m TGC experiments.

Template Image	**Target Images**	**Purity Threshold (%)**	k_{min}	**Parts per Split**
EOSD	10-m MS	95	25,000	2
EOSD	10-m MS+P	95	25,000	2
$EOSD_m$	10-m MS	95	25,000	2
$EOSD_m$	10-m MS+P	95	25,000	2
GFC	10-m MS	99	25,000	2
GFC	10-m MS+P	99	25,000	2
GFC_m	10-m MS	99	25,000	2
GFC_m	10-m MS+P	99	25,000	2

MS = multispectral dataset, MS+P = multispectral and panchromatic dataset,
k_{min} = minimum number of pixels in a cluster.

4.6.2 Orthophoto Classifications

Two orthophotos were used to test high-resolution, small-area classification (Table 4.6). Both classifications used a 2.5-m SPOT MS HRG Pan image as a second target image for clustering.

4.6.3 SVM Baseline Classifications

The SVM algorithm was used to produce baseline classifications of the SPOT 5 MS data. Training data for each ROI were selected manually from SPOT 5 MS images in small, scattered clusters. Land cover was visually interpreted from the SPOT images with assistance

Table 4.6: Experimental parameters for 1-m TGC experiments.

Template Image	**Target Images**	**Purity Threshold (%)**	k_{min}	**Parts per Split**
EOSD	T52-R10-W4, 55412411106301817192B3	95	250,000	2
EOSD	T52-R12-W4, 55412411106301817192B3	95	250,000	2

EOSD = Earth Observation for the Sustainable Development of Forests land-cover map, k_{min} = minimum number of pixels in a cluster.

from the 1-m orthophotos. Training data were optimized by iteratively assessing classification accuracy and adding training data for pixels that were mislabelled in the assessment. Optimization stopped when accuracy gains were not qualitatively apparent. All SVM classifications were performed in ENVI using the same parameters: a radial basis function kernel with $\gamma = 0.25$ and a penalty parameter of 100 (i.e., $C = 100$).

4.7 Classification Accuracy Assessment

Classification accuracy assessment was crucial for evaluating and analysing TGC. Accuracies from the following sources were assessed:

- Experimental TGC classifications,
- EOSD and GFC templates, and
- SVM baseline classifications.

Accuracies were assessed by comparing results with ground-reference data acquired by simple random sampling of individual pixels (Congalton and Green, 2008, p.79).

4.7.1 Ground-reference Samples

Ground-reference data were acquired by visually interpreting the classes of 500 to 700 random pixels within each MS ROI. Samples were generated within ENVI by applying the

Random Sample Using Ground Truth tool (Figure 4.8). For the 10-m experiments, pixels were sampled from the pixel grids output by TGC's MS classifications (i.e., each sample polygon had the same location and dimensions as a pixel in the TGC MS classifications). For the 1-m experiments, pixels were sampled from the 1-m orthophoto pixel grids. Sampled pixels were converted into pixel boundary polygons and saved to shapefiles using ENVI's *Classic ROI to Shapefile* tool.

A sample of 500 pixels were initially extracted from each ROI. However, if the number of sampled forest pixels within an ROI was less than 40, an additional 200 ground-reference pixels were added to the sample.

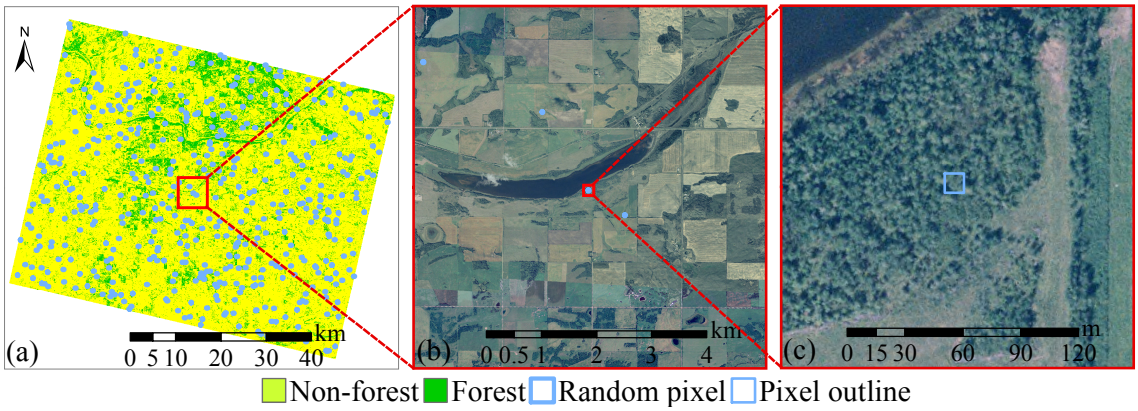

Figure 4.8: Ground-reference data acquired by randomly sampling 500 to 700 pixels within each ROI. Each pixel's class was interpreted using 1-m orthophotos and Google Earth images; (a) a random sample within an experimental ROI, (b) context of a sample pixel within a 1-m orthophoto, and (c) the pixel outline clearly shows that the sample contains trees.

4.7.2 Interpretation

The sample polygons were superimposed on 1-m orthophotos and the contents were examined at high resolution in ArcMap. A class label was written to the attribute table to identify the class of each pixel.

Although orthophotos were the primary source of labelling information, all forest pixels and ambiguous pixels were also examined in Google Earth. For viewing in Google Earth,

each shapefile was converted to Keyhole Markup Language (KML) using the Open Source Geospatial Foundation Tools for Windows (OSGeo4W) tool *ogr2ogr*. An index was added to each shapefile before conversion to KML so that samples in ArcMap were easily matched with those in Google Earth.

A pixel was interpreted as forest if it contained ≥ 30 % tree cover (trees were defined as vegetation with height ≥ 5 m), otherwise, a pixel was interpreted as non-forest. Under this definition, even individual trees were classified as forest, if there was sufficient tree cover.

Differentiation between shrubs and trees was aided by shadows. In images with low solar elevation angles, trees had long shadows and shrubs had short shadows.

In cases where the pixel was interpreted as forest in the orthophoto and as non-forest in Google Earth, the pixel was interpreted as non-forest. In this way, the ground-reference data were made as current as the dates of the Google Earth images (which ranged from 2008 to 2014). To make pixel interpretation reasonably fast, pixels unambiguously interpreted as non-forest in orthophotos were not examined in Google Earth.

4.7.3 Accuracy Assessment Images

Table 4.7 lists the sets of images used for accuracy assessment. Orthorectified 1-m colour

Table 4.7: Accuracy assessment images.

Image Set	**GSD (m)**	**Date**	**Coverage Area**
Orthophotos 1999	1	1999-2001	41,000 km^2
Orthophotos 2011	0.5	August 2011	3,200 km^2
Google Earth	Various	Various	Entire watershed
LiDAR Site	2	2010-7-28	7.5 x 5.25 km^2

aerial photos were obtained that provided very high-resolution coverage of the experimental ROIs. Most of the orthophotos were acquired in the summer of 1999. Many were acquired with low solar elevation angles to more easily identify tree heights by their long shadows. Six townships within the watershed did not have 1-m orthophotos (53-12-W4, 53-13-W4, 54-12-W4, 54-13-W4, 55-12-W4, 55-13-W4).

A smaller set of orthophotos from August 2011 were also obtained for accuracy assessment. This set of 0.5-m orthophotos covers 34 townships within the boundaries of the watershed and the target images, including all townships missing from the set of 1-m orthophotos. These orthophotos were ideal for identifying trees, because of their sharpness and low solar elevation angles.

4.7.4 Mixed Cluster Analysis

The cluster purities of TGC classification of MS EOSD 541-241 ($\phi_F = 0.2$) were measured within a 7.5 x 5.25 km^2 site within the Vermilion River watershed. This site straddled the Vermilion River at 53°29'10"N, 111°21'41"W and was approximately 25 % forest (which was much higher than for any other experimental ROI). Within this site, a high-accuracy forest/non-forest land-cover map was produced from LiDAR data at 2-m spatial resolution.Within each 2-m cell, the last return was subtracted from the first return to produce a Canopy Height Model (CHM). The CHM was thresholded so that values above 5-m were labelled as forest, and otherwise as non-forest. The resulting forest/non-forest map was considered to be ground-reference data because of its high accuracy. The map was then resampled to 10-m spatial resolution using the *Resample* tool in ENVI and used to evaluate the cluster purities of the template-guided clustering of MS 541-241. The cluster purities were determined by calculating the cluster purity image relative to the ground-reference data.

4.7.5 Assessment Software

After production of the ground-reference data, the classification accuracies of the various experiments were calculated. A Python script was written that compared pixel labels in ground-reference shapefiles with classifications obtained by thresholding forest correspondence images with various ϕ_f values. Because the ground-reference samples were gener-

ated for MS ROIs, some samples were outside the boundaries of MS+P ROIs and were ignored by the script when evaluating MS+P classifications. All of the accuracy metrics described in Section 4.7.6 were calculated and saved to files.

4.7.6 Accuracy Metrics

Classification accuracies were measured using statistics calculated from error matrices. The error matrix (Table 4.8), overall accuracy (Equation 4.1), producer's accuracy (Equation 4.2), and user's accuracy (Equation 4.3) were described by Congalton and Green (2008, p. 60). Within the error matrix, f_{ij} represents the number of pixels classified as class i that are actually reference class j. The number of pixels that were classified as class i is f_{i+} and the number of pixels that are actually reference class j is f_{+j}. N is the total number of pixels in the error matrix.

Table 4.8: Error matrix with f_{ij} representing the number of samples classified as land-cover class i that are actually reference land-cover class j, where N is the total number of samples.

		Reference				Row Total
		1	2	$\cdots$	n	
Classified	1	f_{11}	f_{12}	$\cdots$	f_{1n}	f_{1+}
	2	f_{21}	f_{12}	$\cdots$	f_{1n}	f_{2+}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	n	f_{n1}	f_{n2}	$\cdots$	f_{nn}	f_{n+}
Column Total		f_{+1}	f_{+2}	$\cdots$	f_{+n}	N

The overall accuracy (OA) is the proportion of area classified correctly:

$$\text{OA} = \sum_{k=1}^{n} \frac{f_{kk}}{N}. \tag{4.1}$$

The producer's accuracy of class j (PA_j) is the proportion of ground-reference pixels of

class j that are correctly classified:

$$\mathrm{PA_j} = \frac{f_{jj}}{f_{+j}}. \tag{4.2}$$

User's accuracy of class i ($\mathrm{UA_i}$) is the proportion of pixels classified as i that are correct:

$$\mathrm{UA_i} = \frac{f_{ii}}{f_{i+}}. \tag{4.3}$$

The percentage of land cover classified as k ($\%\mathrm{LAND_k}$) is:

$$\%\mathrm{LAND_k} = 100\left(\frac{f_{k+}}{f_{+k}}\right). \tag{4.4}$$

The %LAND metric is useful because it measures the percentage of a particular class within a classification (e.g., the percentage of remnant trees in the Vermilion River watershed). A related metric is the Relative Error of Area for class k ($\mathrm{REA_k}$) (Shao et al., 2003):

$$\mathrm{REA_k} = 100\left(\frac{1}{\mathrm{UA_k}} - \frac{1}{\mathrm{PA_k}}\right), \tag{4.5}$$

that describes the classification error of k relative to the number of correctly classified pixels of class k (i.e., f_{kk}). If there are too many pixels of class k in a classification, $\mathrm{REA_k}$ is positive; if too few, $\mathrm{REA_k}$ is negative.

For convenience, forest metrics were referred to by an "f" subscript and non-forest by an "nf." For example, $\mathrm{PA_f}$ is forest producer's accuracy and $\mathrm{UA_{nf}}$ is non-forest user's accuracy.

4.7.7 Comparing Accuracies of Forest Correspondence Images

The classification derived from a forest correspondence image varies with different values of ϕ_f. Within TGC, the ϕ_f parameter determines the bias towards classifying a pixels as

forest. In practice, it sets the trade-off between PA_f and UA_f. Comparing accuracies of forest correspondence images is complicated by the many trade-offs between PA_f and UA_f that are possible.

The minimum absolute REA_f ($MAREA_f$) classification was chosen as the basis for comparing and evaluating forest correspondence images. This classification is defined as the classification of a forest correspondence image that minimizes the absolute value of REA_f. It is considered the "best" classification of a forest correspondence image, because it balances errors of commission and omission as much as possible so that $\%LAND_f$ is most accurate. The $MAREA_f$ threshold is the ϕ_f that produces the $MAREA_f$ classification as follows:

$$\text{MAREA}_\text{f} \text{ threshold} = \underset{\phi_f}{\operatorname{argmin}} |REA_f|. \tag{4.6}$$

Chapter 5

Results

5.1 10-m TGC Watershed Classifications with Constant ϕ_f

Table 5.1 summarizes the 10-m TGC classification accuracies of the eight parameter combinations of EOSD-based and GFC-based classifications. The PA_f and UA_f values for each template, dataset, and mask combination are graphed in Figure 5.1. All accuracies are tabulated in Appendices A.1 and A.2. These classifications used constant ϕ_f parameters to classify the Vermilion River watershed and were the primary test of TGC's automation and ability to adapt to local conditions without human intervention.

Table 5.1: Constant ϕ_f TGC classification accuracies for the Vermilion River watershed with 10-m spatial resolution. All accuracies from EOSD-based and GFC-based classifications were averaged for each ROI. Overall accuracies of the TGC classifications were high, often exceeding 90 %. However, forest classification accuracies were much lower and highly variable.

Classifications	**Metric (%)**	**538-241**	**538-242**	**541-241**	**541-242**	**542-241J1**	**542-241J6**	**All ROIs**
EOSD and $EOSD_m$ with $\phi_f = 0.2$	$\overline{PA}_f$	66	44	68	33	73	42	54
	$\overline{UA}_f$	58	66	87	85	64	68	71
	$\overline{OA}$	93	94	94	92	87	91	92
	$\overline{REA}_f$	23	-83	-34	-190	19	-89	-59
GFC and GFC_m with $\phi_f = 0.06$	$\overline{PA}_f$	68	43	86	41	85	46	61
	$\overline{UA}_f$	61	68	53	71	55	64	62
	$\overline{OA}$	93	94	87	91	83	91	90
	$\overline{REA}_f$	18	-91	72	-115	66	-61	-19

$\overline{PA}_f$ = average forest Producer's Accuracy, $\overline{UA}_f$ = average forest User's Accuracy, $\overline{OA}$ = average Overall Accuracy, $\overline{REA}_f$ = average forest Relative Error of Area, EOSD = EOSD template, $EOSD_m$ = water-masked EOSD template, GFC = GFC template, GFC_m = water-masked GFC template, ϕ_f = forest correspondence threshold.

Although the overall accuracies for most classifications were greater than 90 %, the TGC forest accuracies were much lower. There were only two classifications for which both PA_f and UA_f were over 70 % (MS EOSD 541-241 and MS $EOSD_m$ 541-241).

5.1.1 Patterns in Classification Accuracy

Classifications frequently had unbalanced forest accuracies, with high errors of commission and low errors of omission (or vice versa). For example, MS GFC_m 541-241 had a PA_f of 88 % and a UA_f of only 53 %; MS+P EOSD 541-242 had a PA_f of only 31 % and a UA_f of 100 %.

In addition to being unbalanced, the forest accuracies were strongly affected by whether the classified ROI was in the north or south (Figure 5.1). All northern ROIs (538-241, 541-241, and 542-241J1) had higher PA_f values than the southern ROIs (538-242, 541-242, and 542-241J6). In other words, forest pixels in the northern ROIs were more likely (up to 55 % more likely) to be classified correctly and false negatives for forest were far more likely in the south.

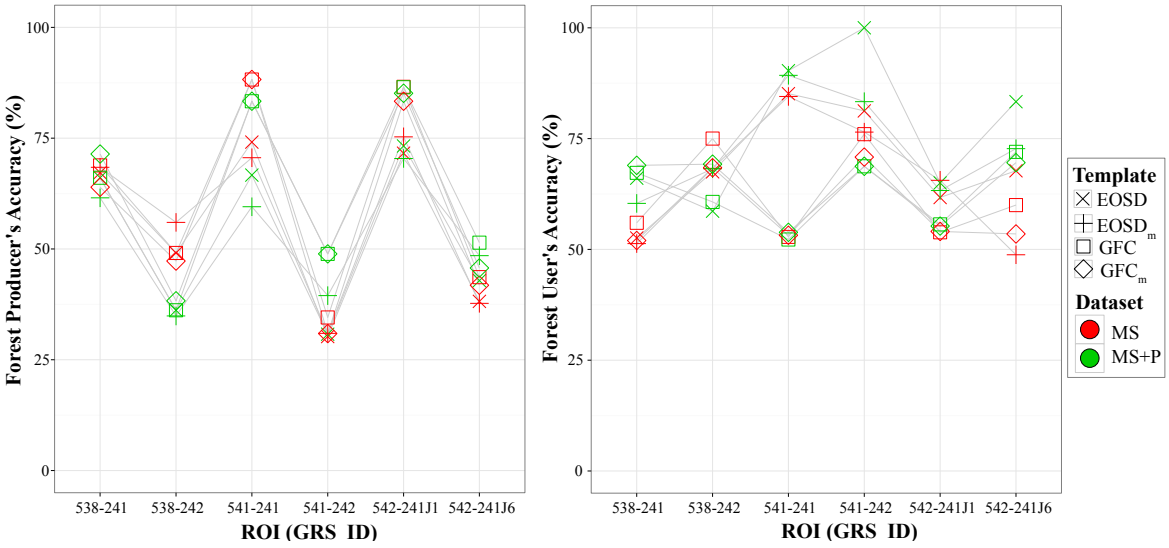

Figure 5.1: The 10-m TGC forest classification accuracies were variable and patterned. The PA_f values (left), were higher in northern ROIs than in corresponding southern ROIs. The UA_f values (right), generally had the opposite pattern, with lower UA_f in the north and higher UA_f in the south.

For user's accuracies, the pattern was less clear (Figure 5.1). However, the general pattern was for a northern ROI to have a lower UA_f than its corresponding southern ROI. For example, MS EOSD 538-241 had a user's accuracy of 53 % whereas its southern neighbour

(MS EOSD 538-242) had a user's accuracy of 68 %. In general, false positives for forest were more likely in the north.

The relatively low percentages of false negatives and high percentages of false positives in northern ROIs meant that classification in northern ROIs was more biased towards forest than the classification in the southern ROIs. This effect can be seen in the seam lines of the land-cover maps in Figure 5.2. Transitions from northern to southern ROIs were frequently marked by a sharp transition from higher forest pixel densities to lower forest pixel densities. In general, northern classifications overestimated the number of forest pixels whereas southern classifications underestimated them (Figure 5.3). Clearly, the TGC classifications were strongly affected by location and not well adapted to local differences.

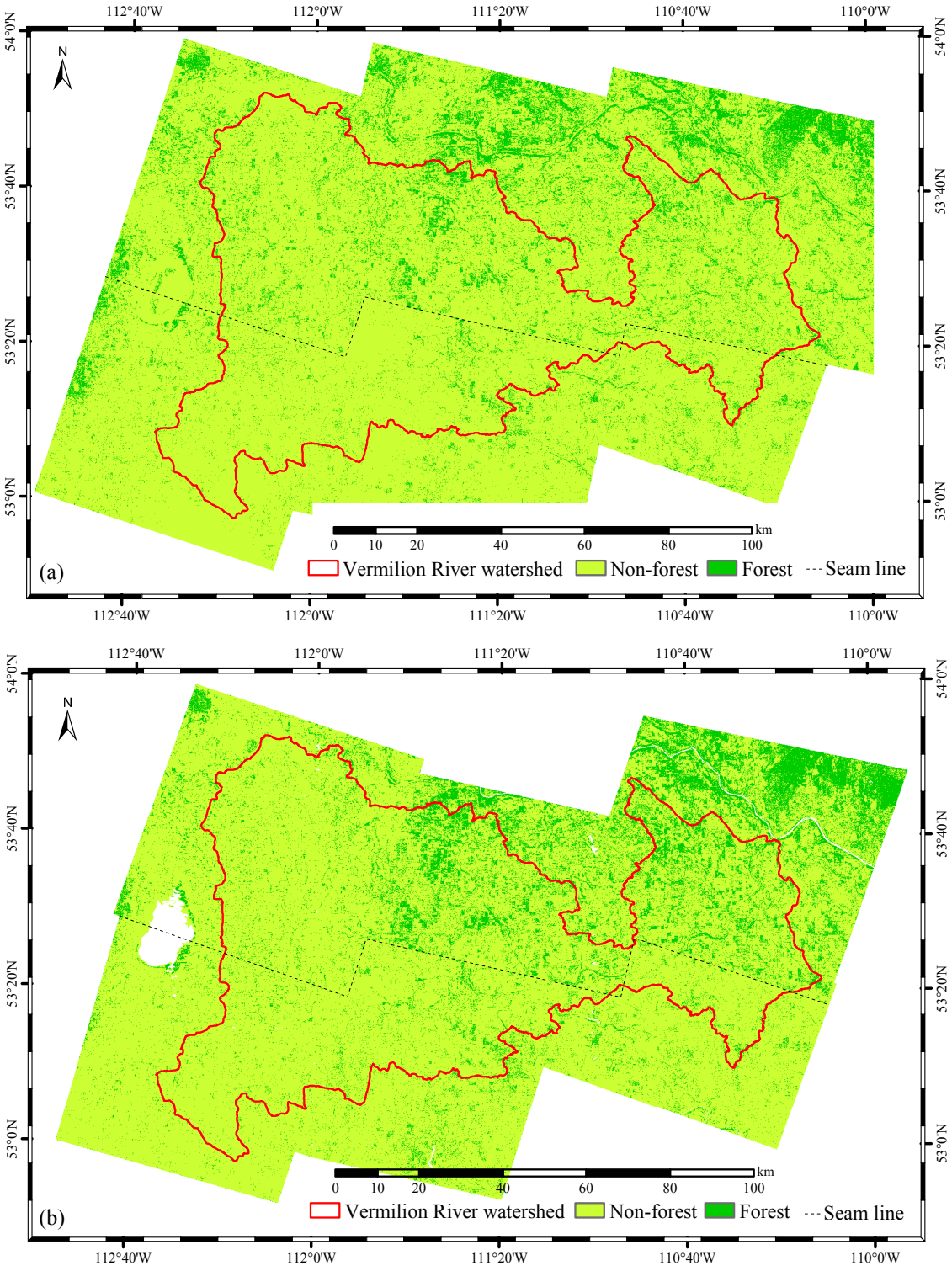

Figure 5.2: Examples of visible seam lines between 10-m TGC classifications showing the differences in biases towards the forest class. The northern images tend to be overforested and the southern images underforested. A classification's apparent relative density of forest pixels along seam lines can be predicted using the REA_f values in Figure 5.3; (a) MS EOSD with $\phi_f = 0.2$ and (b) MS+P GFC_m with $\phi_f = 0.06$.

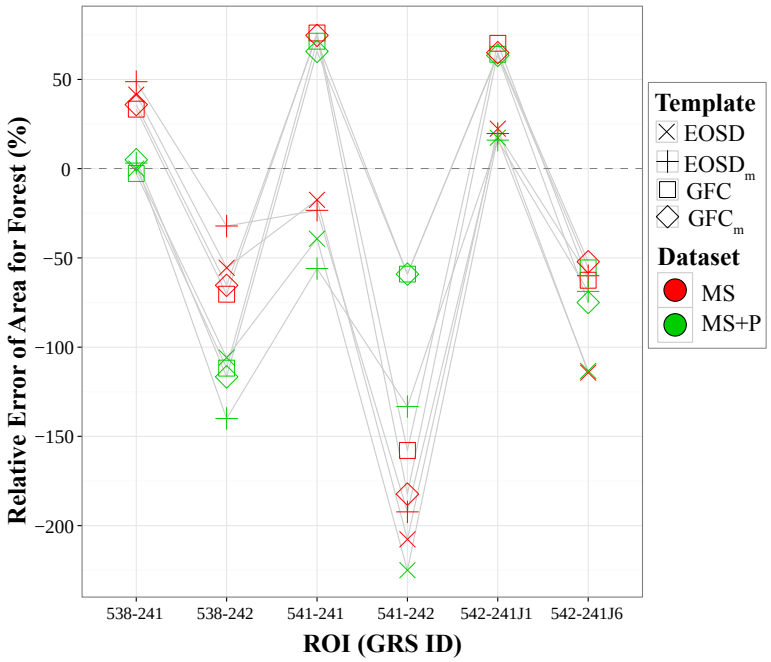

Figure 5.3: Relative error of area (REA) for forest of 10-m TGC classifications (with constant ϕ_f). In general, the number of forest pixels were overestimated in the northern ROIs ($REA_f > 0$) and underestimated in southern ROIs ($REA_f < 0$).

5.1.2 Forest Correspondence Distributions

Figures 5.4 and 5.5 show examples of the distributions of forest correspondence for the ground-reference data. Within some ROIs (i.e., the southern ROIs), a large percentage of forest pixels had forest correspondence values that overlapped those of non-forest pixels.

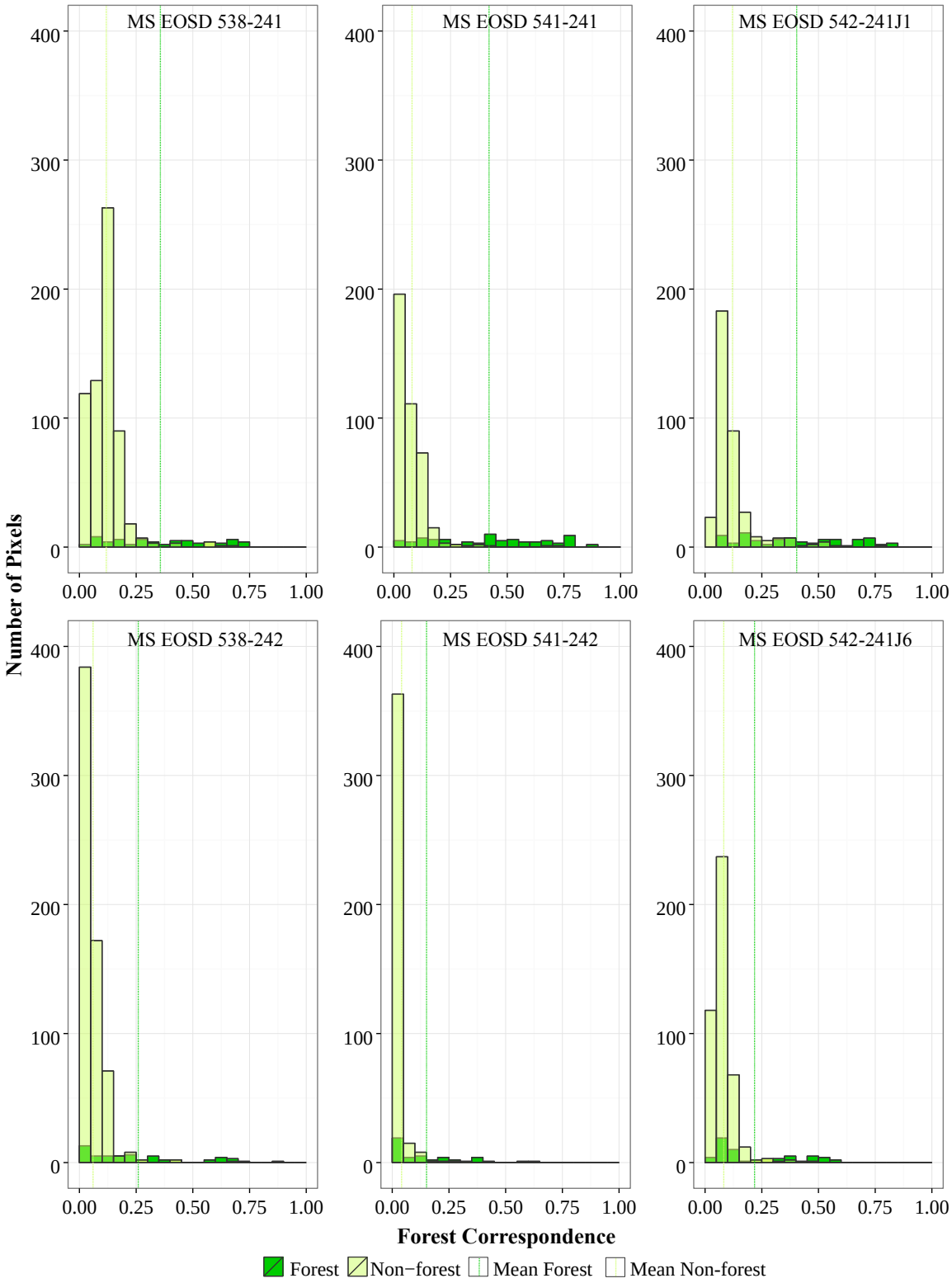

Figure 5.4: MS EOSD forest correspondence histograms of ground-reference samples. Although forest pixels had higher mean forest correspondence, there was substantial overlap of the forest and non-forest distributions.

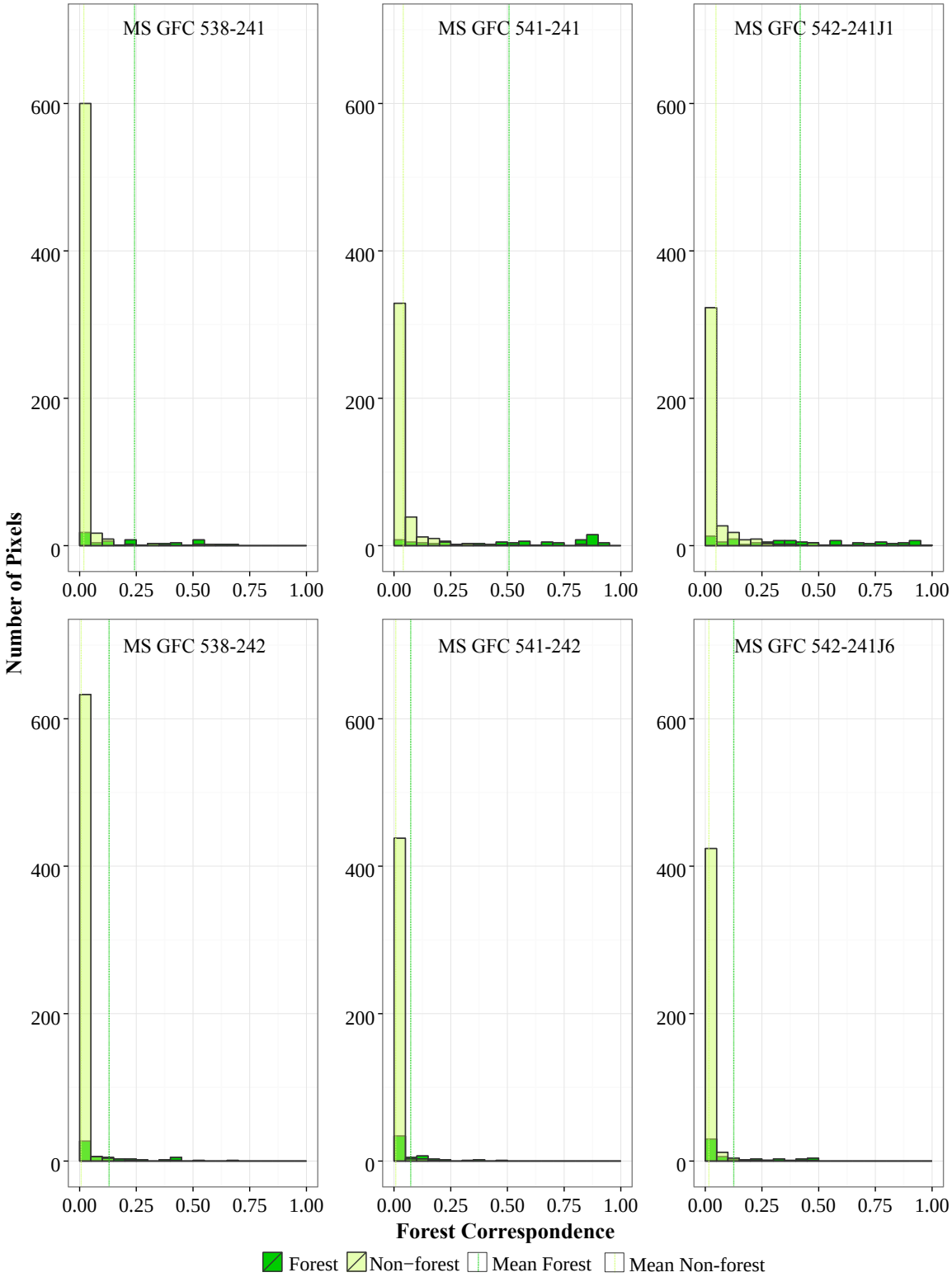

Figure 5.5: MS GFC forest correspondence histograms of ground-reference samples. Although forest pixels had higher mean forest correspondence, there was substantial overlap of the forest and non-forest distributions.

5.1.3 GFC Water Mask Problems

After the TGC results were calculated and analysed, it was discovered that a different water mask was applied to the GFC than was applied to the EOSD. The water mask applied to the GFC had very little detail, showing only a few major waterbodies and rivers (see Figure 5.2b). The origin of the water mask is unknown. As a consequence, results using masked and unmasked GFC templates were sometimes very similar.

5.2 10-m TGC Watershed Classifications with Variable ϕ_f

Figure 5.6 shows examples of TGC classification accuracies for different ϕ_f. All classifications are found in Appendix A.3.

The relationship of ϕ_f to classification accuracy varied with ROI and template. For example, in Figure 5.6, the value of ϕ_f at the point of intersection of the PA_f and UA_f curves (i.e., the $MAREA_f$ threshold) varied greatly between ROIs and templates. According to Figure 5.6, the $MAREA_f$ threshold of MS EOSD 538-241 was approximately 0.24, whereas the $MAREA_f$ threshold using EOSD MS 541-242 was approximately 0.07. When the GFC was used as a template instead of the EOSD, the $MAREA_f$ thresholds changed from 0.24 to 0.09 (538-541) and from 0.07 to 0.03 (541-242). Figure 5.7 shows the PA_f of the $MAREA_f$ classification of each parameter set's forest correspondence image and the corresponding $MAREA_f$ thresholds.

The producer's accuracies of the $MAREA_f$ classifications had the same north/south pattern of variation as the fixed parameter classifications (compare Figures 5.1 and 5.7). The pattern of $MAREA_f$ thresholds was also similar: northern ROIs had higher thresholds than corresponding southern ROIs.

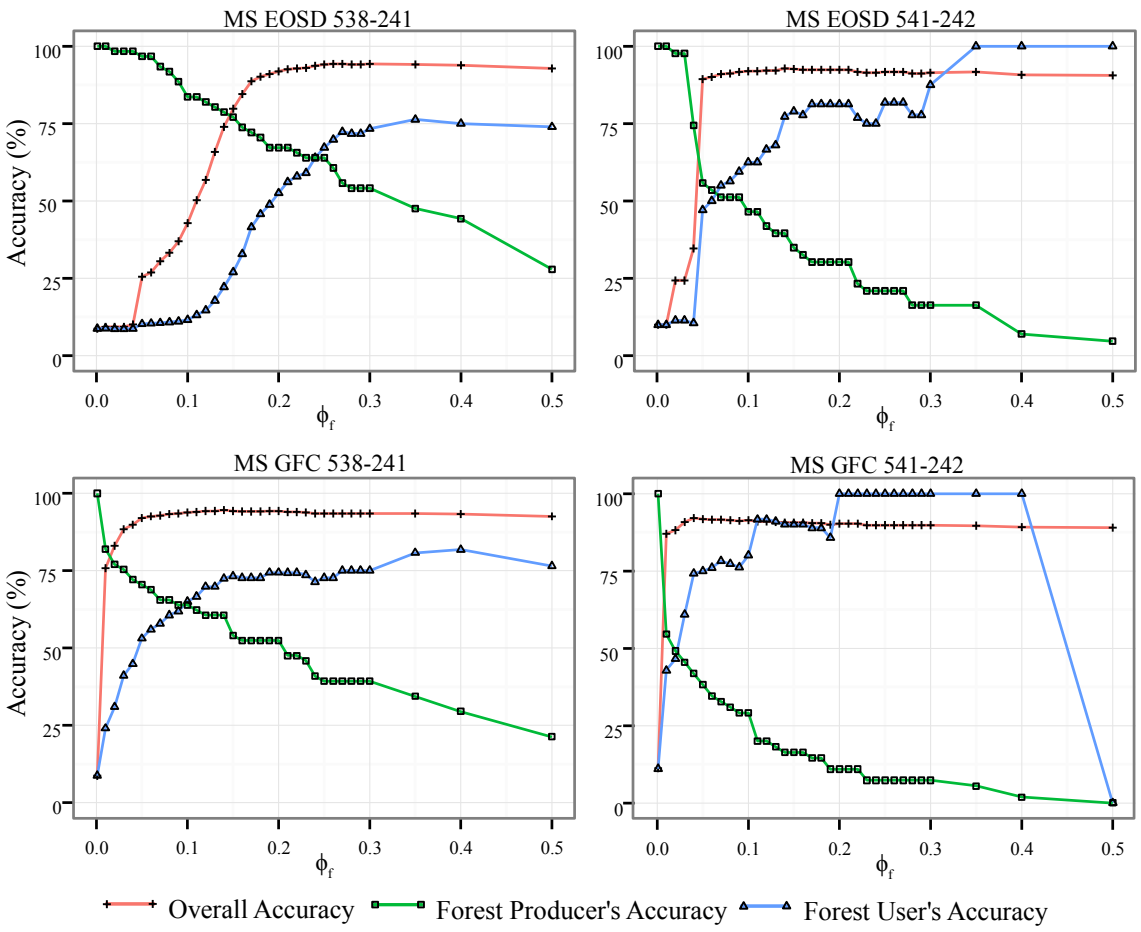

Figure 5.6: TGC classification accuracies changed with ϕ_f. As ϕ_f increased, PA_f decreased and UA_f tended to increase. However, the effect of ϕ_f on accuracies varied with ROI and template.

5.2.1 TGC's Best Classifications

The results in Table 5.2 are the $MAREA_f$ classifications with the highest PA_f of each ROI. Ranging from 58 % to 80 %, they are the best TGC classifications of all combinations of dataset, template, and ϕ_f values.

5.3 1-m TGC Orthophoto Classifications

Table 5.3 shows TGC accuracies of the orthophotos for $\phi_f = 0.2$. Figure 5.8 shows how the TGC classification of T52-10-W4 is a great improvement over the original EOSD template.

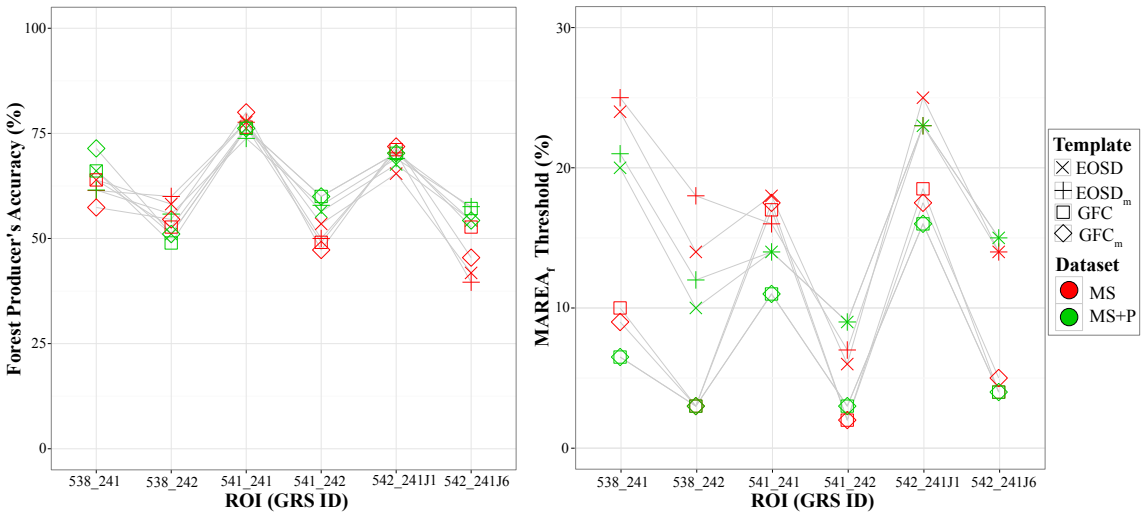

Figure 5.7: TGC's $MAREA_f$ classification accuracies and thresholds had a spatial pattern; (left) the PA_f values were higher in the northern ROIs than in the southern ROIs; (right) the $MAREA_f$ thresholds were also higher in the northern ROIs than in the southern ROIs.

Table 5.2: The best TGC classifications for each ROI (produced by tuning ϕ_f to each ROI).

Property	538-241	538-242	541-241	541-242	542-241J1	542-241J6	All ROIs
PA_f (%)	71	60	80	60	72	58	67
UA_f (%)	69	60	74	59	73	58	66
OA (%)	95	94	92	90	89	91	92
REA_f (%)	5	0	10	4	-3	0	3
Template	GFC_m	$EOSD_m$	GFC_m	GFC/GFC_m	GFC_m	$EOSD_m$	-
Dataset	MS+P	MS	MS+P	MS+P	MS	MS+P	-
ϕ_f (%)	6.5	18	17	3	17.5	15	13

PA_f = forest Producer's Accuracy, UA_f = forest User's Accuracy, OA = Overall Accuracy, REA_f = forest Relative Error of Area, $EOSD_m$ = water-masked EOSD template, GFC = GFC template, GFC_m = water-masked GFC template, MS = 10-m multispectral SPOT data, MS+P = 10-m multispectral and panchromatic SPOT data, ϕ_f = forest correspondence threshold.

Table 5.3: Constant ϕ_f TGC classification accuracies for 1-m orthophotos. The forest classification accuracies were higher than the 10-m accuracies.

Classification	Metric (%)	T52-R10-W4	T52-R12-W4
RGB+P EOSD, $\phi_f = 0.2$	PA_f	86	73
	UA_f	80	90
	OA	91	90
	REA_f	3	-26

PA_f = forest Producer's Accuracy, UA_f = forest User's Accuracy, OA = Overall Accuracy, REA_f = forest Relative Error of Area, RGB+P = colour and panchromatic datasets, EOSD = EOSD template, ϕ_f = forest correspondence threshold.

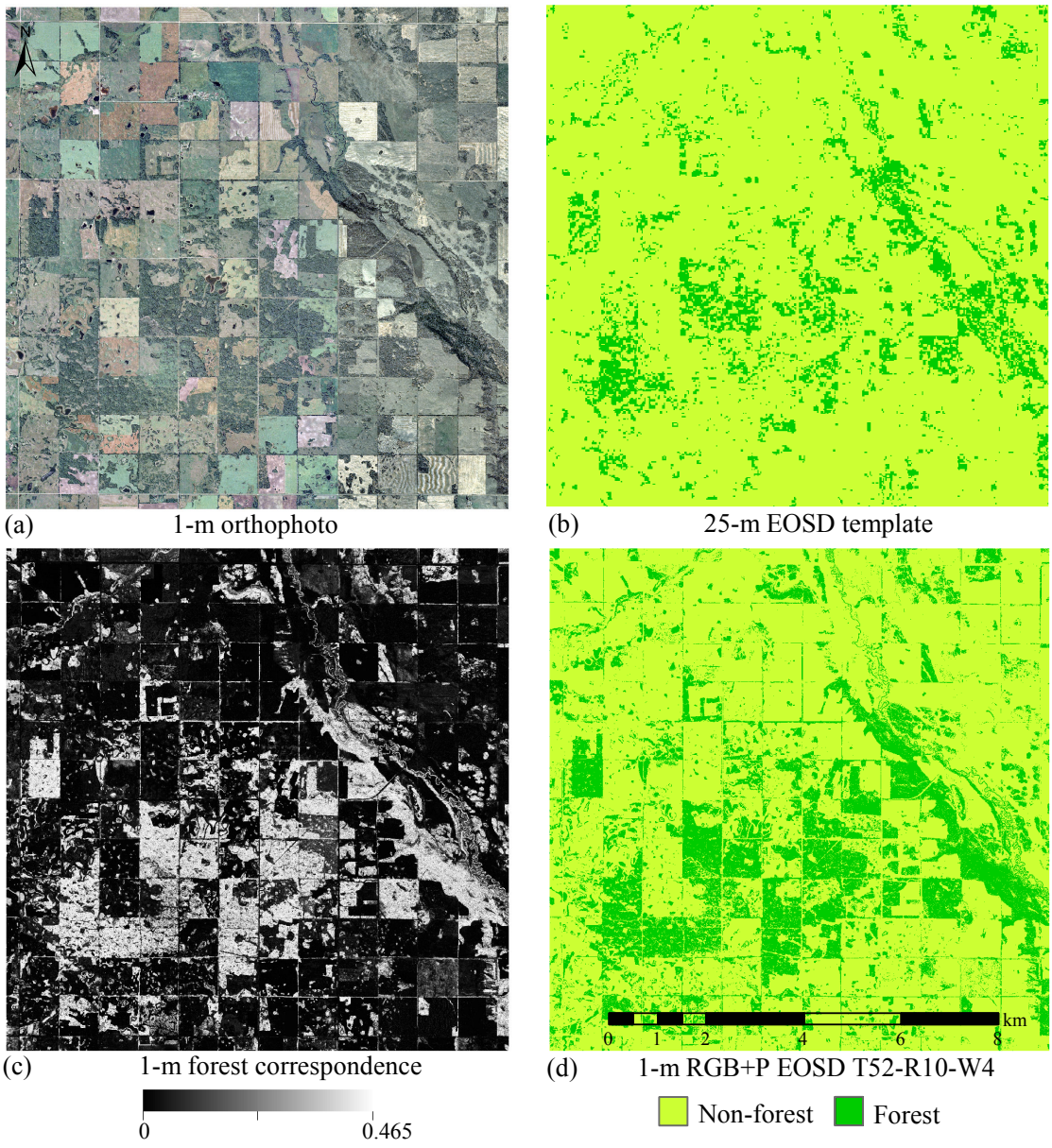

Figure 5.8: Given inputs (a), a 1-m colour orthophoto, a panchromatic band, and (b), the 25-m EOSD template of the same area; TGC calculated (c) the forest correspondence image, and output (d) an accurate 1-m forest/non-forest classification. Note that for this case the TGC classification was more accurate than the original template map and had a much higher spatial resolution.

5.4 SVM Classifications

The results of the SVM classifications are shown in Table 5.4.

Table 5.4: The SVM classification accuracies for each MS ROI. Although the SVM results were more accurate than the TGC results, production of the SVM classifiers was much more labour intensive. Each SVM classification required up to 2 hours of iterating between selection of training data and qualitative accuracy assessment.

Metric (%)	538-241	538-242	541-241	541-242	542-241J1	542-241J6	All ROIs
PA_f	49	67	93	80	91	33	69
UA_f	79	47	95	87	89	67	77
OA	94	91	97	93	94	91	93
REA_f	-77	64	-2	-10	2	-154	-30

PA_f = forest Producer's Accuracy, UA_f = forest User's Accuracy, OA = Overall Accuracy,
REA_f = forest Relative Error of Area.

5.4.1 Comparison with TGC

Table 5.5 shows the classification accuracy differences between constant ϕ_f TGC and the SVM classifications. The SVM PA_f was 14 % higher and the UA_f was 8 % higher on average than the corresponding TGC accuracies.

Table 5.5: Accuracy differences between the constant ϕ_f TGC and SVM classifications. The SVM classifications averaged 14 % higher PA_f and 8 % higher UA_f.

Classification	Subtrahend	Metric	538-241	538-242	541-241	541-242	542-241J1	542-241J6	All ROIs
MS EOSD,	SVM	ΔPA_f	18	-18	-19	-49	-20	5	-14
$\phi_f = 0.2$		ΔUA_f	-26	20	-10	-6	-27	1	-8
		ΔOA	-3	3	-4	-1	-8	0	-2

ΔPA_f = difference in forest Producer's Accuracy, ΔUA_f = difference in forest User's Accuracy, ΔOA = difference in Overall Accuracy,
MS = multispectral 10-m SPOT dataset, EOSD = EOSD template, SVM = Support Vector Machine Classification.

5.5 Template Map Accuracies

The 10-m classification accuracies of the EOSD and GFC template maps are summarized in Table 5.6. Complete results are shown in Figure 5.9 and tabulated in Appendix B.

Template classification accuracies were low. For the EOSD-based templates, the average PA_f and UA_f were below 45 %. The GFC-based templates had very low average

Table 5.6: The average 10-m classification accuracies of EOSD and GFC-based templates.

Template Maps	Metric (%)	538-241	538-242	541-241	541-242	542-241J1	542-241J6	All ROIs
EOSD and $EOSD_m$ templates	$\overline{PA}_f$	41	32	38	33	47	35	38
	$\overline{UA}_f$	27	36	57	56	49	41	44
	$\overline{OA}$	86	90	87	90	80	88	87
	$\overline{REA}_f$	122	-43	-95	-119	-5	-45	-31
GFC and GFC_m templates	$\overline{PA}_f$	22	8	53	14	58	15	28
	$\overline{UA}_f$	75	48	84	78	89	69	74
	$\overline{OA}$	93	92	92	90	90	90	91
	$\overline{REA}_f$	-315	-1387	-73	-590	-61	-539	-494

$\overline{PA}_f$ = average forest Producer's Accuracy, $\overline{UA}_f$ = average forest User's Accuracy, $\overline{OA}$ = average Overall Accuracy, $\overline{REA}_f$ = average forest Relative Error of Area, EOSD = EOSD template, $EOSD_m$ = water-masked EOSD template, GFC = GFC template, GFC_m = water-masked GFC template.

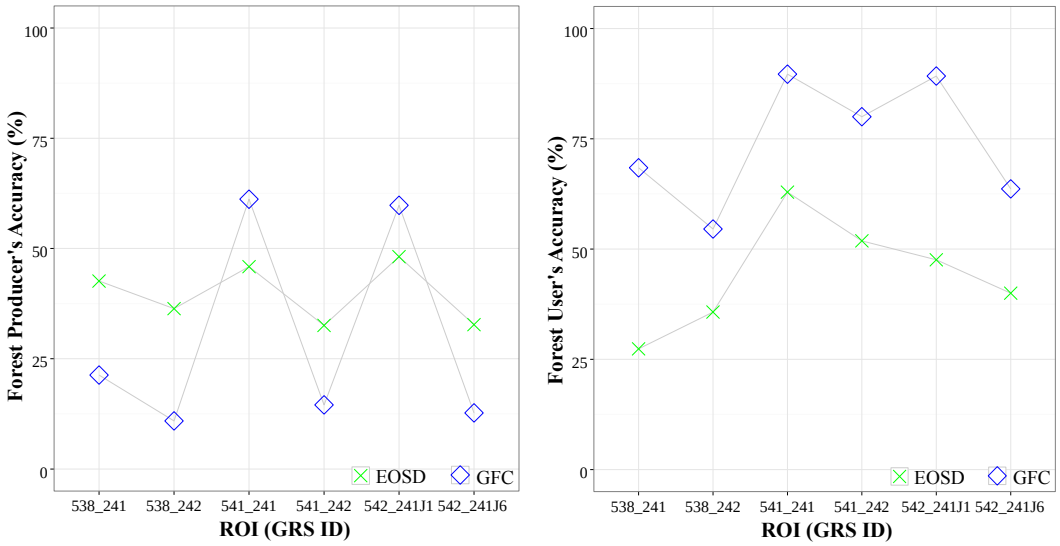

Figure 5.9: 10-m forest classification accuracies of the EOSD and GFC template maps for each MS ROI.

PA_f (28 %) and relatively high average UA_f (74 %). All ROIs (except for 541-241 and 542-241J1 in some GFC-based templates) had PA_f or UA_f below 50 %.

The accuracies showed a clear pattern of spatial dependence where northern ROIs had higher accuracies than corresponding southern ROIs. For example, 541-241 in GFC-based templates had an average PA_f of 53 %, whereas its southern neighbour (541-242) had an average PA_f of only 14 %.

There were also differences between templates. Compared to the EOSD, the GFC had

fewer false positives (instances of non-forest classified as forest) and more false negatives (forest classified as non-forest). In four out of six GFC-based ROIs[1], extremely low average PA_f (8 to 22 %) and relatively high UA_f (48 - 89 %) resulted in the number of forest pixels being severely underestimated. In contrast, the EOSD tended to have more false positives and more true positives. For example, wetlands and dark fields were frequently labelled as forest using the EOSD.

The EOSD template classification accuracies within the orthophoto ROIs are shown in Table 5.7.

Table 5.7: 1-m EOSD template accuracies within orthophoto ROIs. The UA_f values were higher than within the 10-m SPOT ROIs (Table 5.6).

Template	**Metric (%)**	**T52-R10-W4**	**T52-R12-W4**
EOSD	PA_f	33	25
	UA_f	69	86
	OA	80	79
	REA_f	-161	-279

PA_f = forest Producer's Accuracy, UA_f = forest User's Accuracy, OA = Overall Accuracy, REA_f = forest Relative Error of Area, EOSD = EOSD template.

5.5.1 Comparison with TGC

Table 5.8 shows the average classification accuracy differences between the constant ϕ_f TGC and template classifications.

5.5.2 %$LAND_f$

The %$LAND_f$ values of the EOSD and GFC templates for the MS ROIs are shown in Table 5.9 and Figure 5.10. The %$LAND_f$ values of the EOSD were surprisingly close to the percentages of forest estimated from the ground-reference samples. In contrast, the GFC underestimated forest area by up to 400 %. The EOSD's %$LAND_f$ values within the orthophotos are shown in Figure 5.10.

[1] 538-241, 538-242, 541-241, and 542-241J6.

Table 5.8: Average accuracy differences between the TGC classification and the template classification for each ROI.

Classifications and Template	Metric (%)	538-241	538-242	541-241	541-242	542-241J1	542-241J6	All ROIs
EOSD-based	ΔPA_f	25	12	29	-1	25	7	16
	ΔUA_f	30	29	30	30	15	27	27
	ΔOA	7	4	7	2	7	4	5
	ΔREA_f	98	40	-61	71	-24	44	28
GFC-based	ΔPA_f	45	35	33	27	28	31	33
	ΔUA_f	-14	20	-31	-6	-34	-6	-12
	ΔOA	1	2	-5	2	-7	1	-1
	ΔREA_f	-333	-1296	-145	-475	-127	-478	-476

ΔPA_f = difference in forest Producer's Accuracy, ΔUA_f = difference in forest User's Accuracy, ΔOA = difference in Overall Accuracy, ΔREA_f = difference in forest Relative Error of Area, EOSD-based = EOSD and $EOSD_m$ templates, GFC-based = GFC and GFC_m templates.

Table 5.9: $\%LAND_f$ for the EOSD, GFC, and ground reference within each MS ROI.

$\%LAND_f$ Source	538-241	538-242	541-241	541-242	542-241J1	542-241J6	Average
EOSD	15	8	14	6	17	9	12
GFC	3	2	13	2	13	3	3
10-m reference	9	8	17	10	18	11	10

EOSD = EOSD template, GFC = GFC template, $\%LAND_f$ = percentage of forest area within classification.

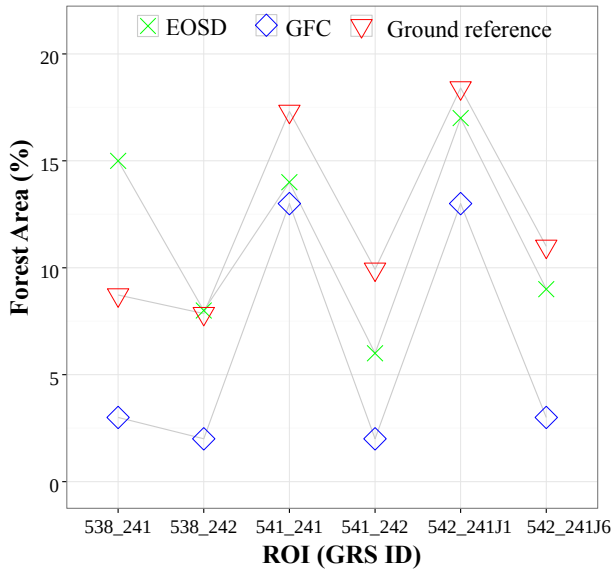

Figure 5.10: $\%LAND_f$ of the EOSD template, GFC template, and ground reference within each MS ROI.

Table 5.10: %$LAND_f$ for the EOSD template within each orthophoto ROI.

%$LAND_f$ Source	**T52-R10-W4**	**T52-R12-W4**
EOSD	12	9
1-m reference	25	27

%$LAND_f$ = percentage of forest area within classification,
EOSD = EOSD template.

5.6 TGC Class Confusion

By visually examining results, it was found that the primary source of class confusion occurred between spectrally similar land-cover. Shallow water, wetlands, and edges of wetlands were frequently confused with forest. Dark cropland was also confused, and was the dominant source of source of error in areas with few wetlands.

Regrettably, water masking did not eliminate the problem of confusion with wetlands, because the wetlands were inadequately masked. For example, the mask removed only a small fraction of the hundreds of wetlands within the area shown in Figure 5.11. Moreover, the masked areas were frequently inaccurate in shape, poorly rasterized, and poorly registered to the target image. Consequently, substantial areas of wetlands were confused with forest (Figure 5.11c).

TGC was frequently confused by pixels on the edges of wetlands and in areas near wetlands. Sometimes, spurious rings of forest pixels surrounded wetlands. In Figure 5.11b, fields containing wetlands sometimes appear as textured variations of dull and bright red that are spectrally similar to forest. Clearly, the wetlands had confusing spectral features.

5.7 TGC Mixed Clusters

Spectrally similar pixels of different classes were frequently found in mixed clusters. For example, some wetlands in Figure 5.12a were falsely labelled as forest by TGC in Figure 5.12b. Figure 5.12c indicates that the clusters of these wetland areas also contain forest pixels. Despite the eighteen clusters used to describe the two small wetland areas within

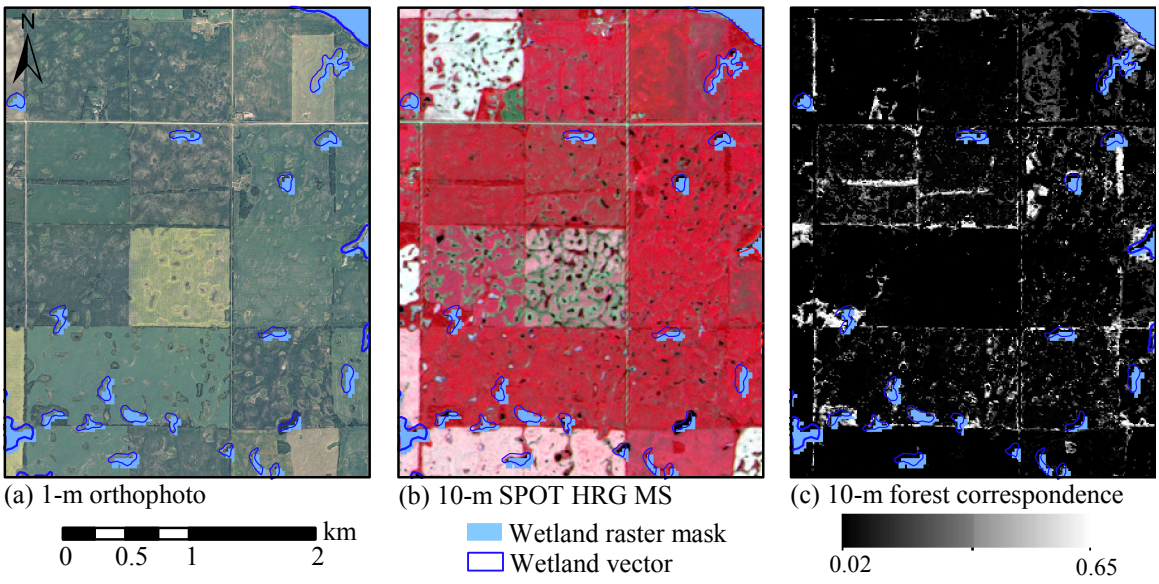

Figure 5.11: The water mask did not eliminate confusion with wetlands. Note the hundreds of unmasked wetlands displayed as (a) small green or soil-coloured patches, (b) black patches or patches that are spectrally similar to forest, (c) bright patches of (spuriously) high forest correspondence (MS $EOSD_m$ 541-242).

this detail, it was not sufficient to separate wetland from forest.

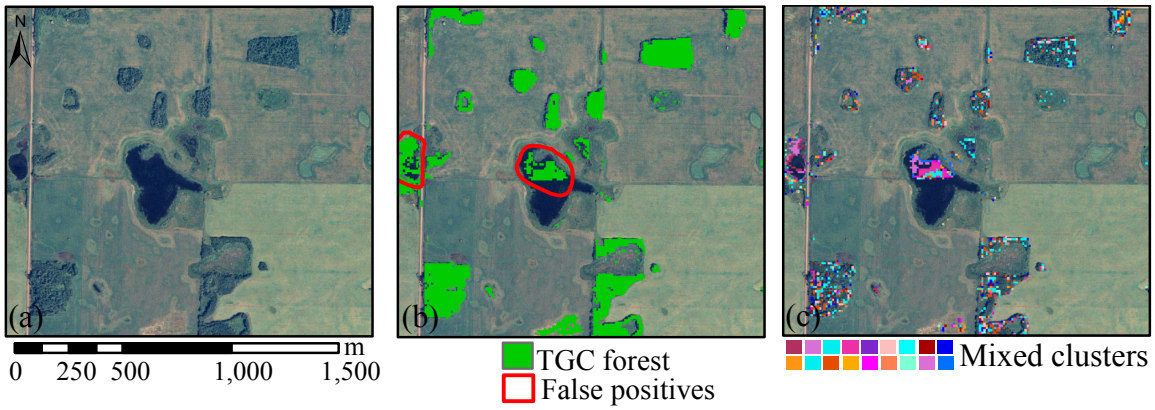

Figure 5.12: Wetland pixels were frequently found within the same clusters as forest pixels; (a) 1-m orthophoto detail showing wetlands and forest, (b) TGC classification showing wetland areas labelled as forest and (c) mixed clusters that correspond to both wetland and forest.

5.7.1 Cluster Purity of the Ground-reference Site

Within the LiDAR ground-reference site, the average cluster purity of each pixel for the MS EOSD 541-241 clustering was 88 % (Figure 5.13). This average cluster purity represents the maximum possible OA for a classification using the clustering. Remarkably, the OA of the MS EOSD 541-241 ($\phi_F = 0.2$) classification within the site was 86 % (Table 5.11). Therefore, the overwhelming majority of error was caused by mixed clusters, not by label error in the template. Low cluster purities were found in areas where forest and non-forest were spectrally similar, namely, in dark photosynthetically-active fields, coniferous tree patches, wetlands, and especially in the Vermilion River's flood plain (Figure 5.13d). However, these cluster purities may not be typical, because the site was 38 % forested (which was much greater than the 8 to 18 % found in the SPOT ROIs).

Table 5.11: The error matrix of the TGC MS EOSD 541-241 ($\phi_f = 0.2$) classification within the LiDAR ground-reference site. The 86 % OA was very close to the limit imposed by the 88 % cluster purity of the site (PA_f = forest producer's accuracy, UA_f = forest user's accuracy, $\%LAND_f$ = percentage of forest area, OA = overall accuracy).

		Ground Reference (Pixels)			Ground Reference (%)		Metrics (%)	
		Non-forest	Forest	Total	Non-forest	Forest		
TGC	Non-forest	243447	50150	293597	96.83	32.27	PA_f	67.73
	Forest	7980	105246	113226	3.17	67.73	UA_f	92.95
	Total	251427	155396	406823	100.00	100.00	$\%LAND_f$	27.83
							OA	85.71

5.8 Template Label Error

The template images were also confused by spectrally similar land cover. For example, wetlands and croplands (Figure 5.14a) were mislabelled by the EOSD template (Figure 5.14b). However, for this particular example, TGC produced a surprisingly good classification despite the template's low classification accuracy (compare Figures 5.14b and 5.14d). Within the forest correspondence image, the wetland areas appear grey or white, indicating their high forest correspondence.

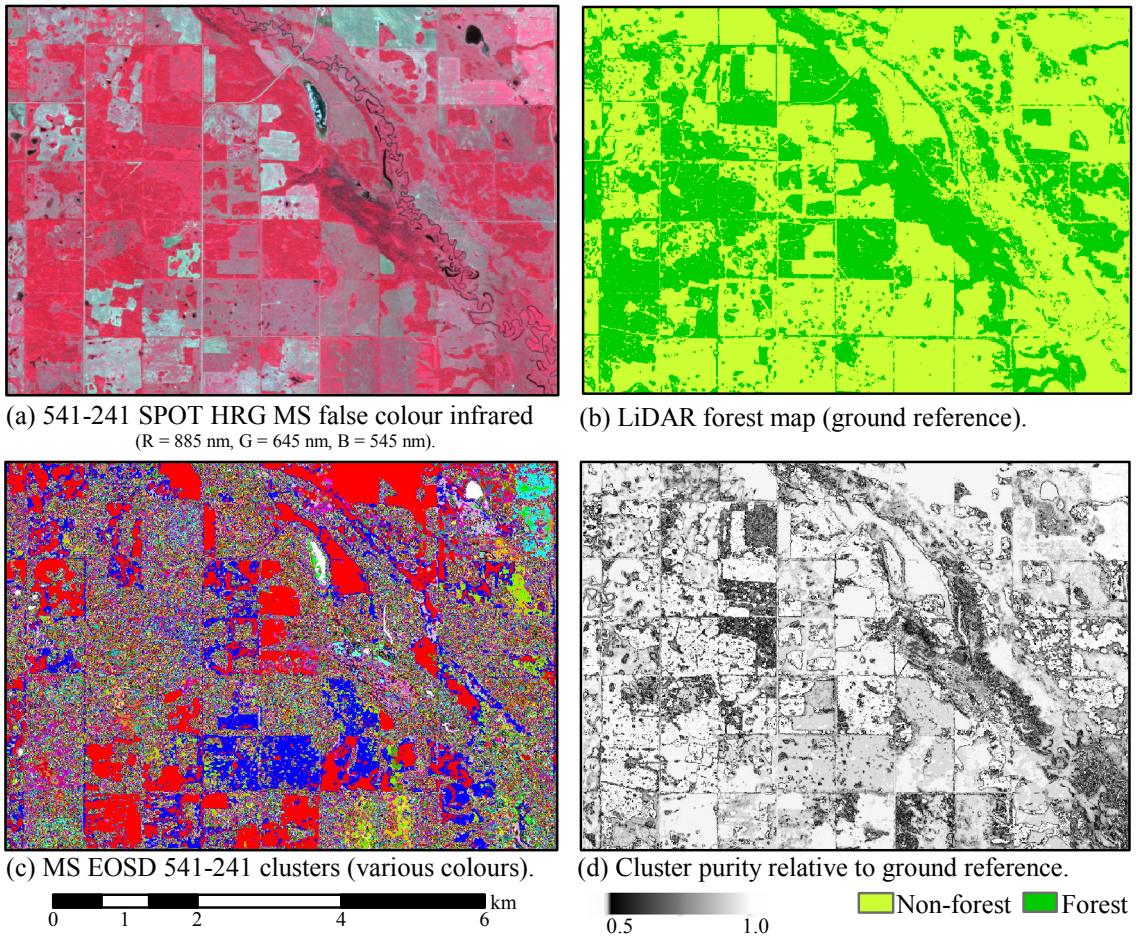

Figure 5.13: Cluster purities of the MS EOSD 541-241 template-guided clustering calculated for a small area using a LiDAR-derived forest map as a ground reference. The 88-% average cluster purity is the maximum possible OA for this clustering.

5.9 Validation Errors

All of the quantitative analysis of classification accuracies relied on accurate ground-reference data. However, this data had several issues that were difficult to quantify. For example, photo interpretation was sometimes ambiguous, particularly for cases where the pixel was near the edge of a tree patch (Figure 5.15) or within patches with low-tree cover. Sometimes it was very difficult to distinguish between shaded ground and tree. In practice, the context of the pixel heavily influenced its interpretation. For example, an indecipherable mix of shadow and vegetation within a forest patch was always interpreted as a forest

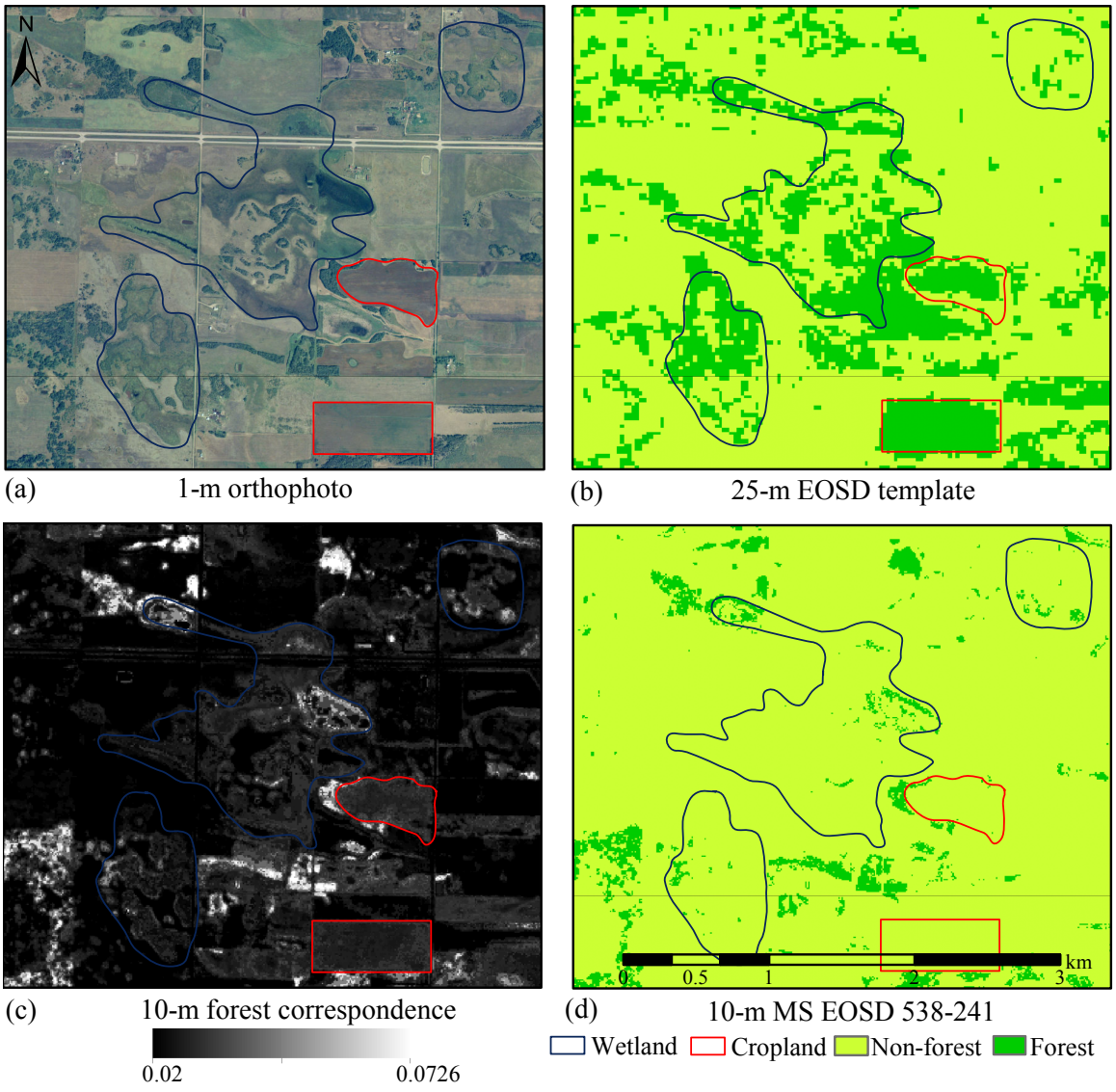

Figure 5.14: (a) Examples of wetland and cropland with spectra similar to forest (i.e., dark), (b) the EOSD confused forest with wetland and cropland, (c) high forest correspondence in wetland and cropland, and (d) despite the confusion of the EOSD, the TGC classification was relatively accurate.

pixel. Although the high-resolution images from Google Earth were helpful in some cases, very often there was no suitable high-resolution image available, or registration error in the Google image prevented the ambiguity from being resolved. Most non-forest pixels (excepting wetlands) were interpreted without difficulty.

Registration errors also caused validation problems. For example, Figure 5.15 shows

how a 10-m pixel that was photo-interpreted as forest in a 1-m orthophoto actually contained non-forest spectra in the 10-m SPOT HRG MS image. Registration errors between the orthophotos and the SPOT images were potentially very serious because of the number of small or narrow forest patches. Even a small misregistration (like that in Figure 5.15) could have caused large changes in classification accuracy.

Misregistration between the SPOT images and orthophotos varied from 0 to 30-m, sometimes within the same SPOT image. Although no registration residuals were calculated for the SPOT images, their automated georegistration was likely to be far less accurate than the GPS-located pixels within the orthophotos. Manual comparison of the SPOT images with orthophotos revealed that 541-241J6 had especially poor registration, with variable errors up to 30-m.

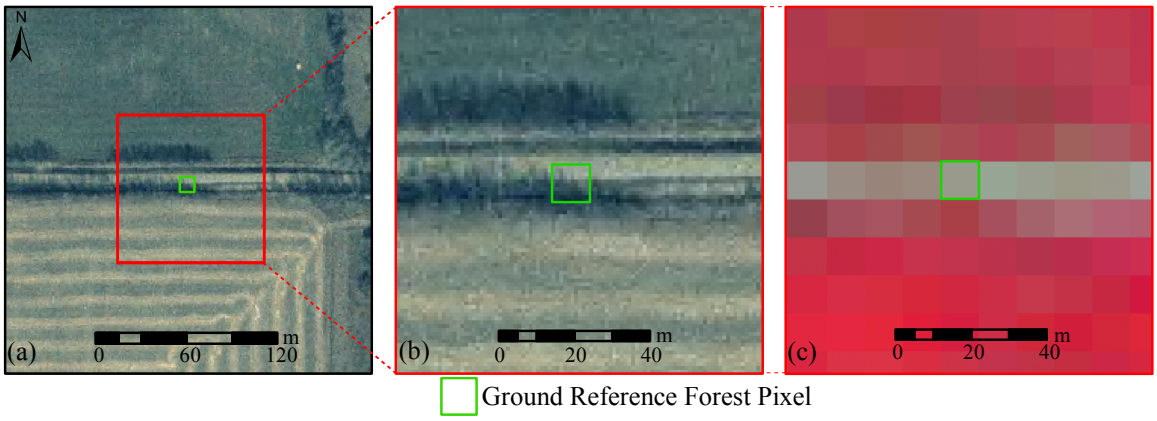

Figure 5.15: Validation accuracy was reduced by registration and photo-interpretation errors; (a) context of a 10-m ground-reference pixel, (b) within the 1-m validation orthophoto, the pixel appeared to contain mostly trees; less than half of the pixel overlapped the road, and (c) within the corresponding 10-m SPOT HRG MS image, the pixel was entirely road.

Chapter 6

Discussion

This chapter investigates TGC's usefulness by analysing the accuracy and automation of the Vermilion River watershed classification experiments. TGC's problematic aspects are analysed and mitigation strategies are suggested.

6.1 Evaluation of TGC Classification Accuracies

6.1.1 10-m Watershed Classifications with Constant ϕ_f

The most important classification experiments were the constant ϕ_f watershed classifications. These classifications were automated; the ROIs were classified without tuning parameters to individual ROIs.

In terms of accuracy, the constant ϕ_f watershed classifications were a success. The OA values averaged 92 %. However, this accuracy is not very impressive because 90 % OA could have been obtained by classifying all pixels as non-forest. Instead, the most important result was that TGC forest classification accuracies were generally higher than the accuracies of the original template classifications (Figure 6.1). With EOSD-based templates, TGC increased PA_f by 16 %, UA_f by 27 %, and OA by 5 %. With GFC-based templates, TGC improved the PA_f values by an average of 33 %. However, UA_f values declined by an average of 12 %. Overall, the experiments demonstrated that TGC improved template map classification accuracies.

In general, TGC increased low forest accuracies (e.g., PA_f values for GFC-based templates) and decreased high forest accuracies (e.g., UA_f values for GFC-based templates). This suggests that TGC may not be very useful when template accuracy is already high.

TGC classification accuracies were not as high as the SVM classification accuracies

(Figure 6.1). The SVM PA_f was approximately 14 % higher and the UA_f was approximately 8 % higher on average than the corresponding accuracies of the constant ϕ_f MS EOSD TGC classifications (Table 5.5). However, the SVM classifications were far more labour intensive.

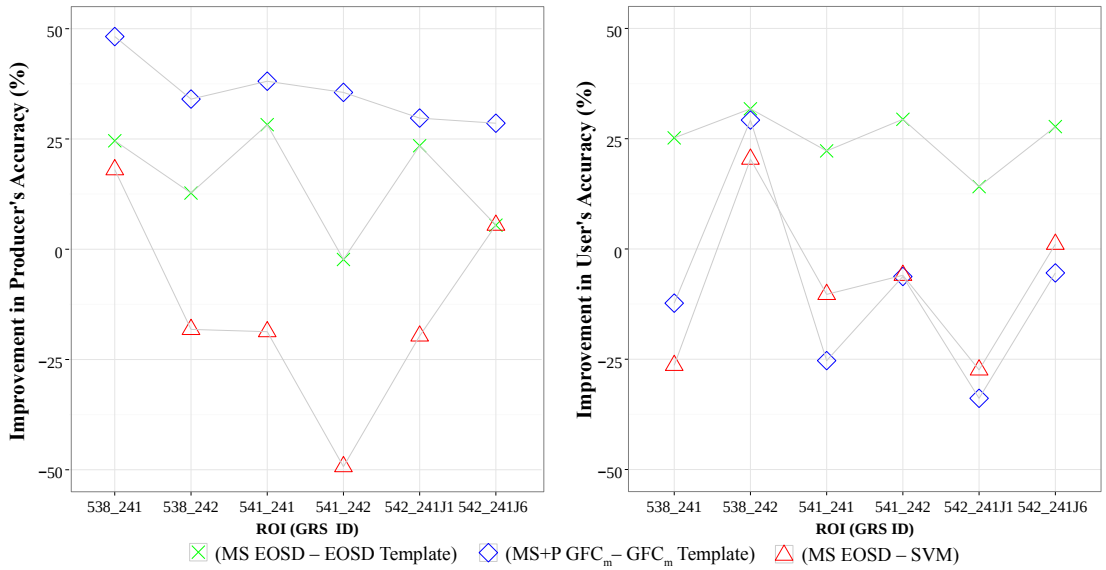

Figure 6.1: In general, TGC classifications had higher accuracies and better balance between forest accuracies than the template map input. With EOSD input, TGC produced large increases in PA_f and UA_f. Although the GFC_m UA_f values were higher than TGC results, TGC maps were an improvement because they had much higher PA_f and more balanced forest accuracies. The SVM averaged 14 % higher than TGC for PA_f and 8 % higher for UA_f. PA_f (left), UA_f (right).

6.1.2 1-m Orthophoto Classifications

TGC's 1-m orthophoto classifications were a great improvement over the original EOSD classifications, having 48 % to 55 % improvement in PA_f, and 4 % to 11 % improvement in UA_f. TGC's success with the 1-m classifications is very promising for future work (although only two experiments were performed).

6.1.3 $MAREA_f$ Threshold Variability

One of the main findings from examining the variable threshold classifications was that the $MAREA_f$ threshold varied greatly between ROIs and between templates. This finding explains why the 10-m constant ϕ_f classifications were biased to have too many forest pixels in the northern ROIs and too few in the southern ROIs; the chosen ϕ_f value was too low for northern ROIs and too high for southern ROIs. The variability of the $MAREA_f$ threshold is extremely problematic because it means a choice between:

- Variable and unbalanced TGC accuracies (when a single ϕ_f is used), or
- Labour intensive classification (when ϕ_f is manually optimized for each ROI and template).

6.1.4 Effects of Datasets, Templates, and Water Masks

The effects of varying data and template parameters on the accuracy of the $MAREA_f$ classifications were difficult to discern. There were PA_f differences of up to 15 % between parameter sets, but no set was clearly superior to the others for all ROIs. For some cases, a single parameter change caused the best classification of an ROI to become the worst classification. For example, the GFC_m template and MS+P dataset produced the highest accuracies for 538-241 and 541-242. However, switching the dataset parameter to MS resulted in the lowest accuracies.

The water-masked templates produced the highest accuracies in all six ROIs. However, water-masked templates also produced the lowest accuracies in four of six ROIs. Water masking was expected to improve the classification accuracies because it reduces confusion between forest and wetlands. However, the improvement was small and inconsistent. There seemed to be little to distinguish between unmasked GFC and unmasked EOSD. Similarly, there was no clear advantage to using MS+P instead of MS data. Presumably,

the dominant errors in TGC were caused by mixed clusters that were not improved by adding a panchromatic band to the target data.

6.1.5 Explaining Patterns in Classification Accuracy

TGC forest classification accuracies were strongly affected by whether an ROI was in the north or south of the watershed. For example, the $MAREA_f$ forest classification accuracies were higher in northern ROIs than in southern ROIs. This accuracy pattern was also found in the template classifications (Figure 5.9) and to a lesser extent in the SVM classifications (Table 5.4). These three different classification methods had similar patterns, showing that there were persistent differences between the ROIs in the north and south that strongly affected classification accuracies.

One of the most obvious differences between north and south was the difference in forest patch sizes. Within the north, there were large forest patches (particularly along the Vermilion and North Saskatchewan Rivers), while in the south the patches were much smaller remnants within agricultural settings. The smaller patches had higher proportions of mixed pixels because of the higher edge-to-area ratios. The higher proportion of mixed pixels in the south made classification more difficult by blending spectra from forest and non-forest classes. This resulted in the lower PA_f and UA_f values seen in the southern experimental results. The effect of patch size on classification accuracy has been described by a variety of studies (Lechner et al., 2009; Smith et al., 2002). For TGC, the relative sizes of the tree patches and image pixels are important.

In addition to patch size, the smaller forest area in the south also caused statistical effects that reduced classification accuracies. With less forest area in the south, each false positive reduced UA_f more in the south than in the north. Even if rates of non-forest classification error had been the same in the north and south, the smaller numbers of forest pixels in the south would have resulted in lower classification accuracies. This helps explain

why the pattern of $\%\text{LAND}_\text{f}$ (Figure 5.10) was similar to the patterns of the MAREA_f and template classification accuracies for each ROI (Figures 5.7 and 5.9). Rare classes are known to have problematic classification accuracies (e.g., Witten and Frank (2011, p. 163)).

In contrast to the southern ROI classifications, the 1-m orthophoto classifications had the following advantages that produced higher forest classification accuracies:

- large forest patches, resulting in a smaller proportion of mixed pixels,
- large percentages of forest area, resulting in reduced effects of false positives and false negatives on forest accuracies, and
- high spatial resolution, resulting in a smaller proportion of mixed pixels.

Although these results are very promising, more classifications need be made to determine TGC's usefulness with very high resolution data.

By inspection of the orthophotos, the southern ROIs were found to have larger numbers of small wetlands. These wetlands had highly variable features that were sometimes similar to those of forest patches, which led to higher rates of class confusion in southern ROIs. In contrast, the northern ROIs contained sections of the Vermilion and North Saskatchewan Rivers that provided external drainage.

Overall, the patterns in classifications accuracy suggested that TGC performs better in ROIs with higher percentages of forest area, fewer wetlands, and with lower percentages of mixed pixels (i.e., larger patches or higher spatial resolution).

6.2 TGC Accuracy and Automation

The experiments showed that TGC classifications are more accurate than the original template maps and less accurate than the SVM classifications. TGC cannot compete with the SVM's accuracy. However, if TGC can reliably improve accuracies on a larger scale

(e.g., provincial or national scale) with minimal or no human intervention, it would be very useful. Although TGC's accuracy and automation on large scales is unknown, the experimental results provide insight into general TGC behaviour. The following sections analyse the problems encountered at the watershed scale to predict the usefulness of TGC's accuracy and automation for general application.

6.3 The Problem of Forest Correspondence Error

The most serious problem encountered in the experiments was forest correspondence error. It occurred when forest pixels had lower forest correspondence than non-forest pixels and meant that no matter what ϕ_f value was selected, some pixels were misclassified. These errors were evident in the overlapping distributions of forest correspondence in Figures 5.4 and 5.5. These errors were also evident in the decreases in UA_f that sometimes occurred as ϕ_f increased (Figure 5.6) and by the fact that the $MAREA_f$ classification accuracies were less than 100 % (Figure 5.7). Forest correspondence errors have two causes: template label error, and mixed clusters.

6.3.1 Template Label Error

Template label error occurs where a template label is not correct. It is a problem because it can arbitrarily increase or decrease the magnitude of forest correspondence and cause forest correspondence error.

In the experiments, wetlands were frequently labelled as forest by the EOSD (e.g., Figure 5.14b), resulting in decreased separability of the forest and non-forest distributions. The EOSD's low UA_f (44 % average) meant that non-forest pixels often had relatively high forest correspondence (Figure 5.4). In contrast, the high UA_f (74 % average) and the low PA_f (28 % average) of the GFC caused both forest and non-forest pixels to have low forest correspondences (Figure 5.5).

6.3.1.1 The Relationship between Template Label Error and Classification Accuracy

This section quantifies how template label error affects TGC classification accuracy using a simple model derived from experimental results. The template label error was calculated for each ROI using ground-reference data and measured using the PA_f and 1-PA_{nf} statistics. By definition, a template's PA_f and 1-PA_{nf} values are equal to the mean forest correspondences of the forest and non-forest pixels. The difference of the means, d, can be used as an approximation of the separability of forest and non-forest pixels:

$$d = PA_f - (1 - PA_{nf}), \tag{6.1}$$

where PA_{nf} is the non-forest producer's accuracy.

The experimental results showed that as d increases, accuracies of the $MAREA_f$ classifications also increases (Figure 6.2). However, the rate of increase was very different between templates. Compared with the GFC, the EOSD accuracies increased more rapidly with small changes in d, whereas the GFC had a much larger range of d values than the EOSD despite having a similar range of accuracies. These differences in behaviour between the EOSD and GFC are problematic for automation, because they make it difficult to predict how increases in template accuracies will affect classification accuracies. However, in general, higher template accuracies produce higher TGC classification accuracies.

6.3.1.2 The Effects of Systematic and Random Template Label Error

Systematic and random variation in template accuracies affected the forest correspondence of all clusters, even when clusters were relatively pure. For example, Figure 6.3 indicates how template label errors, caused by edges and scale effects, reduce forest correspondence in pure clusters.

TGC will always mislabel a cluster if it is entirely mislabelled by the template. This

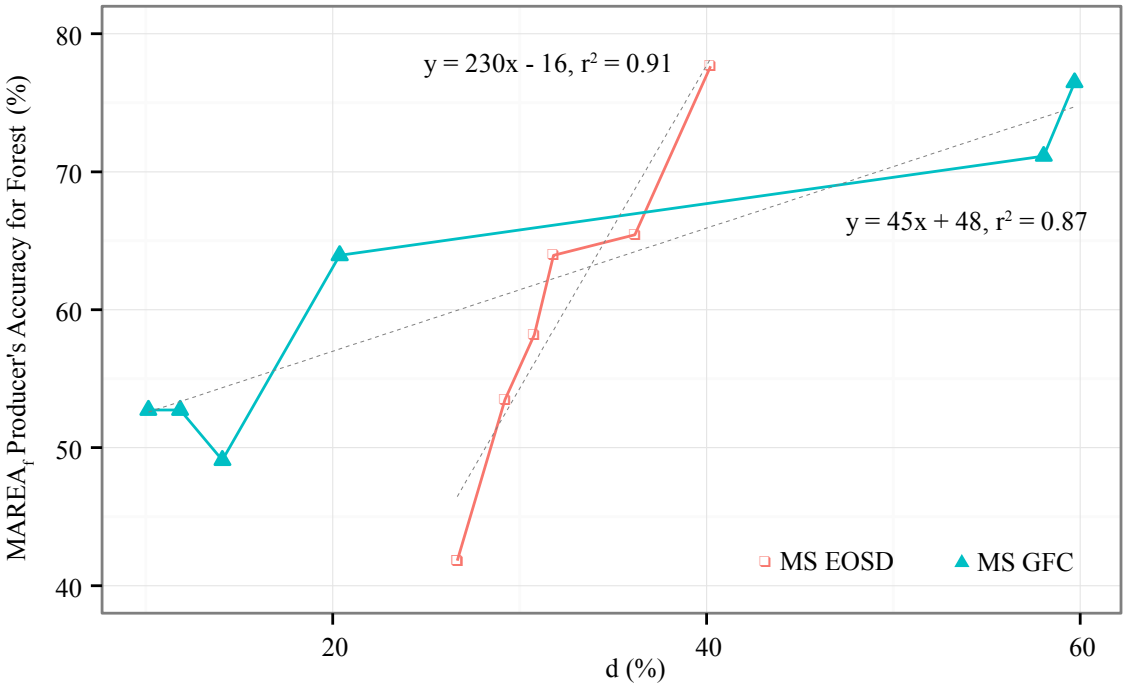

Figure 6.2: PA_f as a function of *d* for each ROI of the MS EOSD and MS GFC $MAREA_f$ classifications. The greater the distance between the template's PA_f and 1-PA_{nf}, the higher the PA_f of the $MAREA_f$ classification. In general, higher template accuracies will produce higher TGC classification accuracies.

problem may occur for spectrally distinct remnant patches that are smaller than the minimum mapping unit of the template (e.g., Figure 6.3e). However, the hope is that small remnant patches will belong to the same clusters as larger, correctly labelled patches (e.g., Figure 6.3f).

The very low PA_f values of the GFC were troubling, because the large errors of omission may include entire spectral classes of forest. For example, patches with low percentages of tree cover may not be well represented by the GFC. Errors of omission that are randomly scattered between different clusters can potentially be corrected, although the forest correspondences will be lower. However, systematic template label errors that are concentrated in problematic clusters will result in classification error.

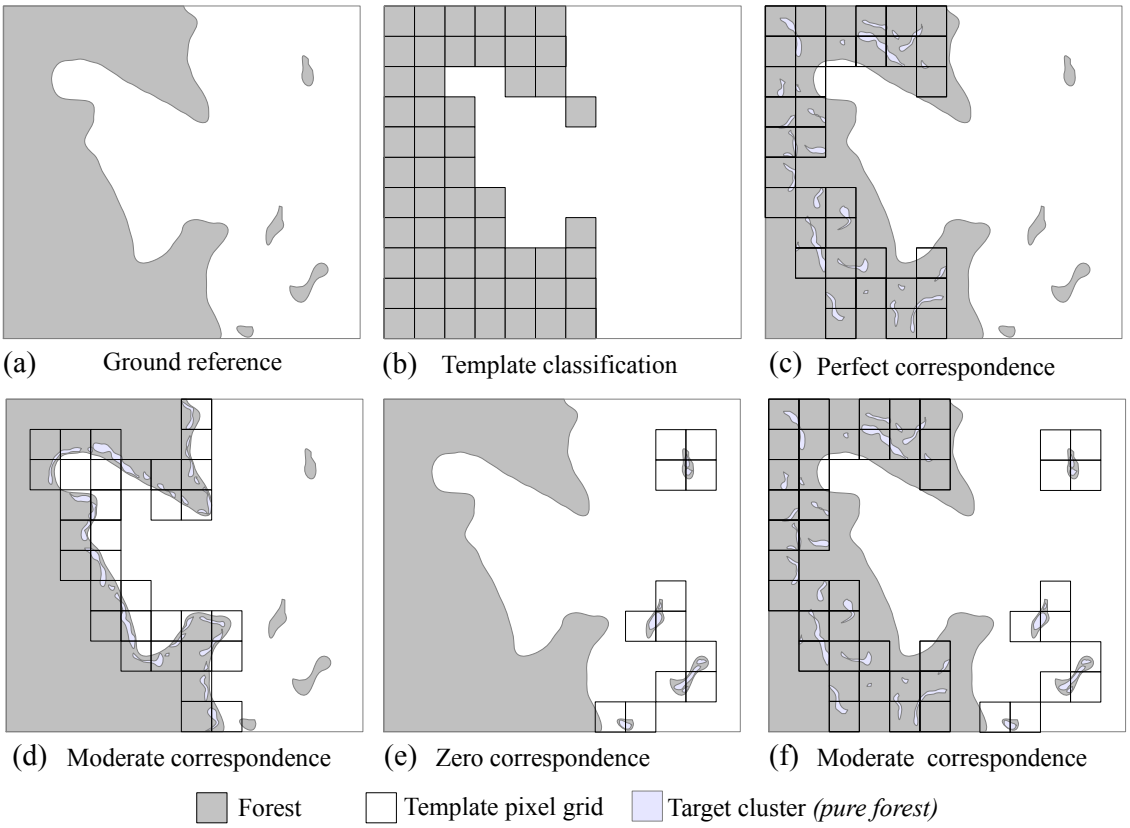

Figure 6.3: Pure forest clusters have lower forest correspondence when they contain small patches or edge pixels; (a) ground reference showing forest, (b) perfect template classification (the label for each template pixel is the dominant class), (c) cluster with perfect forest correspondence to the template, (d) cluster with many edge pixels with moderate correspondence, (e) cluster of small patches with zero forest correspondence, and (f) cluster with some interior pixels and some exterior patches with moderate forest correspondence.

6.3.2 Mixed Cluster Error

The accuracy of TGC was seriously limited by mixed clusters. If the cluster purities in the LiDAR ground-reference site are typical (Table 5.11), then mixed clusters are the most significant problem affecting TGC accuracies. Within the LiDAR site, at least 86 % of error in OA was caused by mixed clusters. Visual inspection of the TGC clusterings revealed that clusters were frequently impure and that spectrally similar pixels of different classes were difficult to separate into different clusters.

Besides limiting classification accuracies, mixed clusters are more likely to be misla-

belled. For example, a mixed cluster that is 70 % forest and 30 % non-forest should be labelled as forest. However, the low forest correspondence expected from the non-forest pixels within the cluster may cause the entire cluster to be mislabelled. With mixed clusters, there is reduced capacity to correct template error.

6.3.3 Template and Cluster Interaction

With mixed clusters having such a large effect on TGC accuracies, the need for purer clusters is obvious. All experiments showed problems with mixed clusters. Fortunately, purer clusters can be produced by changing the stopping parameters of template-guided clustering to allow smaller clusters. However, the problem with using smaller clusters is that they are more likely to be overfitted to the template than larger clusters. As cluster sizes decrease, the importance of structure within the target data decreases and the importance of template label accuracy increases. This raises the following questions:

- How valuable is the structural information in the target?
- How valuable is the label information in the template?
- What is the optimal weighting of template and target information?

The answers to these questions vary according to the classification problem. For example, when the template is 100 % accurate, TGC should ensure that pixels are 100 % pure (e.g., by splitting each cluster until it contains a single pixel). When the template is just slightly more than 50 % accurate, a perfect classification is only possible by splitting the ROI into two (100 % pure) clusters.

The value of information from the template is determined by its accuracy. Within the experiments, template accuracies varied greatly between ROIs; PA_f values for different ROIs varied by a factor of 5 within the GFC. This presents a problem for automatic classification, because variable information quality makes it difficult to select clustering param-

eters that produce pure clusters without overfitting. It seems that TGC has replaced the problem of feature variability with the problem of template variability.

Template variability might be overcome by template maps with data quality layers. For example, the 500-m MLCT land-cover map publishes a "class confidence" layer (Friedl et al., 2010) that might be used to provide per-pixel weighting to class-correspondence calculations. Per-pixel confidence estimates can be easily produced (e.g., Brown et al. (2009)), so all new land-cover maps should publish them to permit error analysis of derived products.

Because template variability is a limitation that cannot be avoided without better template maps, it may be necessary to focus on obtaining target data with better structural information. If data can be more readily clustered into relatively large and pure clusters, template variability would then be less important because error correction would be more robust. Many of the techniques used for improving data separability by supervised classification could be employed to increase cluster purity and size. Image time series contain vast amounts of information and could likely be used to generate large, pure clusters more easily. Similarly, use of different imaging modalities, such as hyperspectral or RADAR imaging, may also improve cluster purity without the need to overfit to the template. Texture and context data may also be helpful.

More advanced clustering algorithms should also be investigated. Minor changes such as using more intelligent seed values (e.g., Arthur and Vassilvitskii (2007)) or more comprehensive changes such as context-sensitive or fuzzy segmentation (Phillips et al., 2012; Lizarazo and Elsner, 2011) could be helpful. Repeating the experiments with standard decision tree algorithms and template map training data would produce interesting results.

6.4 The Problem of Parameter Selection

Automatic parameter selection is complicated by the variability of template accuracy and target structure. In this thesis, the TGC parameters for purity thresholds, minimum cluster sizes, and forest correspondence thresholds were manually selected. However, it is likely that these parameter values, which were tuned to a single ROI, would not perform as well if used with other templates or datasets. The results have shown that the effects of parameters vary according to the template classification accuracies and forest areas within each ROI. The key question is whether TGC parameters that produce improved classifications can be selected automatically (or with minimal effort).

The following sections suggest methods for automating TGC parameter selection. However, although these methods are based on empirical results and models of general TGC behaviour, their general applicability is untested.

6.4.1 Automating Clustering Parameter Selection

There are two clustering parameters that determine how closely clusters are fitted to a template: (i) cluster purity threshold, and, (ii) minimum cluster size.

The experiments showed that the effect of the cluster purity threshold depends on the $\%LAND_f$ value of the template. For example, a 95 % cluster purity threshold applied to GFC 538-241 produced only two clusters, whereas the same cluster purity threshold applied to MS EOSD 538-241 produced 754. This was because the entire 538-241 ROI was already 97 % pure according to the GFC, but only 85 % pure according to the EOSD. This illustrates the basic problem of using a template to guide classification: when the template's accuracy is unknown, metrics derived from the template are only relative measures.

Ideally, clustering parameters would be calculated from template accuracy and target structure information. However, without this information, the parameters for general classification can be calculated using heuristics derived from the experimental results. For

example, the cluster purity threshold might be set automatically by the following equation:

$$\text{Cluster Purity Threshold} = 1 - \frac{\%LAND_f}{3}. \tag{6.2}$$

The cluster purity thresholds calculated from Equation 6.2 using the experimental ROIs are similar to the fixed values used in the experiments (Figure 6.4). The somewhat higher thresholds calculated for the southern ROIs may be beneficial for extracting the smaller forest areas.

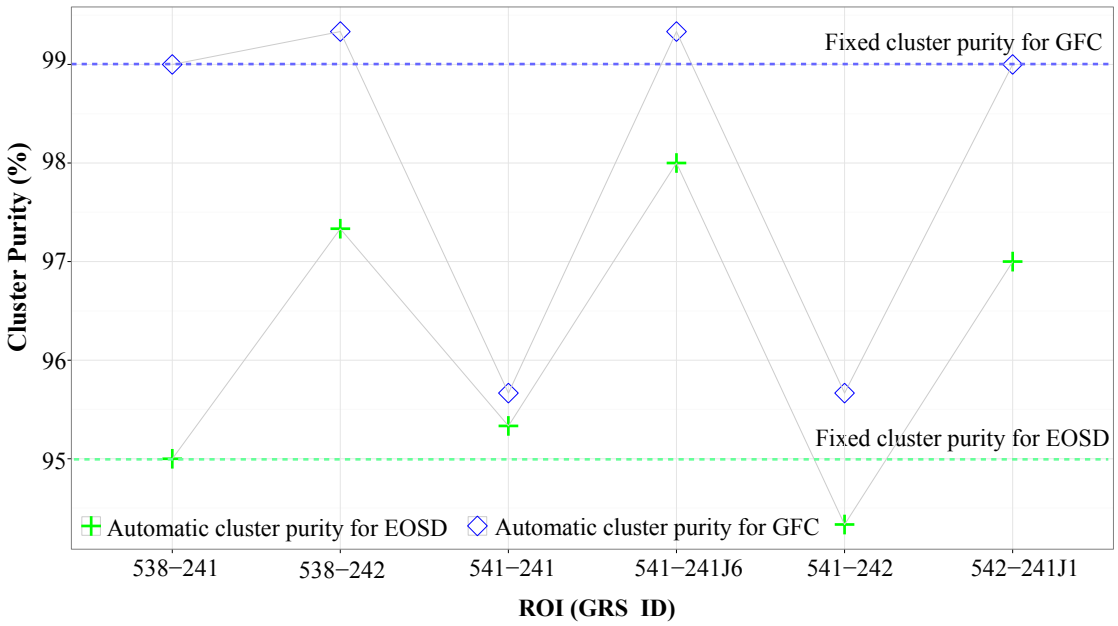

Figure 6.4: The automatic cluster purity thresholds calculated from Equation 6.2 for the experimental ROIs are similar to the fixed values used in the experiments.

The minimum cluster size parameter could be automatically set to 25,000, the same value used in the experiments. For images with very small or very large numbers of pixels, it may be adequate to set the minimum cluster size to 0.05 % of the total number of pixels in the image (i.e., the percentage used by the experiments).

As presented, the automatic TGC clustering parameters are very rudimentary, but they illustrate some basic principles. For example, the cluster purity threshold calculations show

that the number of clusters should be related to the amount of forest data within a template image. The minimum cluster size is an acknowledgement of the limit of information that can be extracted from a template.

6.4.2 Automating ϕ_f Selection

Without classification accuracy or class confidence information for the template, there is no clear solution for selecting ϕ_f. However, if the PA_f for the ROI is known, the $MAREA_f$ threshold can be estimated.

The basic idea is that the $MAREA_f$ threshold is located where there are an equal number of mislabelled pixels above and below the threshold. The experimental results showed that as the number of pixels of the forest class decreased relative to the number of non-forest pixels, the $MAREA_f$ threshold increased. This is because the balancing point between distributions shifted. For example, an ROI that is 90-% forest would likely have a lower $MAREA_f$ threshold than an ROI that is 90-% non-forest. In general, the $MAREA_f$ threshold would vary greatly depending on the proportion of forest to non-forest pixels and the shapes of the forest correspondence distributions.

Equation 6.3 describes how the $MAREA_f$ thresholds of the experiments were related to the PA_f values.

$$\text{Estimated MAREA}_\text{f}\text{ threshold} = \frac{\text{PA}_\text{f}}{2} - k, \tag{6.3}$$

where $k = 0.14$ for EOSD classifications, and $k = 0.22$ for GFC classifications. Figure 6.5 shows the results of applying Equation 6.3 to experimental ROIs. The Willmott's index of agreement (Willmott, 1981) was 0.87 although it would likely be lower if thresholds were predicted for other regions.

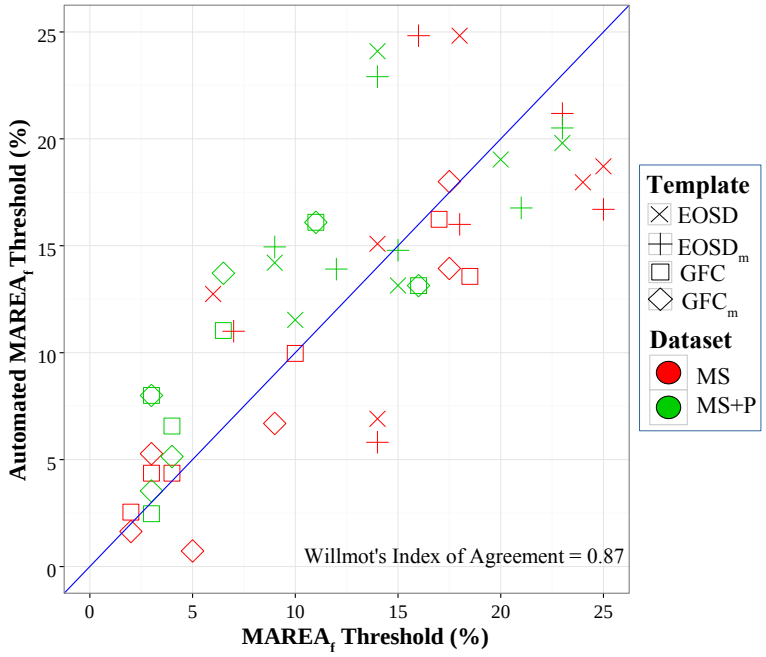

Figure 6.5: MAREA$_f$ thresholds estimated using Equation 6.3 compared with actual MAREA$_f$ values.

6.5 Processing Times

Another problem was that the processing time required to run the *k*-means clustering greatly limited experimentation with different parameters. For example, the MS EOSD clusterings had an average processing time of 4.8 h. This meant that optimizing parameters using a grid search was not practical. Decreased processing times are possible using parallel computing and software optimization (Phillips et al., 2007).

6.6 TGC Usefulness for Automatic Remnant Mapping

The current TGC algorithm is not useful for automatic remnant mapping because of its low classification accuracy. Although TGC results were better than the original templates, they were produced using manually selected parameters. In general, when run automatically with the suggested automatic parameters, TGC is unlikely to produce accurate classifications, even with ideal ϕ_f selection. The main problem is clustering. Mixed clusters greatly

limit the accuracy of TGC classifications. In addition, template label errors reduce the accuracy even more.

Although the current implementation of TGC is not automatic, the hope for successful TGC classifications is found in purer and larger clusters. Multi-temporal, radar, and hyperspectral data can improve separation. Per-pixel information about template-label accuracies can help guide the clustering process to appropriately fit the data. Presumably, better separation of wetlands, dark fields, and forest would have greatly improved experimental accuracies. These expected improvements are tempered by the theoretical limits of clustering, where clusters and class definitions cannot always correspond, and where mixed pixels make pure clusters a theoretical impossibility.

Chapter 7

Conclusions

This thesis focused on reducing the cost of land-cover map production by automating classification. The hypothesis was that land-cover maps contain sufficient information about the spatial distribution of classes to automate land-cover classification by partitioning an image into clusters and labelling each cluster with the class of the land-cover map that it best corresponds to spatially. Although individual experiments produced excellent results, the hypothesis was not met because the experiments showed that general automation would likely result in unacceptably low classification accuracies.

The hypothesis was tested by developing the Template-Guided Classification (TGC) algorithm and applying it to map the forest patches within the Vermilion River watershed in Alberta. Given a satellite image and an overlapping land-cover map as input, TGC partitioned the image into clusters of similar pixels and labelled each cluster with the class from the land-cover map that it best corresponded to spatially. Each cluster was repeatedly split until it was sufficiently pure or could not be split into clusters having more than a specified number of pixels. TGC used clusters to correct errors in the land-cover map; if map labels overlapping a pure cluster were more than 50 % correct, the cluster was labelled with 100 % accuracy. Since template map labels were biased towards the non-forest class, they were weighted to permit correction of even more than 50 % error.

TGC was applied to six SPOT 5 multispectral images at 10-m resolution and two orthoimages at 1-m resolution. Overlapping portions of the 25-m resolution Earth Observation for the Sustainable Development of Forests (EOSD) and the 30-m Global Forest Cover (GFC) maps were used to guide the classifications. TGC parameters were manually optimized for a single image and applied automatically to all images.

TGC forest accuracies were low, averaging 58 % forest producer's accuracy (PA_f) and 67 % forest user's accuracy (UA_f). Although these results were disappointing from the perspective of forest management, they had 25 % fewer errors of omission and 8 % fewer errors of commission than the original land-cover maps. The average PA_f was only 14 % lower than the average PA_f of the SVM classifications, and the average UA_f only 8 % lower. If such improvements could be repeated automatically for large areas, TGC would be very useful. However, analysis indicated that the optimal TGC parameters are highly dependent on spatial variations of land-cover map accuracy and on the relative areas of classes within the target images. Therefore, lower accuracies are expected when TGC is applied over large areas.

In contrast, the 1-m orthophotos were classified very accurately, with PA_f averaging 80 % and UA_f 85 %. The orthophotos had relatively high percentages of forest area (approximately 25 %) that resulted in statistical advantages when calculating forest accuracies.

The analysis of the experimental results found two main sources of error:

1. Mixed clusters caused by low spectral separability of classes and poor choices of clustering parameters, and

2. Template label error caused by error in the initial classification, land-cover changes, spatial resolution effects, and georegistration error.

Mixed clusters were a crucial problem, because the average cluster purity of an ROI is the hard upper limit of its overall classification accuracy. Visual assessment of the clustering results indicated that mixed clusters were a large source of error. This was confirmed within a 5.25 x 7.5 km^2 site of ground-reference data, where more than 80 % of classification errors were caused by mixed clusters.

The template map classification accuracies were low. At 10-m resolution, the EOSD's average PA_f and UA_f were below 45 %; the GFC's average PA_f was 28 % and its average

UA_f was 74 %. Despite the low template accuracies, the LiDAR ground-reference site indicated they were responsible for less than 20 % of TGC's overall classification error.

TGC's optimal clustering depends on template accuracy. As template accuracy increases, the need for error correction decreases. Therefore, with better templates, clustering can be more aggressive, configured to produce smaller (and presumably purer) clusters. In turn, the effect of template error on TGC accuracy depends on clustering. In general, as cluster purity and cluster size increase, the ability to correct template error increases.

From the perspective of automation, the main problem with using a template map to guide classification is the large and unpredictable variation in template accuracies. Template variability makes it difficult to cluster optimally and to determine an appropriate threshold for forest correspondence. Within the different ROIs, the GFC's PA_f values varied by a factor of five; the EOSD's UA_f values varied by a factor of two. Consequently, optimal clustering parameters and the predictive power of template pixels varied unpredictably. TGC used the template as the measure of all things, but the measure itself varied.

To mitigate this variation for general classification, automatic parameters for TGC were suggested that were based upon the experimental results. These parameters have not been tested so their behaviour in general situations is unknown. However, the parameters' equations suggested that it may be useful to investigate the general relationships between the cluster purity threshold and $\%LAND_f$, and between the $MAREAF_f$ threshold and PA_f.

The science question of whether TGC is useful can be answered from two perspectives. When the only consideration is the classification accuracy of remnant trees within the Vermilion River watershed, TGC is not useful because SVM accuracies are typically higher than the TGC accuracies. However, TGC may be useful for automatically improving the accuracies of large-area land-cover maps. Although TGC automatically improved template classification accuracies within the study site, the greater variabilities of land cover and land-cover maps within large areas could cause problems with the automatic selection of

appropriate parameters. Analysis suggested that some of the problems could be mitigated by template accuracy information and that classification accuracies would be improved by using target data with more separable classes. Whether sufficient accuracies can be reliably achieved is an open question.

7.1 Future Work

The goal is to apply TGC to produce spatially high-resolution land-cover maps of Alberta. TGC will be applied to its own output to produce maps with progressively higher spatial resolutions. Multi-temporal and multi-instrument target data will be obtained from archives of 250-m MODIS, 30-m Landsat, 10/20-m Sentinel-2, 20-m Sentinel-1, and 1-m orthophoto images.

TGC will be extended to:

- use multi-temporal and radar data to better distinguish between classes,
- classify more land-cover types, and
- automate TGC parameter selection using class confidence products.

In addition, the use of multiple template maps for generating class confidence values will also be investigated.

References

Aplin, P. (2006). On scales and dynamics in observing the environment. *International Journal of Remote Sensing*, 27(11):2123–2140.

Arthur, D. and Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics.

Ball, G. and Hall, D. (1965). ISODATA, a novel method of data analysis and pattern classification. Technical report, Stanford Research Institute. 61 pages.

Baraldi, A., Bruzzone, L., and Blonda, P. (2005). Quality assessment of classification and cluster maps without ground truth knowledge. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(4):857–873.

Bartholome, E. and Belward, A. S. (2005). GLC2000: a new approach to global land cover mapping from earth observation data. *International Journal of Remote Sensing*, 26(9):1959–1977.

Beauchemin, M. and Thomson, K. P. B. (1997). The evaluation of segmentation results and the overlapping area matrix. *International Journal of Remote Sensing*, 18(18):3895–3899.

Bicheron, P., Defourny, P., Brockmann, C., Schouten, L., Vancutsem, C., Huc, M., Bontemps, S., Leroy, M., Achard, F., Herold, M., Ranera, F., and Arino, O. (2008). GlobCover: products description and validation report. Technical report, Medias France. 47 pages.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press. 358 pages.

Brown, K. M., Foody, G. M., and Atkinson, P. M. (2009). Estimating per-pixel thematic uncertainty in remote sensing classifications. *International Journal of Remote Sensing*, 30(1):209–229.

Bruzzone, L. and Persello, C. (2009). A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(7):2142–2154.

Carroll, Z., Bird, S., Emmett, B., Reynolds, B., and Sinclair, F. (2004). Can tree shelterbelts on agricultural land reduce flood risk? *Soil Use and Management*, 20(3):357–359.

Cihlar, J. (2000). Land cover mapping of large areas from satellites: Status and research priorities. *International Journal of Remote Sensing*, 21(6-7):1093–1114.

Cochrane, M. A. (2000). Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing*, 21(10):2075–2087.

Congalton, R. G. and Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press. 183 pages.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

de Barros Ferraz, S. F., Vettorazzi, C. A., Theobald, D. M., and Ballester, M. V. R. (2005). Landscape dynamics of Amazonian deforestation between 1984 and 2002 in central Rondonia, Brazil: assessment and future scenarios. *Forest Ecology and Management*, 204(1):69 – 85.

De Fries, R. S., Hansen, M., Townshend, J. R. G., and Sohlberg, R. (1998). Global land cover classifications at 8 km spatial resolution: The use of training data derived from Landsat imagery in decision tree classifiers. *International Journal of Remote Sensing*, 19(16):3141–3168.

DeFries, R. and Chan, J. C.-W. (2000). Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing of Environment*, 74(3):503 – 515.

Demir, B., Bovolo, F., and Bruzzone, L. (2013). Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(1):300–312.

DiMiceli, C., Carroll, M., Sohlberg, R., Huang, C., Hansen, M., and Townshend, J. (2011). MODIS vegetation continuous fields (MOD44B) at 250 m spatial resolution for 2010. Collection 5 Percent Tree Cover.

Driver, M. and Higgins, I. (2001). The economic use of remnant vegetation. In Earl, G., Stelling, F., Titcumb, M., and Berwick, S., editors, *Revegetation Guide - Goulburn Broken Catchment Area*, chapter 12, pages 69–74. Department of Natural Resources and Environment, Victoria.

Duda, T. and Canty, M. (2002). Unsupervised classification of satellite imagery: Choosing a good algorithm. *International Journal of Remote Sensing*, 23(11):2193–2212.

Eldridge, D. J. and Freudenberger, D. (2005). Ecosystem wicks: Woodland trees enhance water infiltration in a fragmented agricultural landscape in eastern Australia. *Austral Ecology*, 30(3):336–347.

Fischer, J., Stott, J., and Law, B. S. (2010a). The disproportionate value of scattered trees. *Biological Conservation*, 143(6):1564 – 1567.

Fischer, J., Stott, J., Zerger, A., Warren, G., Sherren, K., and Forrester, R. I. (2009). Reversing a tree regeneration crisis in an endangered ecoregion. *Proceedings of the National Academy of Sciences*, 106(25):10386–10391.

Fischer, J., Zerger, A., Gibbons, P., Stott, J., and Law, B. S. (2010b). Tree decline and the future of Australian farmland biodiversity. *Proceedings of the National Academy of Sciences*, 107(45):19597–19602.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1):185 – 201.

Friedl, M. A. and Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61(3):399–409.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114(1):168 – 182.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Limin, Y., Barnes, C., Herold, N., and Wickham, J. (2011). Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 77(9):858–864.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press. 675 pages.

Gibbons, P. and Boak, M. (2002). The value of paddock trees for regional conservation in an agricultural landscape. *Ecological Management & Restoration*, 3(3):205–210.

Gibbons, P., Lindenmayer, D. B., Fischer, J., Manning, A. D., Weinberg, A., Seddon, J., Ryan, P., and Barrett, G. (2008). The future of scattered trees in agricultural landscapes. *Conservation Biology*, 22(5):1309–1319.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., and Chen, J. (2013). Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 34(7):2607–2654.

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

Hansen, M., Egorov, A., Potapov, P., Stehman, S., Tyukavina, A., Turubanova, S., Roy, D., Goetz, S., Loveland, T., Ju, J., Kommareddy, A., Kovalskyy, V., Forsyth, C., and Bents, T. (2014). Monitoring conterminous United States (CONUS) land cover change

with Web-Enabled Landsat Data (WELD). *Remote Sensing of Environment*, 140(0):466 – 484.

Hansen, M. C. and Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, 122(0):66 – 74.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850–853. Data available on-line from: http://earthenginepartners.appspot.com/science-2013-global-forest.

Harvey, C. and Haber, W. (1998). Remnant trees and the conservation of biodiversity in Costa Rican pastures. *Agroforestry Systems*, 44(1):37–68.

Herold, M., Mayaux, P., Woodcock, C., Baccini, A., and Schmullius, C. (2008). Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, 112(5):2538 – 2556.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J. N., and Wickham, J. (2007). Completion of the 2001 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 73(4):337.

Homer, C., Huang, C., Yang, L., Wylie, B., and Coan, M. (2004). Development of a 2001 National Land-Cover Database for the United States. *Photogrammetric Engineering & Remote Sensing*, 70(7):829–840.

Hsieh, P.-F., Lee, L., and Chen, N.-Y. (2001). Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing. *Geoscience and Remote Sensing, IEEE Transactions on*, 39(12):2657–2663.

Huang, C., Song, K., Kim, S., Townshend, J. R., Davis, P., Masek, J. G., and Goward, S. N. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, 112(3):970 – 985.

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., and Ferreira, L. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1):195 – 213.

Huete, A., Didan, K., Van Leeuwen, W., Jacobson, A., Solanos, R., and Laing, T. (1999). MODIS Vegetation Index (MOD13) algorithm theoretical basis document. 120 pages.

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM Comput. Surv.*, 31(3):264–323.

Jensen, J. R. (1996). *Introductory digital image processing: a remote sensing perspective.* Prentice-Hall Inc., 2nd edition. 316 pages.

Jensen, J. R., Ramsey, E. W., Mackey, H. E., Christensen, E. J., and Sharitz, R. R. (1987). Inland wetland change detection using aircraft MSS data. *Photogrammetric Engineering and Remote Sensing*, 53(5):521–529.

Jiang, H., Strittholt, J. R., Frost, P. A., and Slosser, N. C. (2004). The classification of late seral forests in the Pacific Northwest, USA using Landsat ETM+ imagery. *Remote Sensing of Environment*, 91(3–4):320 – 331.

Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., and Xian, G. (2013). A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sensing of Environment*, 132(0):159 – 175.

Jurskis, V. (2005). Eucalypt decline in Australia, and a general concept of tree decline and dieback. *Forest Ecology and Management*, 215(1-3):1 – 20.

Kelly, P. M. and White, J. M. (1993). Preprocessing remotely sensed data for efficient analysis and classification. In *SPIE Vol. 1963 Applications of Artificial Intelligence 1993: Knowledge-Based Systems in Aerospace and Industry*, volume 1963 of *Applications of Artificial Intelligence*, pages 24–30. Los Alamos National Laboratory, Proc. SPIE.

Knorn, J., Rabe, A., Radeloff, V. C., Kuemmerle, T., Kozak, J., and Hostert, P. (2009). Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote Sensing of Environment*, 113(5):957–964.

Landgrebe, D. (1999). Information extraction principles and methods for multispectral and hyperspectral image data. In Chen, C.-h., editor, *Information Processing for Remote Sensing*, pages 3–38. World Scientific Publishing Co.

Landgrebe, D. A. (2002). Toward a maximally effective means for analysis of hyperspectral data. In *International Symposium on Remote Sensing*, pages 264–269. International Society for Optics and Photonics.

Landsberg, J. and Wylie, F. (1988). Dieback of rural trees in Australia. *GeoJournal*, 17(2):231–237.

Lang, R., Shao, G., Pijanowski, B. C., and Farnsworth, R. L. (2008). Optimizing unsupervised classifications of remotely sensed imagery with a data-assisted labeling approach. *Computers & Geosciences*, 34(12):1877–1885.

Lark, R. M. (1995). A reappraisal of unsupervised classification, I: correspondence between spectral and conceptual classes. *International Journal of Remote Sensing*, 16(8):1425–1443.

Latifovic, R., Zhu, Z.-L., Cihlar, J., Giri, C., and Olthof, I. (2004). Land cover mapping of North and Central America–Global Land Cover 2000. *Remote Sensing of Environment*, 89(1):116 – 127.

Lechner, A. M., Stein, A., Jones, S. D., and Ferwerda, J. G. (2009). Remote sensing of small and linear features: Quantifying the effects of patch size and length, grid position and detectability on land cover mapping. *Remote Sensing of Environment*, 113(10):2194–2204.

Li, C., Wang, J., Wang, L., Hu, L., and Gong, P. (2014). Comparison of classification algorithms and training sample sizes in urban land classification with Landsat Thematic Mapper imagery. *Remote Sensing*, 6(2):964–983.

Lizarazo, I. and Elsner, P. (2011). Segmentation of remotely sensed imagery: moving from sharp objects to fuzzy regions. In Ho, P.-G., editor, *Image Segmentation*, chapter 13. InTech. 550 pages.

Loveland, T. R., Sohl, T. L., Stehman, S. V., Gallant, A. L., Sayler, K. L., and Napton, D. (2002). A strategy for estimating the rates of recent United States land- cover changes. *Photogrammetric Engineering and Remote Sensing*, 68(10):1091–1099.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297. University of California Press, Berkeley, Calif.

Manning, A. D., Fischer, J., and Lindenmayer, D. B. (2006). Scattered trees are keystone structures - implications for conservation. *Biological Conservation*, 132(3):311 – 321.

Maron, M. and Fitzsimons, J. A. (2007). Agricultural intensification and loss of matrix habitat over 23 years in the West Wimmera, south-eastern Australia. *Biological Conservation*, 135(4):587 – 593.

Mayaux, P., Eva, H., Gallego, J., Strahler, A., Herold, M., Agrawal, S., Naumov, S., De Miranda, E., Di Bella, C., Ordoyne, C., Kopin, Y., and Roy, P. S. (2006). Validation of the Global Land Cover 2000 map. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(7):1728–1739.

McIntyre, S. and Hobbs, R. (1999). A framework for conceptualizing human effects on landscapes and its relevance to management and research models. *Conservation Biology*, 13(6):1282–1292.

McIver, D. and Friedl, M. (2002). Using prior probabilities in decision-tree classification of remotely sensed data. *Remote Sensing of Environment*, 81(2–3):253 – 261.

Millsom, D. and Earl, G. (2001). *Goulburn Broken Catchment Revegetation Guide*, chapter 5, pages 36–39. Department of Natural Resources and Environment, Victoria, Australia.

Mountrakis, G., Im, J., and Ogole, C. (2011). Support vector machines in remote sensing: a review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247 – 259.

Musy, R. F., Wynne, R. H., Blinn, C. E., Scrivani, J. A., and McRoberts, R. E. (2006). Automated forest area estimation using iterative guided spectral class rejection. *Photogrammetric engineering and remote sensing*, 72(8):949–960.

NSWA (2005). *State of the North Saskatchewan Watershed Report*, chapter 5. North Saskatchewan Watershed Alliance Society, Edmonton, Alberta. 6 pages.

NSWA (2009). Vermilion River water supply and demand study. Technical report, North Saskatchewan Watershed Alliance, Edmonton, Alberta. 140 pages.

Olthof, I., Butson, C., and Fraser, R. (2005). Signature extension through space for northern landcover classification: A comparison of radiometric correction methods. *Remote Sensing of Environment*, 95(3):290 – 302.

Olthof, I., Latifovic, R., and Pouliot, D. (2009). Development of a circa 2000 land cover map of northern Canada at 30 m resolution from Landsat. *Canadian Journal of Remote Sensing*, 35(2):152–165.

Ozesmi, S. and Bauer, M. E. (2002). Satellite remote sensing of wetlands. *Wetlands Ecology and Management*, 10(5):381–402.

Ozolins, A., Brack, C., and Freudenberger, D. (2001). Abundance and decline of isolated trees in the agricultural landscapes of central New South Wales, Australia. *Pacific Conservation Biology*, 7(3):195.

Parshakov, I., Coburn, C., and Staenz, K. (2014). Automated class labeling of classified Landsat TM imagery using a Hyperion-generated hyperspectral library. *Photogrammetric Engineering & Remote Sensing*, 80(8):797–805.

Phillips, R. D., Watson, L. T., and Wynne, R. H. (2007). Hybrid image classification and parameter selection using a shared memory parallel algorithm. *Computers & Geosciences*, 33(7):875 – 897.

Phillips, R. D., Watson, L. T., Wynne, R. H., and Ramakrishnan, N. (2012). Continuous iterative guided spectral class rejection classification algorithm. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(6):2303–2317.

Pittman, K., Hansen, M. C., Becker-Reshef, I., Potapov, P. V., and Justice, C. O. (2010). Estimating global cropland extent with multi-year MODIS data. *Remote Sensing*, 2(7):1844–1863.

Potapov, P. V., Turubanova, S. A., Hansen, M. C., Adusei, B., Broich, M., Altstatt, A., Mane, L., and Justice, C. O. (2012). Quantifying forest cover loss in Democratic Republic of the Congo, 2000-2010, with Landsat ETM data. *Remote Sensing of Environment*, 122(0):106 – 116.

Price, J. C. (1994). How unique are spectral signatures? *Remote Sensing of Environment*, 49(3):181 – 186.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1):81–106.

Rouse, J. W., Haas, R. H., and Schell, J. A. (1974). Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. Final Report RSC 1978-4, Texas A & M University, College Station, Texas.

Sanchez-Azofeifa, G., Chong, M., Sinkwich, J., and Mamet, S. (2004). Alberta ground cover characterization (AGCC) training and procedures manual. Technical report, Earth Observation Systems Laboratory, Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton. 67 pages.

Schott, J. R. (2007). *Remote Sensing: the Image Chain Approach*. Oxford University Press, New York, second edition. 666 pages.

Scrivani, J. A., Wynne, R. H., Blinn, C. E., and Musy, R. F. (2001). Phase I forest area estimation using Landsat TM and iterative guided spectral class rejection: assessment of possible training data protocols. In *Proceedings of the Second Annual Forest Inventory and Analysis symposium*, pages 11–14.

Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K. M., Channan, S., DiMiceli, C., and Townshend, J. R. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, 6(5):427–448.

Shao, G., We, W., Wu, G., Zhou, X., and Wu, J. (2003). An explicit index for assessing the accuracy of cover-class areas. *Photogrammetric engineering and remote sensing*, 69(8):907–914.

Smith, J. H., Wickham, J. D., Stehman, S. V., and Yang, L. (2002). Impacts of patch size and land-cover heterogeneity on thematic image classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 68(1):65–70.

Song, K. (2010). *Tackling uncertainties and errors in the satellite monitoring of forest cover change*. PhD thesis, University of Maryland. 187 pages.

Song, K., Townshend, J., Kim, S., Davis, P., Clay, R., and Rodas, O. (2005). Improving automated detection of land cover change for large areas using Landsat data. In *Analysis of Multi-Temporal Remote Sensing Images, 2005 International Workshop on the*, pages 39–43.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46:234–240.

Townshend, J. R. (1980). *The spatial resolving power of earth resources satellites: a review*. NASA, Goddard Space Flight Center. 36 pages.

Townshend, J. R., Masek, J. G., Huang, C., Vermote, E. F., Gao, F., Channan, S., Sexton, J. O., Feng, M., Narasimhan, R., Kim, D., Song, K., Song, D., Song, X.-P., Noojipady, P., Tan, B., Hansen, M. C., Li, M., and Wolfe, R. E. (2012). Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges. *International Journal of Digital Earth*, 5(5):373–397.

Tso, B. and Mather, P. (2009). *Classification methods for remotely sensed data*. CRC press, Boca Raton, Florida. 356 pages.

van der Sluijs, J., Hall, R. J., Peddle, D. R., and Schwarz, S. (2014). Influence of spectral endmember variability and modelled background heterogeneity on leading tree species classifications of northern boreal forests, NWT CANADA. In *Proceedings of the 2014 IEEE International Geoscience and Remote Sensing Symposium*. 3 pages.

Vetaas, O. R. (1992). Micro-site effects of trees and shrubs in dry savannas. *Journal of Vegetation Science*, 3(3):337–344.

Vogelmann, J. E., Howard, S. M., Limin, Y., Larson, C. R., Wylie, B. K., and Van Driel, N. (2001). Completion of the 1990s National Land Cover Data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric engineering and remote sensing*, 67(6):650–662.

Wan, Z., Zhang, Y., Zhang, Q., and Li, Z.-l. (2002). Validation of the land-surface temperature products retrieved from Terra moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, 83(1):163–180.

Wayman, J. P., Wynne, R. H., Scrivani, J. A., and Reams, G. A. (2001). Landsat TM-based forest area estimation using iterative guided spectral class rejection. *Photogrammetric Engineering and Remote Sensing*, 67(10):1155–1166.

Willmott, C. J. (1981). On the validation of models. *Physical Geography*, 2(2):184–194.

Witten, I. H. and Frank, E. (2011). *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann Publishers, 3rd edition. 628 pages.

Wulder, M., Cranny, M., and Dechka, J. (2002). An illustrated methodology for land cover mapping of forests with Landsat-7 ETM+ data, version 2. Technical report, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, Canada. 39 pages.

Wulder, M. A., White, J. C., Cranny, M., Hall, R. J., Luther, J. E., Beaudoin, A., Goodenough, D. G., and Dechka, J. A. (2008). Monitoring Canada's forests. part 1: Completion of the EOSD land cover project. *Canadian Journal of Remote Sensing*, 34(6):549–562.

Xian, G., Homer, C., and Fry, J. (2009). Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods. *Remote Sensing of Environment*, 113(6):1133 – 1147.

Yu, L., Wang, J., and Gong, P. (2013). Improving 30 m global land-cover map FROM-GLC with times series MODIS and auxiliary data sets: a segmentation-based approach. *International Journal of Remote Sensing*, 34(16):5851–5867.

Zhang, J., Staenz, K., Rochdi, N., Philippe, T., Ren, X., Lowe, D., and Eddy, P. (2009). Automatic surface reflectance retrieval from SPOT data using MODTRAN-based look-up-tables. In *Proceedings of the 30th Canadian Symposium on Remote Sensing*.

Zhu, X. (2008). Semi-supervised learning literature survey. Technical report, University of Wisconsin-Madison. 60 pages.

Zhu, Z., Woodcock, C. E., and Olofsson, P. (2012). Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment*, 122(0):75 – 91.

Appendices

Appendix A

10-m TGC Watershed Classifications

A.1 Accuracies with Constant ϕ_f

Table A.1: Classification accuracies for TGC with constant ϕ_f for varying templates and datasets. Overall accuracies of TGC classifications were high, often exceeding 90 %. However, forest classification accuracies were lower and highly variable.

Parameters	Metric (%)	538-241	538-242	541-241	541-242	542-241J1	542-241J6	Average
MS EOSD	PA_f	67	49	74	30	72	38	55
with $\phi_f = 0.2$	UA_f	53	68	85	81	62	68	69
	OA	92	94	93	92	87	91	92
	REA_f	41	-56	-17	-208	22	-114	-55
MS+P EOSD	PA_f	66	36	67	31	73	43	53
with $\phi_f = 0.2$	UA_f	66	59	90	100	65	83	77
	OA	94	93	95	92	87	93	93
	REA_f	0	-106	-39	-225	17	-113	-78
MS $EOSD_m$	PA_f	68	56	71	31	75	38	57
with $\phi_f = 0.2$	UA_f	51	68	85	76	66	49	66
	OA	92	94	92	92	88	89	91
	REA_f	49	-32	-23	-192	20	-60	-40
MS+P $EOSD_m$	PA_f	62	35	60	39	70	48	52
with $\phi_f = 0.2$	UA_f	60	68	89	83	63	73	73
	OA	94	94	94	92	86	93	92
	REA_f	3	-140	-56	-133	16	-69	-63
MS GFC	PA_f	69	49	88	35	87	44	62
with $\phi_f = 0.06$	UA_f	56	75	53	76	54	60	62
	OA	93	95	84	92	83	91	89
	REA_f	33	-70	76	-158	70	-63	-19
MS+P GFC	PA_f	66	36	83	49	86	51	62
with $\phi_f = 0.06$	UA_f	67	61	52	69	56	72	63
	OA	94	94	89	91	83	93	91
	REA_f	-3	-112	71	-59	64	-56	-16
MS GFC_m	PA_f	64	47	88	31	83	42	59
with $\phi_f = 0.06$	UA_f	52	68	53	71	54	53	59
	OA	92	94	84	91	83	90	89
	REA_f	36	-65	75	-182	65	-52	-21
MS+P GFC_m	PA_f	71	38	83	49	85	46	62
with $\phi_f = 0.06$	UA_f	69	69	54	69	55	70	64
	OA	95	94	90	91	83	92	91
	REA_f	5	-117	66	-59	63	-75	-19

PA_f = forest Producer's Accuracy, UA_f = forest User's Accuracy, OA = Overall Accuracy,
REA_f = forest Relative Error of Area, ϕ_f = forest correspondence threshold,
MS = 10-m multispectral data, MS+P = 10-m multispectral and panchromatic data,
EOSD =EOSD template, $EOSD_m$ = water-masked EOSD template,
GFC = GFC template, GFC_m = water-masked GFC template,

A.2 Error Matrices with Constant ϕ_f

Table A.2: MS EOSD ($\phi_f = 0.2$) classifications.

538-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	601	20	621
Forest	37	41	78
Total	638	61	699

Ground Reference (%)	
Non-forest	Forest
94.2	32.8
5.8	67.2
100.00	100.00

Metrics (%)	
PA_f	67.2
UA_f	52.6
$\%LAND_f$	11.2
OA	91.8

538-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	632	28	660
Forest	13	27	40
Total	645	55	700

Ground Reference (%)	
Non-forest	Forest
98.0	50.9
2.0	49.1
100.00	100.00

Metrics (%)	
PA_f	49.1
UA_f	67.5
$\%LAND_f$	5.7
OA	94.1

541-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	395	22	417
Forest	11	63	74
Total	406	85	491

Ground Reference (%)	
Non-forest	Forest
97.3	25.9
2.7	74.1
100.00	100.00

Metrics (%)	
PA_f	74.1
UA_f	85.1
$\%LAND_f$	15.1
OA	93.3

541-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	387	30	417
Forest	3	13	16
Total	390	43	433

Ground Reference (%)	
Non-forest	Forest
99.2	69.8
0.8	30.2
100.00	100.00

Metrics (%)	
PA_f	30.2
UA_f	81.2
$\%LAND_f$	3.7
OA	92.4

542-241J1

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	323	23	346
Forest	36	58	94
Total	359	81	440

Ground Reference (%)	
Non-forest	Forest
90.0	28.4
10.0	71.6
100.00	100.00

Metrics (%)	
PA_f	71.6
UA_f	61.7
$\%LAND_f$	21.4
OA	86.6

542-241J6

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	435	34	469
Forest	10	21	31
Total	445	55	500

Ground Reference (%)	
Non-forest	Forest
97.8	61.8
2.2	38.2
100.00	100.00

Metrics (%)	
PA_f	38.2
UA_f	67.7
$\%LAND_f$	6.2
OA	91.2

Table A.3: MS+P EOSD ($\phi_f = 0.2$) classifications.

538-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	597	19	616
Forest	19	37	56
Total	616	56	672

Ground Reference (%)	
Non-forest	Forest
96.9	33.9
3.1	66.1
100.00	100.00

Metrics (%)	
PA_f	66.1
UA_f	66.1
$\%LAND_f$	8.3
OA	94.3

538-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	577	30	607
Forest	12	17	29
Total	589	47	636

Ground Reference (%)	
Non-forest	Forest
98.0	63.8
2.0	36.2
100.00	100.00

Metrics (%)	
PA_f	36.2
UA_f	58.6
$\%LAND_f$	4.6
OA	93.4

541-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	312	14	326
Forest	3	28	31
Total	315	42	357

Ground Reference (%)	
Non-forest	Forest
99.0	33.3
1.0	66.7
100.00	100.00

Metrics (%)	
PA_f	66.7
UA_f	90.3
$\%LAND_f$	8.7
OA	95.2

541-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	302	27	329
Forest	0	12	12
Total	302	39	341

Ground Reference (%)	
Non-forest	Forest
100.0	69.2
0.0	30.8
100.00	100.00

Metrics (%)	
PA_f	30.8
UA_f	100.0
$\%LAND_f$	3.5
OA	92.1

542-241J1

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	262	19	281
Forest	28	52	80
Total	290	71	361

Ground Reference (%)	
Non-forest	Forest
90.3	26.8
9.7	73.2
100.00	100.00

Metrics (%)	
PA_f	73.2
UA_f	65.0
$\%LAND_f$	22.2
OA	87.0

542-241J6

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	304	20	324
Forest	3	15	18
Total	307	35	342

Ground Reference (%)	
Non-forest	Forest
99.0	57.1
1.0	42.9
100.00	100.00

Metrics (%)	
PA_f	42.9
UA_f	83.3
$\%LAND_f$	5.3
OA	93.3

Table A.4: MS EOSD$_{\mathrm{m}}$ ($\phi_f = 0.2$) classifications.

538-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	566	18	584
Forest	37	39	76
Total	603	57	660

Ground Reference (%)	
Non-forest	Forest
93.9	31.6
6.1	68.4
100.00	100.00

Metrics (%)	
PA_f	68.4
UA_f	51.3
%$LAND_f$	11.5
OA	91.7

538-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	558	22	580
Forest	13	28	41
Total	571	50	621

Ground Reference (%)	
Non-forest	Forest
97.7	44.0
2.3	56.0
100.00	100.00

Metrics (%)	
PA_f	56.0
UA_f	68.3
%$LAND_f$	6.6
OA	94.4

541-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	369	25	394
Forest	11	60	71
Total	380	85	465

Ground Reference (%)	
Non-forest	Forest
97.1	29.4
2.9	70.6
100.00	100.00

Metrics (%)	
PA_f	70.6
UA_f	84.5
%$LAND_f$	15.3
OA	92.3

541-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	371	29	400
Forest	4	13	17
Total	375	42	417

Ground Reference (%)	
Non-forest	Forest
98.9	69.0
1.1	31.0
100.00	100.00

Metrics (%)	
PA_f	31.0
UA_f	76.5
%$LAND_f$	4.1
OA	92.1

542-241J1

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	312	20	332
Forest	32	61	93
Total	344	81	425

Ground Reference (%)	
Non-forest	Forest
90.7	24.7
9.3	75.3
100.00	100.00

Metrics (%)	
PA_f	75.3
UA_f	65.6
%$LAND_f$	21.9
OA	87.8

542-241J6

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	396	33	429
Forest	21	20	41
Total	417	53	470

Ground Reference (%)	
Non-forest	Forest
95.0	62.3
5.0	37.7
100.00	100.00

Metrics (%)	
PA_f	37.7
UA_f	48.8
%$LAND_f$	8.7
OA	88.5

Table A.5: MS+P EOSD$_{\mathrm{m}}$ ($\phi_f = 0.2$) classifications.

538-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	561	20	581
Forest	21	32	53
Total	582	52	634

Ground Reference (%)	
Non-forest	Forest
96.4	38.5
3.6	61.5
100.00	100.00

Metrics (%)	
PA_f	61.5
UA_f	60.4
%$LAND_f$	8.4
OA	93.5

538-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	512	28	540
Forest	7	15	22
Total	519	43	562

Ground Reference (%)	
Non-forest	Forest
98.7	65.1
1.3	34.9
100.00	100.00

Metrics (%)	
PA_f	34.9
UA_f	68.2
%$LAND_f$	3.9
OA	93.8

541-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	293	17	310
Forest	3	25	28
Total	296	42	338

Ground Reference (%)	
Non-forest	Forest
99.0	40.5
1.0	59.5
100.00	100.00

Metrics (%)	
PA_f	59.5
UA_f	89.3
%$LAND_f$	8.3
OA	94.1

541-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	284	23	307
Forest	3	15	18
Total	287	38	325

Ground Reference (%)	
Non-forest	Forest
99.0	60.5
1.0	39.5
100.00	100.00

Metrics (%)	
PA_f	39.5
UA_f	83.3
%$LAND_f$	5.5
OA	92.0

542-241J1

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	252	21	273
Forest	29	50	79
Total	281	71	352

Ground Reference (%)	
Non-forest	Forest
89.7	29.6
10.3	70.4
100.00	100.00

Metrics (%)	
PA_f	70.4
UA_f	63.3
%$LAND_f$	22.4
OA	85.8

542-241J6

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	277	17	294
Forest	6	16	22
Total	283	33	316

Ground Reference (%)	
Non-forest	Forest
97.9	51.5
2.1	48.5
100.00	100.00

Metrics (%)	
PA_f	48.5
UA_f	72.7
%$LAND_f$	7.0
OA	92.7

Table A.6: MS GFC ($\phi_f = 0.06$) classifications.

538-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	605	19	624
Forest	33	42	75
Total	638	61	699

Ground Reference (%)	
Non-forest	Forest
94.8	31.1
5.2	68.9
100.00	100.00

Metrics (%)	
PA_f	68.9
UA_f	56.0
$\%LAND_f$	10.7
OA	92.6

538-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	636	28	664
Forest	9	27	36
Total	645	55	700

Ground Reference (%)	
Non-forest	Forest
98.6	50.9
1.4	49.1
100.00	100.00

Metrics (%)	
PA_f	49.1
UA_f	75.0
$\%LAND_f$	5.1
OA	94.7

541-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	339	10	349
Forest	67	75	142
Total	406	85	491

Ground Reference (%)	
Non-forest	Forest
83.5	11.8
16.5	88.2
100.00	100.00

Metrics (%)	
PA_f	88.2
UA_f	52.8
$\%LAND_f$	28.9
OA	84.3

541-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	439	36	475
Forest	6	19	25
Total	445	55	500

Ground Reference (%)	
Non-forest	Forest
98.7	65.5
1.3	34.5
100.00	100.00

Metrics (%)	
PA_f	34.5
UA_f	76.0
$\%LAND_f$	5.0
OA	91.6

542-241J1

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	330	13	343
Forest	72	84	156
Total	402	97	499

Ground Reference (%)	
Non-forest	Forest
82.1	13.4
17.9	86.6
100.00	100.00

Metrics (%)	
PA_f	86.6
UA_f	53.8
$\%LAND_f$	31.3
OA	83.0

542-241J6

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	429	31	460
Forest	16	24	40
Total	445	55	500

Ground Reference (%)	
Non-forest	Forest
96.4	56.4
3.6	43.6
100.00	100.00

Metrics (%)	
PA_f	43.6
UA_f	60.0
$\%LAND_f$	8.0
OA	90.6

Table A.7: MS+P GFC ($\phi_f = 0.06$) classifications.

538-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	598	19	617
Forest	18	37	55
Total	616	56	672

Ground Reference (%)	
Non-forest	Forest
97.1	33.9
2.9	66.1
100.00	100.00

Metrics (%)	
PA_f	66.1
UA_f	67.3
$\%LAND_f$	8.2
OA	94.5

538-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	578	30	608
Forest	11	17	28
Total	589	47	636

Ground Reference (%)	
Non-forest	Forest
98.1	63.8
1.9	36.2
100.00	100.00

Metrics (%)	
PA_f	36.2
UA_f	60.7
$\%LAND_f$	4.4
OA	93.6

541-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	283	7	290
Forest	32	35	67
Total	315	42	357

Ground Reference (%)	
Non-forest	Forest
89.8	16.7
10.2	83.3
100.00	100.00

Metrics (%)	
PA_f	83.3
UA_f	52.2
$\%LAND_f$	18.8
OA	89.1

541-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	318	23	341
Forest	10	22	32
Total	328	45	373

Ground Reference (%)	
Non-forest	Forest
97.0	51.1
3.0	48.9
100.00	100.00

Metrics (%)	
PA_f	48.9
UA_f	68.8
$\%LAND_f$	8.6
OA	91.2

542-241J1

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	244	10	254
Forest	51	64	115
Total	295	74	369

Ground Reference (%)	
Non-forest	Forest
82.7	13.5
17.3	86.5
100.00	100.00

Metrics (%)	
PA_f	86.5
UA_f	55.7
$\%LAND_f$	31.2
OA	83.5

542-241J6

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	300	17	317
Forest	7	18	25
Total	307	35	342

Ground Reference (%)	
Non-forest	Forest
97.7	48.6
2.3	51.4
100.00	100.00

Metrics (%)	
PA_f	51.4
UA_f	72.0
$\%LAND_f$	7.3
OA	93.0

Table A.8: MS GFC_m ($\phi_f = 0.06$) classifications.

538-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	586	22	608
Forest	36	39	75
Total	622	61	683

Ground Reference (%)	
Non-forest	Forest
94.2	36.1
5.8	63.9
100.00	100.00

Metrics (%)	
PA_f	63.9
UA_f	52.0
$\%LAND_f$	11.0
OA	91.5

538-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	607	29	636
Forest	12	26	38
Total	619	55	674

Ground Reference (%)	
Non-forest	Forest
98.1	52.7
1.9	47.3
100.00	100.00

Metrics (%)	
PA_f	47.3
UA_f	68.4
$\%LAND_f$	5.6
OA	93.9

541-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	339	10	349
Forest	66	75	141
Total	405	85	490

Ground Reference (%)	
Non-forest	Forest
83.7	11.8
16.3	88.2
100.00	100.00

Metrics (%)	
PA_f	88.2
UA_f	53.2
$\%LAND_f$	28.8
OA	84.5

541-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	438	38	476
Forest	7	17	24
Total	445	55	500

Ground Reference (%)	
Non-forest	Forest
98.4	69.1
1.6	30.9
100.00	100.00

Metrics (%)	
PA_f	30.9
UA_f	70.8
$\%LAND_f$	4.8
OA	91.0

542-241J1

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	331	16	347
Forest	68	80	148
Total	399	96	495

Ground Reference (%)	
Non-forest	Forest
83.0	16.7
17.0	83.3
100.00	100.00

Metrics (%)	
PA_f	83.3
UA_f	54.1
$\%LAND_f$	29.9
OA	83.0

542-241J6

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	425	32	457
Forest	20	23	43
Total	445	55	500

Ground Reference (%)	
Non-forest	Forest
95.5	58.2
4.5	41.8
100.00	100.00

Metrics (%)	
PA_f	41.8
UA_f	53.5
$\%LAND_f$	8.6
OA	89.6

Table A.9: MS+P GFC_m ($\phi_f = 0.06$) classifications.

538-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	587	16	603
Forest	18	40	58
Total	605	56	661

Ground Reference (%)	
Non-forest	Forest
97.0	28.6
3.0	71.4
100.00	100.00

Metrics (%)	
PA_f	71.4
UA_f	69.0
$\%LAND_f$	8.8
OA	94.9

538-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	555	29	584
Forest	8	18	26
Total	563	47	610

Ground Reference (%)	
Non-forest	Forest
98.6	61.7
1.4	38.3
100.00	100.00

Metrics (%)	
PA_f	38.3
UA_f	69.2
$\%LAND_f$	4.3
OA	93.9

541-241

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	284	7	291
Forest	30	35	65
Total	314	42	356

Ground Reference (%)	
Non-forest	Forest
90.4	16.7
9.6	83.3
100.00	100.00

Metrics (%)	
PA_f	83.3
UA_f	53.8
$\%LAND_f$	18.3
OA	89.6

541-242

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	318	23	341
Forest	10	22	32
Total	328	45	373

Ground Reference (%)	
Non-forest	Forest
97.0	51.1
3.0	48.9
100.00	100.00

Metrics (%)	
PA_f	48.9
UA_f	68.8
$\%LAND_f$	8.6
OA	91.2

542-241J1

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	241	11	252
Forest	51	63	114
Total	292	74	366

Ground Reference (%)	
Non-forest	Forest
82.5	14.9
17.5	85.1
100.00	100.00

Metrics (%)	
PA_f	85.1
UA_f	55.3
$\%LAND_f$	31.1
OA	83.1

542-241J6

	Ground Reference (Pixels)		
	Non-forest	Forest	Total
Non-forest	300	19	319
Forest	7	16	23
Total	307	35	342

Ground Reference (%)	
Non-forest	Forest
97.7	54.3
2.3	45.7
100.00	100.00

Metrics (%)	
PA_f	45.7
UA_f	69.6
$\%LAND_f$	6.7
OA	92.4

A.3 Accuracies with Variable ϕ_f

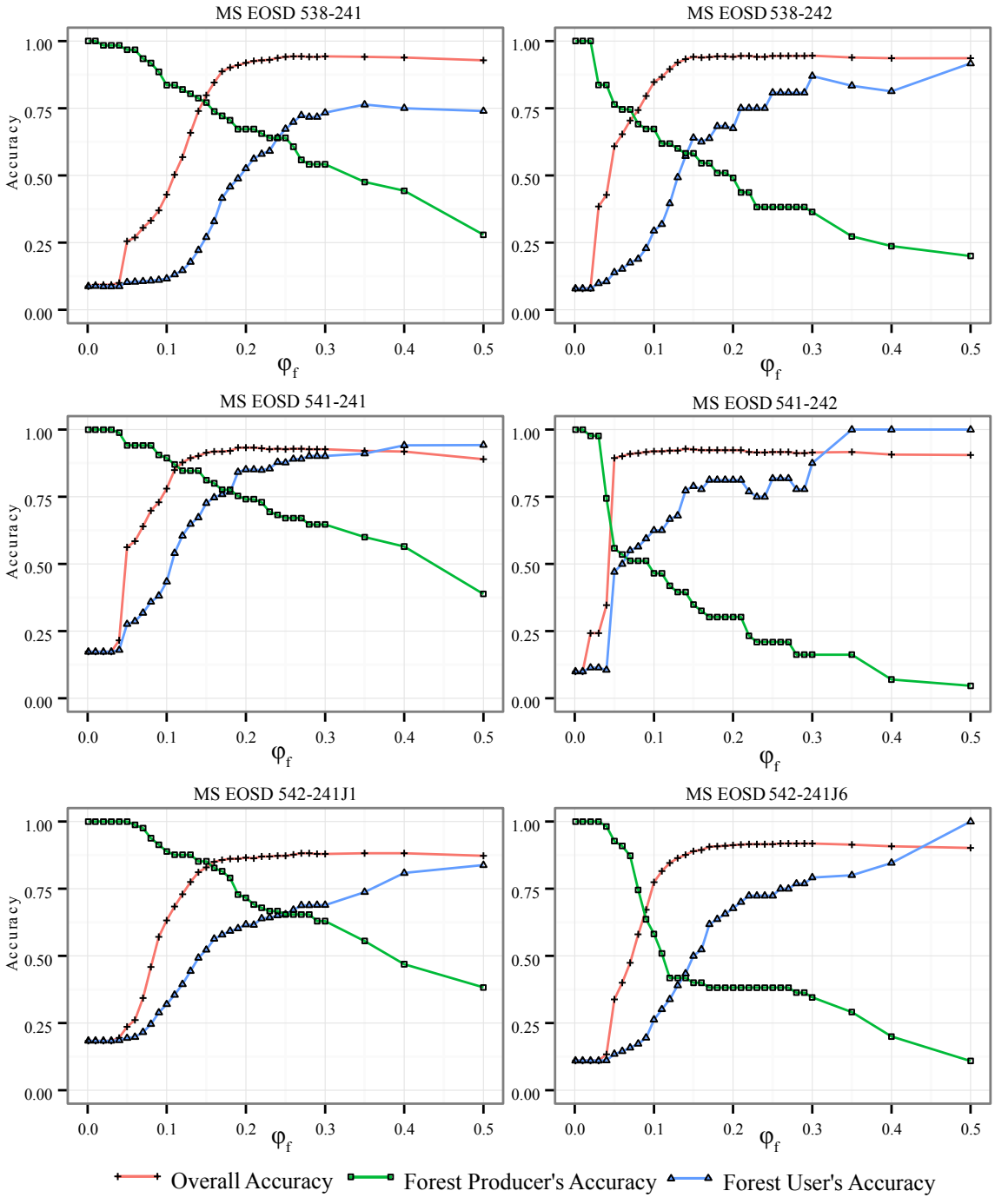

Figure A.1: Accuracies for MS EOSD versus ϕ_f.

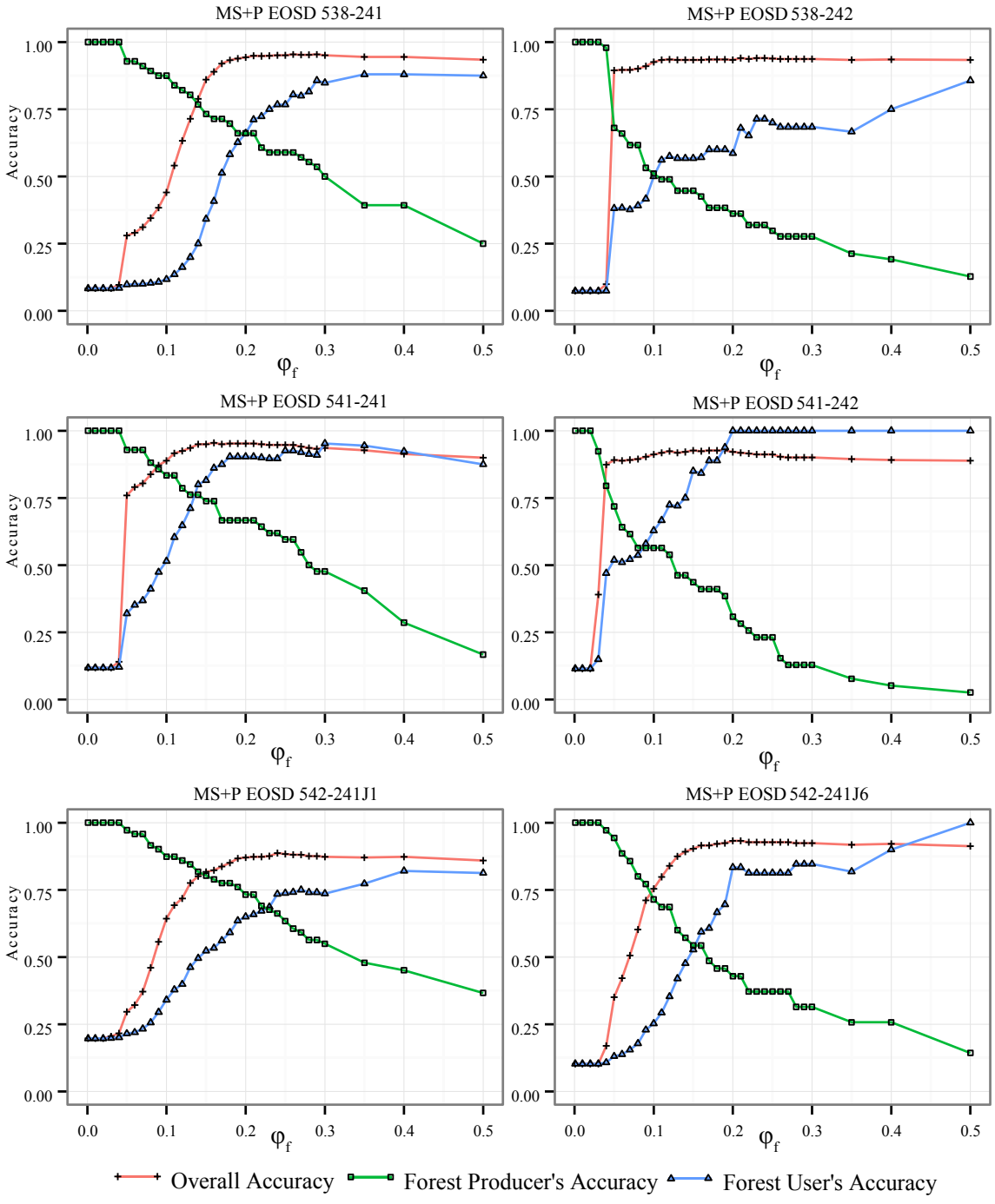

Figure A.2: Accuracies for MS+P EOSD versus ϕ_f.

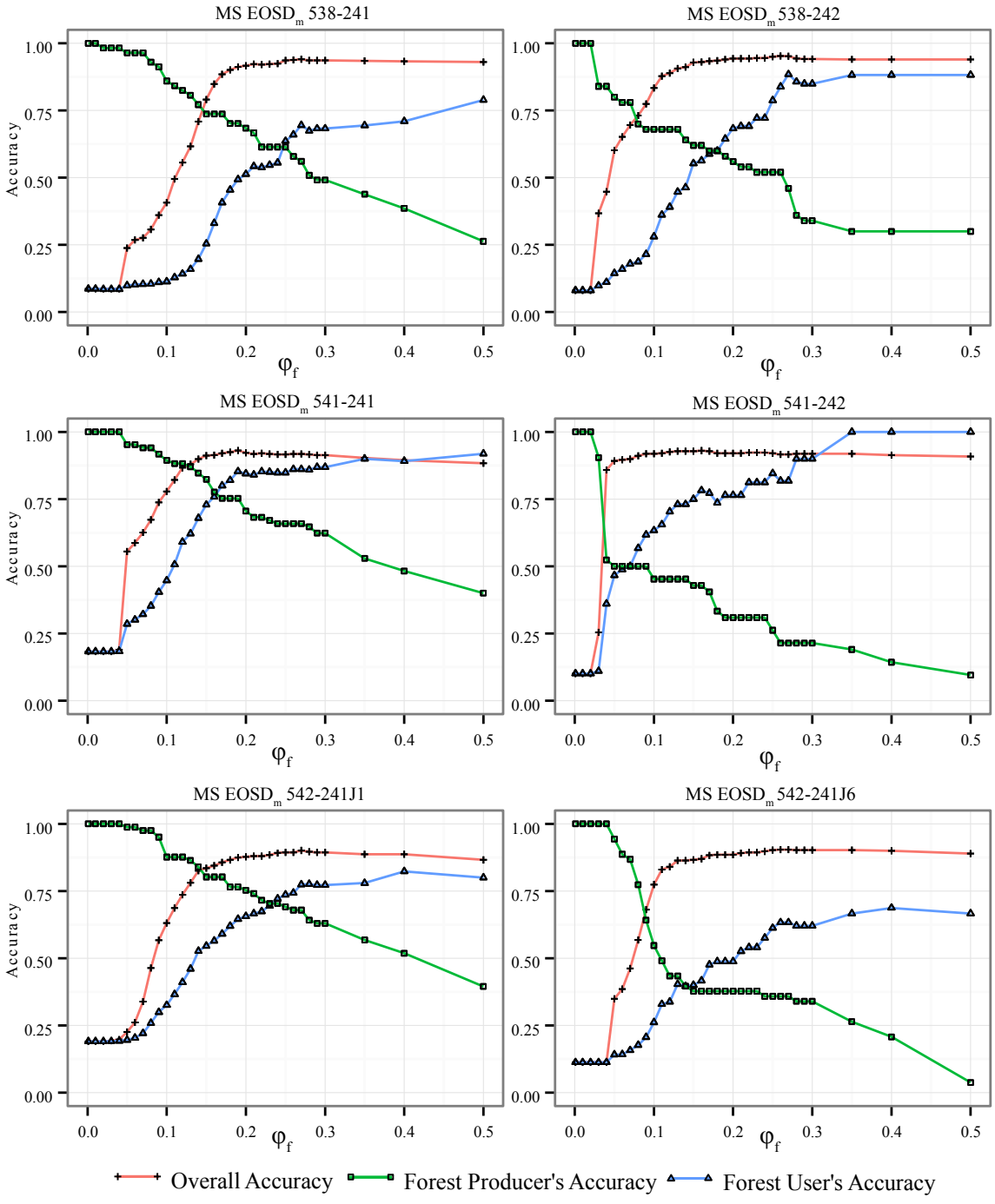

Figure A.3: Accuracies for MS EOSD$_m$ versus ϕ_f.

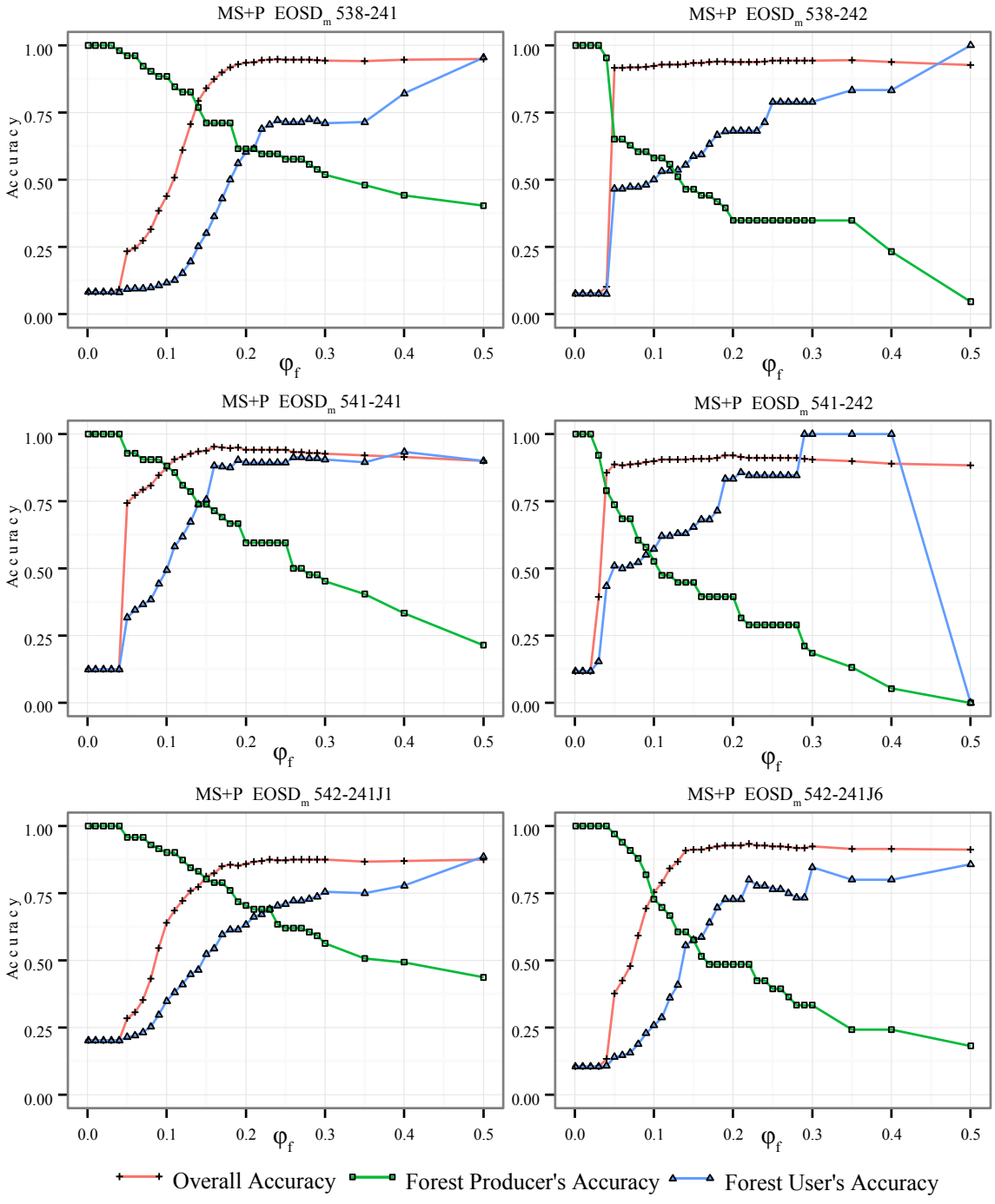

Figure A.4: Accuracies for MS+P EOSD$_m$ versus ϕ_f.

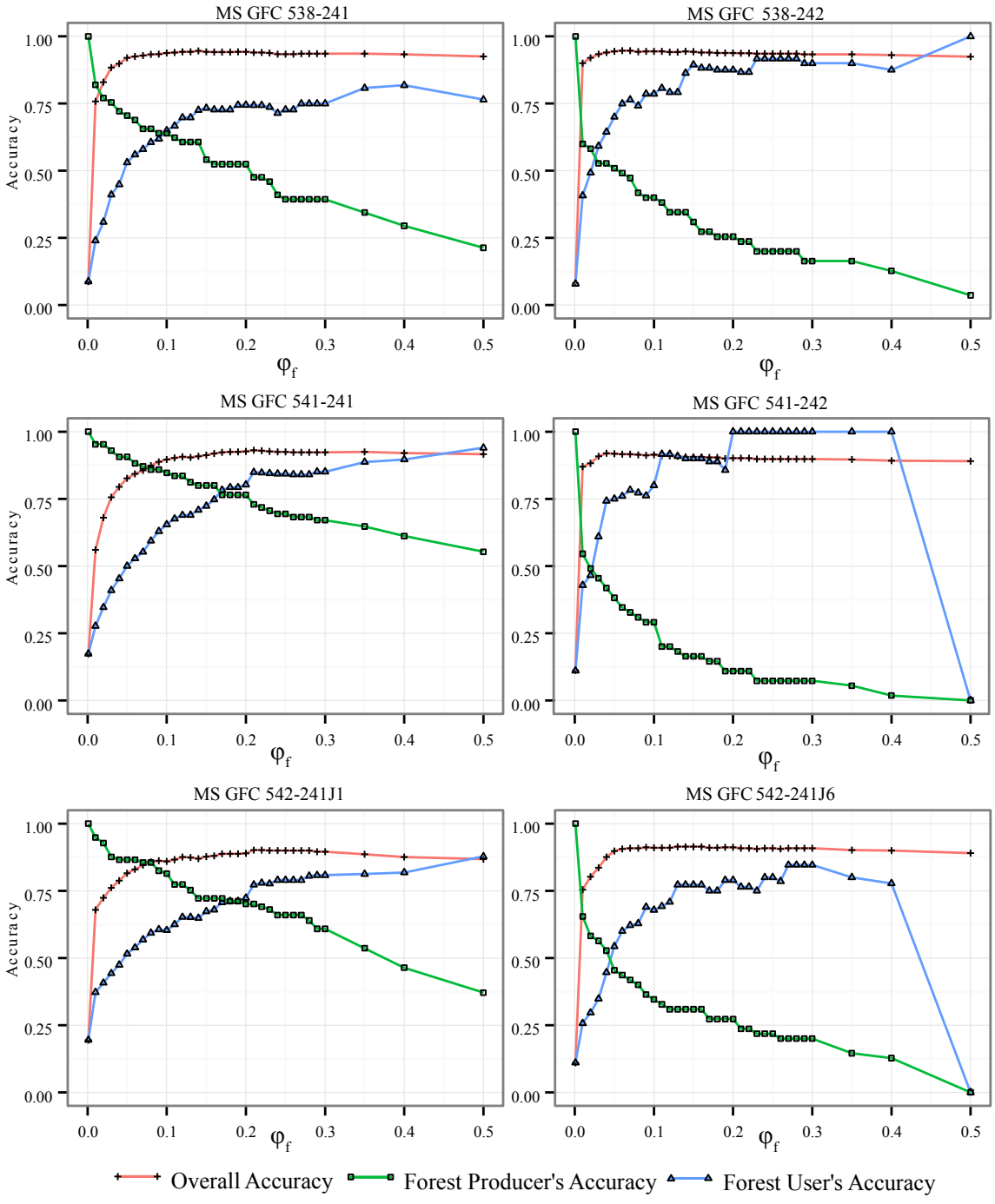

Figure A.5: Accuracies for MS GFC versus ϕ_f.

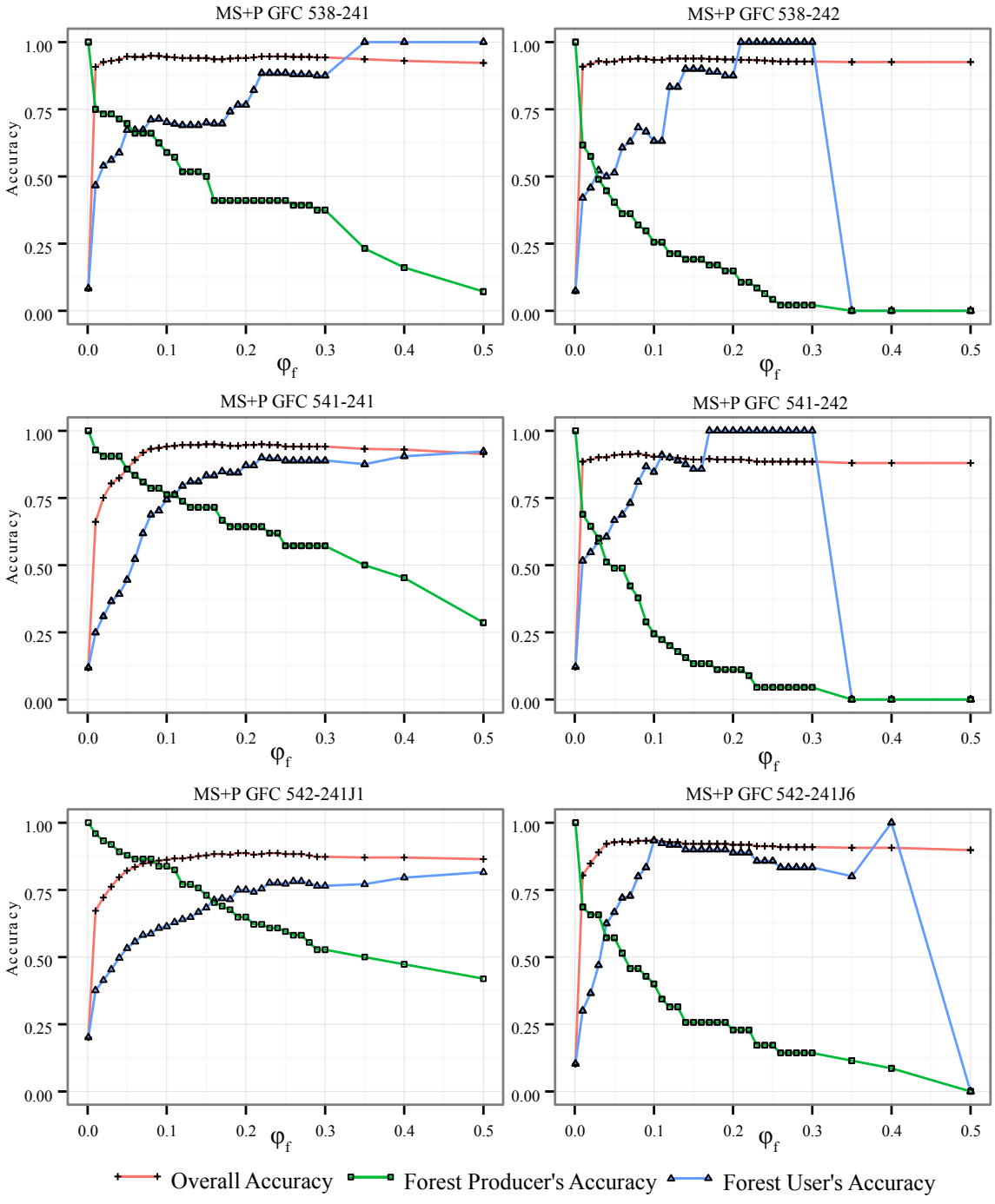

Figure A.6: Accuracies for MS+P GFC versus ϕ_f .

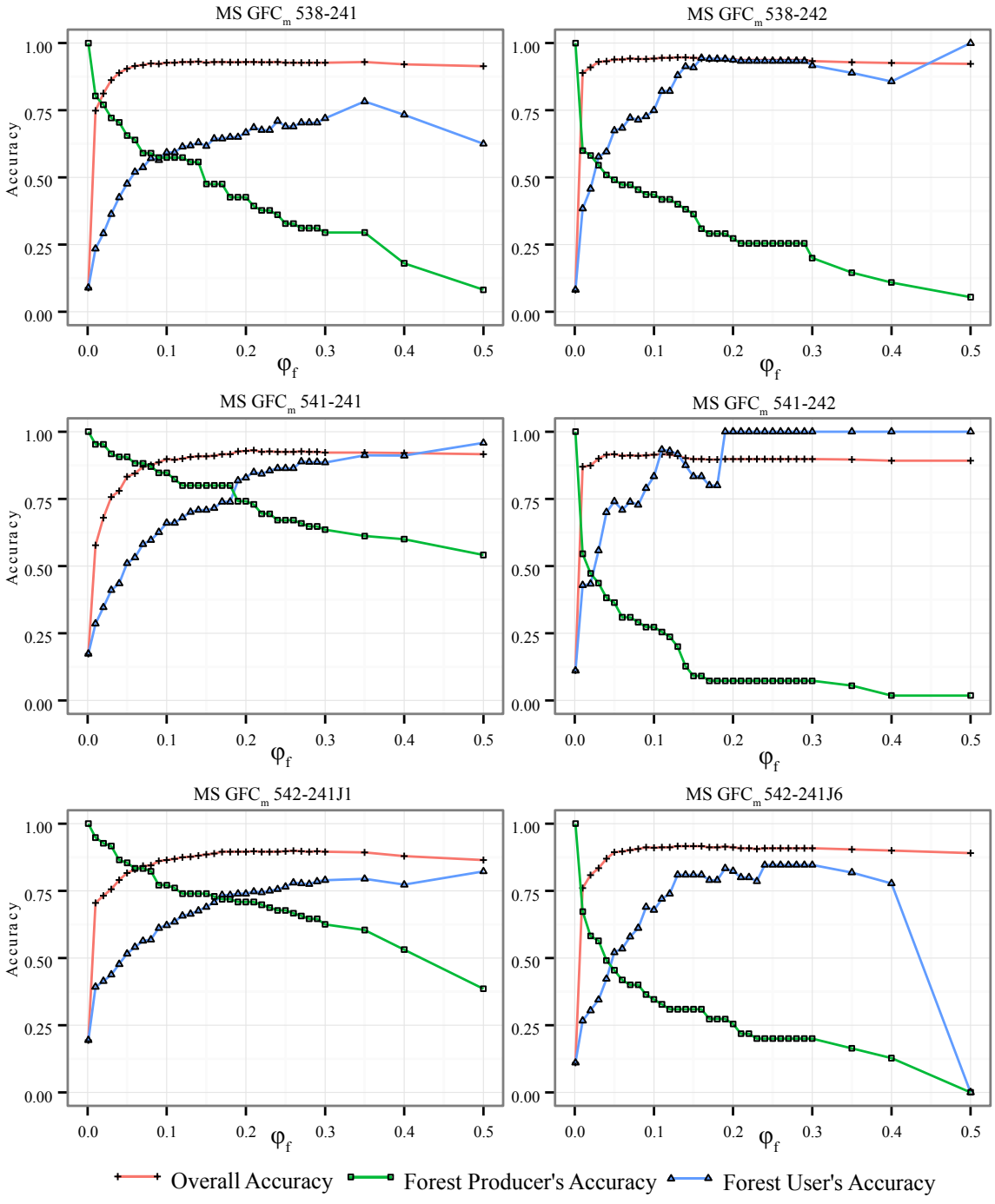

Figure A.7: Accuracies for MS GFC$_m$ versus ϕ_f .

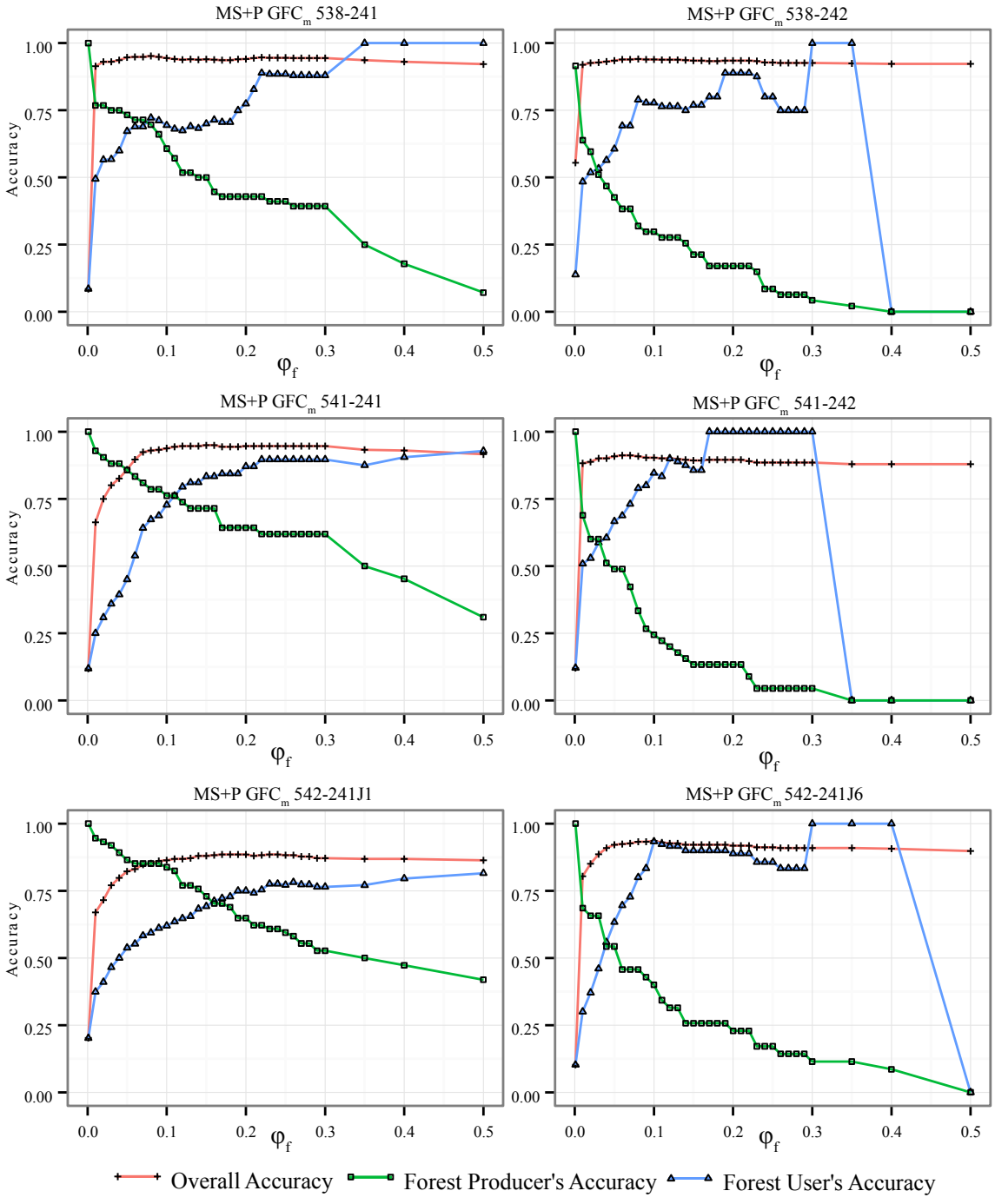

Figure A.8: Accuracies for MS+P GFC_m versus ϕ_f .

Appendix B

10-m Template Classification Accuracies

Table B.1: Template map classification accuracies for each experimental ROI.

Template	Metric (%)	538-241	538-242	541-241	541-242	542-241J1	542-241J6	Average
MS EOSD	PA_f	43	36	46	33	48	33	40
	UA_f	27	36	63	52	48	40	44
	OA	85	90	86	90	81	87	87
	REA_f	131	5	-59	-114	3	-5	-15
MS+P EOSD	PA_f	39	28	33	31	46	40	36
	UA_f	28	38	57	52	49	44	45
	OA	87	91	89	89	80	89	87
	REA_f	100	-100	-123	-131	-9	-21	-47
MS $EOSD_m$	PA_f	42	36	46	33	48	30	39
	UA_f	26	33	63	52	49	38	44
	OA	85	89	85	90	80	87	86
	REA_f	146	22	-59	-107	-3	-69	-12
MS+P $EOSD_m$	PA_f	38	28	31	34	46	36	36
	UA_f	27	39	52	62	49	41	45
	OA	86	91	88	90	80	88	87
	REA_f	110	-100	-131	-131	-12	-33	-49
MS GFC	PA_f	21	11	61	15	60	13	30
	UA_f	68	55	90	80	89	64	74
	OA	92	92	92	90	91	90	91
	REA_f	-323	-733	-52	-563	-55	-629	-392
MS+P GFC	PA_f	23	4	45	13	55	17	26
	UA_f	81	40	79	75	89	75	73
	OA	93	92	92	89	90	91	91
	REA_f	-308	-2100	-95	-617	-68	-450	-606
MS GFC_m	PA_f	21	13	61	15	60	13	30
	UA_f	68	58	90	80	89	64	75
	OA	92	92	92	90	91	90	91
	REA_f	-323	-614	-52	-563	-53	-629	-372
MS+P GFC_m	PA_f	23	4	45	13	55	17	26
	UA_f	81	40	79	75	89	75	73
	OA	93	92	92	89	90	91	91
	REA_f	-308	-2100	-95	-617	-68	-450	-606

PA_f = forest Producer's Accuracy, UA_f = forest User's Accuracy, OA = Overall Accuracy, REA_f = forest Relative Error of Area, $EOSD_m$ = water-masked EOSD template, GFC = GFC template, GFC_m = water-masked GFC template, MS = 10-m multispectral SPOT ROI, MS+P = 10-m multispectral and panchromatic SPOT ROI.