Exploring satellite image analysis methods for characterizing Canterbury shelterbelts, and the application to carbon modelling.

A thesis
submitted in partial fulfilment
of the requirements for the Degree of
Master of Applied Science

at
Lincoln University
by
Lindsay Czerepowicz

Lincoln University
2011

Abstract of a thesis submitted in partial fulfilment of the
requirements for the Degree of Master of Applied Science.

Exploring satellite image analysis methods for characterizing Canterbury shelterbelts, and the application to carbon modelling.

by

Lindsay Czerepowicz

Shelterbelts are a prominent part of Canterbury's agricultural landscape. Despite this, shelterbelts are not particularly well characterized in a spatially explicit manner. This study aimed to develop relatively automated methods for delineating and characterizing shelterbelts using high spatial resolution satellite images, and to then apply these techniques for the purpose of modelling shelterbelt carbon quantities.

First, per-pixel and object-oriented image classification methods for delineating shelterbelts from QuickBird images (0.6 m) were compared. Object-oriented classification with Feature Analyst, a feature extraction software, was the most successful in delineating shelterbelts from the images, with an overall classification accuracy of 92 %.

A statistical modellng method (Random Forests) was then investigated to determine its utility in differentiating shelterbelt tree species using spectral information. Shelterbelt data collected from three study areas (16 km^2 each) within the Hurunui District of Canterbury were used. Shelterbelts could be reliably differentiated into broad species groups to an accuracy of more than 90 %. However, differentiating individual coniferous shelterbelt species proved to be more challenging, with a model accuracy of only about 60 %. Results suggested that blue and red bands are important differentiators of broad species groups, whereas green and near infrared bands are important differentiators of individual coniferous species.

Finally, shelterbelt carbon was modelled as an example application of shelterbelt delineation and species differentiation methods. Shelterbelt carbon was estimated using predetermined allometric equations which utilize field collected measurements. Field-based estimates were then used to model tree biomass using remotely-sensed variables as predictor variables. Regression analysis determined that shelterbelt species, the median value of the red band, shelterbelt width and tree density are all significant predictors of above ground biomass.

Analysis results suggested that 2.6 % of land used for agriculture in the Canterbury Plains are comprised of shelterbelts. This study confirmed that *P.radiata* and *C.macrocarpa* are the major shelterbelt species in Canterbury, comprising 95 % and 3 % of shelterbelts, respectively. This study also suggests that shelterbelts represent a significant carbon reservoir on the Canterbury Plains, sequestering an average of 381 tonnes per hectare of shelterbelt. Currently, only major shelterbelts are accounted for by LUCAS (Land Use and Carbon Analysis System) in the Land Use, Land Use Change, and Forestry (LULUCF) Sector. This study has estimated that shelterbelts contribute at least 9.7 t/ha of carbon to the low producing grassland carbon pool, which is currently estimated at 29 t/ha.

This study has demonstrated that shelterbelts can be successfully delineated and characterized using satellite images. With improvements, the methods presented in this study could be used in the future to semi-automatically delineate and characterize shelterbelts across large agricultural areas of New Zealand, such as the Canterbury Plains. The methods presented here are valuable tools for natural resource management. Potential applications include modelling the effectiveness of shelterbelts as wildlife corridors, assessing the change in shelterbelt landcover resulting from agricultural intensification, and quantifying the shelterbelt carbon pool across a given landscape.

Keywords: shelterbelts, remote sensing, object-oriented classification, Feature Analyst, tree species differentiation, Random Forests, above ground biomass, carbon sequestration.

Acknowledgements

This work has been made possible by the efforts of my supervisors, Bradley Case and Dr Crile Doscher, and by the encouragement of my family and friends.

I thank my supervisors for all of their ideas and comments, particularly Bradley Case who suggested this research topic. I am also grateful to Dr Hamish Rennie, who encouraged me to work towards a Masters degree.

I thank my parents, André and Cheryl Botha, and my sisters, Alison Botha and Nicolette Paul, as well as my best friend, Tania Ngaira, for their continuing support and encouragement. I am also grateful to Alison Botha and Christopher Chick for their contributions as field assistants, particularly to Alison, who helped me in many other ways.

Lastly, I thank my fiancé, Wladek Czerepowicz, for giving me something to look forward to upon completion of this work.

And whatsoever ye do in word or deed, do all in the name of the Lord Jesus, giving thanks to God and the Father by him.

Collosians 3:17

Table of Contents

Abstract ..... **ii**
Acknowledgements ..... **iv**
Table of Contents ..... **v**
List of Tables ..... **viii**
List of Figures ..... **x**
Glossary ..... **xi**
List of Equations ..... **xiv**
Chapter 1 General Introduction **1**
1.1 Aims and Objectives 3
Chapter 2 Literature Review **4**
2.1 Introduction 4
2.2 Shelterbelt Definition 4
2.3 Remote sensing principles 5
2.4 Linear feature extraction from imagery 7
2.5 Differentiation of tree species using image analysis 12
2.6 Estimation of shelterbelt carbon quantities 18
2.7 Literature review summary 23
Chapter 3 General Methods **24**
3.1 Introduction 24
3.2 Study area 24
3.3 Satellite imagery 24
3.4 Image preparation 25
3.5 Field data collection 28
Chapter 4 Image-based Shelterbelt Delineation **31**
4.1 Introduction 31
4.2 Classification methods overview 32
4.3 Methods 35
4.3.1 Per-pixel classification 35
Method 1a: Unsupervised classification 35
Method 1b: Supervised classification 35
4.3.2 Object-oriented image analysis 35
Method 2: Image segmentation and segment labelling 35
Method 3: Feature extraction using Feature Analyst 36
4.3.3 Comparison of methods: Accuracy assessment 40
(i) Shelterbelt identification: area-based assessment 40
(ii) Inclusion of other landcover types: random point-based assessment 42
4.3.4 Application of the Feature Analyst method to all three study areas 42
4.4 Results 43
4.4.1 Comparison of methods for shelterbelt delineation 43
4.4.2 Application of the Feature Analyst method to all three study areas 52
4.5 Discussion 58

4.5.1 Comparison of shelterbelt delineation methods 58
4.5.2 The Feature Analyst Method 59

Chapter 5 Differentiation of Shelterbelt Species 63

5.1 Introduction 63

5.2 Overview of methods used for species discrimination 64

5.3 Methods 66
5.3.1 Spectral and textural information 66
5.3.2 Data exploration 67
5.3.3 The Random Forests algorithm for species discrimination 68

5.4 Results 70
5.4.1 Data collection 70
5.4.2 Data exploration 77
5.4.3 Random Forests 80

5.5 Discussion 87

Chapter 6 Example Application: Modelling Shelterbelt Carbon 90

6.1 Introduction 90

6.2 Methods 91
6.2.1 Field-based carbon estimation 91
6.2.2 Remote sensing-based carbon estimation 95
Data exploration 95
Modelling carbon using regression analysis 95

6.3 Results 99
6.3.1 Field-based carbon estimation 99
6.3.2 Remote sensing-based carbon estimation 105
Data exploration 105
Modelling carbon using regression analysis 107

6.4 Discussion 109

Chapter 7 Discussion 112

Implications for natural resource management 113
Summary of recommendations and future research directions 115
Summary of contributions to shelterbelt research 116

References 117

Appendix A QuickBird Raw Images 127

A.1 Image specifications 127

Appendix B Data Collection in the Field 128

B.1 Illustration of Shelterbelt Measurements 128

B.2 Summary of Shelterbelt and Non-Shelterbelt Observations 129

B.3 Summary of Shelterbelt Observations by Species Type 130

B.4 Physical Characteristics of All Shelterbelts 131

B.5 Physical Characteristics of Shelterbelt Groups 132

Appendix C Shelterbelt Delineation 133

C.1 Per-pixel classification classes 133

C.2 Feature extraction settings 133

C.3 Application of Feature Analyst Model to Other Study Areas 134

C.4 Batch Processing Summary 135

Appendix D Data Exploration 136
D.1 Spectral and Textural Information 136
D.2 Image Interpretation 137
D.3 PCA Correlation Matrices 138
Appendix E Species Differentiation 139
E.1 Random Forest 5000th Tree: Species Groups 139
E.2 Random Forest 5000th Tree: Coniferous Species 140
E.3 Importance of Variables for Differentiating Species Groups 141
E.4 Importance of Variables for Differentiating Coniferous Species 142

List of Tables

Table 2.1 Summary of literature describing methods for extracting linear vegetation features. 10
Table 2.2 Summary of feature and image characteristics that affect classification accuracy. 11
Table 2.3 Basic variables used for manual species identification from aerial photographs. 13
Table 2.4 Summary of methods for classifying tree species in forests and shelterbelts. 16
Table 2.5 Summary of feature and image characteristics that affect species classification accuracy. 17
Table 3.1 Spectral and spatial resolutions of the prepared image. 26
Table 3.2 Prepared images for study areas 1, 2 and 3. 26
Table 3.3 Measurements and observations recorded for shelterbelts visited during field data collection. 30
Table 4.1 Summary of tools used to remove clutter from Feature Analyst intermediate results. 38
Table 4.2 Summary of training, intermediate and mask features used for shelterbelt delineation by Feature Analyst. 38
Table 4.3 Equations used to calculate true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) for methods delineating shelterbelts. 41
Table 4.4 Comparison of area-based classification accuracies of methods delineating shelterbelts. 44
Table 4.5 Comparison of random point-based classification accuracies of methods delineating shelterbelts. 46
Table 4.6 Classification accuracy of Feature Analyst shelterbelt delineation by batch processing. 53
Table 4.7 Subjective comparison of methods showing the superiority of Feature Analyst. 62
Table 5.1 Spectral and textural variables extracted for digitized and Feature Analyst shelterbelts. 67
Table 5.2 Summary of shelterbelt species sampled in study areas 1, 2 and 3 by shelterbelt number and length. 72
Table 5.3 Summary of physical characteristics of field-sampled shelterbelts by study area, and for all three study areas combined. 73
Table 5.4 Proportion of variance explained by principal components. 77
Table 5.5 PCA Loadings indicating the importance of variables. 78
Table 5.6 Accuracy of differentiating conifer and broadleaved shelterbelts using Random Forests. 81
Table 5.7 Accuracy of differentiating different species of coniferous shelterbelts using Random Forests. 82
Table 5.8 Overall accuracy of differentiating shelterbelt species type using Random Forests. 83
Table 5.9 Summary of shelterbelt species as predicted by Random Forests in study areas 1, 2 and 3. 83
Table 6.1 Correlation matrix showing correlations between tree biomass (ln AGB) and spectral (A) and physical (B) shelterbelt variables. 98
Table 6.2 Correlation matrix showing correlations between shelterbelt tree density and spectral (A) and physical (B) shelterbelt variables. 98
Table 6.3 Summary of above ground biomass (AGB) per average tree and shelterbelt carbon by species type. 101
Table 6.4 Difference in biomass for pruned and unpruned coniferous shelterbelts. 103
Table 6.5 Difference in heights for pruned and unpruned coniferous shelterbelts. 103

Table 6.6 Summary of delineated shelterbelt species by shelterbelt area (ha), proportion of study areas (%), carbon per area of shelterbelt (t/ha), and total shelterbelt carbon (t) for study areas. 104
Table 6.7 Proportion of variance explained by principal components. 106
Table 6.8 PCA Loadings indicating the importance of variables. 106
Table 6.9 Summary of the linear regression model predicting above ground biomass (AGB) for the average tree in a shelterbelt. 107
Table 6.10 Summary of the linear regression model predicting tree density (AGB) for the average tree in a shelterbelt. 108

List of Figures

Figure 3.1 Map of three study areas in relation to image availability at 1 March 2010. 25
Figure 3.2 Flowchart showing sequence of steps taken in the ArcGIS Image Analysis software to prepare the two satellite images for further analyses. 27
Figure 4.1 Summary of methodologies used for shelterbelt delineation. 34
Figure 4.2 Flowchart showing initial learning and hierarchical learning steps used by the Feature Analyst Learner to delineate shelterbelts. 39
Figure 4.3 Intersect analysis for area-based accuracy assessment of methods delineating shelterbelts. 41
Figure 4.4 Unsupervised per-pixel classification showing 'shelterbelt' and 'shadow' classes. 48
Figure 4.5 Supervised per-pixel classification showing 'shelterbelt' and 'shadow' classes. . 49
Figure 4.6 Object-oriented classification by image segmentation and labelling showing 'shelterbelt' and 'other' classes. 50
Figure 4.7 Feature extraction by Feature Analyst showing only the 'shelterbelt' class. No 'shadow' class is shown because features of interest are extracted by Feature Analyst without classifying the rest of the image. 51
Figure 4.8 Shelterbelts extracted by Feature Analyst batch processing in study area 1. 55
Figure 4.9 Shelterbelts extracted by Feature Analyst batch processing in study area 2. 56
Figure 4.10 Shelterbelts extracted by Feature Analyst batch processing in study area 3. 57
Figure 5.1 Flowchart showing the training of Random Forests algorithms to differentiate species groups (A) and coniferous species (C), and their subsequnt application to predict the species group (B) and coniferous species (D) of Feature Analyst shelterbelts. 69
Figure 5.2 Summary of field-sampled mixed shelterbelts by species proportions and length. 73
Figure 5.3 Shelterbelts sampled during field data collection in study area 1. 74
Figure 5.4 Shelterbelts sampled during field data collection in study area 2. 75
Figure 5.5 Shelterbelts sampled during field data collection in study area 3. 76
Figure 5.6 Biplot of PC1 and PC2 of image-derived variables for *P.radiata* (PCA 1). 79
Figure 5.7 Biplot of PC1 and PC2 of image-derived variables for *C.macrocarpa* (PCA 1). . 79
Figure 5.8 Shelterbelt species types in study area 1 as predicted by Random Forests. 84
Figure 5.9 Shelterbelt species types in study area 2 as predicted by Random Forests. 85
Figure 5.10 Shelterbelt species types in study area 3 as predicted by Random Forests. 86
Figure 6.1 Steps for calculating field-based and remote sensing-based estimates for above ground biomass (AGB) and carbon (C). 94
Figure 6.2 Comparing the above ground biomass (AGB) of the average tree by shelterbelt species. 102
Figure 6.3 Comparing shelterbelt carbon by shelterbelt species. 102
Figure 6.4 Comparing above ground biomass (AGB) and carbon for pruned and unpruned coniferous shelterbelts, as estimated by the field-based method. 104
Figure 6.5 Biplot of PC1 and PC2 of image-derived and physical variables for coniferous shelterbelts (PCA 2). 105

Glossary

Above-ground biomass	The weight of the above-ground portion of the tree, when oven-dried. Usually measured in mass per area, e.g. kg/ha (Jenkins *et al.*, 2003).
Allometry	The relationship between changes in shape and overall size, e.g. the relationship between tree biomass and stem diameter (Gayon, 2000).
Carbon sequestration	The removal of carbon dioxide from the atmosphere through photosynthesis, and the subsequent carbon storage in plant cells (Johnson, 2008).
Delineation	The manual or automated process of outlining features (Campbell, 2007).
Digitization	The manual process of outlining and labelling features using a mouse and cursor (Campbell, 2007).
Fuzzy (soft) classification	Each pixel/object has membership grade values that describe the proportion of each land cover type found within that pixel/object (Jensen, 2005).
Hard classification	Each pixel/object is assigned to a single land cover class (Jensen, 2005).
Multivariate data	A dataset which comprises of observations on different variables for a number of objects or individuals (e.g. observations for different characteristics of shelterbelts within a study area). Multivariate analysis usually involves finding relationships between not only variables, but also the individuals, through calculating averages (Chatfield & Collins, 1980).
NDVI	This is the Normalized Difference Vegetation Index, which takes the difference between near infrared (NI) and red (R) reflectance (Jensen, 2005): NDVI = (NIR – R) / (NIR + R)
Object-oriented classification	Pixels with similar spectral and spatial attributes are grouped into homogenous segments/objects. Objects are then labelled using sets of spectral/spatial criteria for each land cover type (Goetz *et al.*, 2003; Jensen, 2005; Johansen *et al.*, 2007; Miller *et al.*, 2009; Wiseman *et al.*, 2009).
Overall accuracy	The total area correctly classed in all categories divided by the total area in all categories (Congalton, 1991).

Panchromatic band	A band or image in black and white mode, which is sensed at a higher spatial resolution than its corresponding multispectral image (colour bands) (Lillesand & Kiefer, 1994).
Per-pixel classification	Each pixel in an image is assigned to a class according to a certain set of criteria or spectral signatures (ERDAS, 2008).
Producer's accuracy	The area correctly classed in a category divided by the actual area in that category. This is referred to as the producer's accuracy because the producer of the classification is interested in how well a certain category can be classified (Congalton, 1991).
Random Forests	A hierarchical tree classification method. Random Forests (Breiman, 2001) is an iterative learning method which generates many classification trees and aggregates the results (Liaw & Wiener, 2002). Random Forests constructs classification or regression trees for predicting discrete or continuous variables, respectively.
Raster format	When an image is stored in raster format, it means that it is composed of many grid cells, or pixels. The attributes of each grid cell is linked to that specific location. E.g. satellite images are stored in raster format. The value of each pixel is the spectral response of the landcover at that specific location (Schuurman, 2004).
Regression analysis	An investigation of the relationship between one or more *x* (response/dependent) variable(s) and *y* (predictor/ independent) variable(s) (Everitt, 2005; Johnson & Wichern, 2007). A line of best fit can be used to describe this relationship. This study used multiple regression analysis where several *x* variables were used to describe a *y* variable.
Remote sensing	The process of recording from a distance the electromagnetic radiation reflected or emitted from the earth's surface (Campbell, 2007; Cracknell & Hayes, 2007; Jensen, 2005).
Spatial resolution	The smallest area that can be separately recorded as an entity on an image, i.e. pixel (Campbell, 2007).
Spectral resolution	The number and size of specific wavelength intervals captured by a sensor (Jensen, 2005).
Spectral signature	A set of spectral responses across different wavelengths is a spectral signature (Campbell, 2007).

Texture	The frequency of tonal change (smoothness/coarseness) on an image (Lillesand & Kiefer, 1994). It measures the variablity of grey values and the spatial dependence of this variability (Buddenbaum *et al.*, 2005). Standard deviation is an example of a simple measure of texture (Jensen, 2005).
User's Accuracy	The area correctly classed in a category divided by the total area classed in that category. This is referred to as the user's accuracy because the user is interested in the reliability of the classification (Congalton, 1991).
Vector format	When an image is stored in vector format, it means that it is composed of points, lines and polygons. E.g. Shelterbelts can be stored as polygons. Each shelterbelt polygon can have multiple attributes linked to it, such as tree height, tree species and tree density (Schuurman, 2004).
Vegetation Index	Used to indicate relative abundance and activity of green vegetation by sensing cholophyl absorption within a canopy; e.g. NDVI (Jensen, 2005).

List of Equations

EQUATION PARAMETERS (A)

NDVI = Normalized Difference Vegetation Index

NIR = Near infrared band

R = Red band

N = Total number of trees

D = Shelterbelt tree density (trees/m^2)

L = Shelterbelt length (m)

A = Shelterbelt area (m^2)

S_{Ri} = Tree spacing for row *i* (m)

EQUATIONS (A)

(1) $$NDVI = \frac{NIR - R}{NIR + R}$$

(2) $$\text{Tree density (D)} = \frac{N}{A}$$

(3) $$\text{Total number of trees (N)} = \frac{L}{S_{R1}} + \frac{L}{S_{R2}} \ldots + \frac{L}{S_{Ri}}$$

EQUATION PARAMETERS (B)

D = Shelterbelt tree density (trees/km^2)

N = Total number of trees

n_{Ri} = Trees in row *i*

S_{Ri} = Tree spacing for row *i* (km)

A = Shelterbelt area (km^2)

L = Shelterbelt length (km)

W = Shelterbelt width (km)

H_{Ri} = Height for row *i* (m)

DBH_{Ri} = DBH for row *i* (cm)

H = Height for the average tree (m)

DBH = Diameter for the average tree (cm)

AGB = Above ground biomass

C = Above ground carbon

EQUATIONS (B)

(4) $$\text{Area} = W \times L$$

(5) $$\text{Number of trees in row } i \ (n_{Ri}) = \frac{L}{S_{Ri}}$$

(6) $$\text{Total number of trees (N)} = n_{R1} + n_{R1} \ldots + n_{Ri}$$

(7) Average Tree Height (H) = $(H_{R1} \times n_{R1}) + (H_{R2} \times n_{R2}) + ... + (H_{Ri} \times n_{Ri})$

(8) Average tree diameter (DBH) = $(DBH_{R1} \times n_{R1}) + (DBH_{R2} \times n_{R2}) + ... + (DBH_{Ri} \times n_{Ri})$

AGB for Coniferous spp. (kg/tree)

(9) $\ln AGB = -0.9069 + 1.2273 \ln DBH + 0.1411 (\ln DBH)^2 - 0.0078 \ln H + 0.0840 (\ln H)^2$

AGB for *Eucalyptus spp.* (kg/tree)

(10) $\ln AGB = \ln ((1.22 DBH^2 \times H) \times 10^{-4})$

AGB for *Populus spp.* (kg/tree)

11) $\ln AGB = -2.763 + 2.524 (\ln DBH)$

AGB per Row *i* (kg/row)

(12) $AGB_{Ri} = AGB \times n_{Ri}$

AGB per Shelterbelt (t/shelterbelt)

(13) $AGB_{SH} = AGB_{R1} + AGB_{R2} + ... + AGB_{Ri} \div 1000$

AGB per Area of Shelterbelt (t/ha)

(14) $AGB_{SHA} = (AGB_{SH} / A) \div 100$

AGB for the Average Tree in a Shelterbelt (kg/tree)

(15) $AGB_{tree} = (AGB_{SH} / N) \times 1000$

Carbon per Shelterbelt (t/shelterbelt)

(16) $C = AGB_{SH} \times 0.50$

Carbon per Area of Shelterbelt (t/ha)

(17) $C_A = AGB_{SHA} \times 0.50$

Carbon per Length of Shelterbelt (t/km)

(18) $C_L = C / L$

REMOTE SENSING-BASED MODEL FOR PREDICTING CARBON

EQUATION PARAMETERS (C)

D = Shelterbelt tree density (trees/m^2) R = Red median

L = Shelterbelt length (m) W = Shelterbelt width (m)

LW = L : W (m) A = Shelterbelt area (m^2)

AGB = Above ground biomass C = Above ground carbon

X_S = 0.7341 for *P.radiata*

0.6690 for *C.macrocarpa*

0.7090 for *P.radiata* / *C.macrocarpa*

0.6755 for *P.radiata* / other coniferous spp.

0.7432 for other coniferous spp.

0.0000 for a mixture of other coniferous species

EQUATIONS (C)

Density

(19) $\log D = -1.3665 - 0.8179 \log W - 0.19894 \log LW$

AGB for the Average Tree in a Shelterbelt (kg/tree)

(20) $\log (\ln AGB_{tree\ S}) = 1.0226 - 0.0024 R + 0.0982 \log W - 0.0687 D + X_S$

Carbon per Shelterbelt (t/shelterbelt)

(21) $C = AGB_{tree} \times 0.5 \times D \times A \div 1000$

Chapter 1
General Introduction

Linear vegetation features, such as shelterbelts and hedges, are a prominent part of New Zealand's modern agricultural landscape. Prior to human settlement, forest covered most of New Zealand (Cameron, 1964). Destruction of forest in favour of agriculture began during the Maori agricultural revolution, and continued after European settlement (Cameron, 1964). Canterbury was largely a treeless plain by the time of British settlement, and shelterbelts and hedges were subsequently planted under the guidance of the Canterbury Association (Price, 1993).

Linear vegetation patches, such as shelterbelts, have significant ecological and economic value despite their small sizes (Lechner *et al.*, 2009). Shelterbelts have many beneficial functions, from increasing agricultural productivity, to increasing biodiversity and species richness (Gregory, 1995; Ministry of Forestry, 1992; Stringer, 1977): Shelterbelts protect crops, pasture and livestock from adverse weather conditions (Gregory, 1995; Ministry of Forestry, 1992; Stringer, 1977), while adding to the amenity value of agricultural land. Shelterbelts provide habitat for native fauna, such as birds, spiders and other invertebrates ((McLachlan & Wratten, 2003), as well as providing connectivity between habitats (Kristensen & Casperson, 2002). Lastly, shelterbelts provide many ecosystem services (Ministry of Forestry, 1992; Stringer, 1977), such as supplying a sustainable source of timber/firewood, creating favourable conditions for pollination, diminishing soil erosion, reducing water requirements by decreasing evaporation, and finally sequestering carbon. Carbon sequestration is arguably the most important ecosystem service provided by shelterbelts, particularly in view of global warming: shelterbelts remove carbon dioxide from the atmosphere, and therefore mitigate greenhouse gas emissions (De Brauw, 2006). Despite this, not much is known about shelterbelt carbon sequestration.

Accurate mapping of shelterbelts is important for their management as natural resources. However, this can be problematic due to their small areal extent and often fragmented nature (Lechner *et al.*, 2009). The advent of remote sensing and image analysis technologies have made the accurate mapping of these features possible (Cracknell, 1998; Hengl, 2006). Remotely-sensed data refers to data gathered about an object from a distance, such as a satellite image or an aerial photograph (Cracknell & Hayes, 2007). Sensors record electromagnetic radiation reflected or emitted from the earth's surface (Jensen, 2005).

Remotely-sensed imagery not only provides spectral information, but also information about the shape and texture of features (Johansen *et al.*, 2007). Image analysis of remotely-sensed imagery therefore enables both the identification and further classification of linear vegetation features. For example, high correlations have been found between spectral bands and vegetation parameters such as biomass (Lu, 2006).

Remote sensing of the earth from aircraft and satellites is already being applied to a number of environmental science areas. New applications are continually being developed as remote sensing technologies improve and images with higher resolutions become available (Cracknell & Hayes, 2007). One such application is tracking the extent and condition of trees outside of forests, the importance of which is widely recognized (Liknes *et al.*, 2010). Large-scale accounting of agroforestry plantings, such as shelterbelts, could be used as a planning tool for the management of carbon and other ecosystem services (Liknes *et al.*, 2010). For example, accurate mapping of shelterbelts could contribute to carbon accounting efforts in New Zealand. The United Nations Framework Convention on Climate Change requires New Zealand (and other signatories) to submit an annual greenhouse gas inventory which covers all human-induced greenhouse gas emissions and removals (Ministry for the Environment, 2009). The Land Use and Carbon Analysis System (LUCAS) was established in 2005 to measure carbon removal and emissions from land-use change and forestry (Ministry for the Environment). Shelterbelts do not have their own land-use category, but major shelterbelts are included in the grassland category (Ministry for the Environment, 2009). The shelterbelt carbon stock therefore remains unquantified as a separate entity.

Tracking the extent of trees outside of forest is also important for ecological applications. For example, the effectiveness of shelterbelts as wildlife corridors could be assessed in terms of shelterbelt attributes, such as width and length (Lechner *et al.*, 2009). Reliable mapping of linear features will also improve the detail and accuracy of rural and urban area habitat maps. Freeman and Buck (2003) identified the mapping of linear features as a problem area when creating an ecological map of the city of Dunedin.

A method for accurately mapping and characterizing shelterbelts is therefore needed to fully understand and utilize the many benefits of shelterbelts in Canterbury and New Zealand. This would facilitate the recognition and management of shelterbelts as a valuable natural resource and provider of ecosystem services.

1.1 Aims and Objectives

The overall aims of this research are (i) to explore and test methods for delineating and classifying shelterbelt characteristics using satellite imagery, and (ii) to explore the application of results for modelling shelterbelt carbon across a portion of the Canterbury Plains. The thesis is comprised of the following chapters:

Chapter 2 (Literature Review) provides an overview of remote sensing-based methods for delineating shelterbelts, differentiating tree species, and estimating carbon sequestration.

Chapter 3 (General Methods) describes the study area and satellite imagery used for shelterbelt delineation and analyses. It also outlines image preparation steps and field data collection.

Chapter 4 (Image-based Shelterbelt Delineation) explores methods for delineating shelterbelts from high-resolution multispectral imagery. This was achieved by (i) comparing per-pixel and object-oriented classification methods, including a method utilizing specialist feature extraction software; and (ii) investigating whether batch processing affects classification accuracy.

Chapter 5 (Differentiation of Shelterbelt Species) investigates whether shelterbelt species can be differentiated using remotely-sensed spectral information. This was achieved by (i) exploring which spectral variables explain shelterbelt variability; and (ii) using a Random Forests classification method to differentiate shelterbelt species using image-derived variables.

Chapter 6 (Example Application: Modelling Shelterbelt Carbon) provides an example application of the methods presented in Chapters 4 and 5: shelterbelt carbon sequestration is modelled using remotely-sensed data. This was achieved by (i) investigating which spectral variables are likely predictors of shelterbelt physical characteristics; (ii) estimating shelterbelt carbon sequestration using a field-based method; and (iii) using regression analysis to model the relationship between shelterbelt carbon and shelterbelt physical and spectral characteristics.

Chapter 7 (Discussion) provides a general discussion on the success of characterizing shelterbelts using satellite image analysis, as well as potential applications. It discusses the overall contributions of this thesis to current knowledge and makes recommendations for future research.

Chapter 2
Literature Review

2.1 Introduction

This chapter will provide an overview of previous research studies investigating shelterbelt delineation, tree species differentiation and above-ground biomass (carbon) estimation.

2.2 Shelterbelt Definition

A shelterbelt is defined as a band of single or multi-rowed trees that acts as a windbreak for protecting livestock or crops (Price, 1993). This is one of five types of woodland shelter for protecting livestock, all of which fall under the term agroforestry (Gregory, 1995). Shelterbelts in Canterbury are dominated by two Californian tree species, Monterey pine (*Pinus radiata*) and Monterey cypress (*Cupressus macrocarpa*) (Price, 1993). In contrast, a shadebelt is a single row of deciduous trees which provide shade. In New Zealand shadebelts generally consist of an east-west line of poplar trees that give shade to the south (Gregory, 1995). On the contrary, other forms of woodland shelter are non-linear: woodlot blocks provide shelter from adverse weather directly under their canopy (Gregory, 1995); forest grazings are larger than woodlot blocks, and are used for over-wintering livestock (Gregory, 1995); whereas trees-on-pasture are areas of established pasture overplanted with pine trees (Gregory, 1995).

Objective criteria for what defines a shelterbelt vary between studies. For example, David and Rhyner (1999) defined windbreaks as having an association with agricultural land, a minimum length of 180 m, a maximum width of 18 m, and a maximum gap length of 18 m within a given windbreak. In contrast, Kristensen and Casperson (2002) defined shelterbelts as having a minimum length of 10 m, and a maximum gap length of 10 m within a shelterbelt.

Both shelterbelts and shadebelts will be considered as shelterbelts in this research, as both are linear belts of trees that provide shelter. Hedgerows will also be considered in this chapter. A hedgerow can be thought of as the European equivalent of a shelterbelt, and is defined as a row of bushes or low trees growing close together which have been managed to maintain a linear barrier (Baudry *et al.*, 2000). In contrast, hedges in New Zealand are narrow bands of low dense shrubs that separate fields, and will therefore be excluded. Hedges in Canterbury consist predominantly of gorse, an English hedge-plant (Price, 1993).

2.3 Remote sensing principles

The term "remote sensing", in its broadest sense, means the gathering of information about an object or the environment from a distance (Campbell, 2007; Cracknell & Hayes, 2007; Jensen, 2005). More specifically, it refers to the measuring of radiation in one or more regions of the electromagnetic spectrum, reflected or emitted from the earth's surface (Campbell, 2007). Remote sensing is both a science and an art. As a science, it uses mathematical and statistical algorithms to extract valuable information from data, as well as involving cartography and Geographic Information Systems (GIS). As an art, it involves visual photo or image interpretation (Jensen, 2005).

Electromagnetic radiation is measured using passive or active sensors which are satellite- or airborne (Cracknell & Hayes, 2007). Passive sensors record energy reflected or emitted by the earth at particular wavelengths. Reflected energy refers mainly to electromagnetic radiation in the visible and near-infrared spectrums; whereas emitted energy refers mainly to the far infrared spectrum which reveals thermal information about the earth. In contrast, active sensors generate their own energy and record radiation reflected back from the earth (Campbell, 2007; Cracknell & Hayes, 2007).

Passive sensors record images in either photographic or digital form (Campbell, 2007; Jensen, 2005). Photographic sensors record radiation in the visible, near-infrared or ultraviolet spectrums by refraction of light. Aerial photography is the most practical, inexpensive and widely used method of remote sensing (Campbell, 2007). In contrast, digital sensors record the patterns of image brightness in one or more spectral channels as an array of numbers in digital mode. This makes digital images easier to process and analyze than photographic images. Digital images consist of pixels, which each represent the brightness of a small region on the earth's surface (Campbell, 2007).

Remote sensing images can be analyzed using visual interpretation or computer-based classification techniques. Image analysis techniques exploit spectral and spatial information (Campbell, 2007; Richards & Jia, 2006). Spectral information refers to the spectral response (colour) of pixels/features. A set of spectral responses across different wavelengths is a spectral signature (Campbell, 2007). Image algebra can also be used for spectral-based image analysis, such as creating band ratios and vegetation indices. The latter is used to monitor vegetation health and productivity, by sensing the chlorophyll absorption within a canopy (Jensen, 2005). The Normalized Difference Vegetation Index (NDVI) (developed by Rouse *et*

al. (1974)) is most commonly used (Jensen, 2005), and takes the difference between near infrared (NIR) and red (R) reflectance:

$$NDVI = \frac{NIR - R}{NIR + R} \quad (1)$$

In contrast, spatial information refers to the shape, size, orientation and context of earth features captured by an image (Richards & Jia, 2006). The latter is the spatial relationship of each pixel to its neighbours (Jensen, 2005). An example of a context measure is texture, which is the frequency of tonal change (i.e. smoothness/coarseness) within a feature (Jensen, 2005; Lillesand & Kiefer, 1994).

The detail of spectral and spatial information depends on the spectral and spatial resolutions of the image. Spectral resolution refers to the number and size of specific wavelength intervals known as bands or channels (Jensen, 2005). For example, multispectral images usually have three to four bands, plus an extra band in black and white (panchromatic) mode. In contrast, spatial resolution refers to the smallest area that can be separately recorded as an entity on an image, i.e. the size of a pixel for a digital image (Campbell, 2007).

Optical satellite images commonly used in New Zealand are moderate spatial resolution images from Landsat (30 m) and SPOT (10 m) satellites, and high-resolution images from the SPOT-5 (5 m) satellite (Landcare Research). High-resolution QuickBird II imagery (DigitalGlobe Corporate) is currently becoming freely available in New Zealand through the KiwImage Project (NZ Defence Force Geospatial Intelligence Oganisation), and will be used in this research. It is an example of high-resolution multispectral imagery with bands in the visible and near infrared spectrum. The imagery is recorded by a satellite-borne passive sensor which records reflected radiation in digital format. QuickBird II red, green, blue and near infrared bands have a spatial resolution of 2.4 m, while the panchromatic band has a spatial resolution of 0.6 m.

2.4 Linear feature extraction from imagery

There is a vast amount of literature concerning the mapping of linear land cover features, such roads, railways, rivers and shelterbelts (Quackenbush, 2004). The majority of techniques, however, focus on the extraction of road and rail networks (Thornton *et al.*, 2007). Linear vegetation features, such as shelterbelts, have been extracted from a wide variety of images sources, including land cover maps (Kristensen & Casperson, 2002), aerial photographs (David & Rhyner, 1999; Kristensen & Casperson, 2002), and digital imagery (Aksoy *et al.*, 2010; Lennon *et al.*, 2000; Liknes *et al.*, 2010; Pankiw *et al.*, 2009; Tansey *et al.*, 2009; Thornton *et al.*, 2006, 2007; Vannier & Hubert-Moy, 2008; Wiseman *et al.*, 2009).

The availability of very high-resolution imagery has made the accurate mapping of linear vegetation features possible (Cracknell, 1998; Hengl, 2006; Quackenbush, 2004). High, moderate and coarse resolutions refer to spatial resolutions of less than 10 m, 10 to 100 m, and greater than 100 m, respectively (Lu, 2006). Prior to the advent of high-resolution imagery, mapping of linear vegetation features produced limited results because at least two pixels are required to represent the width of narrow objects (Hengl, 2006). Moderate-resolution imagery, for example, had a resolution that was too coarse to differentiate narrow bands of vegetation (Johansen *et al.*, 2007). Moderate to course resolution imagery has pixels larger than most features of interest, such that more than one land cover type may be present within each pixel. This results in a noisy spectral response and a poor classification result (Johansen *et al.*, 2007). In contrast, high-resolution imagery enables the differentiation of narrow features by providing more detailed spectral and spatial information (Johansen *et al.*, 2007).

Linear vegetation feature extraction techniques range from manual to automated methods (Table 2.1). Manual extraction is the most basic feature extraction method, using human image interpretation and digitizing techniques (David & Rhyner, 1999), where the user outlines and labels features using a mouse and cursor (Campbell, 2007). Manual extraction has the advantage of utilizing the interpretation skills of the user, but can be time-consuming and therefore expensive (Quackenbush, 2004). Its success is also limited by the spatial resolution of the imagery, with insufficient resolutions leading to inaccuracies in length and position (David & Rhyner, 1999; Kristensen & Casperson, 2002). In addition to this, manual extraction techniques are no longer sufficient to cope with increases in the amount of available data (Quackenbush, 2004). Automated or semi-automated extraction techniques are therefore needed. Semi-automated methods are optimal, because there is a compromise between the speed and accuracy of a computer algorithm, and the interpretation skills of the

user (Gruen & Li, 1997). Semi-automated methods therefore produce more reliable results (Baumgartner *et al.*, 1999; Gruen & Li, 1997).

The majority of linear feature extraction techniques reported in the literature focus on the extraction of roads (Thornton *et al.*, 2007). There are four main reasons why techniques used to extract roads tend to be unsuitable for the extraction of linear vegetation features (Aksoy *et al.*, 2010): (i) Roads have paired edges as boundaries, which are collinear parallel line segments. In contrast, linear vegetation features often produce lots of small, irregular line segments within and along feature boundaries (Aksoy *et al.*, 2010); (ii) Roads tend to be greater than 10 m in width, whereas linear vegetation features are often less than 5 m in width (e.g. hedgerows) (Thornton *et al.*, 2007); (iii) Linear vegetation features tend to have more directional variation than roads along their length, especially where they follow natural or man-made boundaries, such as streams or minor roads (Thornton *et al.*, 2007); and (iv) Linear vegetation features may have additional irregularities such as breakpoints, localized changes in size, and embedded land cover objects (such as a large oak tree in a pruned pine shelterbelt) (Thornton *et al.*, 2007).

Techniques which have been used to extract linear vegetation features, such as shelterbelts and hedgerows, include pixel swapping (a type of per-pixel classification) (Thornton *et al.*, 2006, 2007), object-oriented classification (Aksoy *et al.*, 2010; Lennon *et al.*, 2000; Liknes *et al.*, 2010; Pankiw *et al.*, 2009; Tansey *et al.*, 2009; Vannier & Hubert-Moy, 2008; Wiseman *et al.*, 2009), and mathematical morphology (Aksoy *et al.*, 2010; Lennon *et al.*, 2000; Tansey *et al.*, 2009; Thornton *et al.*, 2006, 2007; Vannier & Hubert-Moy, 2008; Wiseman *et al.*, 2009). The majority of published studies present methods for extracting hedgerows (Table 2.1).

Until recently, linear features on images were most often classified using per-pixel classification techniques, such as pixel swapping. Per-pixel classification methods work by assigning each pixel to a land cover class according to a set of spectral signatures (Blaschke *et al.*, 2000; ERDAS, 2008; Jensen, 2005). Pixel swapping is a type of fuzzy classification, which means that each pixel is assigned to multiple land cover classes, and then each class is graded according to the proportion of that class in the pixel (Jensen, 2005). Pixel swapping involves creating sub-pixels, assigning each sub-pixel to a single class using a distance-weighted function of neighbouring pixels, and swapping sub-pixels to increase spatial correlation of the sub-pixels (Thornton *et al.*, 2007). Sub-pixel mapping is an alternative to object-oriented mapping techniques (discussed below), and has the advantage that features with a smaller width than the pixel size can be mapped (Thornton *et al.*, 2006). Pixel swapping algorithms can also be specially adapted for extracting linear features. Thornton *et*

al. (2007), for instance, found that an anisotropic pixel swapping algorithm extracted up to 5 % more hedgerows than a non-anisotropic algorithm. Anisotropic pixel swapping extracted 94 % of all hedgerows on average.

However, per-pixel classification is usually not adequate for extracting linear features, as only spectral information is utilized (Quackenbush, 2004). In contrast, object-oriented classification utilizes both spectral and spatial information from high-resolution images (Goetz *et al.*, 2003; Johansen *et al.*, 2007). This often involves an image segmentation step prior to classification, which involves grouping pixels into relatively spectrally homogenous objects using complex algorithms (Miller, 2009; Wiseman *et al.*, 2009). Object-oriented classification improves classification accuracy when mapping linear vegetation features (Johansen *et al.*, 2007).

Object-oriented classification is therefore the most commonly reported method for extracting shelterbelts and hedgerows (Table 2.1). The majority of studies used Definiens eCognition, a specialist feature extraction software (Liknes *et al.*, 2010; Pankiw *et al.*, 2009; Tansey *et al.*, 2009; Vannier & Hubert-Moy, 2008; Wiseman *et al.*, 2009). Definiens eCognition mimics the process of human perception (Tansey *et al.*, 2009), whereby objects are first defined by colour, shape, size, texture, patterns and context criteria, and then classified using a fuzzy classification approach (Tansey *et al.*, 2009). These studies reported high classification accuracies, extracting between 80 and 96 % of all shelterbelts. Wiseman *et al.* (2009) had the highest classification accuracy, with 96 % of all shelterbelts identified using eCognition. Object attributes relating to shape and area were found to be important predictors of shelterbelts. This differs from Liknes *et al.* (2010) who found that texture was the most important predictor. This difference may be because Liknes *et al.* (2010) extracted shelterbelts as part of a general class for agricultural tree cover.

Mathematical morphology, the third technique for extracting linear vegetation features, is only reported as having been used in combination with pixel swapping (Thornton *et al.*, 2006) or object-oriented classification (Aksoy *et al.*, 2010; Lennon *et al.*, 2000). Mathematical morphology is used to analyse the shape and form of objects. It is based on logical relationships between pixels. An image is considered as a set of points, and a suitable structuring shape is used as a probe (Shih, 2009). For example, Aksoy *et al.* (2010) used a morphological filter to eliminate features that were too wide or too narrow to be hedgerows.

Classification accuracy not only depends on the method used, but also on the quality of the imagery, and feature characteristics (Table 2.2). Vannier and Hubert-Moy (2008) confirmed

that extraction from moderate resolution images produces poor results, whereas extraction from high-resolution images with rich spectral information produces the best results. Thornton *et al.* (2006) found that the width of hedgerows affect classification accuracy. For example, 94 % of 20-23 m wide hedgerows were extracted compared to only 47 % of 1.8-3 m wide hedgerows. Vannier and Hubert-Moy (2008) found that classification accuracy increased with hedgerow density; and Pankiw *et al.* (2009) found that classification accuracy was good for field and farmyard shelterbelts, whereas roadside shelterbelts had an unacceptable number of false positives.

These studies show that semi-automated extraction of linear vegetation features can produce good results, extracting between 63 % and 96 % of all shelterbelts (Table 2.1).

Table 2.1 Summary of literature describing methods for extracting linear vegetation features.

Method	Study	Study Area	Feature	Scale or Spatial Resolution	Accuracy
Manual extraction	**David & Rhyner (1999)**	USA	Shelterbelts	1 : 20,000	Positional inaccuracy of about 10 m
	Kristensen & Casperson (2002)	Europe	Shelterbelts	1 : 20,000 1 : 25,000 1: 10,000	Positional inaccuracy of about 35 m
Pixel swapping & mathematical morphology	**Thornton *et al.* (2006)**	Europe	Hedgerows	2.6 m	63 % of all hedgerows
Pixel swapping (linearised)	**Thornton *et al.* (2006, 2007)**	Europe	Hedgerows	2.5 m 5 m	94 % of all hedgerows
Object-oriented classification & mathematical morphology	**Lennon *et al.* (2000)**	Europe	Hedgerows	2 m	Not reported
	Aksoy *et al.* (2010)	Europe	Hedgerows	0.6 m	59 % True + 35 % False +
Object-oriented classification with eCognition	**Vannier & Hubert-Moy (2008)**	Europe	Hedgerows	5 m	87 % (overall)
	Tansey *et al.* (2009)	Europe	Hedgerows	0.25 m	Not reported
	Wisemen *et al.* (2009)	Canada	Shelterbelts	0.625 m	96 % of all shelterbelts
	Pankiw *et al.* (2009)	Canada	Shelterbelts	2.5 m	80 % of all shelterbelts
	Liknes *et al.* (2010)	USA	Agricultural treecover	1 m	85 % (overall)

Table 2.2 Summary of feature and image characteristics that affect classification accuracy.

	Characteristic		Effect on Accuracy	Reference
Shelterbelt	**Width**	Narrow	–	Thornton *et al.* (2006)
		Wide	+	
	Density	Sparse	–	Vannier & Hubert-Moy (2008)
		Dense	+	
	Type	Roadside	–	Pankiw *et al.* (2009)
Image	**Spatial Resolution**	Moderate	–	Vannier & Hubert-Moy (2008)
		High	+	
	Spectral Richness	Low	–	Vannier & Hubert-Moy (2008)
		High	+	

2.5 Differentiation of tree species using image analysis

Once shelterbelts are identified, it is highly useful to know which species comprise different shelterbelts. For example, differentiating shelterbelt tree species is important for estimating shelterbelt biomass (Wiseman *et al.*, 2009), and therefore modelling shelterbelt carbon sequestration (see section 2.6).

Tree species have been identified from a variety of image types (Table 2.4), including aerial photographs (Erikson, 2004; Lillesand & Kiefer, 1994), passive digital imagery (Brandtberg, 2002; Buddenbaum *et al.*, 2005; Hájek, 2008; Katoh *et al.*, 2009; Leckie *et al.*, 2003; Olofsson *et al.*, 2006; Wiseman *et al.*, 2009), active digital imagery (Holmgren & Persson, 2004; Orka *et al.*, 2009; Suratno *et al.*, 2009), and a mixture of passive and active digital imagery (Holmgren *et al.*, 2008; Ke *et al.*, 2010; Waser *et al.*, 2011). A variety of techniques have been used, ranging from manual interpretation (Lillesand & Kiefer, 1994), to automated classification (Brandtberg, 2002; Buddenbaum *et al.*, 2005; Erikson, 2004; Förster & Kleinschmit, 2006; Hájek, 2008; Holmgren & Persson, 2004; Holmgren *et al.*, 2008; Katoh *et al.*, 2009; Key *et al.*, 2001; Leckie *et al.*, 2003; Olofsson *et al.*, 2006; Orka *et al.*, 2009; Suratno *et al.*, 2009; Waser *et al.*, 2011).

The most basic method for identifying tree species from images is manual interpretation of aerial photographs. Several variables relating to tree crown characteristics and context can be used (Table 2.3) (Lillesand & Kiefer, 1994). However, interpretation accuracy, consistency, timeliness, and the volume of data to be processed all limit the practicality of such an approach. For example, accuracies for 1:10 000 to 1:20 000 scale photographs are generally 70-85% for the main species in a stand, but can be lower (Hay *et al.*, 2005).

Image analysis efficiency and classification can be dramatically improved by using more sophisticated image analysis techniques which compromise between the speed and accuracy of a computer algorithm and the interpretation skills of the user (Gruen & Li, 1997). Further spectral and spatial variables, such as additional spectral channels (e.g. band ratios), image transforms (e.g. vegetation indices), and texture transformations can then be utilized (Förster & Kleinschmit, 2006; Hájek, 2008; Jensen, 2005). However, this, too, may not be straightforward; spectral and spatial variables of two species may not be significantly different from each other, as variables often have broad ranges and spectral signatures may overlap (Förster & Kleinschmit, 2006; Hájek, 2008).

Table 2.3 Basic variables used for manual species identification from aerial photographs.

Shape --------	Of the tree crown;
Size ----------	Of the tree crown;
Pattern -------	The spatial arrangement of tree crowns;
Tone (hue) --	The relative brightness of different stands/crowns;
Texture ------	The frequency of tonal change (i.e. smoothness/coarseness) inside the tree crown;
Shadows ----	This is useful if the profile of the tree is shown;
Site ----------	Different tree species are more likely to be found at certain geographic or topographic locations.

There is a substantial body of literature on tree species identification from remotely-sensed imagery. Most studies to date focus on classifying individual tree crowns into species, so that forest composition can be estimated (Table 2.4). This is important in forest management, for purposes such as estimating timber volume (Hájek, 2008). This review will focus on studies from the last decade, but will not consider the various methods for extracting tree crown or tree stand objects. Instead, the focus will be on how those objects were classed into species-type after extraction.

Automated methods for identifying tree species range from per-pixel classification (Buddenbaum *et al.*, 2005; Key *et al.*, 2001) to object-oriented classification (Table 2.4). The majority of previous studies used an object-oriented approach, with methods ranging from fuzzy (Brandtberg, 2002; Förster & Kleinschmit, 2006; Hájek, 2008) to hard classification techniques (Erikson, 2004; Holmgren & Persson, 2004; Holmgren *et al.*, 2008; Katoh *et al.*, 2009; Leckie *et al.*, 2003; Olofsson *et al.*, 2006; Orka *et al.*, 2009; Suratno *et al.*, 2009; Waser *et al.*, 2011). Fuzzy classification techniques allow each tree crown object to have degrees of membership in multiple sets, where a set is a species type (Brandtberg, 2002). In contrast, hard classification techniques allow each tree crown object to belong to only one class (Jensen, 2005). Overall, fuzzy classification techniques produce poorer results with overall accuracies ranging between 67 and 75 % for the reviewed literature; compared to 68 to 96 % for hard classification techniques (Table 2.4).

Tree crown objects have been classified into species type using various image-derived spectral and spatial variables with or without ancillary data. Spectral and spatial variables range from spectral means (Brandtberg, 2002; Olofsson *et al.*, 2006), to vegetation indices and band ratios (Hájek, 2008), to tree crown area (Brandtberg, 2002), to textural information (Hájek, 2008). Ancillary data refers to data acquired by means other than remote sensing (Campbell, 2007), such as soil, silvicultural and digital contour maps (Förster & Kleinschmit, 2006; Hájek, 2008), as well as digital terrain models (DTMs) (Förster & Kleinschmit, 2006). Classification methods using a combination of spectral, spatial and ancillary information are

more accurate than methods using only spectral information (Table 2.5) (Buddenbaum *et al.*, 2005; Förster & Kleinschmit, 2006).

Tree species classification accuracy not only depends on the variables used, but also on image and feature characteristics (Table 2.5). Image factors such as spatial resolution and the timing of capture affect the capacity for species differentiation (Erikson, 2004; Holmgren *et al.*, 2008; Key *et al.*, 2001). The latter is due to seasonal changes in foliage (Key *et al.*, 2001; Lillesand & Kiefer, 1994). Images captured in autumn, as well as multitemporal imagery, are the best for differentiating species (Holmgren *et al.*, 2008; Key *et al.*, 2001). Feature characteristics, such as variability in tree age and tree density, also affect classification accuracy (Förster & Kleinschmit, 2006; Holmgren & Persson, 2004). This is because spectral and spatial signatures are affected by tree age and the distance between tree crowns (Förster & Kleinschmit, 2006; Holmgren & Persson, 2004). Lastly, classification accuracy is affected by species abundance (Katoh *et al.*, 2009; Waser *et al.*, 2011), and whether or not tree crowns are classified at an individual, stand or plot level (Suratno *et al.*, 2009). This is because spectral and spatial signatures may not be representative for rare species, and stands/plots may contain a mixture of tree species (Katoh *et al.*, 2009; Waser *et al.*, 2011).

Tree species classification can also be improved by using imagery acquired from active sensors (Table 2.5). For example, airborne laser scanning (ALS) has recently become a key technology for creating and updating forest inventories by providing important information about canopy structure (Suratno *et al.*, 2009). ALS-derived information include digital surface (DSMs) and digital elevation models (DEMs) (Ke *et al.*, 2010), as well as variables relating to height, canopy shape, crown reflectance properties, and stem density (Holmgren *et al.*, 2008; Suratno *et al.*, 2009). A number of studies have shown that tree species can be identified from ALS data, which is commonly referred to as LIDAR (Light Detection and Ranging) data (Table 2.4) (Holmgren & Persson, 2004; Holmgren *et al.*, 2008; Ke *et al.*, 2010; Orka *et al.*, 2009; Suratno *et al.*, 2009; Waser *et al.*, 2011). Combining LIDAR data with spectral information from satellite imagery can increase species classification accuracy by up to 5 % (Holmgren *et al.*, 2008; Ke *et al.*, 2010). Aerial imagery may be more efficient at differentiating coniferous from deciduous trees, while LIDAR data may be more efficient at differentiating coniferous species (Holmgren *et al.*, 2008). This may be because ALS-derived crown information can be used to differentiate species with similar spectral signatures (Holmgren & Persson, 2004) using additional information about crown structure (Orka *et al.*, 2009). ALS is also particularly valuable for eliminating shadow (Ke *et al.*, 2010).

Unfortunately, using LIDAR data is not always a feasible option, as LIDAR data is expensive to acquire.

On the whole, classification into general coniferous and broadleaved species groups has been met with better success than classification into specific species. For example, Brandtberg (2002) found that further classification of species groups into specific species types decreased accuracy by up to 20 %; Erikson (2004) found that accuracy was decreased by 14 % (Table 2.5). Classification accuracy also varies depending on which tree species are being classified. For example, Olofsson *et al.* (2006) found that classification accuracies for different species ranged between 77 and 98 %; whereas Katoh *et al.* (2009) found that classification accuracies varied from 35 % for rare coniferous species to 97 % for the broadleaved species group. Unfortunately most work has been done on North American and European tree species. No work has thus far been done on differentiating *P.radiata* and *C.macrocarpa*, the major shelterbelt species of Canterbury.

Species classification of shelterbelt objects is rare in the literature. It is reasonable to suggest that classification of shelterbelts rather than tree crowns may produce less accurate results. For instance, useful information relating to individual crown shape and structure may be reduced given that shelterbelts may have high tree densities, are often pruned, and may be a mixture of species or age classes (Table 2.5). For example, Suratno *et al.* (2009) found that classifying tree species at plot level as opposed to tree crown level decreased accuracy by up to 27 %. One study investigating the classification of shelterbelts into species type (Wiseman *et al.*, 2009) used Principal Component Analysis (PCA) and Multiple Discriminate Analysis (MDA) to determine if shelterbelt objects could be classified into species type using spectral and spatial characteristics derived from multispectral imagery. Mean red reflectance, object density, object length and width were most important for differentiating species. However, Wiseman *et al.* (2009) did not apply this classification to the shelterbelt objects. This method therefore remains untested.

Table 2.4 Summary of methods for classifying tree species in forests and shelterbelts.

		Study Area	Data	Species Classes		Spatial Resolution	Overall Accuracy
				Conifer	Broadleaf		
Tree Crown: Fuzzy Object-based	**Brandtberg (2002)**	Sweden	Aerial imagery	All Norway spruce Scots pine	All Birch Aspen	0.10 m	87 % 67 %
	Förster & Kleinschmit (2006)	Germany	QuickBird, DTM, soil & forestry maps.	Norway Spruce Larch	Beech Black alder Sycamore	0.6 m	77 %
	Hájek (2008)	Czech Republic	Aerial imagery	Young Mature	Young Mature	0.5 m	75 %
Tree Crown: Hard Object-Based	**Leckie *et al.* (2003)**	Canada	Aerial imagery	Douglas fir Grand fir Amabilis fir Western red cedar Western hemlock	All	0.6 m	93 %
	Erikson (2004)	Sweden	Aerial photography (converted to digital format)	Norway spruce Scots pine All	Birch Aspen All	0.03 m 0.10 m 0.03 m	77 % 71 % 91 %
	Holmgren & Persson (2004)	Sweden	LIDAR	Norway spruce Scots pine		N/A	95 %
	Olofsson *et al.* (2006)	Sweden	Aerial imagery	Scots pine Norway spruce	All	0.29 m	89 %
	Holmgren *et al.* (2008)	Sweden	Aerial imagery & LIDAR	Norway spruce Scots pine	All	0.1 m	96 %
	Katoh *et al.* (2009)	Japan	Aerial imagery	Japanese cypress Japanese red pine Japanese larch Japanese cedar Dawn redwood	All	0.5 m	78 %
	Orka *et al.* (2009)	Norway	LIDAR	Norway spruce	Birch	N/A	88 %
	Suratno *et al.* (2009)	USA	LIDAR	Douglas fir Ponderosa pine Lodgepole pine Western larch		N/A	68 %
	Waser *et al.* (2011)	Switzer-land	Aerial imagery & LIDAR	White fir Norway spruce	Alder Ash Beech Birch Maple	0.25 m 0.5 m	76 %
Per-Pixel	**Key *et al.* (2001)**	USA	Multi-temporal aerial imagery		Yellow poplar White oak Red oak Red maple	0.06 m	74 %
	Buddenbaum *et al.* (2005)	Germany	Hyperspectral data	Norway spruce Douglas fir		5 m	78 %
Stand: Object-based	**Hájek (2006)**	Czech Republic	IKONOS & DEM	Spruce Larch	Oak Birch Maple	4 m 1 m	83 % 83 %
	Wiseman *et al.* (2009)	Canada	Aerial imagery		Acute willow American elm Caragana Green ash Mixed x 2	0.625 m	Not reported
	Ke *et al.* (2010)	USA	QuickBird & LIDAR	Norway spruce Pine Hemlock Larch	Deciduous	0.6 m	94 %

Table 2.5 Summary of feature and image characteristics that affect species classification accuracy.

	Characteristic		Effect on Accuracy	Reference
Tree	**Abundance of species**	Rare (small sample) or in mixed stands	–	Katoh *et al.* (2009) Waser *et al.* (2011)
	Age (e.g. crown structure changes with age)	High variability	–	Homgren & Persson (2004)
		Same age class	+	
	Level of classification	Tree level	+	Suratno *et al.* (2009)
		Plot level (dominant species)	–	
Data	**Timing of capture**	Autumn	+	Holmgren *et al.* (2008) Key *et al.* (2001)
	Multitemporal data	Increased # of dates	+	Key *et al.* (2001)
	Type of sensor	Pasive (Aerial imagery)	+	Holmgren *et al.* (2008) Ke *et al.* (2010)
		Active (LIDAR)	–	
		Aerial imagery + LIDAR	++	
	Spatial resolution	Moderate	–	Erikson (2004)
		High	+	
	Spectral richness	Increased # of bands	+	Ke *et al.* (2010)
	Ancillary data	Imagery only	–	Förster & Kleinschmit (2006)
		Imagery + ancillary data	+	
	Image-derived variables	Spectral only	–	Buddenbaum *et al.* (2005)
		Spectral + stem density	+	
		Spectral + textural	++	
		Spectral + stem density + textural	++	

2.6 Estimation of shelterbelt carbon quantities

Carbon content can be estimated from the above-ground biomass of a tree or a shelterbelt (Carswell *et al.*, 2009; Coomes *et al.*, 2002; Kort & Turnock, 1999)). Tree carbon content ranges from 45 to 55 % of tree biomass depending on species, age and surrounding vegetation (Thenkabail *et al.*, 2004). A reasonable assumption is therefore that 50 % of tree biomass consists of carbon (Carswell *et al.*, 2009; Coomes *et al.*, 2002; Kort & Turnock, 1999) (Thenkabail *et al.*, 2004). Traditionally, biomass estimation has relied on field data collection. In recent times however, remotely-sensed data has been shown useful for estimating biomass (Kalaitzidis & Zianis, 2009).

Shelterbelt species information and age are critical factors in modelling of shelterbelt biomass (Wiseman *et al.*, 2009). Spectral information may be useful not only for telling species apart, but also for estimating shelterbelt age (Wiseman *et al.*, 2009). For example, old shelterbelts appear heterogeneous while young shelterbelts appear homogenous. This is because old shelterbelts have complex crown structures with shadow intermixed, while young shelterbelts have simple crown structures with much less shadow intermixed (Wiseman *et al.*, 2009). In theory, carbon sequestration could therefore be modelled using image-derived variables. In fact, remote sensing techniques for estimating above-ground biomass has become prevalent in recent years, with most work being done on coniferous forests (Lu, 2006).

There are three different approaches for estimating above-ground biomass: field measurement-based, GIS-based and remote sensing-based methods (Lu, 2006). Field-based methods are based on direct allometric relationships between a tree's radial growth and its biomass accumulation. Tree diameter at breast height (DBH) and tree height are both good predictors of biomass, and are therefore the most commonly used dimensions (Jenkins *et al.*, 2004). Field-based methods involve either destructive or dimensional analysis. Destructive analysis involves analyzing wood samples from representative trees to determine the relationship between a given tree parameter and tree biomass. For example, (Kort & Turnock, 1999) estimated carbon content by chopping down and weighing trees. In contrast, dimensional analysis involves calculating carbon storage from allometric equations already found in the literature (Brandle *et al.*, 1992). Dimensional analysis is most often used (Jenkins (Jenkins *et al.*, 2004), as destructive analysis is costly and impractical in most circumstances.

Field-based estimations provide the most accurate way for estimating biomass (Lu, 2006), because site- and species-specific models are developed (Jenkins *et al.*, 2004; Kort & Turnock, 1999). However field-based methods are time consuming and labour intensive, often

unpractical to implement, and unsuitable for analysis across a landscape (Lu, 2006; Zheng *et al.*, 2004). GIS- and remote-sensing based methods, on the other hand can be used for analysis across a landscape. These methods rely on field-based biomass estimates (which may or may not be site specific) as a basis for modelling the relationship between biomass and environmental or remotely-sensed data.

GIS-based methods model relationships between biomass and environmental data such as elevation, slope, soil and climate data (Lu, 2006; Magcale-Macandog *et al.*, 2006). For example, Magcale-Macandog *et al.* (2006) determined a relationship between above-ground forest biomass and environmental factors using regression analysis. However, GIS-based methods have not been used extensively, as relationships between above-ground biomass and environmental data are often weak, and it is frequently difficult to obtain good quality ancillary data (Lu, 2006).

Remote sensing-based methods, on the other hand, are becoming increasingly popular as there are high correlations between spectral bands and vegetation parameters (Lu, 2006). Remote sensing-based methods of biomass estimation model relationships between field-based biomass estimations and variables derived from remotely-sensed data (Kalaitzidis & Zianis, 2009; Lu, 2006). The digital format of remotely-sensed data means that large quantities of data can be processed relatively quickly (Lu, 2006). Remote-sensing based methods are also very useful for remote areas (Lu, 2006). Lu (2006) and Kalaitzidis and Zianis (2009) provide particularly good reviews of remote sensing-based biomass estimation.

Various relationships between biomass and spectral responses are reported in the literature. For example, Phua and Saito (2003) estimated biomass from crown diameter which was derived from blue reflectance. Steininger (2000) reported a relationship between canopy reflectance and regrowth biomass, with middle-infrared and near infrared having the highest correlations and bands in the visible spectrum having the lowest, and Thenkabail *et al.* (2004) reported a linear relationship between biomass and spectral reflectivity and/or a normalized difference vegetation index (NDVI). Thenkabail *et al.* found that the best models involved the red band or indices using red and blue bands.

Spectro-biophysical relationships are modelled directly or indirectly from remotely-sensed data (Lu, 2006) using parametric (Castillo-Santiago *et al.*, 2010; Dong *et al.*, 2003) or non-parametric methods (Foody *et al.*, 2003; Tomppo *et al.*, 2002). Parametric methods are the most common, and involve developing empirical relationships between biophysical properties and remotely-sensed data using regression analysis (Kalaitzidis & Zianis, 2009). The main

disadvantage of parametric methods is that they are only applicable to the specific landscape on which they were developed (Kalaitzidis & Zianis, 2009). This, however, is also true for non-parametric methods (Foody *et al.*, 2003), such as regression trees, which find the natural groupings in the data without assuming a normal distribution (Afifi *et al.*, 2004; Chatfield & Collins, 1980)

Biomass estimation models often use spectral variables or vegetation indices directly derived from remotely-sensed data as predictor variables (Lu, 2006; Steininger, 2000; Thenkabail *et al.*, 2004) (Zheng *et al.*, 2004). Vegetation indices are useful to remove spectral variability caused by environmental conditions (Lu, 2006). Interestingly, NDVI is the most commonly used vegetation index, even though other vegetation indices may be more accurate for biomass predictions (Foody *et al.*, 2003). The three major disadvantages with direct relationships are that they become saturated after a certain age, relationships may differ between sites, and relationships differ between species type (Foody *et al.*, 2003; Steininger, 2000; Thenkabail *et al.*, 2004; Zheng *et al.*, 2004). For example, Steininger (2000) found that the canopy reflectance-biomass relationship for Brazilian forest stands plateaued after 15 years of age. Foody *et al.* (2003) found that the importance of spectral variables for predicting biomass varied significantly for different sites, and Zheng *et al.* (2004) found that separating types of forest improved biomass estimates. Estimates for hardwood forests were strongly related to near-infrared reflectance, whereas estimates for pine forests were strongly related to a corrected normalized difference vegetation index (NDVIc).

In contrast, indirect biomass estimations can be more reliable than direct estimations, especially for small samples (Phua M & Saito, 2003). Indirect estimations are made from canopy parameters, which are derived from remotely-sensed data using multiple regression analysis or canopy reflectance models (Lu, 2006). For example, Popescu *et al. (2003)* derived crown diameter from LIDAR data, and found that it alone explained 78 % of the variance associated with biomass.

Forest biomass has been estimated from fine, medium and coarse resolution remotely-sensed data (Lu, 2006). For example, high-resolution IKONOS imagery has been used to estimate biomass and carbon stock levels of African oil palms at a landscape level (Thenkabail *et al.*, 2004). Estimations were made for two age categories of oil palm. However, the ability to differentiate age groups was limited by significant intermixing of classes. This indicates that four bands may not always be sufficient for biomass estimations. In contrast, medium spatial resolution imagery has been used to estimate biomass at local and regional scales (Lu, 2006). For example, Zheng *et al.* (2004) estimated biomass for hardwood and pine forests across a

landscape from Landsat 7 ETM data. While Mickler *et al.* (2002) estimated carbon storage for forests across the Southern United States using Landsat TM coupled with a physiologically based productivity model. On the other end of the scale, course resolution imagery is used for national, continental and global scale biomass analysis (Lu, 2006).

High-resolution imagery does not necessarily produce better results for forests than lower resolution imagery: high spectral variation and shadows caused by topography and landcover features make creating biomass models difficult (Lu, 2006). However, spectral and spatial information can be combined to improve biomass estimations (Lu, 2006; Nichol & Sarker, 2011). For example, Lu (2005) used a combination of spectral and textural information to combat shadow. This is useful because biomass is poorly correlated to spectral responses but highly correlated to texture in forest sites with significant shadow or complex vegetation structures (e.g. tropical forests). The converse is true for forest sites without shadow or with simple vegetation structures (e.g. coniferous forests) (Lu, 2005, 2006).

Radar and LIDAR data can also be used to improve biomass estimates (Kalaitzidis & Zianis, 2009; Orka *et al.*, 2009). LIDAR and radar sensors can be used to measure components of canopy structure, such as height and volume (Kalaitzidis & Zianis, 2009; Lu, 2006). The advantage of radar data is that information can be collected irrespective of weather or light conditions (Lu, 2006). However, the saturation problem, as discussed for optical sensors, is also common in radar data (Lu, 2006; Luckman *et al.*, 1997).

Unfortunately remote sensing-based methods for estimating biomass cannot yet replace field-based methods, as the latter is usually more accurate (Kalaitzidis & Zianis, 2009). Nevertheless, one of the major advantages of using remote sensing for biomass modelling is that it can provide information at a range of spatial and temporal scales in a consistent manner (Foody *et al.*, 2003). Unfortunately remote-sensing-based models also have problems with site transferability. More advances in remote sensing-based methods are therefore required. Meanwhile, field measurements must be used to calibrate and validate models that utilize remotely-sensed data.

Shelterbelt biomass estimation from remotely-sensed imagery has not previously been investigated. Direct biomass estimation methods are likely to be more appropriate than indirect methods. This is because shelterbelts are often pruned, and it may be difficult to differentiate individual crowns in very dense shelterbelts. It is also likely that estimates from high-resolution imagery would be more accurate than estimates from lower resolutions due to the small spatial extent of shelterbelts (Cracknell, 1998). Lastly, it is probably reasonable to

suggest that textural information is likely to be less important than spectral information as the vegetation structure of shelterbelts is simple.

Studies that estimate shelterbelt carbon sequestration are not very common. The amount of carbon sequestered in shelterbelts has been estimated for areas in the United States (Brandle *et al.*, 1992), Canada (Kort & Turnock, 1999), and China (Wang & Feng, 1995). However, these studies used methods based on field measurements. Brandle *et al.* (1992) estimated that the above-ground carbon content for a single row shelterbelt (20 year-old trees) was 0.68 t/km for shrubs, 5.41 t/km for hardwoods, and 9.14 t/km for conifers. Kort and Turnock (1999) estimated that carbon content ranged from 11 t/km for blackthorn (type of shrub), to 105 t/km for hybrid poplars (type of hardwood). Conifer shelterbelts lay in between, ranging from 24 to 41 t/km. The carbon estimates from Kort and Turnock (1999) are substantially greater, due to a difference in tree age. The estimations from Brandle *et al.* (1992) were for twenty year-old shelterbelts, whereas the estimations from Kort and Turnock (1999) were based on 17 to 90 year old shelterbelts. Twenty year-old trees have only reached 30-60 % of mature height, and subsequently have not reached the maximum potential for carbon storage.

The tree species of interest to this research are *P.radiata* and *C.macrocarpa*. Remote-sensing based methods have not yet been used to estimate biomass and carbon spatially for these species. GIS-based and field-based biomass modelling for *P.radiata* has, however, been done in New Zealand (Bi *et al.*, 2010; Coops *et al.*, 1998; Madgwick, 1994; Moore, 2010). The C_change model developed by Beets *et al.* (1999) is currently used to predict carbon stock for New Zealand plantation forests as part of the Land Use and Carbon Analysis System (LUCAS) (Moore, 2010). This model relies on knowledge about stand parameters, such as age and height, as well as management parameters, such as rotation, pruning and harvesting operations (Beets *et al.*, 1999).

2.7 Literature review summary

A number of conclusions can be made from the literature reviewed:

1. Studies using remote sensing image analysis to delineate shelterbelts, differentiate species comprising shelterbelts, and estimate shelterbelt biomass are uncommon, representing a gap in knowledge.

2. Object-oriented classification with specialist feature extraction software is likely to produce the best results for shelterbelt extraction, because object-oriented methods utilize both spectral and spatial information; and specialist feature extraction software utilizes the interpretation skills of the user, as well as being relatively fast and accurate.

3. It is well established that coniferous tree species can be differentiated from broadleaved species. However, no work has been done thus far to determine whether *P.radiata* can be differentiated from *C.macrocarpa*. High-resolution multispectral imagery has been shown to be sufficient for differentiating at least some coniferous species, and will therefore be used in this research.

4. No work has been done on estimating shelterbelt biomass and/or carbon from remotely-sensed data. This research will therefore use a combination of field collected data to estimate coniferous shelterbelt biomass so that the relationship between biomass and remotely-sensed data can be modelled.

Chapter 3
General Methods

3.1 Introduction

This chapter describes the study areas and satellite imagery used for image analysis, as well as describing image preparation steps, and field data collection. Image preparation was a necessary precursor to all subsequent image analyses in the thesis, i.e. shelterbelt delineation (see Chapter 4, Image-based Shelterbelt Delineation), species differentiation (see Chapter 5, Differentiation of Shelterbelt Species), and carbon modelling (see Chapter 6, Example Application: Modelling Shelterbelt Carbon). Field data was collected for species differentiation (see Chapter 5) and carbon modelling (see Chapter 6).

3.2 Study area

Three study areas (16 km^2 each) were chosen within the Hurunui District of North Canterbury to provide a representative example of Canterbury shelterbelts (Figure 3.1). All three study areas were selected from QuickBird images captured on the same date so that atmospheric correction would not be required (see section 3.4, Satellite imagery). Agricultural areas with large numbers of shelterbelts were chosen. Study areas also had to have sufficient road access, as only shelterbelts parallel to public roads were sampled during field data collection (see section 3.5, Field data collection).

3.3 Satellite imagery

QuickBird II imagery in the New Zealand Transverse Mercator (NZTM) projection was chosen for use in this research, as it has a fine spatial resolution suitable for analysing narrow features such as shelterbelts. It also has a reasonable spectral resolution with red, green, blue and near infrared bands available, as well as a panchromatic band. This makes it suitable for species differentiation and biomass estimation.

QuickBird II multispectral (2.4 m) and panchromatic (0.6 m) images were obtained for the study area (Appendix A.1). Two images captured on the same date (27 February 2008) were chosen, as this made atmospheric correction unnecessary. Atmospheric correction is needed when extracting biophysical information from vegetation across different images captured at different times, as it corrects subtle differences in reflectance caused by differing atmospheric conditions (Jensen, 2005).

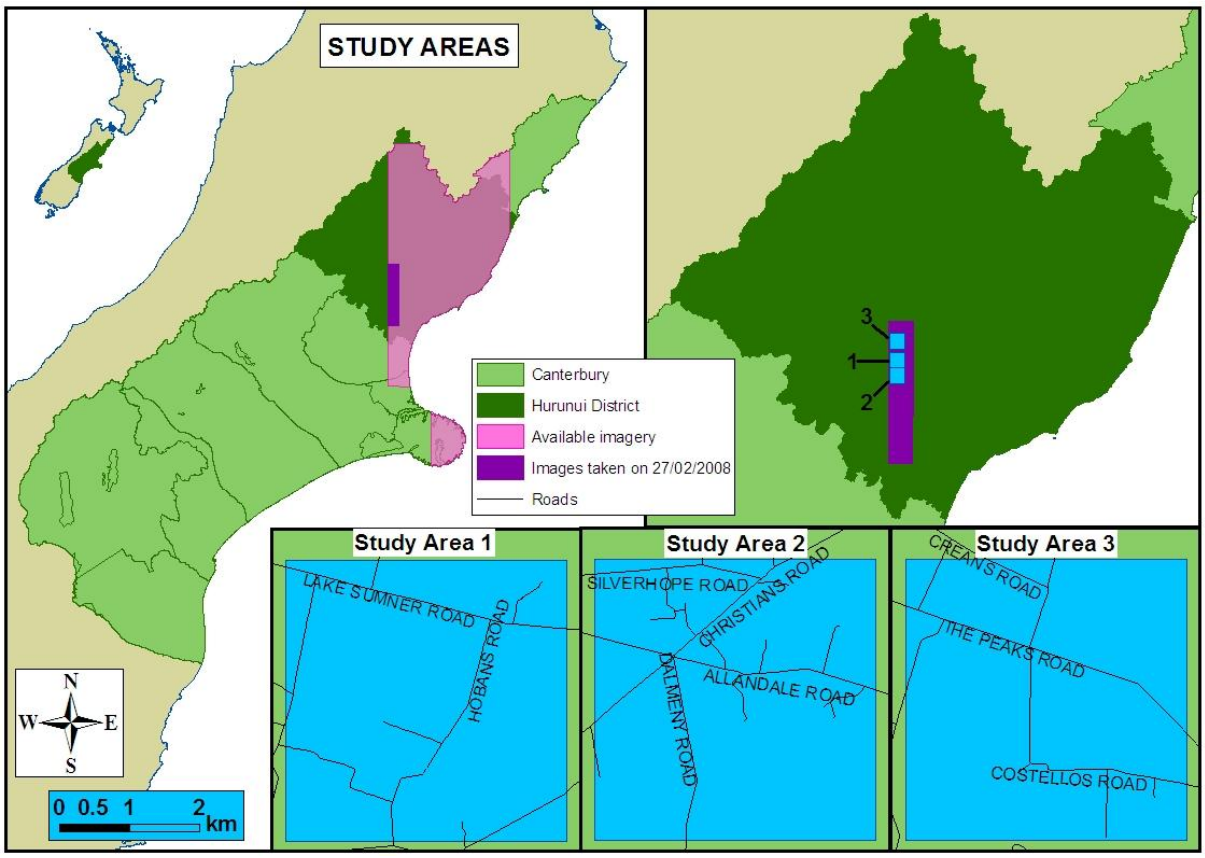


Figure 3.1 Map of three study areas in relation to image availability at 1 March 2010.

3.4 Image preparation

The image was prepared so that spectral information could be extracted from the three study areas for the analyses described in Chapters 4 (Image-based shelterbelt delineation), 5 (Differentiation of Shelterbelt Species) and 6. Image preparation steps included pansharpening and mosaicing two images, cutting the resulting image to the three study areas, and calculating a NDVI for each study area, as described below (Figure 3.2). These steps were carried out using ArcGIS 9.3(ESRI, 2009a).

The two QuickBird images captured on the same date were pansharpened, which is the process of fusing a high-resolution panchromatic image with a lower-resolution multispectral image to create a multispectral image with the same resolution as the panchromatic image (ESRI, 2009b). The IHS (Intensity, Hue and Saturation) method was used. Red, green and blue bands (bands 3, 2 and 1 respectively) were given a weighting of 0.3, while the near-infrared band (band 4) was given a weighting of 0.1. These weightings were selected using the default weightings for a 3-band image as guidelines: 0.334, 0.333, and 0.333 for red,

green, and blue, respectively (ESRI, 2009b). This resulted in two pansharpened multispectral images with a spatial resolution of 0.6 m.

Next, the resulting images were mosaiced to form one image. Shapefiles of each study area, as well as a shapefile encompassing all three study areas, were then used to clip out subsets of the mosaiced image.

Lastly, a Normalized Difference Vegetation Index (NDVI) was calculated for each subset, using the Image Analyst Extension (ERDAS, 2009) (Tables 3.1 & 3.2) which applies the formula (developed by Rouse *et al.* (1974)):

$$NDVI = \frac{NIR - R}{NIR + R} \qquad (1)$$

Where: NIR = Near infrared
R = Red

This resulted in a 4-band image and a NDVI for each subset (Tables 3.1 and 3.2), which were used for all subsequent analyses.

Table 3.1 Spectral and spatial resolutions of the prepared image.

Imagery	Band		Spectral resolution	Spatial Resolution
Pansharpened Multispectral	1	Red	630-690 nm	0.6 m
	2	Green	520-600 nm	0.6 m
	3	Blue	450-520 nm	0.6 m
	4	Near-Infrared	760-900 nm	0.6 m
NDVI	1	Black & White	N/A	0.6 m

Table 3.2 Prepared images for study areas 1, 2 and 3.

Study Area	Vector file	Pansharpened image	NDVI
1	BV24.shp	R1G2B3NI4.img	NDVI.img
2	BV24_2.shp	R1G2B3NI4_2.img	NDVI_2.img
3	BV24_3.shp	R1G2B3NI4_3.img	NDVI_3.img
1,2, & 3	Area_all3.shp	R1G2B3NI4_all.img	NDVI_all.img
Uncut	N/A	Append.img	NDVI_whole.img

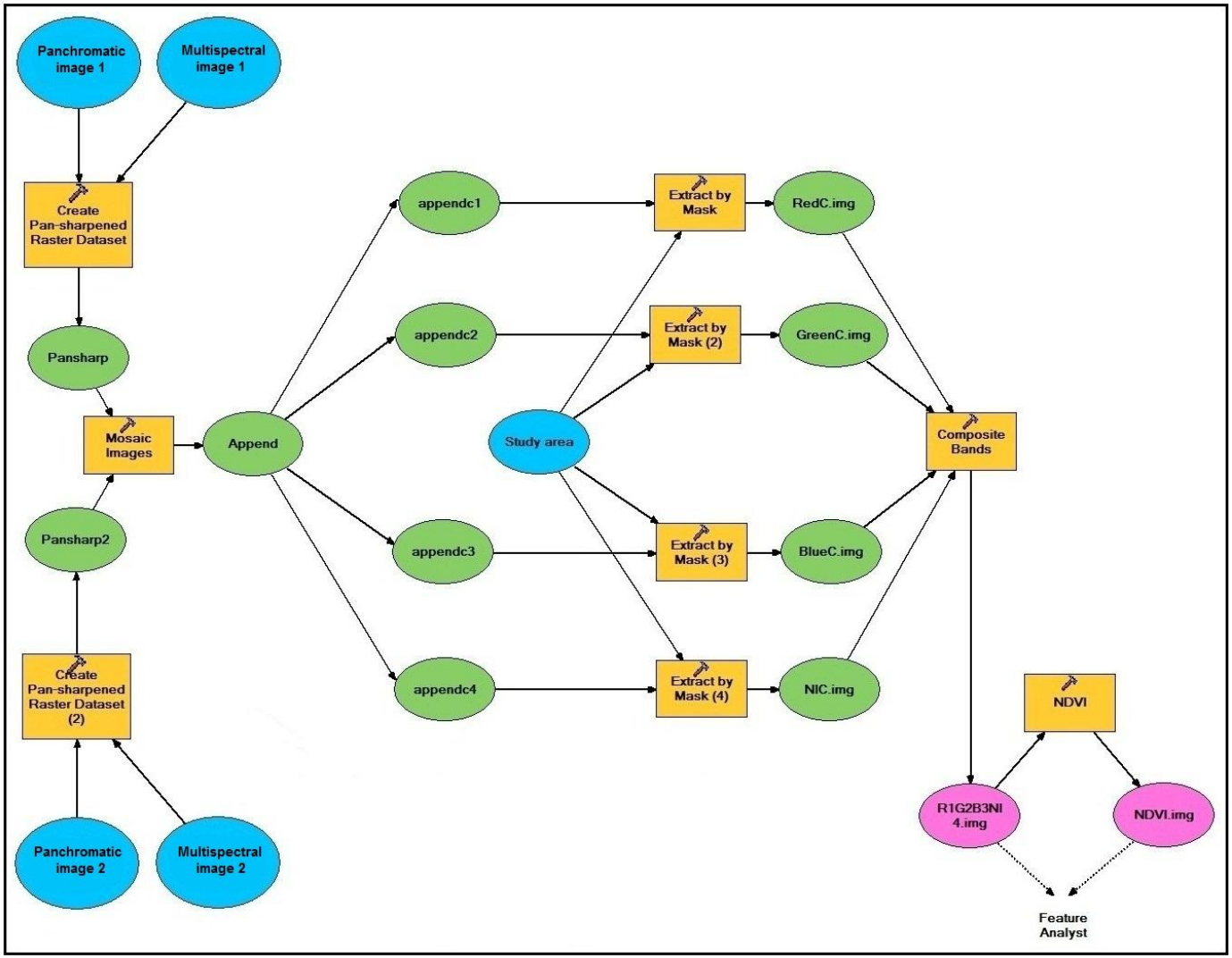


Figure 3.2 Flowchart showing sequence of steps taken in the ArcGIS Image Analysis software to prepare the two satellite images for further analyses.

3.5 Field data collection

Shelterbelt data was collected in the field to provide information for shelterbelt characterization. Collected data on shelterbelt species and shelterbelt dimensions were used for species differentiation and carbon estimation from QuickBird images in Chapters 5 and 6, respectively. The results from field data collection are therefore presented in Chapter 5.

For this work, a shelterbelt was defined as:

- A band of trees acting as a windbreak or to provide shade for livestock or crops (Gregory, 1995; Price, 1993),
- having a linear shape,
- having an association with agricultural land,
- being at least 10 m long,
- having a spacing between trees of no more than 5 m for shelterbelts with a single row, or having a spacing between trees of no more than 8 m in the densest row for shelterbelts with multiple rows, and
- having a height of at least 2 m (tallest row) for shelterbelts with mature trees, i.e. DBH > 12 cm.

Field work was carried out in all three study areas. A map of the three study areas was loaded into ArcPad 7.0 (ESRI, 2005) on a handheld pocket computer. This map contained basic information about roads, study area boundaries, and shelterbelt location. The latter was provided by the shelterbelt class of an unsupervised classification (see Chapter 4, Image-based Shelterbelt Delineation), which was converted to vector format. These shelterbelt shapefiles were created for study areas 1, 2 and 3, because raster images (e.g. R1G2B3NI4.img) were too large to import into ArcPad. Shelterbelts extracted by Feature Analyst software (Overwatch, 2010) (see Chapter 4) were not used, as the software had not yet been purchased.

All shelterbelts parallel to a public road within the three study areas were sampled. This was decided upon so that no permission was needed to access shelterbelts on private land. Sampling was completed after seven days. It was decided that the sample (n = 133) was adequate as it contained a reasonable number of *P.radiata* (n = 60) and *C.macrocarpa* (n = 31) shelterbelts. Unfortunately, there were only a few broadleaved shelterbelts in the sample

areas, and therefore, broadleaved shelterbelts made up only a small proportion of the sample (n = 15).

Shelterbelt observations and dimensions were recorded by hand on a printed-out spreadsheet, as this was more efficient than recording everything digitally in ArcPad. Data was later typed into a Microsoft Excel spreadsheet. Observations about other vegetation types were also sometimes recorded as a matter of interest, and to aid image interpretation.

ArcPad was used in combination with a blue-tooth Geographic Positioning System (GPS) to record shelterbelt position. A point was drawn to mark the beginning of each shelterbelt, and a line was then drawn to the position at which the shelterbelt ended. This information was used later when digitizing shelterbelts.

A rangefinder and clinometer were used to record height-related measurements (Table 3.3; Appendix B.1). The rangefinder was used to measure the horizontal distance between the user and the shelterbelt. The clinometer was used to measure the angle (in degrees) between the ground and the treetops.

Shelterbelt width, tree diameter, and spacing between trees and rows were measured with a tape measure (Table 3.3; Appendix B.1). These were recorded to one decimal place for meters, and zero decimal places for centimetres. Dimensions were estimated without a tape measure when taking exact measurements was unpractical. Estimates were recorded in a category, such as 0 – 3 meters wide. All measurements and estimates should be considered as rough approximations only.

All shelterbelts in the sample visible on the satellite image were digitized as polygons. Shelterbelt area (m^2) was calculated for digitized shelterbelts using the geometry calculator in ArcMap (ESRI, 2009a). The length of digitized shelterbelts was then calculated from area and width (Table 3.3). The length of undigitized shelterbelts (i.e. not visible on the image) was estimated by calculating the length of the lines digitized during field work.

Lastly, tree density was calculated for each shelterbelt:

$$\text{Density (trees/m}^2\text{)} = \quad D = \frac{N}{A} \qquad (2)$$

Where: $A =$ shelterbelt area (m^2) $\quad L =$ shelterbelt length (m)

$$N = \text{number of trees} = \frac{L}{S_{R1}} + \frac{L}{S_{R2}} \ldots + \frac{L}{S_{Ri}} \qquad (3)$$

$S_{Ri} =$ spacing for row *i*

Table 3.3 Measurements and observations recorded for shelterbelts visited during field data collection.

PARAMETER	DESCRIPTION (See Appendix B.1)
Shelterbelt ID	Each shelterbelt/sample was given a unique number.
Type	Each sample was recorded as either a shelterbelt (Yes) or another type of planting/vegetation (No). The type of shelterbelt and type of other vegetation was also recorded: Yes: • Single species • Mixed species • Chopped down No: • Hedge • Woodlot block • Group of trees • Willow trees • Other (e.g. planting of trees that does not fit into any category).
Species of tree	The species recorded included *P.radiata*, *C.macrocarpa*, *Populus* spp., and *Eucalyptus* spp.. All other species were grouped into a general class: 'other conifer', 'native', 'other deciduous', or 'willow'.
Width (m)	Estimate categories: 0 – 3 m; 3 – 5 m ; 5 – 10 m; 10 – 15 m; 15 – 20 m; > 20 m
Height (m)	Calculated from rangefinder (distance) and clinometer measurements (angle).
Length (m)	Calculated from digitized shelterbelt area (Area/Width);
Diameter at breast height DBH (cm)	A representative tree for each age class was chosen. Circumference was measured and diameter calculated, or diameter was estimated. Estimate categories: < 6 cm (seedling); 6 – 12 cm (sapling); 12 – 30 cm; 30 – 60 cm; 60 – 90 cm; 90 – 120 cm; > 120 cm.
Crown ratio (%)	Estimated to 1 significant figure.
Number of rows	
Spacing in each row (m)	Two consecutive trees with a representative gap between them were chosen for each row. Spacing was measured between the centres of two consecutive tree trunks. Estimate categories: < 1 m; 1 – 2 m; 2 – 3 m; 3 – 5 m; > 5 m
Number of trees	Trees (n) were counted for very short shelterbelts, or calculated using length (L) and spacing in each row (S): $n = (L/S_1) + (L/S_2) + \ldots + (L/S_i)$ Where *i* = number of rows
Other information	Health categories: • Good; • Mediocre (e.g. some trees dead); or • Bad (e.g. lots of trees dead/dying). Management categories: • Pruned • Unpruned • Half pruned (e.g. sides are pruned but tops are unpruned)

Chapter 4
Image-based Shelterbelt Delineation

4.1 Introduction

Shelterbelts are a prominent part of New Zealand's agricultural landscape, particularly of Canterbury (Cameron, 1964; Price, 1993): the Canterbury Plains have one of the most impressive displays of shelterbelts and hedges in the world (Price, 1993). In addition to being aesthetically pleasing, shelterbelts have many economic and ecological benefits (Gregory, 1995; Ministry of Forestry, 1992; Stringer, 1977). Despite this, shelterbelts are not particularly well characterized in a spatially explicit manner. Methods that are accurate, spatially explicit and automated are therefore needed to identify shelterbelts across a landscape.

The advent of remote sensing technology has made the accurate mapping of shelterbelts possible by providing high spatial resolution images (Cracknell, 1998; Hengl, 2006). Shelterbelt delineation can be problematic as they have such a small aerial extent and are often fragmented (Lechner *et al.*, 2009). A handful of studies outside of New Zealand have reported delineating shelterbelts from high-resolution images (Aksoy *et al.*, 2010; Lennon *et al.*, 2000; Liknes *et al.*, 2010; Pankiw *et al.*, 2009; Tansey *et al.*, 2009; Thornton *et al.*, 2006, 2007; Vannier & Hubert-Moy, 2008; Wiseman *et al.*, 2009). However, no work of this nature has been done thus far in New Zealand.

This chapter explores methods for delineating shelterbelts across an area of the Canterbury Plains from high resolution multispectral imagery. Per-pixel and object-oriented classification methods, including a method utilizing feature extraction software, are compared. This chapter also determines whether batch processing affects classification accuracy.

4.2 Classification methods overview

There are two main approaches that can be taken to extract shelterbelts from high-resolution imagery: per-pixel classification and object-oriented classification. Per-pixel classification methods work by assigning each pixel to a land cover class according to a set of spectral signatures (Blaschke *et al.*, 2000; ERDAS, 2008; Jensen, 2005). Per-pixel classification can be carried out automatically (unsupervised) or with input from the user (supervised).

Unsupervised per-pixel classification clusters pixels into groups which are inherent in the data. The groupings are based on statistical criteria using an iterative self-organizing data analysis technique. Once the classification is complete, the user labels each class using simple image interpretation skills (ERDAS, 2008; Jensen, 2005).

Supervised per-pixel classification relies on both the classification software and the user's pattern recognition skills: (i) the user identifies training sites; (ii) the software calculates multivariate statistical parameters for each training site; and (iii) all remaining pixels are evaluated by the software, based on the training-based parameters. Each pixel is finally assigned to the class to which it is most likely to belong (ERDAS, 2008; Jensen, 2005).

Until recently, most images were classified using per-pixel classification (Blaschke *et al.*, 2000; Jensen, 2005). However, using spectral information alone usually does not result in high classification accuracies when extracting linear features (Quackenbush, 2004). Object-oriented classification is an alternative to per-pixel classification which utilizes both spectral and spatial information (Jensen, 2005). This typically involves an image segmentation phase and an object classification phase. Image segmentation creates objects by dividing the image into areas which are spatially and spectrally homogenous (Jensen, 2005). These objects can then be classified or labelled.

Another object-oriented approach is feature extraction (Miller *et al.*, 2009): features are recognized and outlined by a human operator or an automated system (Guyon & Elisseeff, 2006; Overwatch Ltd., 2010; Quackenbush, 2004). Manual feature extraction utilizes human skills, but it is time-consuming, expensive, and therefore unpractical for large projects (Quackenbush, 2004). Unsupervised automated extraction, on the other end of the scale, is fast but not necessarily reliable (Quackenbush, 2004). A happy medium is semi-automated feature extraction or supervised machine learning. This aims to reproduce human learning capabilities by incorporating user input with an automated system (Guyon & Elisseeff, 2006).

Commercial feature extraction softwares are available, such as Definiens eCognition, SAGA, ENVI Feature Extraction, and Overwatch Feature Analyst (Blaschke, 2010). Feature Analyst (Overwatch, 2010) employs a semi-automated approach: the software "learns" from user-identified examples using a pixel-based computerized-learning classification method, as described below (Miller *et al.*, 2009):

(i) The Learner analyses the training set according to user-defined settings to discover overall spectral and spatial characteristics of the training features (Miller *et al.*, 2009; Overwatch Ltd., 2010);

(ii) The Learner then classifies every pixel in the image according to what it has learnt from the training set (Overwatch Ltd., 2010);

(iii) The output is converted from raster to vector by aggregating all neighbouring pixels with the same classification. This results in a vector output of distinct shapes (Miller *et al.*, 2009);

(iv) The user can provide feedback at this stage by identifying examples of false positives and false negatives. This leads to more cycles of learning which remove clutter and retrieve missed features. This is known as hierarchical learning (Overwatch Ltd., 2010).

In this chapter, three approaches for identifying shelterbelts were investigated. Per-pixel classification methods were undertaken as this could be achieved using GIS sofware already licensed by Lincoln University, namely ArcGIS 9.3 (ESRI, 2009a). An object-oriented method employing image segmentation freeware and per-pixel classification results was employed, as this would also not incur any extra costs. Lastly, a second object-oriented method was undertaken which utilized the specialist feature extraction software, Feature Analyst. This method was chosen as the other methods produced unsatisfactory results.

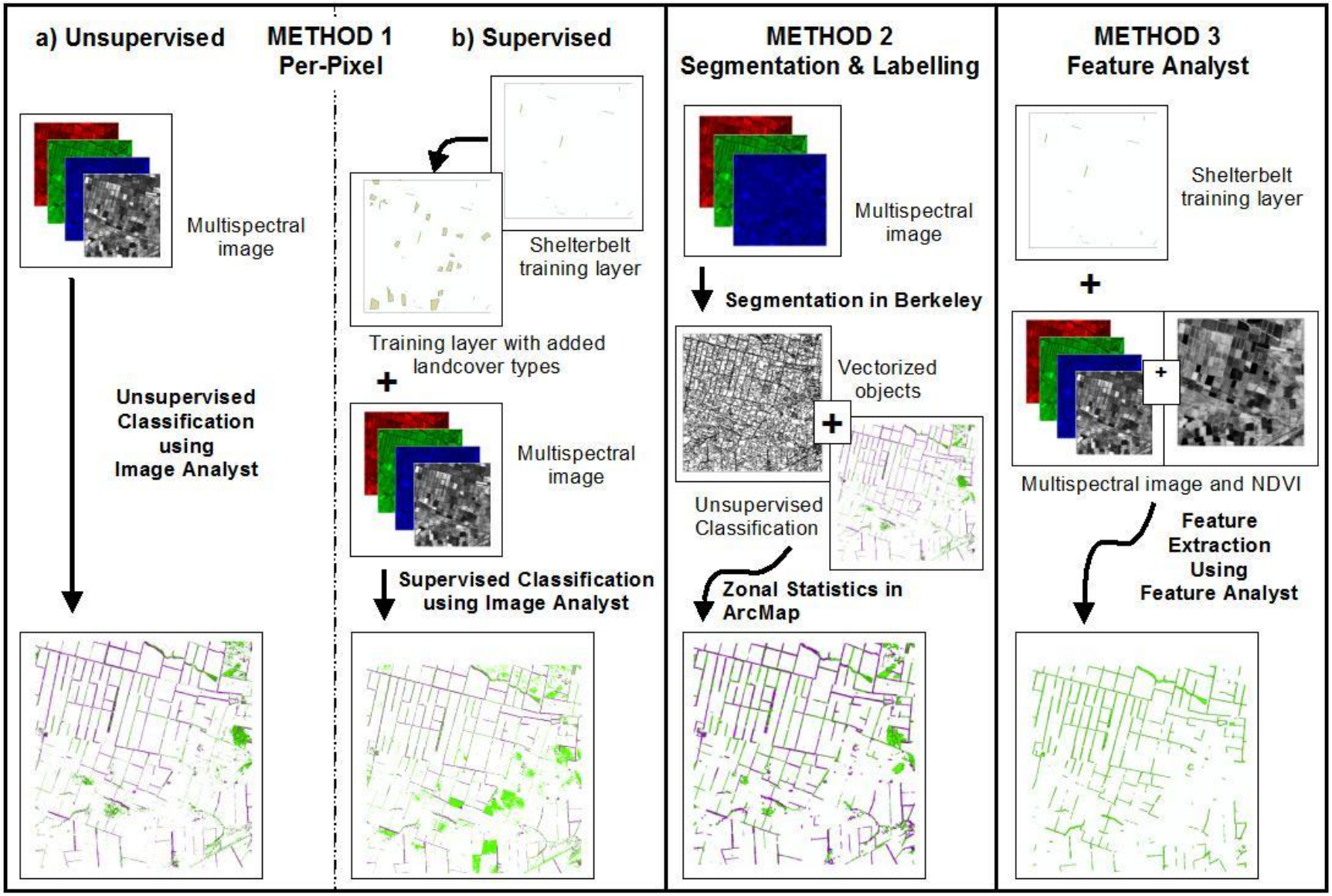


Figure 4.1 Summary of methodologies used for shelterbelt delineation.

4.3 Methods

Three approaches were trialled on study area 1 (Figure 4.1): per-pixel classification (Method 1), object-oriented classification with image segmentation and labelling (Method 2), and object-oriented feature extraction with Feature Analyst (Method 3). The best method was applied to all three study areas.

4.3.1 Per-pixel classification

Per-pixel classification (Method 1) can be carried out in ArcMap (ESRI, 2009a) using the Image Analysis extension for ArcGIS 9.3 (ERDAS, 2009). The Image Analysis classification tools operate by sorting pixels into the number of classes specified by the user. Each pixel is assigned to a class according to a certain set of criteria or spectral signatures (ERDAS, 2008). The Image Analysis extension can perform both unsupervised and supervised classifications:

Method 1a: Unsupervised classification

Unsupervised classifications of QuickBird imagery of study area 1 were performed. The total number of classes were varied between 10 and 25 for each classification run. Each result was examined and compared to the original satellite image to determine which classes represented shelterbelts and shadows cast by shelterbelts. The best result (20 classes) was identified. It was then reclassified, based on visual inspection, to three classes: 'shelterbelt', 'shadow', and 'other' (Appendix C.1).

Method 1b: Supervised classification

A maximum-likelihood supervised classification was carried out for study area 1 using a training shapefile. This consisted of 16 'shelterbelt' polygons, 16 'shadow' polygons and 20 'other' polygons (Appendix C.1). This same training shapefile was later used for Feature Analyst, but without 'shadow' and 'other' polygons. The 'other' landcover class had the highest number of training sites and the largest area, because spectral characteristics varied considerably for this class.

4.3.2 Object-oriented image analysis

Method 2: Image segmentation and segment labelling

Two freeware applications, SPRING 5.1 (Brazil's National Institute for Space Research (INPE), 2009) and BerkeleyImageSeg 1.0rc8 (Berkeley Environmental Technology, 2010), were investigated for image segmentation. SPRING is freeware that was developed by Brazil's National Institute for Space Research (INPE). It is a GIS and remote-sensing image processing system with an object-oriented data model (Câmara *et al.*, 1996; DPI/INPE, 2009).

SPRING uses a region-growing algorithm to segment images (Câmara *et al.*, 1996). Image segmentation by SPRING was trialled on study area 1. However, this approach was abandoned because SPRING was difficult to use (e.g. SPRING help files were not comprehensive), and difficulties were experienced when trying to export results out of SPRING into ArcMap. The main advantage of SPRING is that it can be downloaded for free (www.dpi.inpe.br/spring/).

BerkeleyImageSeg (BIS) is an image segmentation application which uses a region-merging algorithm. Contiguous objects (initially every pixel is an object) are merged depending on spectral and shape heterogeneity (Clinton, 2010). The user controls three parameters. The first parameter, the threshold, determines the number of merging cycles. Therefore, the higher the threshold, the larger the resulting objects. The second parameter, shape, specifies the weightings of shape and spectral heterogeneity. The third parameter, compactness, specifies the weightings of shape compactness and shape smoothness (Clinton, 2010).

A major advantage of BIS is that a 30-day, full-featured and unrestricted trial version can be downloaded for free (www.berkenviro.com/berkeleyimgseg/). Learning how to use BIS is also easy, with the creator of BIS being accessible and helpful. In addition, the output format is immediately compatible with ArcGIS.

BIS was used for image segmentation of study area 1. Trials were performed on part of study area 1 using the BerkeleyImageSeg Wizard to determine which values for threshold, shape, and compactness would be used. BerkeleyImageSeg Command Line Automation was used to segment a three-band image (RGB) of study area 1. The parameters used for threshold, shape and compactness were 100, 0.7 and 0.8 respectively. The resulting objects were then imported into ArcMap for segment classification.

Segments were labelled by overlaying results from the unsupervised classification with the image segments. Each segment was classified according to which unsupervised classification class the majority of pixels within the segment belonged (majority statistic). This produced a table containing the majority values for each object, which was joined to the segmented objects. This resulted in the classification of all objects as either ‘shadow’, ‘shelterbelt’, or ‘other’.

Method 3: Feature extraction using Feature Analyst

Trials were conducted on a small portion (1 km^2) of study area 1 to determine the best settings for the Feature Analyst Learner, and to experiment with training shapes and hierarchical learning.

The trial runs indicated that (i) best results were reached when using the 4-band multispectral image and the NDVI as inputs, a good training layer, and two cycles of clutter removal; (ii) using the NDVI as well as the multispectral image improved exclusion of shadow; (iii) ensuring that the training layer was representative of the whole image and that it excluded any 'ground' pixels decreased initial false positives and false negatives; (iv) the default settings for 'Narrow Linear Feature' produced good results; and (iv) two cycles of hierarchical learning were optimal, i.e. adding more cycles of clutter removal (Table 4.1), or adding missed features did not improve results. A good training shapefile was therefore the most important for reducing final numbers of false positives and false negatives; and

The best settings were applied to the whole of study area 1 (Appendix C.2). More training shapes had to be added to the trial training shapefile to make it representative of the whole of study area 1. The training shapefile was modified until it produced satisfactory initial results. This file was used for training during supervised classification as the 'shelterbelt' class.

Experimentation on the whole image confirmed that two cycles of clutter removal resulted in the best output (Figure 4.2). The 'begin removing clutter' tool was used for each hierarchical learning cycle: a copy of the previous shapefile is created in which the user can identify examples of correctly and incorrectly identified features (Table 4.1). The Feature Analyst Learner uses this information as a type of mask in the next cycle of hierarchical learning (Table 4.2).

The last step in the feature extraction process was an aggregation step. In this step, objects smaller than 50 m^2 were eliminated, as these objects were considered to be too small to be shelterbelts (Figure 4.2). The feature analyst method produced three intermediates and one final output in vector format (Table 4.2). Intermediate and output shapefiles illustrate how the classification is refined during hierarchical learning, i.e. the number of polygons and total area decrease as clutter (false positives) is removed.

Table 4.1 Summary of tools used to remove clutter from Feature Analyst intermediate results.

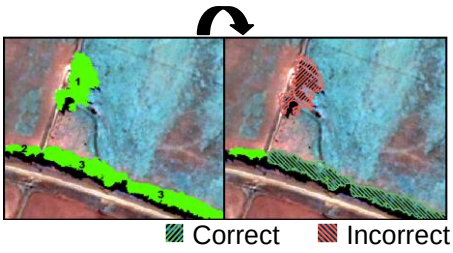
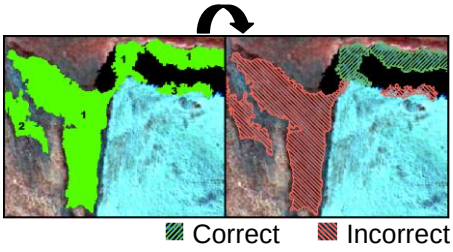
Tool	Function	Illustration
Select correct features	Identifies examples of correct polygons	
Select incorrect features	Identifies examples of incorrect polygons	
Digitize correct features	Identifies correct portion of a polygon which also contains an incorrect area.	
Digitize incorrect features	Identifies incorrect portion of a polygon which also contains a correct area.	

Table 4.2 Summary of training, intermediate and mask features used for shelterbelt delineation by Feature Analyst.

Type		Polygons	Total area m^2
Training		16	14540
Intermediate		1277	852371
Mask	Correct	39	70072
	Incorrect	65	11718
Intermediate		980	708608
Mask	Correct	43	95165
	Incorrect	61	9690
Intermediate		394	593279
Output		328	591807

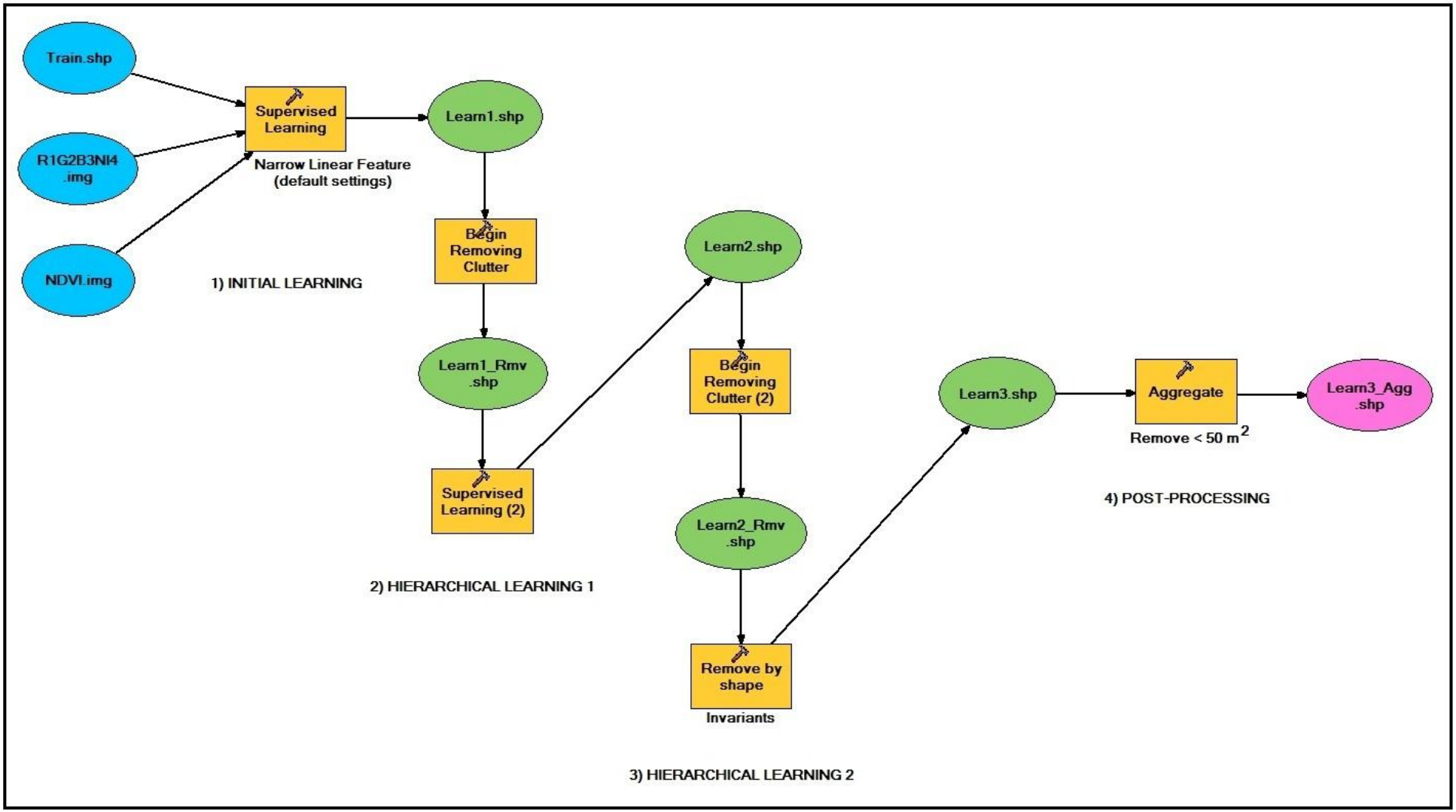


Figure 4.2 Flowchart showing initial learning and hierarchical learning steps used by the Feature Analyst Learner to delineate shelterbelts.

4.3.3 Comparison of methods: Accuracy assessment

The accuracy assessment aimed to answer two questions: (i) how good was each method at identifying shelterbelts (i.e. the producer's accuracy)? and (ii) to what degree did each method include other unwanted (non-shelterbelt) landcover classes in the classification (i.e. the user's accuracy)?

(i) Shelterbelt identification: area-based assessment

To answer the first question regarding the accuracy of methods in identifying shelterbelts correctly, an assessment method was developed that assessed the area of true positives.

Fifty random points were created using the 'create random points' tool (ArcGIS data management tool). A buffer zone with a radius of 30 m was then created for each point. This created an area on which shelterbelt assessment could be based. Each circular area was then moved to be located over its nearest shelterbelt, unless it was already over one. Shelterbelts within these areas were digitized as polygons directly off the image (Figure 4.3, a). The image objects identified as shelterbelts by each method were then clipped to the circular areas (Figure 4.3, b). This involved first converting the raster results of per-pixel classifications to vector format (polygons). Intersect analysis was performed to combine the clipped shelterbelts produced by each method and the digitized shelterbelt polygons (Figure 4.3, c).

The percentages of true and false positives (TP, FP), true and false negatives (TN, FN), and overall error were then calculated (Table 4.3). The overlapping areas ($Area_{intersect}$) resulted in polygons for each method which represented the area of shelterbelts correctly classed (TP). The area incorrectly classed as shelterbelts (FP) was calculated by subtracting the total area classed as shelterbelt ($Area_{method}$) from the overlapping area ($Area_{intersect}$). The area correctly classed as other land cover (TN) was calculated by subtracting FP from the area of other landcover ($Area_{other}$). The area incorrectly classed as other land cover (FN) was calculated by subtracting TP from the area of digitized shelterbelt ($Area_{digitized}$).

This assessment addressed the first question regarding the ability of each method to correctly identify shelterbelts by assessing true positives. It also partially addressed the second question in that it assessed the area of false positives in the vicinity of shelterbelts. The area of false positives in the vicinity of shelterbelts is expected to be higher for methods which did not differentiate well between shelterbelts and shadow. However, false positives may be underestimated, as this method did not assess areas at some distance from shelterbelts.

Table 4.3 Equations used to calculate true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) for methods delineating shelterbelts.

		Classification: Shelterbelt	Classification: Other*	Total Area (m^2)	*PRODUCER'S ACCURACY*
Actual	**Shelterbelt**	TP = ($Area_{intersect}$)	FN = ($Area_{digitized}$) – TP	($Area_{digitized}$)	TP / ($Area_{digitized}$)
Actual	**Other***	FP = ($Area_{method}$) – TP	TN = ($Area_{other}$) – FP	($Area_{other}$) = ($Area_{circle}$) - ($Area_{digitized}$)	TN / ($Area_{other}$)
Total Area (m^2)		$Area_{method}$	$Area_{circle}$	$Area_{total}$= ($Area_{method}$) + ($Area_{circle}$)	
USER'S ACCURACY		TP/ ($Area_{method}$)	TN ($Area_{circle}$)		
%	***TRUE***	TP/ ($Area_{method}$)	TN / ($Area_{circle}$)	***OVERALL ACCURACY*** = $\frac{(TP + TN)}{Area_{total}}$	
%	***FALSE***	FP/ ($Area_{method}$)	FN / ($Area_{circle}$)		
		POSITIVE	***NEGATIVE***		

* Includes 'shadow' class for Methods 1 and 2, and represents the background class for Method 3.

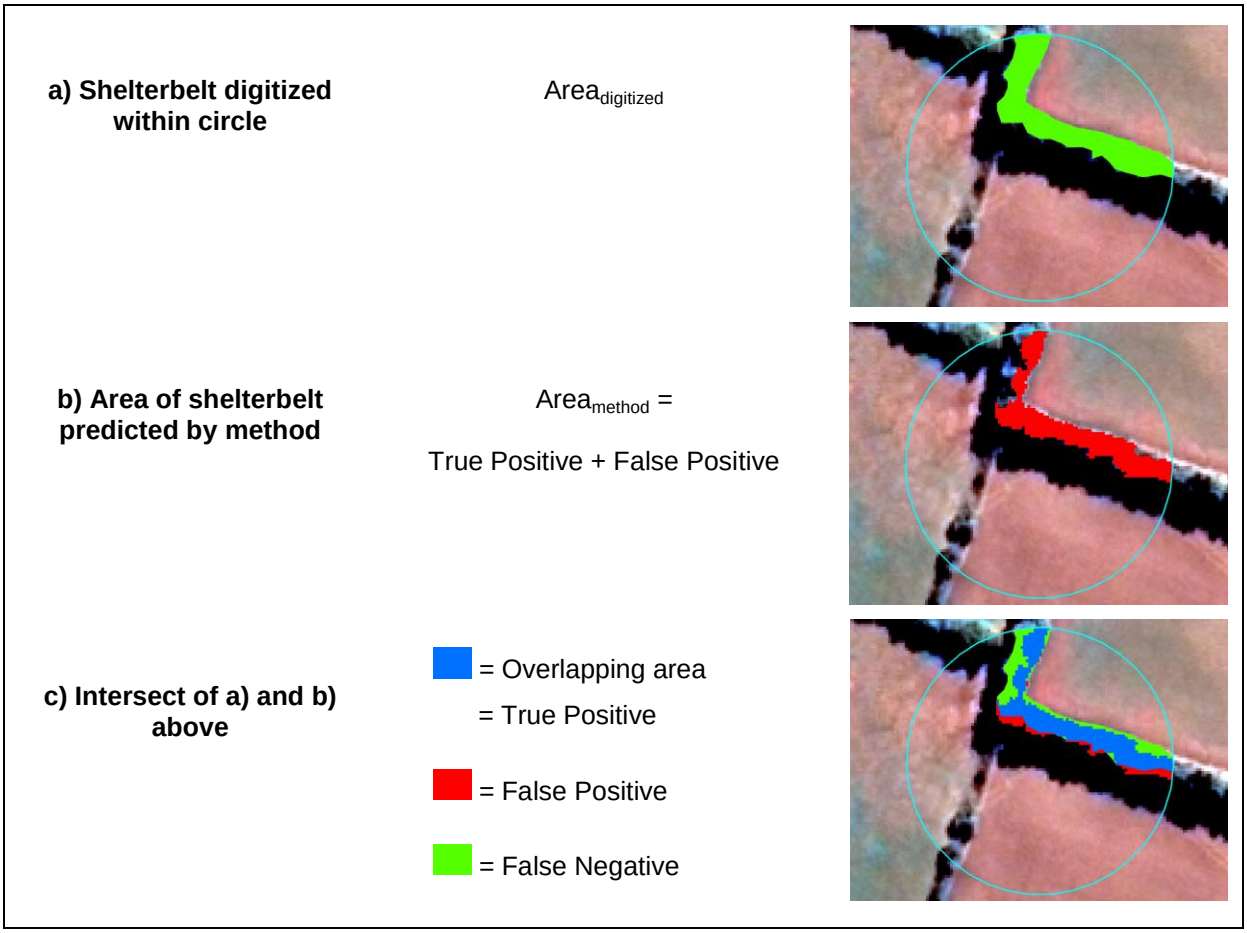


Figure 4.3 Intersect analysis for area-based accuracy assessment of methods delineating shelterbelts.

(ii) Inclusion of other landcover types: random point-based assessment

To answer the second question regarding the inclusion of other unwanted landcover classes in the shelterbelt classification, an assessment method was developed that assessed the number of false positives at some distance from shelterbelts.

The 'create random points' tool was used to generate 150 random points. The actual landcover class of the pixel on which the point fell was recorded, as well as the class predicted by each method. Landcover was recorded as either 'shelterbelt' or 'other'. The latter included the 'shadow' class for Methods 1a, 1b and 2, and the background class for Method 3. In contrast to the first accuracy assessment which was area-based, this assessment was pixel- or point-based. Eight of the random points fell onto shelterbelts, while 142 points fell onto other landcover types. The percentage of false positives were then calculated, as well as the overall error.

This assessment addressed the second question by assessing the number of false positives away from shelterbelts, such as areas of paddock and other tree cover.

4.3.4 Application of the Feature Analyst method to all three study areas

As initial results indicated the superiority of the Feature Analyst method (Method 3), the Batch Processing tool in Feature Analyst was used to apply the saved shelterbelt model to an area encompassing all three study areas (Appendices C.3 and C.4). This provided a means to determine how the method developed on study area 1 is generally applicable to other areas on the image.

A brief examination of the result identified some incorrect polygons due to an artifact on the satellite image: A line could be seen across the image where two image tiles were joined (see image preparation). This was corrected by deleting 15 polygons; and modifying eight polygons which were both artifact and shelterbelt.

The accuracy of batch processing classification was assessed using the area-based method described in the previous section (4.3.3 Comparison of methods).

4.4 Results

4.4.1 Comparison of methods for shelterbelt delineation

Unsupervised per-pixel (Method 1a) and supervised per-pixel (Method 1b) classifications identified 0.84 km^2 of shelterbelt area each, while the object-oriented classification (Method 2) and Feature Analyst (Method 3) identified 0.71 km^2 and 0.59 km^2, respectively (Figures 4.4 to 4.7).

Per-pixel classification methods produced the results with the lowest accuracy according to the area-based accuracy assessment (Table 4.4). Unsupervised classification identified only 53 % of actual shelterbelt area while supervised classification identified 86 %. Object-oriented image segmentation and labelling did not improve the unsupervised per-pixel classification, with only 52 % of actual shelterbelt area identified. Feature Analyst identified 71 % of shelterbelt by area, which was less than the supervised classification. However, the supervised classification method produced the largest proportion of false positives (58 %), while Feature Analyst identified the smallest proportion (39 %). Feature Analyst therefore had the highest overall classification accuracy (91 %).

The random point-based accuracy assessment (Table 4.5) showed that the unsupervised per-pixel classification method falsely identified non-shelterbelt features as shelterbelts in 44 % of the random point sample. Supervised per-pixel classification produced a similar result to the unsupervised classification, with a 42 % false positive rate. Object-oriented image segmentation and labelling did not improve per-pixel classification results, with a 43 % false positive rate. Feature Analyst had the lowest false positive rate, with only 11 % of non-shelterbelt features identified incorrectly as shelterbelt. Feature Analyst therefore had the highest overall classification accuracy (99 %).

Table 4.4 Comparison of area-based classification accuracies of methods delineating shelterbelts.

(1a) PER-PIXEL: UNSUPERVISED		Classification: Shelterbelt	Classification: Other	Total	*PRODUCER'S ACCURACY*
Actual	**Shelterbelt**	**8331**	7402	15733	**53 %**
	Other	10518	**114510**	125028	92 %
Total		18849	121912	140761	
USER'S ACCURACY		44 %	94 %		
%	***TRUE***	**44 %**	94 %	***OVERALL ACCURACY = 87 %***	
	FALSE	56 %	6 %		
		POSITIVE	***NEGATIVE***		

(1b) PER-PIXEL: SUPERVISED		Classification: Shelterbelt	Classification: Other	Total	*PRODUCER'S ACCURACY*
Actual	**Shelterbelt**	**13605**	2128	15733	**86 %**
	Other	18989	**106039**	125028	85 %
Total		32594	108167	140761	
USER'S ACCURACY		42 %	98 %		
%	***TRUE***	**42 %**	98 %	***OVERALL ACCURACY = 84 %***	
	FALSE	58 %	2 %		
		POSITIVE	***NEGATIVE***		

(2) OBJECT-ORIENTED		Classification: Shelterbelt	Classification: Other	Total	*PRODUCER'S ACCURACY*
Actual	**Shelterbelt**	**8147**	7586	15733	**52 %**
	Other	10808	**114220**	125028	91 %
Total		18955	121806	140761	
USER'S ACCURACY		43 %	94 %		
%	***TRUE***	**43 %**	94 %	***OVERALL ACCURACY = 87 %***	
	FALSE	57 %	6 %		
		POSITIVE	***NEGATIVE***		

(Table 4.4 continued)

(3) FEATURE ANALYST		Classification		Total	*PRODUCER'S ACCURACY*
		Shelterbelt	**Other**		
Actual	**Shelterbelt**	**11230**	4503	15733	**71 %**
	Other	7321	**117707**	125028	94 %
Total		18551	122210	140761	
USER'S ACCURACY		61 %	96 %	***OVERALL ACCURACY* = 91 %**	
%	***TRUE***	**61 %**	96 %		
	FALSE	39 %	4 %		
		POSITIVE	***NEGATIVE***		

Table 4.5 Comparison of random point-based classification accuracies of methods delineating shelterbelts.

(1a) PER-PIXEL: UNSUPERVISED		Classification: Shelterbelt	Classification: Other	Total	*PRODUCER'S ACCURACY*
Actual	**Shelterbelt**	**5**	3	8	63 %
	Other	4	**138**	142	97 %
Total		9	141	150	
USER'S ACCURACY		**56 %**	98 %		
%	***TRUE***	56 %	98 %	***OVERALL ACCURACY = 95 %***	
	FALSE	**44 %**	2 %		
		POSITIVE	***NEGATIVE***		

(1b) PER-PIXEL: SUPERVISED		Classification: Shelterbelt	Classification: Other	Total	*PRODUCER'S ACCURACY*
Actual	**Shelterbelt**	**7**	1	8	98 %
	Other	5	**137**	142	96 %
Total		12	138	150	
USER'S ACCURACY		**58 %**	99 %		
%	***TRUE***	58 %	99 %	***OVERALL ACCURACY = 96 %***	
	FALSE	**42 %**	1 %		
		POSITIVE	***NEGATIVE***		

(2) OBJECT-ORIENTED		Classification: Shelterbelt	Classification: Other	Total	*PRODUCER'S ACCURACY*
Actual	**Shelterbelt**	**4**	4	8	50 %
	Other	3	**139**	142	98 %
Total		7	143	150	
USER'S ACCURACY		**57 %**	97 %		
%	***TRUE***	57 %	97 %	***OVERALL ACCURACY = 95 %***	
	FALSE	**43 %**	3 %		
		POSITIVE	***NEGATIVE***		

(Table 4.5 continued)

(3) FEATURE ANALYST		Classification		Total	*PRODUCER'S ACCURACY*
		Shelterbelt	**Other**		
Actual	**Shelterbelt**	**8**	0	8	100 %
	Other	1	**141**	142	99 %
Total		9	141	150	
USER'S ACCURACY		**89 %**	100 %	***OVERALL ACCURACY = 99 %***	
%	***TRUE***	89 %	100 %		
	FALSE	**11 %**	0 %		
		POSITIVE	***NEGATIVE***		

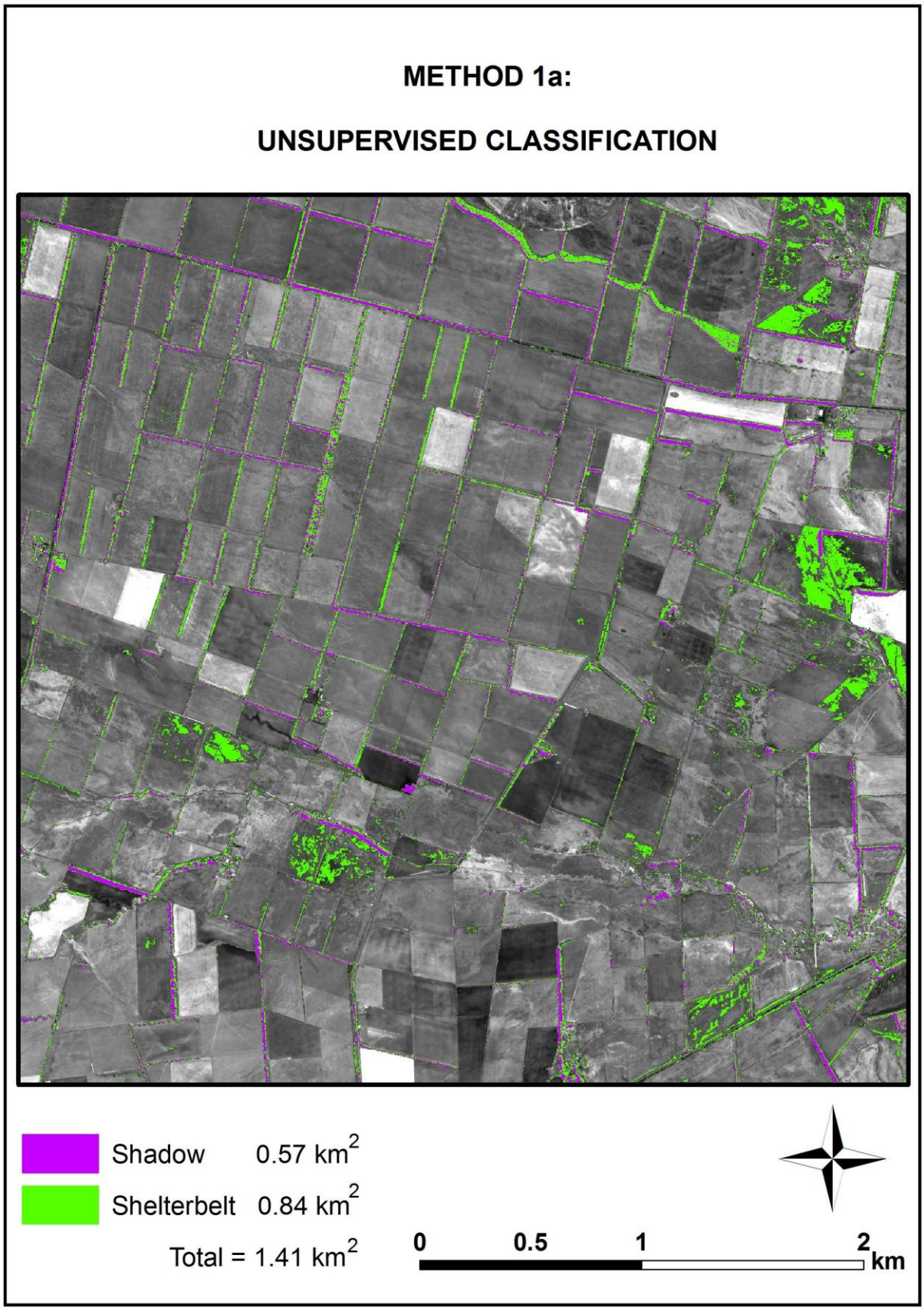


Figure 4.4 Unsupervised per-pixel classification showing 'shelterbelt' and 'shadow' classes.

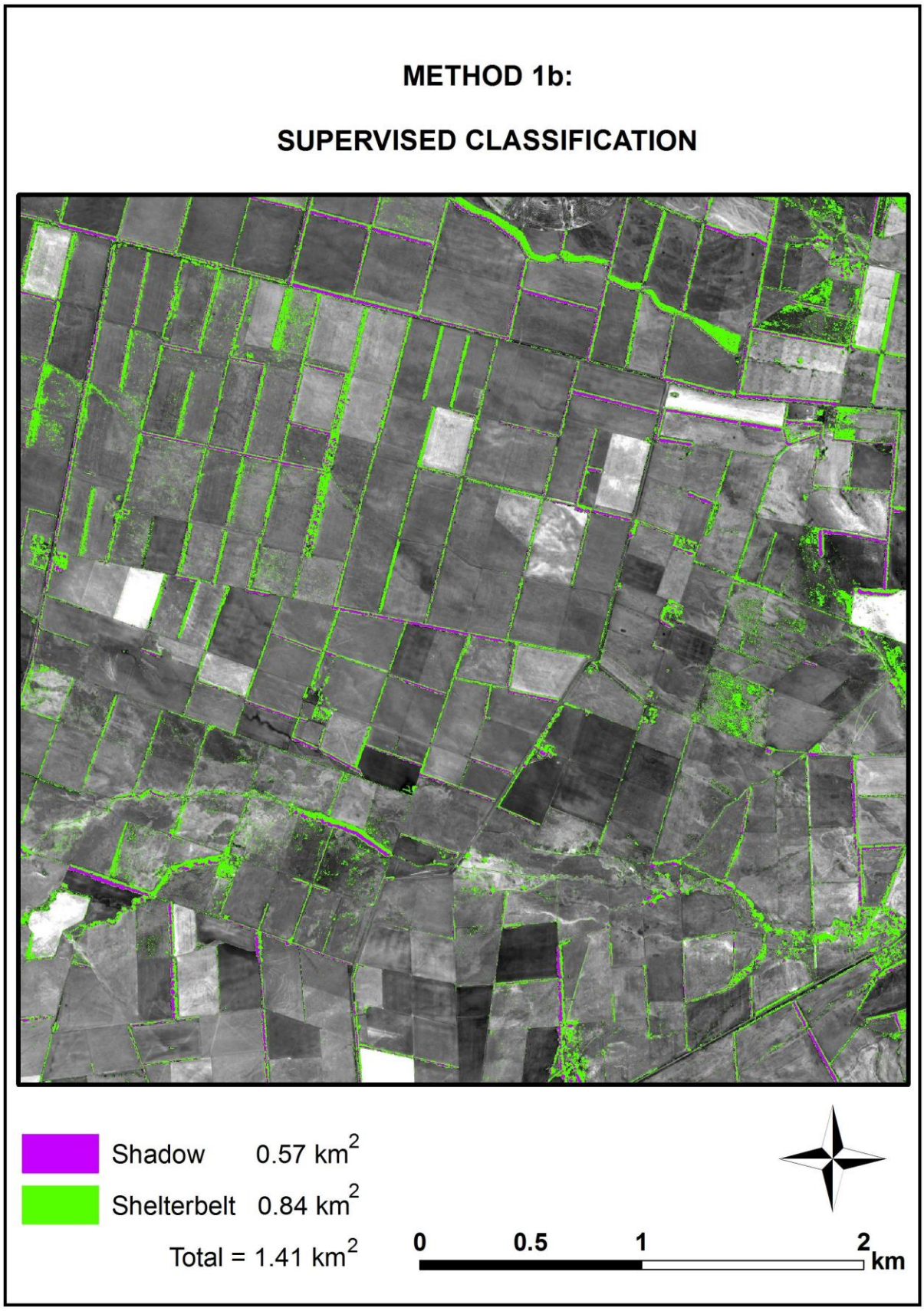


Figure 4.5 Supervised per-pixel classification showing 'shelterbelt' and 'shadow' classes.

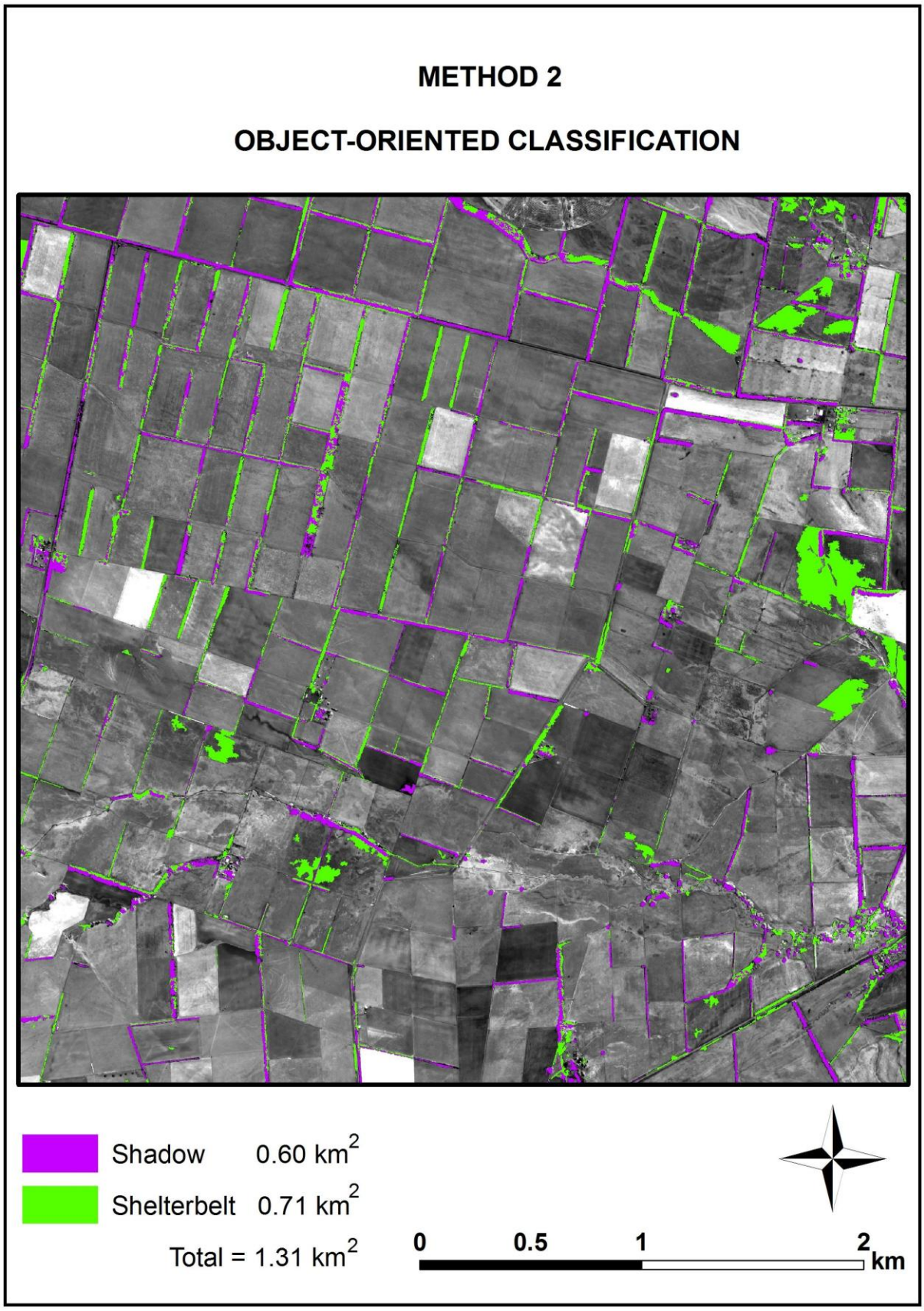


Figure 4.6 Object-oriented classification by image segmentation and labelling showing 'shelterbelt' and 'other' classes.

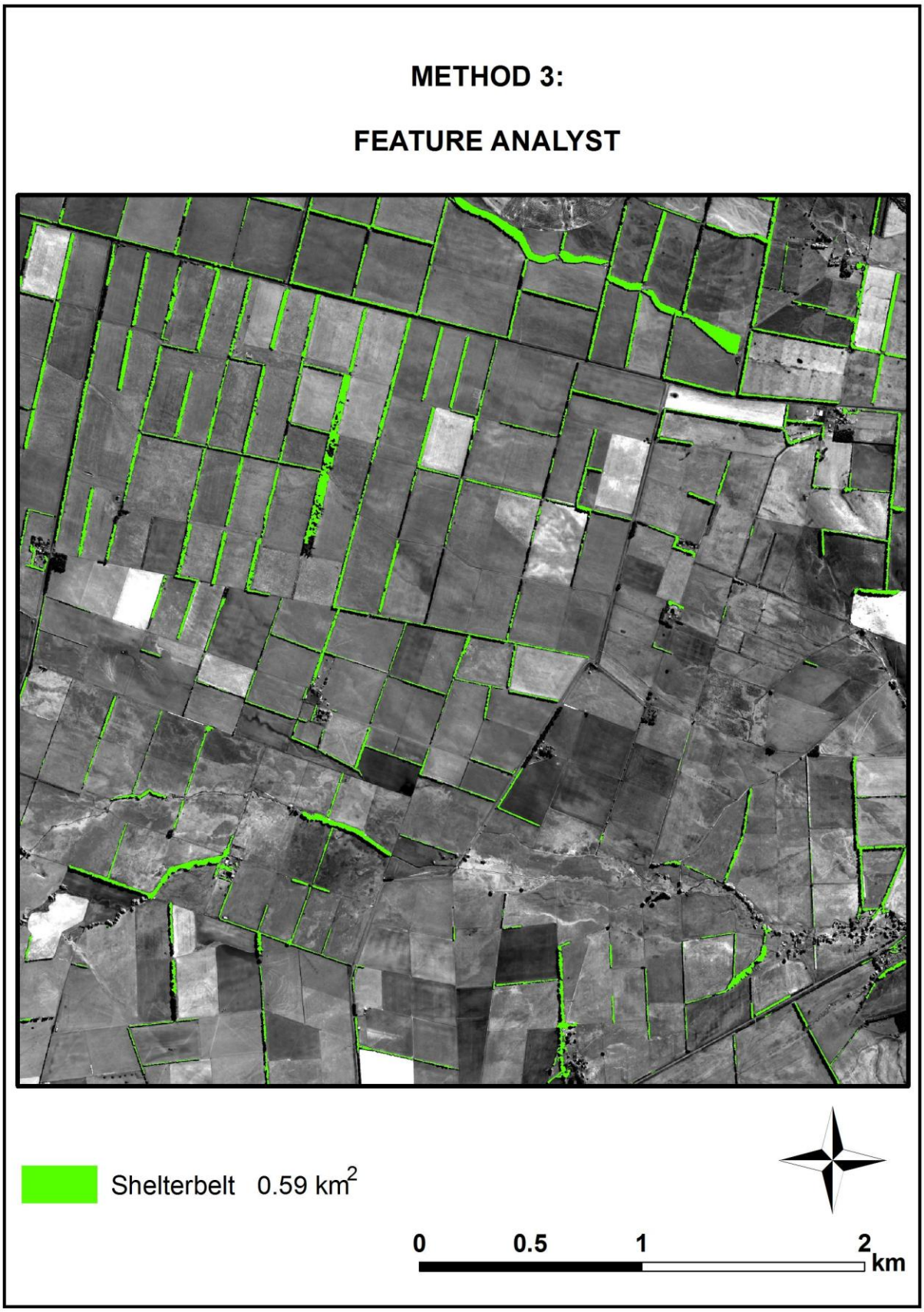


Figure 4.7 Feature extraction by Feature Analyst showing only the 'shelterbelt' class. No 'shadow' class is shown because features of interest are extracted by Feature Analyst without classifying the rest of the image.

4.4.2 Application of the Feature Analyst method to all three study areas

The batch processing tool identified 997 shelterbelt polygons across the three study areas. This gave a total shelterbelt area of 1.23 km^2 out of a 48 km^2 area. Shelterbelts therefore made up 2.6 % of the study areas. (Figures 4.8 to 4.10).

In comparing batch processing to individual processing of study area 1, batch processing extracted 0.01 km^2 of shelterbelt less than individual processing. However, overall accuracy for study area 1 remained approximately the same for individual and batch processing results.

Batch processing overall classification accuracy was approximately the same for all study areas (92 – 93 %) (Table 4.6). Overall, Feature Analyst correctly identified 73 % of shelterbelt area, with an overall classification accuracy of 92 %.

Table 4.6 Classification accuracy of Feature Analyst shelterbelt delineation by batch processing.

ALL THREE STUDY AREAS		Classification: Shelterbelt	Classification: Other	Total	PRODUCER'S ACCURACY
Actual	Shelterbelt	**40253**	15207	55460	73 %
	Other	17732	**349091**	366823	95 %
Total		57985	364298	422283	
USER'S ACCURACY		69 %	96 %		
%	TRUE	69 %	96 %	OVERALL ACCURACY = 92 %	
	FALSE	31 %	4 %		
		POSITIVE	NEGATIVE		

STUDY AREA 1		Classification: Shelterbelt	Classification: Other	Total	PRODUCER'S ACCURACY
Actual	Shelterbelt	**11260**	4473	15733	72 %
	Other	7298	**117730**	125028	94 %
Total		18558	122203	140761	
USER'S ACCURACY		61 %	96 %		
%	TRUE	61 %	96 %	OVERALL ACCURACY = 92 %	
	FALSE	39 %	4 %		
		POSITIVE	NEGATIVE		

STUDY AREA 2		Classification: Shelterbelt	Classification: Other	Total	PRODUCER'S ACCURACY
Actual	Shelterbelt	**15574**	6150	21724	72 %
	Other	5492	**113545**	119037	95 %
Total		21066	119695	140761	
USER'S ACCURACY		74 %	95 %		
%	TRUE	74 %	95 %	OVERALL ACCURACY = 92 %	
	FALSE	26 %	5 %		
		POSITIVE	NEGATIVE		

(Table 4.6 continued)

STUDY AREA 3		Classification		Total	*PRODUCER'S ACCURACY*
		Shelterbelt	**Other**		
Actual	**Shelterbelt**	**13419**	4584	18003	75 %
	Other	4942	**117816**	122758	96 %
Total		18361	122400	140761	
USER'S ACCURACY		73 %	96 %	***OVERALL ACCURACY* = 93 %**	
%	***TRUE***	73 %	96 %		
	FALSE	27 %	4 %		
		POSITIVE	***NEGATIVE***		

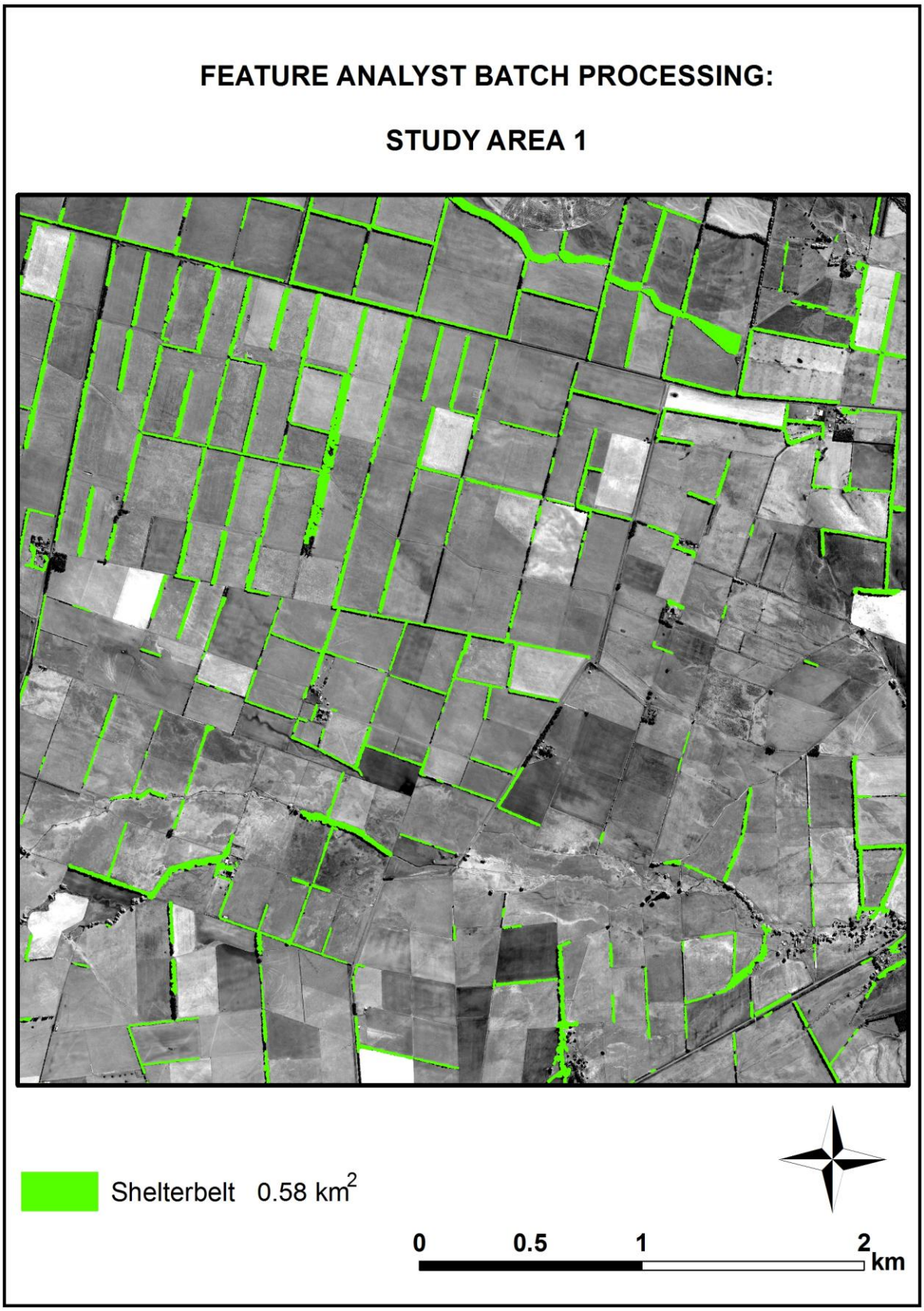


Figure 4.8 Shelterbelts extracted by Feature Analyst batch processing in study area 1.

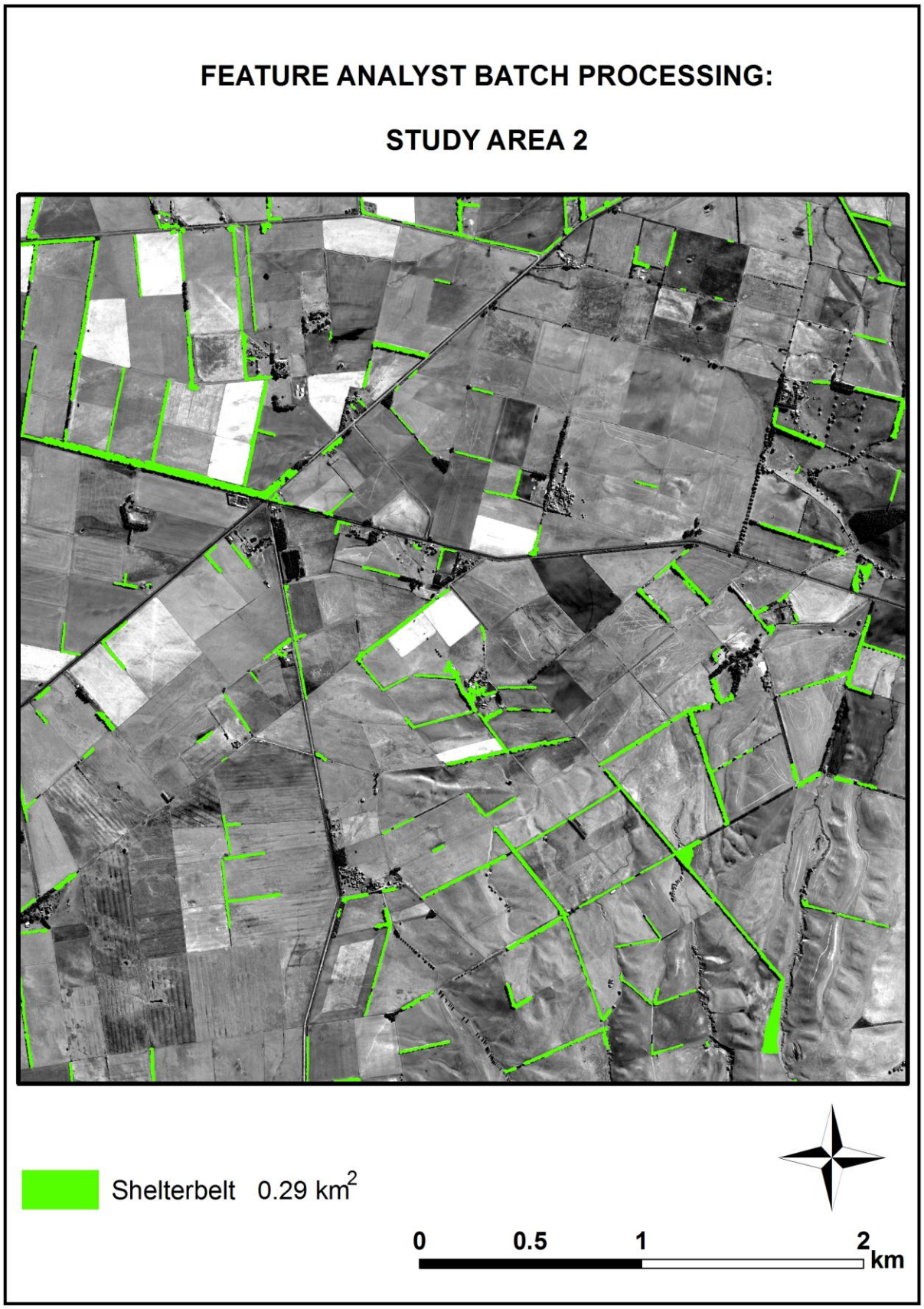


Figure 4.9 Shelterbelts extracted by Feature Analyst batch processing in study area 2.

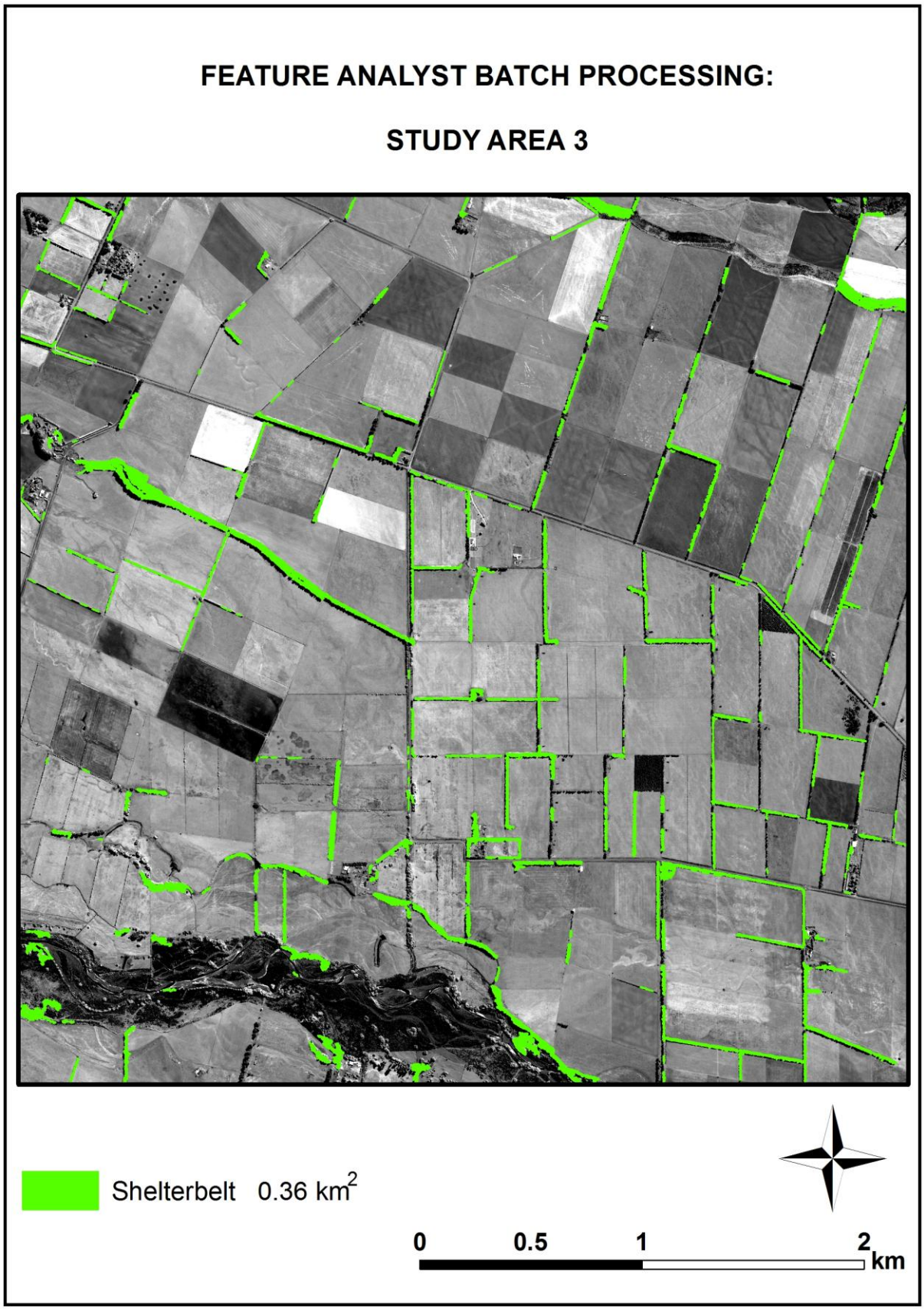


Figure 4.10 Shelterbelts extracted by Feature Analyst batch processing in study area 3.

4.5 Discussion

4.5.1 Comparison of shelterbelt delineation methods

Shelterbelt delineation was the most successful using the Feature Analyst method (Method 3), while per-pixel unsupervised and supervised classification methods (Methods 1a and 1b, respectively) and the object-oriented classification by image segmentation and labelling (Method 2) produced similar results to each other.

The area-based accuracy assessments indicated that Methods 1a and 2 were not particularly good at identifying shelterbelts, with only about 50 % of shelterbelt area correctly identified. Methods 2 and 3 were much better at identifying shelterbelts, with 86 % and 71 % of shelterbelt area correctly identified, respectively.

Both point-based and area-based accuracy assessments indicated that Method 3 was very good at excluding other land cover classes from the shelterbelt classification compared to Methods 1a, 1b and 2. The area-based and point-based accuracy assessments showed that the proportion of correctly identified shelterbelts in the shelterbelt classification was 17 to 33 % higher for Method 3 than other methods, respectively. Methods 1a, 1b and 2 did not differentiate between shelterbelts and their shadows, or between shelterbelts and non-shelterbelt trees. In addition, areas of paddock with tree-like spectral characteristics were included in the 'shelterbelt' class.

The actual classification accuracies of the methods are likely to lie between those reported by area-based and point-based assessments. This is because the area-based assessment only considered areas near shelterbelts and not the rest of the image, whereas the point-based assessment focused on the rest of the image. Method 3 (Feature Analyst) therefore had an overall classification accuracy of between 91 and 99 %. Both accuracy assessments show that the Feature Analyst method produced the best results, with a 4 to 7 % higher overall classification accuracy than other methods.

The Feature Analyst method was also superior to other methods according to a subjective comparison (Table 4.7). Feature Analyst was the most effective and user-friendly for several reasons. Feature Analyst is convenient to use, because it operates as an extension in ArcMap, has several built-in default settings for extracting different types of features, and produces the final output in vector (rather than raster) format. In addition, Feature Analyst has a feature modeler which can be used to view and save feature extraction models, as well as applying models to larger areas or similar images. Feature Analyst is therefore easy and comparatively fast to use. Learning how to use Feature Analyst was also easy as the reference manual is

comprehensive, and Overwatch technical support staff are accessible and helpful. Perhaps the most important benefit of Feature Analyst is that it is a semi-automatic feature extraction system, i.e. user feedback is used to refine results during hierarchical steps. The major disadvantage of using Feature Analyst was the cost of the software.

The superior performance of Feature Analyst confirms what is already well established in the literature: object-oriented classification improves per-pixel classification of linear features significantly (Johansen *et al.*, 2007; Quackenbush, 2004). However, the poor performance of object-oriented classification by image segmentation and labelling was disappointing: object labelling was only as good as the per-pixel classification used. Other methods for object labelling, such as data mining, are likely to improve results (Clinton, 2010). The BIS User's Guide advises using Weka, an open source data mining software (Clinton, 2010).

4.5.2 The Feature Analyst Method

Delineation of shelterbelts from high-resolution imagery by Feature Analyst is comparable to those methods in the literature using eCognition, the first feature extraction software to be commercially avaliable (Blaschke, 2010): eCognition is reported in the literature as extracting 80 % (Pankiw *et al.*, 2009) to 96 % (Wiseman *et al.*, 2009) of all shelterbelts, while Feature Analyst extracted 73 % of all shelterbelts; eCognition is reported as having an overall classification accuracy of 85 % (Liknes *et al.*, 2010) to 87 % (Vannier & Hubert-Moy, 2008), while Feature Analyst had an overall classification accuracy of 92 %. According to this, Feature Analyst may perform better than eCognition when extracting shelterbelts. However, this would need to be confirmed by using both software applications to classify the same image. The difference in accuracy may also be due to differences in shelterbelt characteristics between sites, such as shelterbelt width (Thornton *et al.*, 2006), and tree density (Vannier & Hubert-Moy, 2008), and perhaps differences in shelterbelt and hedgerow composition.

The accuracy of the Feature Analyst method can be improved by combining QuickBird imagery with active remotely-sensed data, such as LIDAR data. LIDAR data provides additional information about feature height, canopy reflectance and structure, and stem density (Ke *et al.*, 2010). It would therefore be valuable in extracting shelterbelts, particularly from flat landscapes such as the Canterbury Plains, eliminating shadow (Ke *et al.*, 2010), and for any subsequent analyses. For example, LIDAR data can be used for tree species differentiation (Holmgren *et al.*, 2008; Ke *et al.*, 2010), characterizing shelterbelt structure such as stem density (Ke *et al.*, 2010), and for estimating biomass (Kalaitzidis & Zianis, 2009; Orka *et al.*, 2009). Unfortunately acquiring LIDAR data was not feasible for this

research, as LIDAR data is expensive. However, LIDAR data may be worth the cost for large scale projects, such as carbon accounting.

Batch processing a larger area of the image using Feature Analyst did not decrease classification accuracy. This was unexpected, as one would expect the results to change due to a greater spectral and spatial variation of features within a larger area. For example, (Aksoy *et al.*, 2010) found that site specific training produced higher overall accuracies.

The method presented here shows promise for classification across a landscape despite having been developed and applied to a relatively small area. This is because there is potential to automate at least some parts of the methodology (Liknes *et al.*, 2010).

However, the major limitation of applying such a method across a landscape would be the need to correct images captured on different dates for differing atmospheric conditions. Atmospheric correction procedures can be used, but are often complicated and inaccurate if used incorrectly (Chavez, 1996). Absolute atmospheric correction is the most accurate but also the least practical method. In situ measurements of the atmospheric profile tend to be impossible to obtain in practice, especially if historical satellite images are used, as in this research (Chavez, 1996; Wu *et al.*, 2005). The alternative is relative (image-based) correction which normalizes the intensities of bands in images from different dates to a standard scene (Jensen, 2005). The images used in this research were captured on the same date, and were therefore not corrected. A method for correcting QuickBird imagery for atmospheric conditions is, however, needed for the future, but was beyond the scope of this research.

These results show that shelterbelts of a New Zealand landscape can be successfully extracted from high spatial resolution imagery in a semi-automated way. Batch processing results show promise for automated image processing and automatically updating databases, both of which are a recent focus of object-oriented image analysis research (Blaschke, 2010).

Semi-automatic shelterbelt delineation, as presented here has several ecological and economic applications, such as modelling the effectiveness of shelterbelts as wildlife corridors and habitat (Lechner *et al.*, 2009), improving the detail of rural habitat maps (Freeman & Buck, 2003), and quantifying the shelterbelt carbon pool across a given landscape (Wiseman *et al.*, 2009) (see Chapter 6, Example Application: Modelling Shelterbelt Carbon). Shelterbelt delineation could also be used to assess the need for shelterbelt-related agro-environmental policies, by mapping changes in shelterbelt land coverage. Shelterbelt-related policies and programs will be discussed further in Chapter 7 (Discussion).

In conclusion, the method presented here for shelterbelt delineation could be a valuable tool for natural resource management. However, to fully utilize this tool, methods to characterize shelterbelts are also needed. A method for classifying shelterbelts into species type is presented in Chapter 5 (Differentiation of Shelterbelt Species), with an example application of shelterbelt delineation and species differentiation provided in Chapter 6 (Example Application: Modelling Shelterbelt Carbon).

Table 4.7 Subjective comparison of methods showing the superiority of Feature Analyst.

	APPROACH	METHOD	SOFTWARE			OUTPUT	GOODNESS OF CLASSIFICATION			
			Software	Cost	User-friendly		False Positives			False negatives
							Shadow	Ground	Other trees	
1	**Per-pixel classification**	**a) Unsupervised** **b) Supervised**	ArcGIS & Image Analyst (IA)	No extra (have licence already) ★★★★	Yes ★★★☆	Raster file ★☆☆☆	Significant ★☆☆☆	Significant (some areas of paddock) ★☆☆☆	Cannnot distinguish ★☆☆☆	Significant (in shadow class) ★☆☆☆
2	**Object-oriented classification**	**Image segmentation followed by labelling with majority statistics**	Berkeley Image Segmentation (BIS); ArcGIS & IA	BIS: Unrestricted free trial; AUS$1150 for single seat licence ★★☆☆	Yes – but time-consuming to get segmentation parameters correct ★★☆☆	Vector file; Compatible with ArcGIS ★★★★	Significant ★☆☆☆	Significant (some areas of paddock) ★☆☆☆	Cannnot distinguish ★☆☆☆	Significant (in shadow class) ★☆☆☆
3	**Feature extraction with specialist software**	**Feature Analyst**	Feature Analyst (FA)	US$1000 for single seat licence ★☆☆☆	Yes – Functions as an extension to ArcGIS ★★★★	Vector file; Compatible with ArcGIS ★★★★	Not significant ★★★★	Not significant ★★★★	Can distinguish to a reasonable level ★★★★	Not significant ★★★★

Chapter 5
Differentiation of Shelterbelt Species

5.1 Introduction

Once shelterbelts are identified, it is often necessary to know which species comprise different shelterbelts. For example, differentiating shelterbelt tree species is important for estimating shelterbelt biomass and carbon (Wiseman *et al.*, 2009). Species information may also be useful for habitat suitability modelling, such as considering native versus exotic tree species (Buytaert, 2007; Lechner *et al.*, 2009).

Automated methods that can differentiate shelterbelt tree species using information derived from remotely-sensed data are therefore needed. Species classification of shelterbelt objects is rare in the literature. Most studies differentiate tree species from remotely-sensed imagery at an individual tree crown level for the purpose of estimating forest composition (Brandtberg, 2002; Erikson, 2004; Förster & Kleinschmit, 2006; Hájek, 2008; Holmgren & Persson, 2004; Holmgren *et al.*, 2008; Katoh *et al.*, 2009; Leckie *et al.*, 2003; Olofsson *et al.*, 2006; Orka *et al.*, 2009; Suratno *et al.*, 2009; Waser *et al.*, 2011). No work has, however, been done on differentiating the major shelterbelt species of Canterbury, *P.radiata* and *C.macrocarpa*. Differentiating these coniferous species may be problematic as spectral and spatial variables of similar species often overlap (Förster & Kleinschmit, 2006; Hájek, 2008). Differentiating shelterbelt tree species may be difficult because information about individual crown shape and structure is lost, as well as considering that shelterbelts are often pruned and may be a mixture of species or age classes.

This chapter investigates whether shelterbelt species can be differentiated using remotely-sensed spectral information. The results from field data collection (see Chapter 3, General Methods) are presented in this chapter, as this dataset was used for the analyses conducted in this chapter. This chapter explored shelterbelt spectral variability to determine which spectral variables are likely to explain shelterbelt variability. This chapter also presents a Random Forests method for shelterbelt species differentiation.

5.2 Overview of methods used for species discrimination

Many different image-derived variables can be used for species classification. Selected variables, as well as the importance of variables, vary depending on the species being differentiated and the imagery being used (see Chapter 2, Literature Review). Analysis techniques which examine all variables simultaneously must be used to uncover patterns and key features within the data (Everitt, 2005).

The interpretation of remotely-sensed datasets require multivariate analysis techniques, because remotely-sensed datasets usually contain multiple variables. Interpretation of multivariate data can be assisted using data reduction or simplification techniques, such as principal component analysis (PCA) (Johnson & Wichern, 2007). PCA explains the structure of the dataset through a few linear combinations or principal components, which are orthogonal (not correlated). In this way, PCA reduces the dimensiality of the data (Chatfield & Collins, 1980; Johnson & Wichern, 2007). Principal components are derived in decreasing order of importance, so that the first principal component accounts for the largest proportion of variation within the data (Chatfield & Collins, 1980). Outputs of PCA include: a correlation matrix, which provides correlation values between variables; loadings which measure the correlation between principal components and variables (i.e. variable importance); and biplots of principal components which reveal groupings of variables (Chatfield & Collins, 1980). PCA is mainly used as an exploratory analysis, but principal components can also be used as inputs to multiple regression or cluster anlalysis (Chatfield & Collins, 1980; Johnson & Wichern, 2007).

There are several possible ways to model relationships between variables in multivariate data. For example, regression analysis, discriminant analysis, and cluster analysis. Other techniques, such as data mining, require appropriate software and fast computers (Johnson & Wichern, 2007). Multiple regression analysis investigates the relationships between various responses (e.g. species classes) and a set of predictor variables (Everitt, 2005; Johnson & Wichern, 2007): a model is fit graphically, and then that fit is evaluated (Chatterjee & Hadi, 2006). In contrast, multiple discriminant analysis and classification procedures construct classification rules (from a training set) that predict the class of an object (Carroll & Green, 1997; Everitt, 2005; Johnson & Wichern, 2007). Discriminant analysis describes graphically or algebraically the differential features of objects from two or more known classes (e.g. different tree species); whereas classification sorts new objects into two or more classes using rules derived from objects within known classes (Johnson & Wichern, 2007). Regression analysis and discriminant analysis assume that the data are normally distributed; whereas

classification techniques, such as cluster analysis, do not (Chatfield & Collins, 1980). In addition, regression analysis for species classification may be problematic if there are multiple species classes, and because model linearity is often difficult to assess for high dimensional data (Chatterjee & Hadi, 2006).

In contrast to regression and discrimant analysis, cluster analysis finds the natural groupings of objects: similar objects are allocated to the same group, e.g. hierarchical trees (Afifi *et al.*, 2004; Chatfield & Collins, 1980). Hierarchical trees work in successive steps by splitting off objects which are the most dissimilar to the remaining objects (Afifi *et al.*, 2004). Classification trees predict the classes of a categorical variable, whereas regression trees predict the value of a continuous variable.

An example of a hierarchical tree method is the Random Forests method. Random Forests (Breiman, 2001) is an iterative learning method which generates many classification trees and aggregates the results (Liaw & Wiener, 2002). Random Forests works by independently constructing each tree using a random sample of the data set. Each tree node is split using the best among a randomly chosen subset of predictor variables (Liaw & Wiener, 2002).

The Random Forests method was selected for shelterbelt species differentiation, because it does not assume a normal distribution, it is iterative, it is resilient to overfitting due to the large number of trees used, and it validates the model using a cross validation procedure which does not require new data. Random Forests is also easy to use as it is not particularly sensitive to the two parameters set by the user (the number of predictor variables used, and the number of trees). In addition, the Random Forests method has performed very well compared to other classification procedures (Breiman, 2001; Liaw & Wiener, 2002).

5.3 Methods

Differentiation of shelterbelt objects into tree species type was carried out in three parts. First, spectral information was extracted for digitized shelterbelt objects for which field data were collected (see Chapter 3, General Methods). Secondly, Principal Component Analysis (PCA) was used to explore the data. Thirdly the Random Forests algorithm was used to to create a series of classification trees, which were then used to predict the species class of shelterbelts delineated by the Feature Analyst Method (see Chapter 4, Image-based Shelterbelt Delieation).

5.3.1 Spectral and textural information

Two spectral channels and a measure of texture additional to the red, green, blue and near-infrared bands of the satellite imagery were calculated. Additional spectral channels were a near infrared/green band ratio (NI/G), as this is useful for tree species differentiation (Hájek, 2008); as well as the NDVI produced during image preparation (see Chapter 3, General Methods). A simple measure of texture based on standard deviation was calculated from the panchromatic image using the Image Analysis extension for ArcGIS 9.3 (ERDAS, 2009). The standard deviation was calculated for pixels inside a moving 3 x 3 rectangular window.

The Zonal Statistics tool in ArcMap (ESRI, 2009a) was used to extract spectral and textural information for digitized shelterbelts and Feature Analyst shelterbelts (Table 5.1; Appendix D.1). Values relating to spectral mean and variety were extracted from red (R), green (G), blue (B), and near infrared (NI) bands, as well as the NDVI, NI/G band ratio, and texture.

The information extracted for digitized shelterbelts (see Chapter 3, General Methods) was used for data exploration and to train the Random Forests model. This is because digitized shelterbelts had corresponding field data, which included species information. The Random Forests model was then used to predict the species class of Feature Analyst shelterbelts using the spectral and textural information extracted from them.

Table 5.1 Spectral and textural variables extracted for digitized and Feature Analyst shelterbelts.

		SUMMARY STATISTICS EXTRACTED FOR EACH SHELTERBELT OBJECT								
		Min	Max	Range	Mean	Std	Variety	Majority	Minority	Median
BAND	Red	√	√	√	√	√	√	√	√	√
	Green	√	√	√	√	√	√	√	√	√
	Blue	√	√	√	√	√	√	√	√	√
	NI	√	√	√	√	√	√	√	√	√
OTHER	NDVI	√	√	√	√	√				
	NI/G	√	√	√	√	√				
	Texture	√	√	√	√	√				

Selected variables for PCA 1 with reduced variables

5.3.2 Data exploration

Exploration of remotely-sensed data for digitized shelterbelts was carried out to provide insight into the structure of the data (e.g. intercorrelation between variables), and to investigate whether the structure of the data differed for *P.radiata* and *C.macrocarpa* shelterbelts.

The image was first examined to see if species could easily be differentiated visually. Of particular interest was whether or not the difference between broadleaved and coniferous shelterbelts could be seen.

Next, the statistical software package R (The R Foundation for Statistical Computing, 2010) was used to carry out Principal Component Analysis (PCA) to explore the variability within the data. PCA was conducted on single species *P.radiata* shelterbelts (n = 58) and single species *C.macrocarpa* shelterbelts (n = 26) separately. PCA was initially conducted using all spectral and textural variables. However, this analysis was difficult to interpret due to a high colinearity among variables. PCA was then conducted on selected variables (n = 18), which were means and standard deviations for spectral bands, band ratios and texture; as well as variety values for spectral bands (PCA 1) (Table 5.1). These variables were chosen as they described the average value and variation of values for spectral channels and the measure of texture. PCA results (i.e. correlation matrices, loadings, and biplots) were compared for *P.radiata* and *C.macrocarpa* to determine if any structural differences within the data were

present, and to determine which variables were the most important in describing each PCA axis. These variables were selected for PCA in Chapter 6 (Example Application: Modelling Shelterbelt Carbon) which investigated which image-derived variables are likely to be important predictors of shelterbelt physical characteristics (PCA 2).

5.3.3 The Random Forests algorithm for species discrimination

The Random Forests algorithm (Breiman & Cutler, 2010) was run in R using information extracted from digitized shelterbelts (with species information). All spectral and textural variables (n = 51) were included as predictor variables, as collinearity among variables is not an issue with Random Forests (Breiman, 2001). Random Forests was set to run 5000 trees so that the importances of all 51 variables would be stable (Breiman, 2002). The number of variables randomly selected at each node to determine the split was experimented with to find the optimal number (Breiman, 2001, 2002). Initially it was set at 7 (squareroot of 51), then 3 (half of 7), 14 (double 7) and 25 (half of 51). The optimal number of variables sampled at each node were determined to be 14 and 7 for differentiation of species groups and coniferous species, respectively.

Random Forests was used to create models to predict whether digitized shelterbelts were broadleaved or coniferous (Figure 5.1 A), and the species type of coniferous digitized shelterbelts (Figure 5.1 C). These Random Forests models were then used to predict whether shelterbelts delineated using the Feature Analyst method (Method 3 in Chapter 4) were broadleaved or coniferous (Figure 5.1 B), and to predict the species type of coniferous Feature Analyst shelterbelts (Figure 5.1 D). Random Forests was trained using digitized shelterbelts rather than Feature Analyst shelterbelts, because Feature Analyst (Overwatch, 2010) did not identify all shelterbelts for which field data were available, and the area covered by Feature Analyst shelterbelts was not always consistent with the area covered by the digitized shelterbelts (e.g. Feature Analyst sometimes included areas of shadow in shelterbelt objects).

The predicted class of Feature Analyst shelterbelts for which field data was available (see Chapter 3, General Methods), was compared to their actual class to estimate the accuracy of the prediction. This estimated accuracy was then compared to the Random Forests "out of the bag" (OOB) error rate to determine whether species classification of automatically delineated shelterbelts is possible. OOB error is estimated by predicting the class of the objects not in the random ("out of the bag") sample at each iteration. The errors estimated at each iteration are then aggregated to get an overall estimate of error (Liaw & Wiener, 2002).

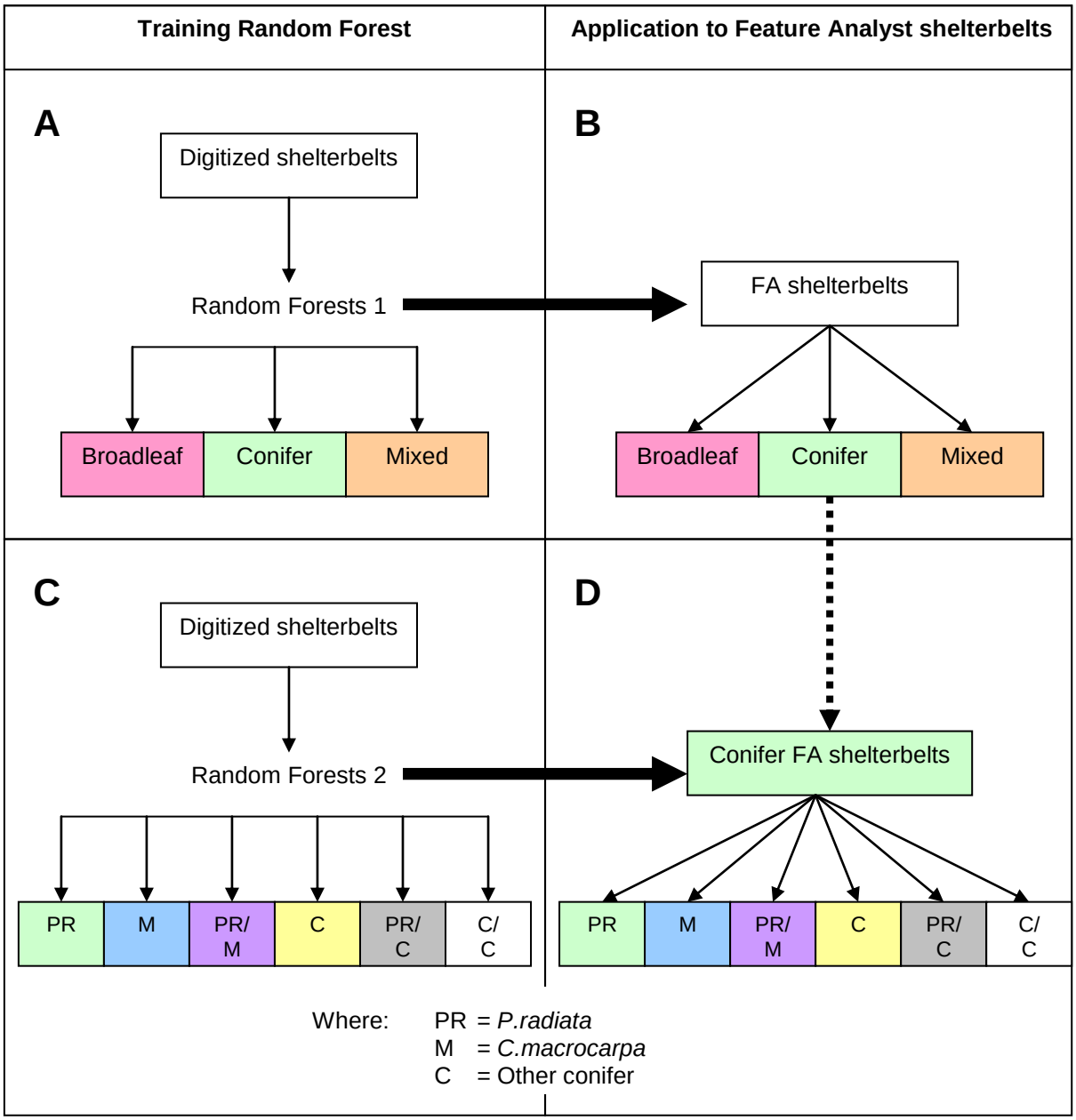


Figure 5.1 Flowchart showing the training of Random Forests algorithms to differentiate species groups (A) and coniferous species (C), and their subsequnt application to predict the species group (B) and coniferous species (D) of Feature Analyst shelterbelts.

5.4 Results

Results from field data collection (see Chapter 3, General Methods), PCA data exploration, and Random Forests are presented in this section.

5.4.1 Data collection

A total of 182 observations were recorded in the three study areas. This included a sample of 133 shelterbelts parallel to the road. The remaining 49 observations were of vegetation features, such as hedges and groups of trees (Appendix B.2). These were recorded to aid image interpretation (e.g. during accuracy assessment), and were not used for image analysis.

Unfortunately 17 out of the 133 shelterbelts were not clearly visible on the satellite image and could therefore not be used for tree species differentiation and carbon estimation. This had a major impact on the number of broadleaved shelterbelts for analysis, with half of all broadleaved shelterbelts eliminated (Appendix B.3). Shelterbelts not visible on the images were significantly different to visible shelterbelts (Appendix B.4): Visible shelterbelts were wider by 3-5 m, longer by 7-19 m, taller by 6-8 m[*], had a larger tree diameter by 23-27 cm, and had 15-20 less trees per 100 m^2. Therefore, it is reasonable to conclude that shelterbelts not visible on the images may be more recently planted, and therefore too small to see on the image. In fact, some shelterbelts (i.e. seedling and sapling) would not have been planted at the time of image capture. Only the 116 visible shelterbelts are considered further.

Results from field work support that shelterbelts in Canterbury are dominantly *P.radiata* and *C.macrocarpa* (Price, 1993): 50% of the sample was *P.radiata*, 22 % was *C.macrocarpa*, and a further 10 % was a mix between *P.radiata* and *C.macrocarpa*. Coniferous shelterbelts made up 93 % of shelterbelt length, with only 4 % of shelterbelt length being broadleaf (Table 5.2).

Shelterbelts had an average width of 8 m, length of 158 m, height of 12 m, and average tree diameter of 42 cm (Table 5.3). Shelterbelts had an average of 1.6 rows of trees with a density of 14 trees/100m^2. Almost all shelterbelts were in good health, and 87 % of shelterbelts had a crown ratio of 90-100 %. About half of shelterbelts were pruned, with half being unpruned. Shelterbelts did not vary significantly among study areas.

Coniferous shelterbelts differed significantly from broadleaved shelterbelts in length, height, tree diameter, management and crown ratio (Appendix B.5): Broadleaved shelterbelts were shorter by 72-84 m, taller by 2-6 m, and had a smaller tree diameter by 12-16 cm.

[*] Shelterbelt height in this chapter refers to the height of the tallest row. The tallest height was considered in this chapter, as it was assumed that the tallest row would have the most influence on spectral characteristics.

Broadleaved shelterbelts were all unpruned, compared to 46 % of coniferous shelterbelts being pruned, and only 25 % of broadleaved shelterbelts had a crown ratio of 90-100 %, compared to 93 % of coniferous shelterbelts.

Mixed species shelterbelts varied in composition: e.g. the percentage of *P.radiata* in *P.radiata* and *C.macrocarpa* mixed (PR/M) shelterbelts varied between 20 and 95 %, with only about a third of PR/M shelterbelt length consisting of an equal proportion of species (Figure 5.2).

The greatest number of shelterbelts were visited in study area 1 (n = 43), followed by study areas 2 (n = 38) and 3 (n = 37). Broadleaved shelterbelts were not equally represented in all study areas, with none being sampled in study area 2, only poplar in study area 1, and only eucalyptus in study area 3 (Table 5.2; Figures 5.3 to 5.5).

Table 5.2 Summary of shelterbelt species sampled in study areas 1, 2 and 3 by shelterbelt number and length.

	SPECIES	STUDY AREA 1		2		3		TOTAL		PERCENTAGE	
		#	km	#	km	#	Km	#	km	#	km
Conifer	*P.radiata* (P)	24	4.67	14	1.47	21	3.41	58	9.55	50 %	52 %
	C.macrocarpa (M)	8	0.62	15	2.58	4	0.40	26	3.59	22 %	20 %
	P.radiata & *C.macrocarpa* *****	4	1.27	4	0.71	4	0.74	12	2.72	10 %	15 %
	P.radiata & Other Conifer *****	1	0.41	2	0.22			3	0.63	3 %	3 %
	Other Conifer (C)			1	0.06	3	0.51	4	0.57	3 %	3 %
	TOTAL CONIFER	**37**	**6.96**	**36**	**5.04**	**32**	**5.06**	**103**	**17.06**	**89 %**	**93 %**
Broadleaf	*Eucalyptus spp.* (E)					3	0.42	3	0.42	3 %	2 %
	Populus spp. (P)	2	0.09					2	0.09	2 %	< 1 %
	Mixed Native	1	0.03					1	0.03	1 %	< 1 %
	Populus spp. & Other Broadleaf *****	1	0.04					1	0.04	1 %	< 1 %
	Other Broadleaf (B)					1	0.13	1	0.13	1 %	1 %
	TOTAL BROADLEAF	**4**	**0.17**	**0**	**0**	**4**	**0.54**	**8**	**0.71**	**7 %**	**4 %**
Mixed	Conifer & Broadleaf spp. *****	2	0.39	2	0.15	1	0.03	5	0.56	4 %	3 %
	TOTAL MIXED	**2**	**0.39**	**2**	**0.15**	**1**	**0.03**	**5**	**0.56**	**4 %**	**3 %**
	TOTAL	**43**	**7.52**	**38**	**5.19**	**37**	**5.62**	**116**	**18.33**		

Table 5.3 Summary of physical characteristics of field-sampled shelterbelts by study area, and for all three study areas combined.

SUMMARY STATISTICS		PHYSICAL CHARACTERISTICS					
		Width (m)	Length (m)	Rows	Height of tallest row (m)	DBH (cm)	Density (trees/ 100m^2)
Study area 1	Mean	7.0	178	1.6	13.0	43.2	12.7
	Range	15.5	722	5	25.3	138	48.9
	Std	3.2	168	1.0	6.0	26.6	10.2
	Sample size	43					
Study area 2	Mean	9.0	146	1.8	10.0	44.4	15.9
	Range	29.5	389	8	35.1	62.8	118.5
	Std	8.3	105	1.6	8.8	17.7	22.3
	Sample size	38					
Study area 3	Mean	6.6	152	1.5	12.1	37.2	12.4
	Range	9.8	572	2	22.9	68.2	34.5
	Std	3.1	132	0.7	6.7	15.3	7.8
	Sample size	37					
All study areas	Mean	7.6	158	1.6	11.8	41.9	13.7
	Range	29.5	733	8	35.1	143.2	1.32
	Std	5.5	141	1.1	7.4	21.3	14.98
	Sample size	116					

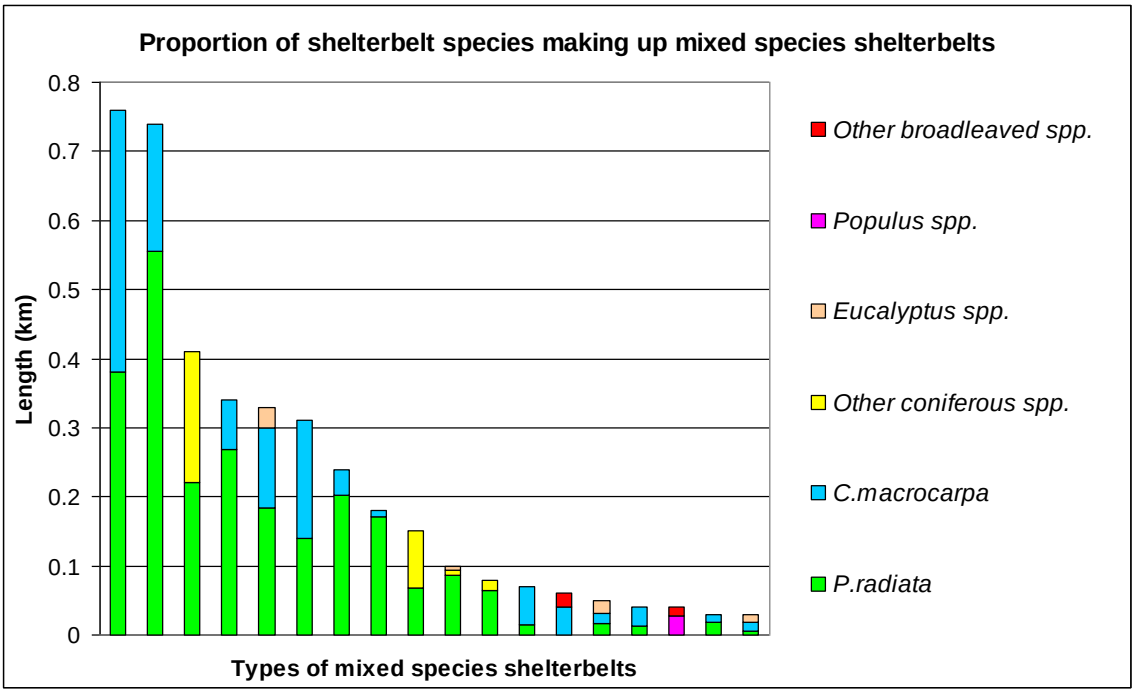


Figure 5.2 Summary of field-sampled mixed shelterbelts by species proportions and length.

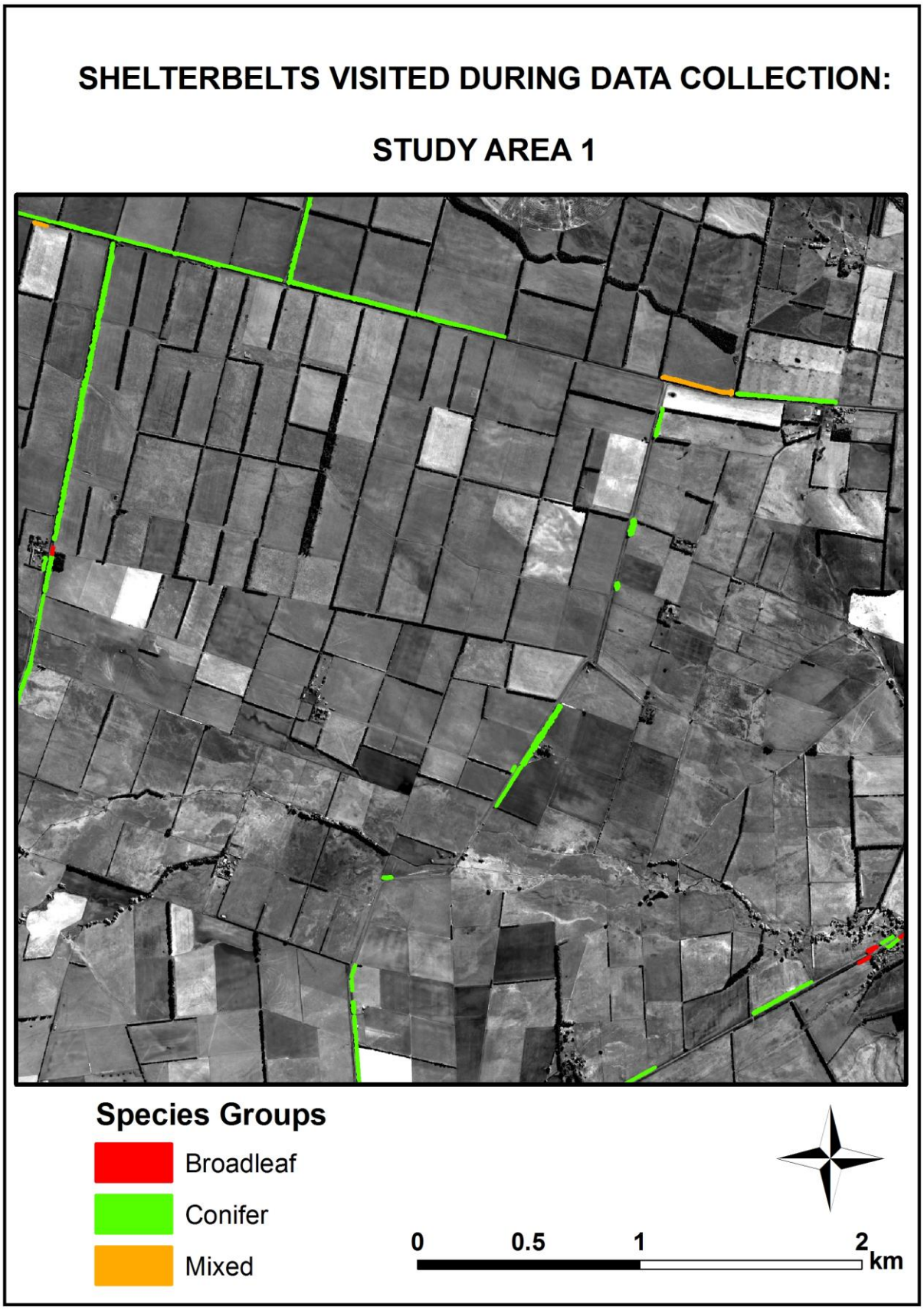


Figure 5.3 Shelterbelts sampled during field data collection in study area 1.

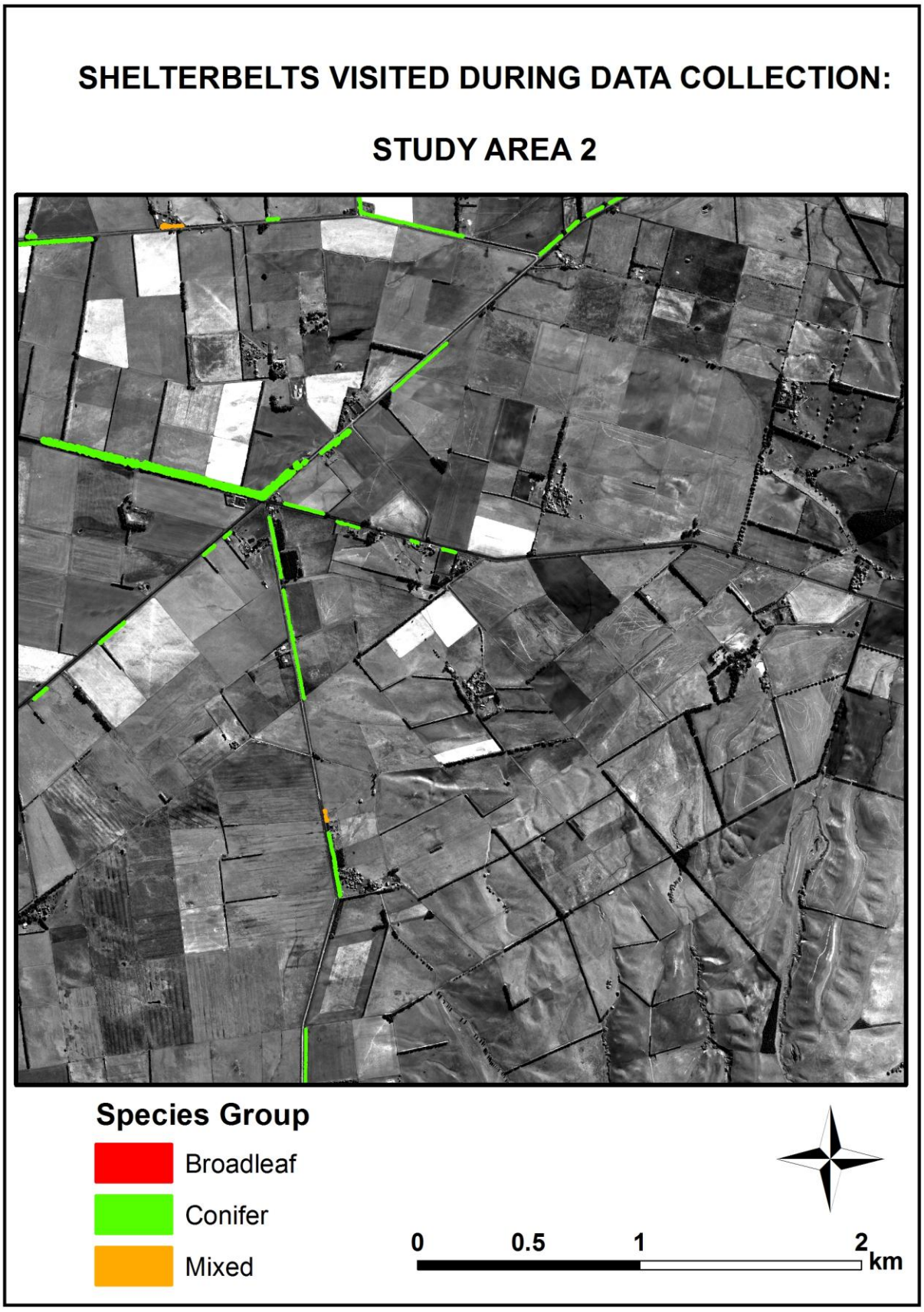


Figure 5.4 Shelterbelts sampled during field data collection in study area 2.

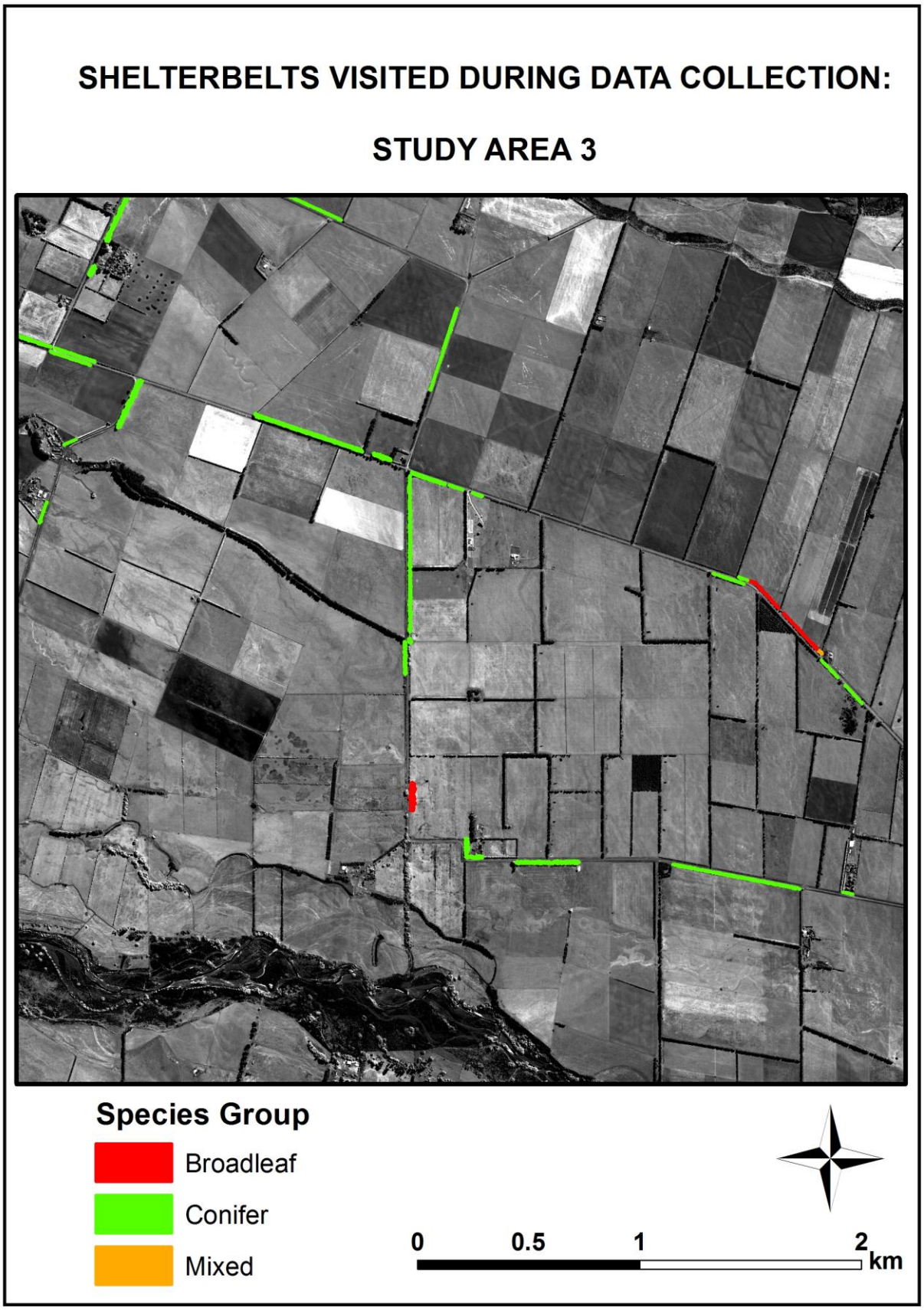


Figure 5.5 Shelterbelts sampled during field data collection in study area 3.

5.4.2 Data exploration

Visual examination of the image indicated that broadleaved shelterbelts could not always be easily differentiated from coniferous shelterbelts. It appeared that *Eucalyptus spp*. looked the most different compared to other broadleaf species. However, *C.macrocarpa* shelterbelts appeared to vary in appearance, and could be confused with broadleaved shelterbelts as well as *P.radiata* shelterbelts (Appendix D.2).

PCA 1 indicated that variability within spectral data could not easily be explained by only a few variables. Loadings were not remarkably high for any of the variables, indicating that there is a high degree of colinearity among image-derived variables (Table 5.5). The first two principal component axes explained 66 and 70 % of spectral variability within *C.macrocarpa* and *P.radiata* shelterbelts, respectively (Table 5.4).

The structure of spectral data was very similar for *P.radiata* and *C.macrocarpa*. Spectral variables were reduced to three non-correlated principal component (PC) axes (Figures 5.6 and 5.7), the third of which differed between species: (i) PC axis 1 described spectral means; (ii) PC axis 2 described spectral variation; and (iii) PC axis 3 described texture for *P.radiata*, and NI/G values for *C.macrocarpa*. NI/G values for *P.radiata* were included in PC axes 1 and 2; while texture values for *C.macrocarpa* were included in PC axis 2. The differences in the third axis provide an indication that NI/G and texture values may be important in differentiating these two coniferous species.

Table 5.4 Proportion of variance explained by principal components.

	PCA 1					
	IMAGE-DERIVED VARIABLES ONLY					
	P.radiata			***C.macrocarpa***		
	PC1	**PC2**	**PC3**	**PC1**	**PC2**	**PC3**
Proportion	53 %	17 %	12 %	43 %	22 %	12 %
TOTAL	53 %	70 %	82 %	43 %	66 %	77 %

Table 5.5 PCA Loadings indicating the importance of variables.

		PCA 1 IMAGE-DERIVED VARIABLES ONLY					
		P.radiata			*C.macrocarpa*		
		PC1	PC2	PC3	PC1	PC2	PC3
R	Mean	0.24	-0.37	0.10	-0.20	0.38	-0.08
	Std	-0.22	-0.30	0.11	0.26	0.30	-0.07
	Variety	-0.25	-0.18	0.14	0.27	0.19	0.06
G	Mean	0.23	-0.37	-0.10	-0.22	0.37	-0.07
	Std	-0.30	-0.14	-0.05	0.32	0.17	-0.05
	Variety	-0.30	-0.09	0.05	0.32	0.06	0.08
B	Mean	0.23	-0.33	-0.14	-0.21	0.37	-0.06
	Std	-0.31	-0.05	-0.07	0.33	0.04	-0.17
	Variety	-0.31	-0.04	0.00	0.33	0.03	0.08
NI	Mean	0.14	-0.26	-0.49	-0.23	0.33	0.03
	Std	-0.27	-0.11	0.09	0.24	0.04	-0.08
	Variety	-0.30	-0.03	0.10	0.28	-0.06	0.21
NDVI	Mean	-0.18	0.22	-0.48	0.06	-0.22	0.25
	Std	-0.21	-0.24	0.21	0.28	0.19	0.07
NI/G	Mean	-0.16	0.12	-0.55	0.01	0.26	0.53
	Std	-0.27	-0.06	-0.20	0.00	0.27	0.53
T	Mean	-0.02	-0.32	-0.21	0.17	0.20	-0.35
	Std	-0.05	-0.41	-0.08	0.10	0.19	-0.36

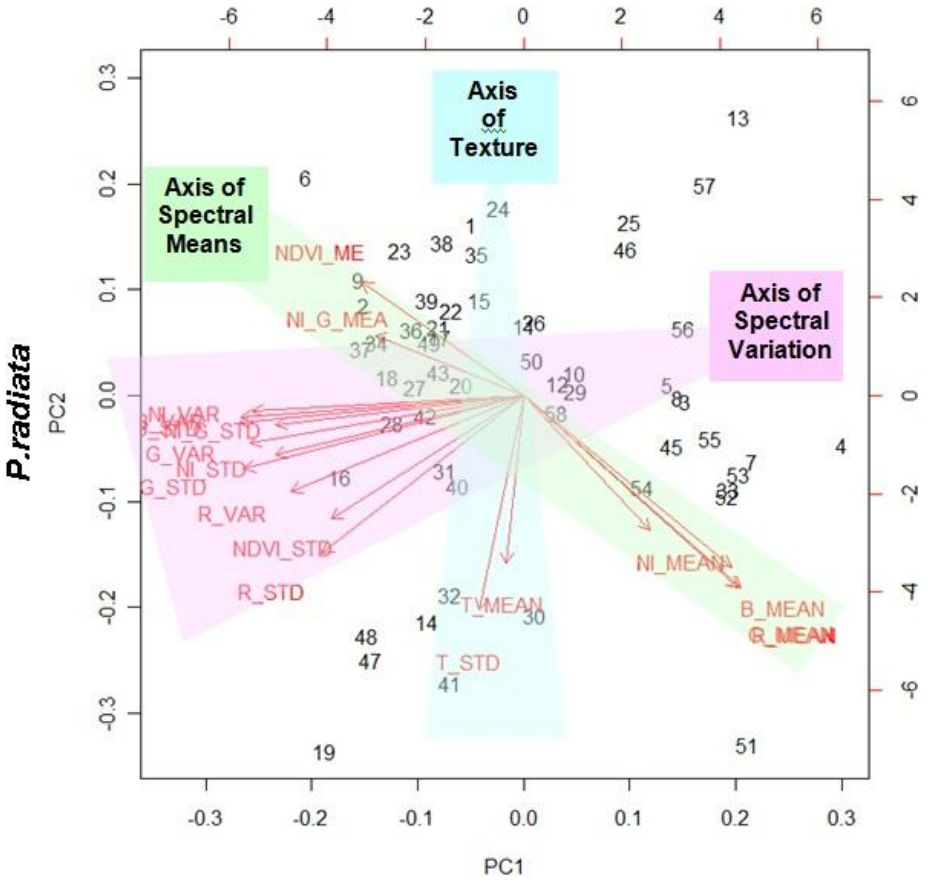


Figure 5.6 Biplot of PC1 and PC2 of image-derived variables for *P.radiata* (PCA 1).

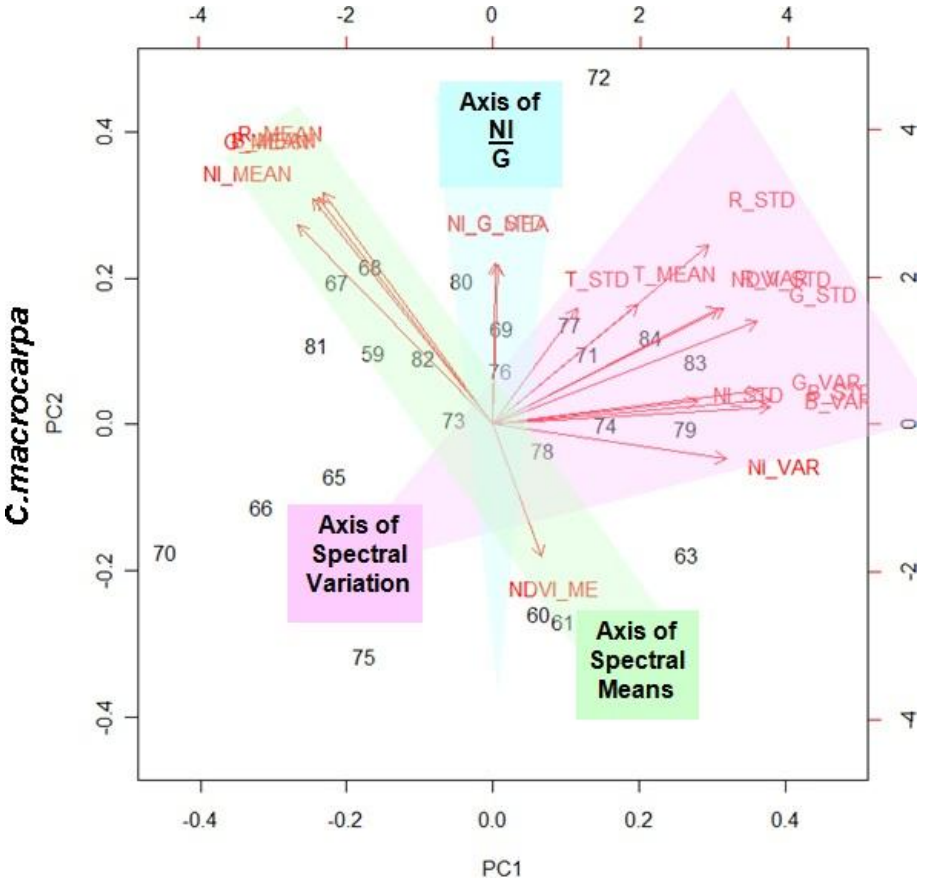


Figure 5.7 Biplot of PC1 and PC2 of image-derived variables for *C.macrocarpa* (PCA 1).

5.4.3 Random Forests

Differentiation of field-identified shelterbelts into 'broadleaf', 'conifer', and 'mixed' species groups produced good results, with an out of the bag (OOB) error rate of only 9 % (Table 5.6). Application of the species group model to Feature Analyst shelterbelts resulted in the identification of 1027 coniferous, 83 broadleaved, and two mixed shelterbelts. Application of the model to Feature Analyst shelterbelts resulted to an even lower error rate (6 %) than the error estimated during Random Forest training.

The classification of the 'conifer' class into species type resulted in an OOB error rate of 39 % (Table 5.7). The error rate increased by 5 % when the model was applied to Feature Analyst shelterbelts (Table 5.7). Random Forests were good at classifying *P.radiata* shelterbelts, with only 2 % of *P.radiata* shelterbelts classed as a different species. However, an estimated 46 % of shelterbelts identified as *P.radiata* actually belonged to another species class. For example, 67 % of *C.macrocarpa* shelterbelts were wrongly classified as *P.radiata*. The classification of mixed species shelterbelts was even poorer, with only one out of 1027 coniferous shelterbelts being classed as *P.radiata*/*C.macrocarpa*. This classification identified 936 *P.radiata*, 88 *C.macrocarpa*, one mixed species (*P.radiata*/*C.macrocarpa*) and two other coniferous species shelterbelts.

Blue, red, and near infrared bands, the NDVI, as well as texture were important for differentiating conifer, broadleaf and mixed species groups (Appendix E.3). The most important predictors were blue majority, maximum, median and range; majority values for red and near infrared; NDVI standard deviation; and texture mean. This was indicated by the increase in the OOB error rate when these variables were omitted during iterations.

In contrast, green and near infrared bands, as well as texture were the most important for differentiating coniferous species (Appendix E.4). The most important predictors were: green mean, median, maximum, minority and standard deviation; near infrared mean, median, maximum, majority and variety; texture mean and standard deviation.

An example of a classification tree in each forest is provided in Appendix E.1 and E.2. The overall classification accuracy of classifying feature analyst objects into broadleaved, mixed and various coniferous species classes was estimated at 57 % (Table 5.8).

Maps of shelterbelts showing predicted tree species suggest that *P.radiata* shelterbelts are indeed in the majority, while broadleaved shelterbelts are in the minority in the study areas (Figures 5.8 to 5.10). Random Forests predicted that 2.42 % and 0.07 % of the study areas consist of *P.radiata* and *C.macrocarpa* shelterbelts, respectively. Only 0.04 % and 0.02 % of

the study areas consist of broadleaved and mixed (broadleaved/coniferous) shelterbelts, respectively (Table 5.9).

Table 5.6 Accuracy of differentiating conifer and broadleaved shelterbelts using Random Forests.

DIGITIZED SHELTERBELTS		Classification			Total	*PRODUCER'S ACCURACY*
		Broadleaf	**Conifer**	**Mixed**		
Actual	**Broadleaf**	**4**	4	0	8	50 %
	Conifer	1	**102**	0	103	99 %
	Mixed	0	5	**0**	5	0 %
	Total	5	111	0	116	
	USER'S ACCURACY	80 %	92 %	0 %		**91 %**
				OOB ESTIMATE OF ERROR		**8.62 %**

FEATURE ANALYST SHELTERBELTS		Classification			Total	*PRODUCER'S ACCURACY*
		Broadleaf	**Conifer**	**Mixed**		
Actual	**Broadleaf**	**3**	2	0	5	40 %
	Conifer	1	**88**	0	89	1 %
	Mixed	0	3	**0**	3	100 %
	Total	4	93	0	97	
	USER'S ACCURACY	75 %	95 %	0 %		**94 %**

Table 5.7 Accuracy of differentiating different species of coniferous shelterbelts using Random Forests.

DIGITIZED SHELTERBELTS		Classification						Total	*PRODUCER'S ACCURACY*
		PR	**M**	**PR/M**	**PR/C**	**C**	**C/C**		
Actual	***P.radiata* (PR)**	**53**	5	0	0	0	0	58	9 %
	***C.macrocarpa* (M)**	15	**10**	1	0	0	0	25	62 %
	PR/M	12	0	**0**	0	0	0	12	100 %
	PR/C	3	0	0	**0**	0	0	3	100 %
	Other conifer (C)	2	1	0	0	**0**	0	3	100 %
	Mixed (C/C)	0	1	0	0	0	**0**	1	100 %
	Total	85	17	1	0	0	0	103	
	USER'S ACCURACY	62 %	59 %	0 %	0 %	0 %	0 %		**61 %**
						OOB ESTIMATE OF ERROR			**39 %**

FEATURE ANALYST SHELTERBELTS		Classification						Total	*PRODUCER'S ACCURACY*
		PR	**M**	**PR/M**	**PR/C**	**C**	**C/C**		
Actual	***P.radiata* (PR)**	**44**	1	0	0	0	0	45	98 %
	***C.macrocarpa* (M)**	16	**8**	0	0	0	0	24	33 %
	PR/M	14	0	**0**	0	0	0	14	0 %
	PR/C	3	0	0	**0**	0	0	3	0 %
	Other conifer (C)	0	2	0	0	**0**	0	2	0 %
	Mixed (C/C)	0	0	0	0	0	**0**	0	0 %
	Broadleaf spp.	2	0	0	0	0	0	0	0 %
	Mixed (Broadleaf & Conifer spp.)	3	0	0	0	0	0	3	0 %
	Total	82	11	0	0	0	0	93	
	USER'S ACCURACY	54 %	73 %	0 %	0 %	0 %	0 %		**56 %**

Table 5.8 Overall accuracy of differentiating shelterbelt species type using Random Forests.

FEATURE ANALYST SHELTERBELTS		Classification									
		CONIFER (Co)						OTHER			
		PR	M	PR/ M	PR/ C	C	C/C	B	Mx	Total	*PRODUCER'S ACCURACY*
Actual	***P.radiata* (PR)**	**44**	1	0	0	0	0	0	0	45	98 %
	***C.macrocarpa* (M)**	16	**8**	0	0	0	0	0	0	24	33 %
	PR/M	14	0	**0**	0	0	0	0	0	14	0 %
	PR/C	3	0	0	**0**	0	0	0	0	3	0 %
	Other conifer (C)	0	2	0	0	**0**	0	1	0	3	0 %
	Mixed (C/C)	0	0	0	0	0	**0**	0	0	0	0 %
	Broadleaf spp. (B)	2	0	0	0	0	0	**3**	0	5	60 %
	Mixed (Mx)	3	0	0	0	0	0	0	**0**	3	0 %
	Total	82	11	0	0	0	0	4	0	97	
	USER'S ACCURACY	54 %	73 %	0 %	0 %	0 %	0 %	75 %	0 %		**57 %**

Table 5.9 Summary of shelterbelt species as predicted by Random Forests in study areas 1, 2 and 3.

RANDOM FOREST PREDICTION OF SHELTERBELT SPECIES		FEATURE ANALYST SHELTEBELTS				
		Study area 1	Study area 2	Study area 3	All three study areas	% of study areas
SHELTERBELT SPECIES	***P.radiata***	54.76 ha	27.22 ha	33.95 ha	115.93 ha	**2.42 %**
	C.macrocarpa	1.60 ha	1.24 ha	0.28 ha	3.12 ha	**0.07 %**
	P.radiata/ C.macrocarpa	0 ha	0.02 ha	0 ha	0.02 ha	**0.00 %**
	Other Conifer	0 ha	0 ha	0.02 ha	0.02 ha	**0.00 %**
	Broadleaf	0.50 ha	0.26 ha	1.34 ha	2.10 ha	**0.04 %**
	Mixed (Broadleaf/ Conifer)	1.00 ha	0 ha	0 ha	1.00 ha	**0.02 %**
				TOTAL	122.19 ha out of 4800 ha	**2.55 %**

Figure 5.8 Shelterbelt species types in study area 1 as predicted by Random Forests.

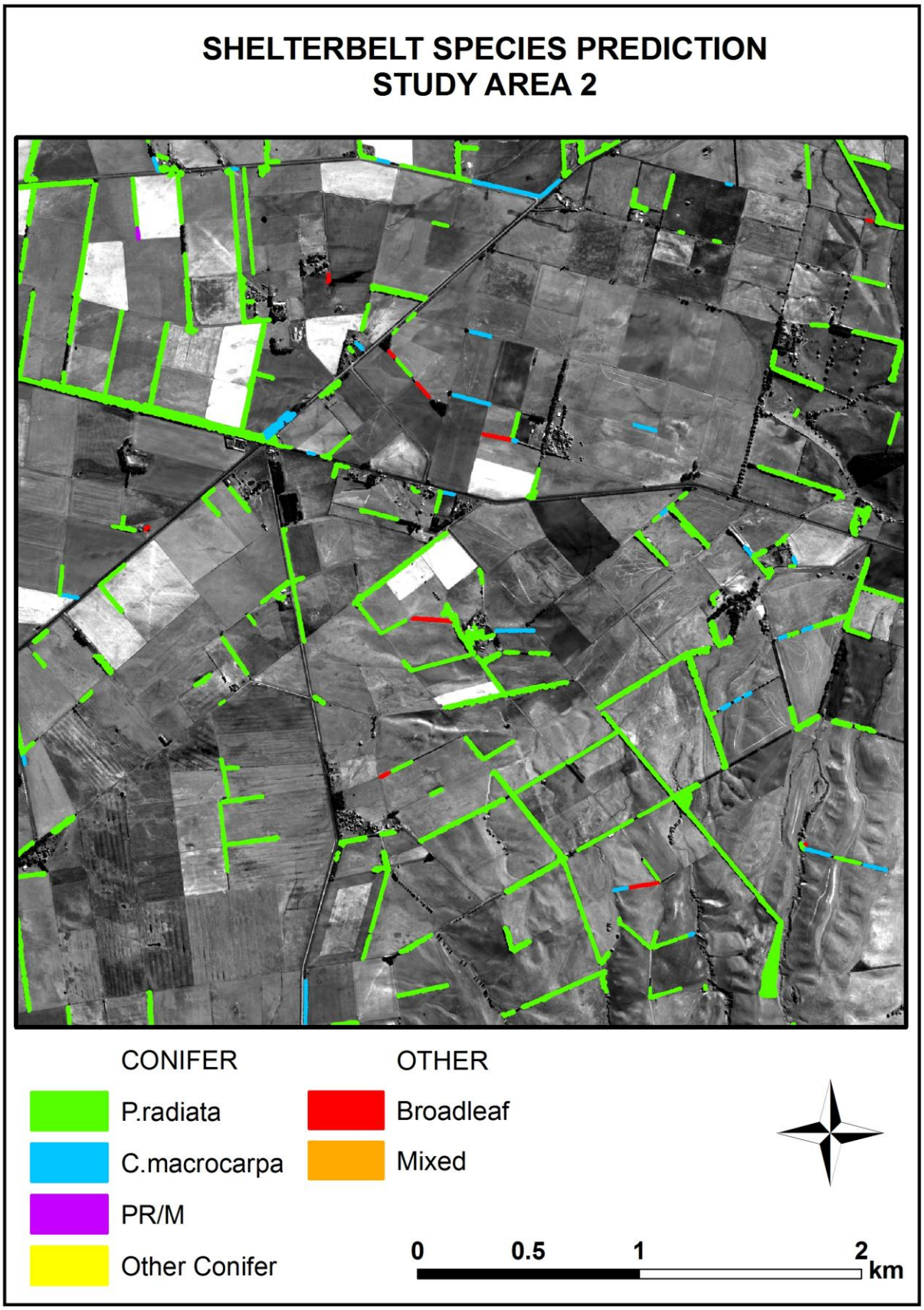


Figure 5.9 Shelterbelt species types in study area 2 as predicted by Random Forests.

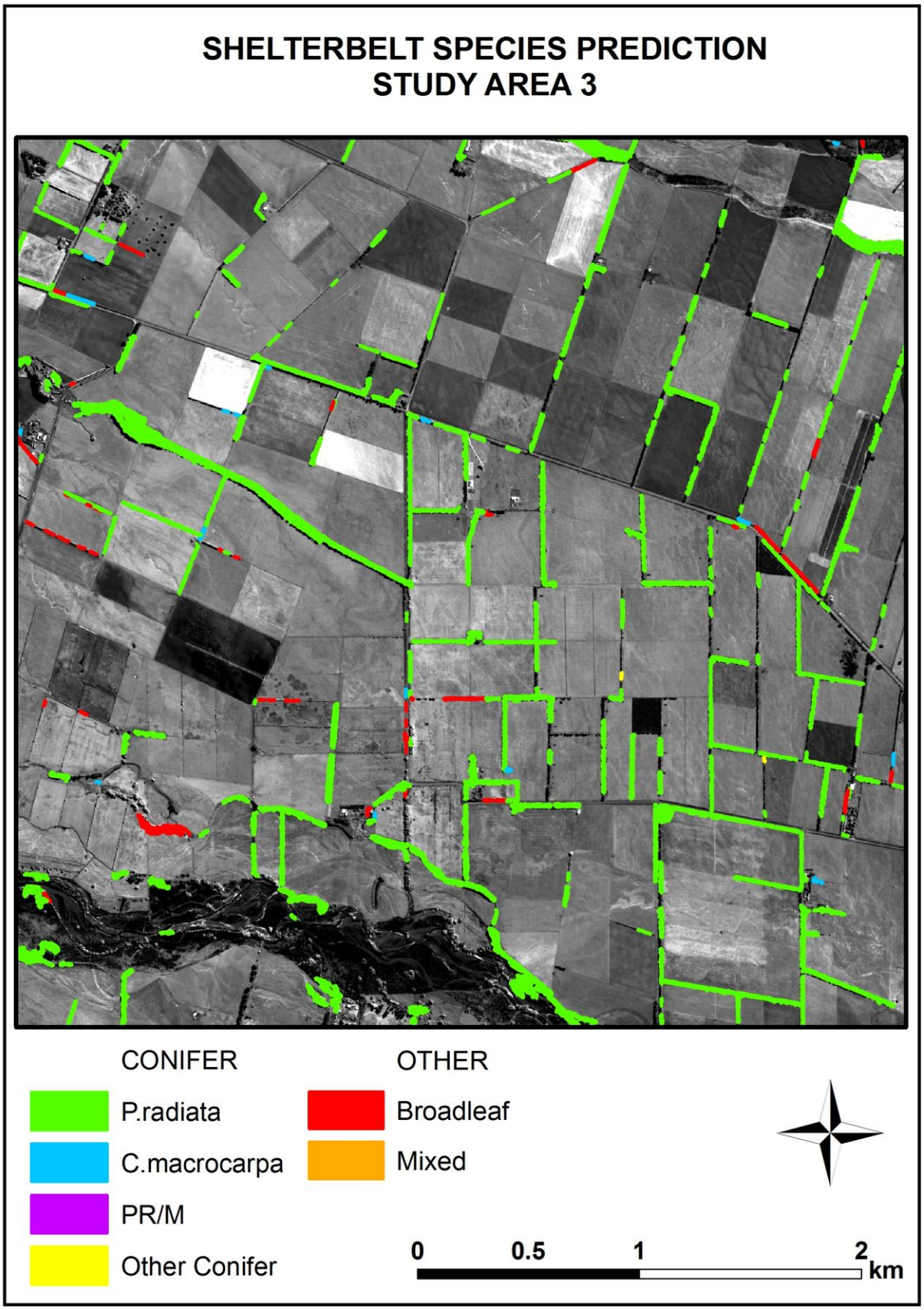


Figure 5.10 Shelterbelt species types in study area 3 as predicted by Random Forests.

5.5 Discussion

This study showed that shelterbelts could be reliably classified by species groups to an accuracy of greater than 90 %. However, the classification of coniferous shelterbelts into species type proved to be more challenging, with a classification accuracy of approximately 60 %. These results suggest that classification into broad species groups is much more reliable than classification into individual species.

The findings of this chapter are in agreement with works by Brandtberg (2002) and Erikson (2004) which determined that the further classification of species groups into individual species decreased classification accuracy by 14 to 20 % (Brandtberg, 2002; Erikson, 2004). In addition, it is well known that differentiating coniferous species is a difficult task due to coniferous species having similar crown structures and colour, and therefore similar spectral response patterns (Buddenbaum *et al.*, 2005).

The application of the coniferous species model (trained on digitized shelterbelts) to Feature Analyst shelterbelts resulted in a decrease in classification accuracy to 56 %. This 5 % decrease was not surprising: the predictions of the model were expected to be more accurate when applied to the training data; digitized shelterbelts may differ to Feature Analyst shelterbelts in how many shadow pixels are included; and a single Feature Analyst polygon often contained multiple shelterbelts. The latter issue was minimized by cutting Feature Analyst polygons so that shelterbelts were separated by orientation. It was assumed that a new type of shelterbelt began at each direction change.

These results indicate that the method presented in this chapter will need some improvements before it can be used for automatic identification of shelterbelt species. At this stage, the model does not have a high enough accuracy, and the error rate increases when it is applied to automatically delineated shelterbelt objects. However, using Feature Analyst shelterbelts instead of user-identified shelterbelts for training may improve results, as differences in the spectral characteristics of Feature Analyst and digitized shelterbelt objects would not pose a problem.

The classification accuracies of Random Forests for different species varied considerably, and were 0 % for mixed species classes, 54 % for *P.radiata*, 73 % for *C.macrocarpa*, and 75 % for broadleaved shelterbelts. The accuracy for *P.radiata* was much lower than expected, given that it was the majority of the sample. This is because the model grouped a high proportion of *C.macrocarpa* and mixed species shelterbelts into this class. The latter is to be expected as *P.radiata* is the dominant species in about two-thirds of *P.radiata*/*C.macrocarpa* mixed

shelterbelts, as estimated from field collected data. The fact that a high proportion of *C.macrocarpa* shelterbelts were classed as *P.radiata* indicates that the spectral signatures for these two species overlap. This is not surprising given the PCA results, and because it is common for coniferous species to have similar spectral responses (Buddenbaum *et al.*, 2005).

The range in species classification accuracy for the study areas was 75 %. This is quite large compared to what is reported in the literature for species classification of individual tree crowns. Katoh *et al.* (2009) reported a range of 62 %, while Olofsson *et al.* (2006) reported a range of only 21 %. However, the inclusion of mixed species classes (some rare) and the generic class for other less common coniferous species explains why differentiation of coniferous species was not very accurate and why accuracies varied considerably (Katoh *et al.*, 2009; Waser *et al.*, 2011).

The Random Forests method identified the green and near infrared bands as being the most important predictors of coniferous species type, with texture being third most important. This is interesting in light of PCA results which suggested that the NI/G band ratio, rather than separate near infrared and green bands, would be the most important predictor, along with texture. These results cannot be compared to any results published in the literature, as no work for *P.radiata* and *C.macrocarpa* is available. The single known study which classified shelterbelt objects into species type in Canada (Wiseman *et al.*, 2009) determined that the mean of the red band was the most important spectral variable for differentiating species type. Variable importance cannot, however, be compared between these two studies, as completely different tree species were involved.

The method presented here has a lower classification accuracy than published methods for classifying tree crown objects and forest stands, which range from 67 % to 96 % (Brandtberg, 2002; Buddenbaum *et al.*, 2005; Erikson, 2004; Förster & Kleinschmit, 2006; Hájek, 2008; Holmgren & Persson, 2004; Holmgren *et al.*, 2008; Katoh *et al.*, 2009; Key *et al.*, 2001; Leckie *et al.*, 2003; Olofsson *et al.*, 2006; Orka *et al.*, 2009; Suratno *et al.*, 2009; Waser *et al.*, 2011). This is not surprising given that useful information relating to individual crown shape and structure is lost since tree crowns within a shelterbelt are no longer considered separately. Shelterbelt density and pruning will also affect crown characteristics. In addition to this, the variability in shelterbelt age and tree age within shelterbelts also decreases accuracy, because tree age affects spectral response (Holmgren & Persson, 2004).

The Random Forests method for species classification may be improved by using different imagery. First, QuickBird images captured in autumn instead of summer could be used:

Holmgren *et al.* (2008) found that using images captured in autumn instead of summer increased the classification accuracy for two coniferous species by 4 %. Secondly, a combination of QuickBird and LIDAR imagery may be necessary for significantly better species discrimination. LIDAR data provides additional information about canopy shape and structure, crown reflective properties, and stem density (Holmgren *et al.*, 2008; Suratno *et al.*, 2009). LIDAR data may therefore be more efficient at differentiating coniferous species with similar spectral signatures (Holmgren & Persson, 2004; Holmgren *et al.*, 2008). As discuassed in the previous chapter, LIDAR data is unfortunately expensive to acquire.

Further improvements may also be made using a more sophisticated measure of texture, such as geostatistical measures of texture. These measures give the relationship between similarity and distance. For example, a semivariogram can be used to calculate the mean sum of squares of differences between pairs which are separated by a given distance (Buddenbaum *et al.*, 2005). In contrast, the method presented in this chapter used only a very simple measure of texture, the standard deviation of neighbouring pixels. Buddenbaum *et al.* (2005) found that combining spectral information with a geostatistical measure of texture improved classification results by upto 8 %.

In addition, the reliability of these results can be improved by including shelterbelts from a larger area of Canterbury in the field sample, and by increasing the number of broadleaved shelterbelts. At this stage, results support that differentiation of broadleaved and coniferous species is very accurate (Brandtberg, 2002; Erikson, 2004; Hájek, 2008; Holmgren *et al.*, 2008; Katoh *et al.*, 2009; Ke *et al.*, 2010; Leckie *et al.*, 2003; Olofsson *et al.*, 2006). However, these results will have to be confirmed.

In conclusion, results presented here show promise for differentiating *P.radiata* and *C.macrocarpa* shelterbelts, particularly since differentiating coniferous species can be difficult (Buddenbaum *et al.*, 2005). The accuracy of this method will, however, need to be improved before it can be used for carbon accounting projects and habitat suitability modelling. In addition, the ability of this method to differentiate shelterbelt species groups needs to be confirmed. An example application of shelterbelt delineation and species classification is provided in the following chapter.

Chapter 6
Example Application: Modelling Shelterbelt Carbon

6.1 Introduction

Shelterbelt delineation and characterization of structure and species type can be used for various applications, such as modelling the suitability of shelterbelts as wildlife corridors and habitat (Lechner *et al.*, 2009), and modelling shelterbelt carbon (Wiseman *et al.*, 2009). This chapter will present a method for modelling shelterbelt carbon as an example application of the methods described in Chapter 4 (Image-based Shelterbelt Delineation) and Chapter 5 (Differentiation of Shelterbelt Species).

Carbon content can be estimated from above-ground biomass estimations (Carswell *et al.*, 2009; Coomes *et al.*, 2002; Kort & Turnock, 1999) because approximately 50 % of tree biomass consists of carbon (Carswell *et al.*, 2009; Coomes *et al.*, 2002; Kort & Turnock, 1999; Thenkabail *et al.*, 2004). Carbon sequestration is one of the major ecosystem services that shelterbelts provide. Shelterbelts can therefore be utilized for offsetting carbon and carbon trading, as explained below.

The Kyoto Protocol is an international treaty which aims to mitigate human-induced global warming. It was agreed upon in 1997 and implemented in 2004 by 37 countries, including New Zealand. Signatory countries agreed to slow the increase of greenhouse gas emissions by reducing carbon dioxide emissions to 95% of the 1990 emission level by 2008 to 2012 (De Brauw, 2006). The mechanism for reducing global greenhouse gas emissions is called carbon trading. This has created a global carbon market in which carbon emissions are a marketable commodity (Miller, 2009).

The C_change model (Beets *et al.*, 1999) is currently used to fulfil New Zealand's reporting requirements for *P.radiata* plantations under the Kyoto Protocol (Moore, 2010). Unfortunately, it could not be implemented in this research, as the model requires detailed knowledge about stand and management not collected for this research. In addition, this model is designed specifically for plantation forests, and therefore is not necessarily suited to shelterbelts.

This chapter investigates the potential for shelterbelt carbon using (i) a field-based method, and (ii) a remote sensing-based method. Average shelterbelt carbon is estimated for shelterbelt species groups using published allometric biomass equations which utilize field

data (see Chapter 3, General Methods). These estimates are used to model the relationship between carbon and remotely-sensed spectral information using regression analysis.

6.2 Methods

Biomass can be estimated using field-based, GIS-based, and remote sensing-based methods (Lu, 2006). The latter two methods model the relationship between field-estimated carbon and remotely-sensed or environmental data. Tree carbon is first estimated using biomass regression equations that utilize field collected measurements, such as tree diameter and height (Jenkins *et al.*, 2004). Remote sensing and GIS-based methods then relate field-based estimates to remotely-sensed and environmental data, respectively, to predict carbon density across large areas (Lu, 2006; Zheng *et al.*, 2004).

6.2.1 Field-based carbon estimation

Field-based biomass regressions are usually both site and species specific (Jenkins *et al.*, 2004; Kort & Turnock, 1999). Therefore, ideally a different regression is used for each species at each different site. Unfortunately species specific equations are not always available for a given site. For example, New Zealand-specific biomass equations are available for *P.radiata* (Bi *et al.*, 2010; Moore, 2010), but not for *C.macrocarpa*.

Allometric equations estimating above-ground tree biomass were sourced from the literature. Biomass equations were selected to be species- and New Zealand-specific when possible, and to utilize predictor variables recorded during field data collection (see Chapter 3, General Methods). For example, biomass equations presented by Bi (2010) for *P.radiata* were unsuitable, as tree age was a required input, and was not measured in the field in this study. Equations containing both tree height and diameter at breast height were preferred, as the height variable helped to compensate for pruning (see later).

The New Zealand-specific biomass equation presented by Moore (2010) for *P.radiata* trees was selected for coniferous species (Figure 6.1 Equation 9). It included two parameters which were recorded during field data collection: tree height and tree diameter at breast height (DBH). This equation was applied to *P.radiata* and *C.macrocarpa*, as no species-specific equation for *C.macrocarpa* is available in the literature. It was also applied to shelterbelts composed of other less common conifer species (e.g. *Pseudotsuga menziesii*).

The New Zealand-specific biomass equation presented by Senelwa and Sims (1998) for *Eucalyptus spp.* were selected for *Eucalyptus spp.* shelterbelts (Figure 6.1 Equation 10). It also included tree height and diameter at breast height. In contrast, a New Zealand-specific

equation was not available in the literature for *Populus spp.*. An equation developed for *Populus tremuloides* in Canada by Case and Hall (2008) was therefore selected (Figure 6.1 Equation 11). Accuracy of the biomass estimation is not necessarily decreased by using a non-site-specific equation. For example, Specht and West (2003) found that biomass estimations for Australian plantations were more accurate when using an allometric equation from Canada, than using a region-specific equation. The equation for *Populus spp*. did not include tree height. This, however, was not important as all broadleaved shelterbelts were unpruned (see results for data collection in Chapter 5). The *Eucalyptus spp.* equation (Figure 6.1 Equation 10) was also applied to other broadleaved shelterbelts, as this is a New Zealand-specific equation, and the Senelwa and Sims (1998) recommended its use for other short rotation tree species. The single native shelterbelt in the sample was excluded from biomass and carbon calculations.

Shelterbelt carbon was calculated in a number of steps (Figure 6.1), because field collected measurements were for the average tree in a shelterbelt or shelterbelt row. Above ground biomass (AGB) (kg/tree) was first calculated for each tree (Figure 6.1 Equations 9 – 11), and then for each row (kg/row) and/or tree type within each shelterbelt (Figure 6.1 Equation 12). This enabled shelterbelt biomass (t/shelterbelt) to be calculated (Figure 6.1 Equations 13 – 14). The AGB for an average tree within each shelterbelt was then calculated (Figure 6.1 Equation 15), the natural logarithm of which was used in regression analysis (see section 6.2.2, Remote sensing-based carbon estimation). Lastly, carbon was calculated per shelterbelt (t/shelterbelt), per shelterbelt area (t/ha) and per shelterbelt length (t/km) (Figure 6.1 Equations 16 – 18). The latter was calculated for the purpose of comparing shelterbelt carbon in this study to shelterbelt carbon in other studies.

Biomass estimates for pruned and unpruned coniferous shelterbelts were then compared to determine to what extent the "height" parameter in the allometric formulas was able to account for the effect of pruning. This was done by comparing biomass and heights of unpruned and pruned trees for different DBH categories. Shelterbelts recorded as being half pruned were included in the unpruned category. This is because half pruned shelterbelts had pruned sides but unpruned tops. Biomass of half pruned shelterbelts is therefore unlikely to be affected to such a large degree.

The average carbon per hectare (t/ha) for each shelterbelt species was then applied to the total area of that species for the three study areas, as predicted in Chapter 5 (Differentiation of Shelterbelt Species) from satellite imagery. These estimates were then used to give a rough estimate for shelterbelt carbon sequestration in the Canterbury Plains.

STEP 1: CALCULATE SHELTERBELT PARAMETERS (where necessary)

Width (W)		km	
Length (L)		km	
Area (A)	= W x L	km^2	*(4)*
Spacing in Row$_i$ (S_{Ri})		km	
Number of trees in Row$_i$ (n_{Ri})	= L / S_{Ri}	Trees	*(5)*
Total number of trees (N)	= $n_{R1} + n_{R2} + ... + n_{Ri}$	Trees	*(6)*
Height (H)	= ((H_{R1} x n_{R1}) + (H_{R2} x n_{R2}) + … + (H_{Ri} x n_{Ri})) / N	m	*(7)*
Diameter (DBH)	= ((DBH_{R1} x n_{R1}) + (DBH_{R2} x n_{R2}) + … + (DBH_{Ri} x n_{Ri})) / N	cm	*(8)*

STEP 2: CALCULATE BIOMASS FOR EACH TYPE OF TREE IN EACH SHELTERBELT

Conifers	ln AGB = -0.9069 + 1.2273 ln DBH + 0.1411 $(\ln DBH)^2$ – 0.0078 ln H + 0.0840 $(\ln H)^2$	kg/tree	*(9)*
Eucalyptus and other broadleaves	ln AGB = ln ((1.22 DBH^2 x H) x 10^{-4})	kg/tree	*(10)*
Poplar	ln AGB = -2.763 + 2.524 (ln DBH)	kg/tree	*(11)*

STEP 3: CALCULATE BIOMASS FOR EACH ROW

Row *i*	AGB_{Ri} = AGB x n_{Ri}	kg/row	*(12)*

STEP 4: CALCULATE BIOMASS FOR EACH SHELTERBELT

AGB per shelterbelt	AGB_{SH} = (AGB_{R1} + AGB_{R2} + ... + AGB_{Ri}) ÷ 1000	t/shelterbelt	*(13)*
AGB per area of shelterbelt	AGB_{SHA} = (AGB_{SH} / A) ÷ 100	t/ha	*(14)*

STEP 5: CALCULATE BIOMASS FOR THE AVERAGE TREE IN EACH SHELTERBELT

Average tree	AGB_{tree} = AGB_{SH} / N x 1000	kg/tree	*(15)*

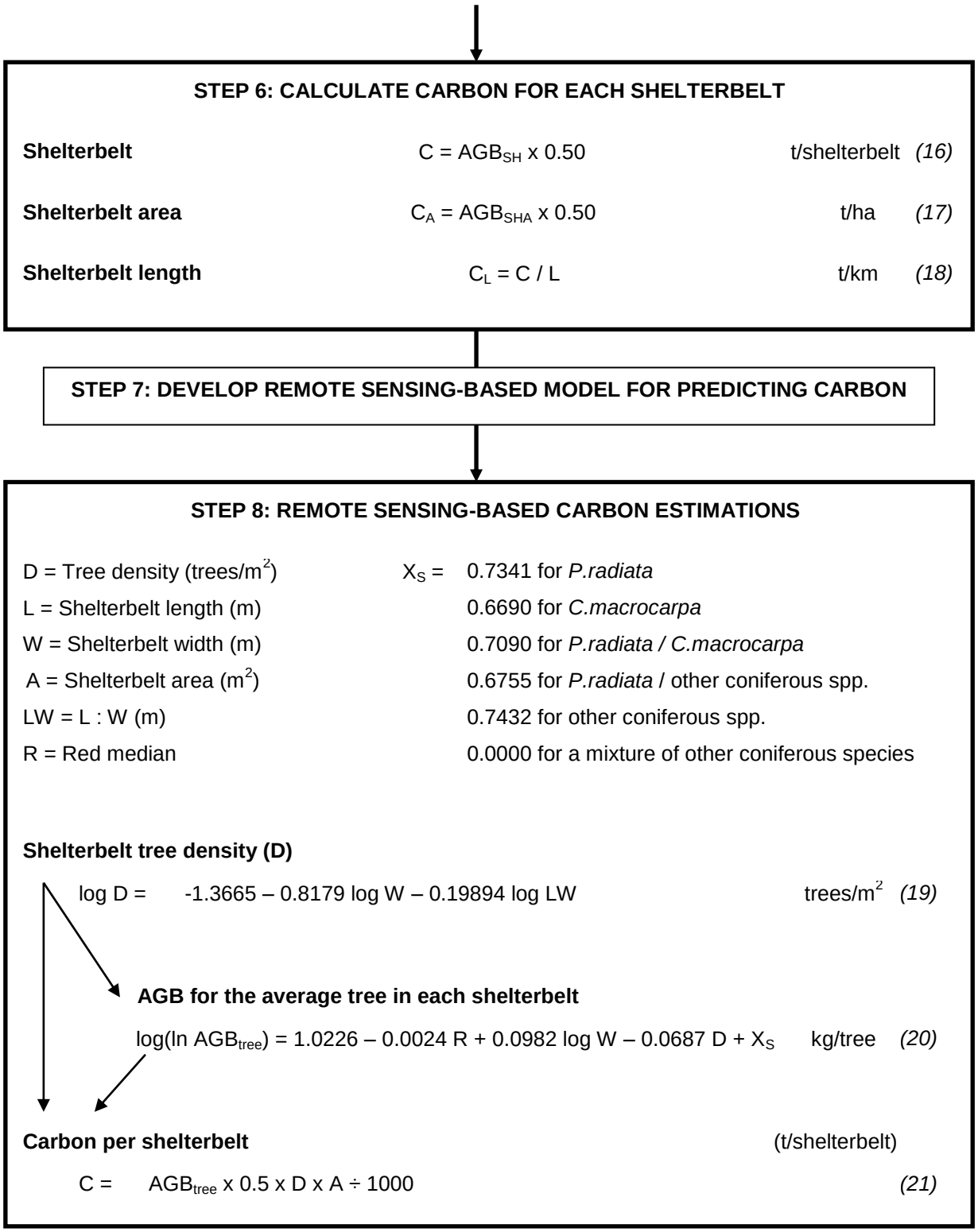


Figure 6.1 Steps for calculating field-based and remote sensing-based estimates for above ground biomass (AGB) and carbon (C).

6.2.2 Remote sensing-based carbon estimation

Only coniferous shelterbelts (n = 103) are considered in this section, due to the insufficient number of broadleaved shelterbelts (n = 8). Carbon modelling was performed using the whole coniferous shelterbelt sample as training data, so as to get a model with the best fit. This should therefore be considered as preliminary analysis only, which shows that modelling carbon from remotely sensed data is possible. The model presented was therefore not validated against independent data.

Data exploration

Remotely-sensed (see Chapter 5, Differentiation of Shelterbelt Species) and field-collected data (see Chapter 3, General Methods) for digitized shelterbelts were explored to investigate which image-derived variables are likely to be important predictors of shelterbelt physical characteristics. The software package R (The R Foundation for Statistical Computing, 2010) was used to carry out principal component analysis (PCA) on coniferous shelterbelts. This PCA (PCA 2) used the log of three physical variables and spectral variables selected from PCA 1 (Chapter 5, Shelterbelt Species Differentiation). Physical variables were transformed to give them normal distributions. The selected physical variables were height of the tallest row, tree diameter at breast height (DBH), and tree density, as these are important for shelterbelt carbon estimation. The height of the tallest row was used here as it was assumed that the tallest row had more influence on spectral characteristics. Ten out of 18 image-derived variables were selected from PCA 1 to reduce variable redundancy. The most important variables were selected by comparing PCA loadings (Table 5.5), biplots (Figures 5.6 and 5.7) and correlation matrices (Appendix D.3) of the two tree species. Selected variables were red, near infrared, NDVI, NI/G and texture means; green, near infrared, NDVI and texture standard deviations; as well as blue variety. Results from PCA 2 were examined to determine which image-derived variables are likely to be important predictors of shelterbelt carbon.

Modelling carbon using regression analysis

Multiple regression analysis was performed in R (The R Foundation for Statistical Computing, 2010) to model the relationship between carbon and remotely-sensed data. This was achieved using a stepwise backward selection approach. This involved performing the regression on several variables to start with, eliminating the most insignificant variables and re-running the regression, and repeating this process until a model with only significant variables was obtained.

Multiple regression analysis was performed on remotely-sensed (see Chapter 5, Differentiation of Shelterbelt Species) and field collected data (see Chapter 3, General Methods) for digitized shelterbelts, as described below. Above ground biomass for the average tree (ln ABG/tree) was modelled first, followed by shelterbelt density. Average tree biomass was modelled instead of shelterbelt biomass. This is because field-based estimates, on which the model is based, were calculated using measurements of the average tree in each shelterbelt (Figure 6.1). Modelling shelterbelt biomass may therefore cause some uncertainty (error) in the model. Shelterbelt density was modelled, as shelterbelt carbon can be estimated from tree biomass (AGB_T), tree density (D) and shelterbelt area (A):

$$\text{Shelterbelt Carbon (t/shelterbelt)} = AGB_T \times 0.5 \times D \times A \div 1000 \qquad (21)$$

The first step was to select variables highly correlated to ln AGB. A correlation matrix, which included all spectral variables extracted in Chapter 5 (n = 51), and five physical variables, was created (Table 6.1). Shelterbelt width, length, area, tree density, as well as a length to width ratio, were selected as physical variables as they could potentially be derived from satellite imagery. Shelterbelt dimensions could be calculated from delineated shelterbelt objects, and tree density can be modelled using shelterbelt length and width (see later). Physical variables and ln AGB were transformed to give them normal distributions, and because it increased the correlation between ln AGB and other variables. Transforming variables also improve the fit of the model (Afifi *et al.*, 2004).

Spectral and physical variables with the highest correlations to tree biomass were selected (Table 6.1). One spectral variable explaining spectral means (red median), one spectral variable explaining spectral variability (near infrared variety), and one textural variable (texture minimum) were selected (see axes of PCA 1, Chapter 5). All physical variables except shelterbelt length were also selected.

Lastly, shelterbelt pruning versus no pruning and shelterbelt species were also included as variables in the regression analysis to see whether or not these variables are signficant in biomass modelling.

Regression analysis by stepwise backward selection was performed to eliminate insignficant variables systematically. The *p* values of the t-statistic were used to identify insignificant variables (Afifi *et al.*, 2004). A small *p* value indicates that it is very likely that the slope of the relationship between the predictor variable and biomass is significantly greater than zero. Area and near infrared variety were eliminated first as they had the highest *p* values. The length to width ratio was eliminated next, followed by pruning and texture minimum.

The remaining variables (species, red median, width and density) were all significant. No more variables were eliminated, as this reduced the adjusted R^2 value of the model. Eliminating more variables therefore reduced the amount of biomass variability explained by the model. The adjusted R^2 value was used to indicate the fit of the model, because it is corrected for the number of variables included in the model. The unadjusted R^2 value automatically increases when more variables are added to the model, regardless of whether or not more variability is explained (Afifi *et al.*, 2004).

A similar approach to modelling tree biomass was taken for modelling shelterbelt tree density. Spectral variables did not have high correlations to tree density, therefore only the red median was selected (Table 6.2). Shelterbelt width, the length to width ratio, and area were also selected, with width having the highest correlation to density.

Regression analysis by stepwise backward selection was then performed. Red minimum, the length to width ratio, and area were all non-significant predictors of density, and were therefore eliminated. The final model contained only shelterbelt width and length.

Table 6.1 Correlation matrix showing correlations between tree biomass (ln AGB) and spectral (A) and physical (B) shelterbelt variables.

Log (ln AGB)	(A) VARIABLES FOR SPECTRAL AND TEXURAL CHANNELS								
	Min	Max	Range	Mean	Std	Variety	Majority	Minority	Median
RED	-0.421	-0.134	0.102	-0.580	0.190	0.218	-0.412	0.075	-0.581
GREEN	-0.352	-0.159	0.117	-0.536	0.326	0.314	-0.301	-0.365	-0.512
BLUE	-0.343	-0.069	0.195	-0.518	0.369	0.323	-0.327	-0.340	-0.481
NI	-0.541	-0.121	0.367	-0.429	0.368	0.391	-0.355	-0.541	-0.401
NDVI	-0.072	0.292	0.245	0.375	0.202				
NI/G	-0.39	-0.038	-0.038	-0.019	-0.038				
Texture	-0.406	-0.121	-0.040	-0.189	-0.175				

Log (ln AGB)	(B) LOG OF PHYSICAL VARIABLES
Length	0.013
Width	0.621
L:W	0.362
Area	0.321
Density	-0.522

= selected for regression analysis

Table 6.2 Correlation matrix showing correlations between shelterbelt tree density and spectral (A) and physical (B) shelterbelt variables.

Log (Density)	(A) VARIABLES FOR SPECTRAL AND TEXURAL CHANNELS								
	Min	Max	Range	Mean	Std	Variety	Majority	Minority	Median
RED	0.345	0.085	-0.104	0.310	-0.209	-0.192	0.141	0.062	0.275
GREEN	0.312	0.066	-0.157	0.286	-0.297	-0.274	0.053	0.299	0.232
BLUE	0.326	0.039	-0.201	0.316	-0.279	-0.272	0.138	0.318	0.254
NI	0.377	0.146	-0.216	0.248	-0.194	-0.234	0.174	0.376	0.208
NDVI	0.047	-0.225	-0.177	-0.201	-0.190				
NI/G	0.288	0.030	0.029	0.020	0.030				
Texture	0.227	0.082	0.037	0.204	0.212				

Log (Density)	(B) LOG OF PHYSICAL VARIABLES
Length	0.081
Width	-0.757
L:W	-0.532
Area	-0.316

6.3 Results

6.3.1 Field-based carbon estimation

The mean above ground biomass (AGB) per tree was the highest for coniferous shelterbelts at 797 kg/tree, compared to a mean of 214 kg/tree for broadleaved shelterbelts (Table 6.3). AGB per tree was not significantly different for the different coniferous species types (Figure 6.2). However, *C.macrocarpa* had the highest mean AGB at 992 kg/tree, compared to a mean of 785 kg/tree for *P.radiata*. The mean AGB for mixed *P.radiata/C.macrocarpa* species was even lower at 722 kg/tree. *Populus spp.* had the highest mean AGB per tree for broadleaved shelterbelts at 432 kg/tree, compared to only 1.9 kg/tree for *Eucalyptus spp*. Biomass estimates for *Eucalyptus spp.* were very low compared to other species, probably because all *Eucalyptus* shelterbelts consisted of relatively young trees.

Likewise, mean shelterbelt carbon was the highest for coniferous shelterbelts at 338 t/ha, compared to a mean of 266 t/ha for broadleaved shelterbelts (Table 6.3). Shelterbelt carbon was also not significantly different for the different coniferous species types (Figure 6.3). However, in contrast to AGB per tree, mixed *P.radiata/C.macrocarpa* species had the highest mean carbon content at 400 t/ha, compared to 389 t/h and 257 t/ha for *P.radiata* and *C.macrocarpa* shelterbelts, respectively. This is because *P.radiata/C.macrocarpa* shelterbelts had higher tree densities (mean of 15 trees/100m^2) than *P.radiata* (mean of 15 trees/100m^2) and *C.macrocarpa* (11 trees/100m^2) shelterbelts. Likewise, *Populus spp.* shelterbelts had a very high tree density (33 trees/100 m^2), and therefore had the highest carbon content (808 t/ha) out of all shelterbelts. However, broadleaved shelterbelts still had a lower mean carbon content than coniferous shelterbelts, because *Populus spp.* shelterbelts made up only a small part of the sample (n = 2), and *Eucalyptus spp.* shelterbelts (n = 3) had the lowest carbon content (0.6 t/ha) out of all shelterbelts.

Comparisons between pruned and unpruned coniferous shelterbelts determined that field-based biomass estimations accounted for pruning to at least some extent, because height was a parameter in the formula. Pruned trees had 22 to 31 % less biomass than pruned trees of the same diameter (Table 6.4; Figure 6.4), and were 48 to 57 % shorter (Table 6.5). However, shelterbelt carbon (per hectare) of pruned and unpruned shelterbelts overlap (Figure 6.4), most likely due to variation in tree density.

Application of carbon estimates to results from shelterbelt delineation (see Chapter 4, Image-base Shelterbelt Delineation) and shelterbelt species differentiation (see Chapter 5,

Differentiation of Shelterbelt Species) indicate that shelterbelts in Canterbury store at least 9.7 tonnes of carbon per hectare of land used for agriculture in Canterbury (Table 6.6).

Table 6.3 Summary of above ground biomass (AGB) per average tree and shelterbelt carbon by species type.

		ALL SHELTER-BELTS	SPECIES GROUP			CONIFER				BROADLEAF	
			Conifer	Broadleaf	Mixed	*P.radiata* (PR)	*C.macrocarpa* (M)	PR/M	Other Mixed	*Eucalyptus spp.*	*Populus spp.*
AGB for the average tree (kg/tree)	Minimum	1	9	0.1	66	37	22	118	9	1.7	121
	Maximum	10186	10186	743	958	7524	10186	2949	2949	2.2	743
	Mean	**741**	**797**	**214**	**341**	**785**	**992**	**722**	**612**	**1.9**	**432**
	Std dev	1278	1336	326	376	1082	2052	789	715	0.2	439
Shelterbelt carbon (t/shelterbelt)	Minimum	0.01	0.09	0.01	0.5	0.5	0.1	6	0.9	0.01	2.7
	Maximum	1512	1512	9.4	92	1512	292	137	137	0.12	7.4
	Mean	**52**	**57**	**2.8**	**22**	**75**	**34**	**42**	**21**	**0.06**	**5.1**
	Std dev	158	166	4.0	39	214	68	40	37	0.06	3.2
Shelterbelt carbon (t/ha)	Minimum	0.22	12	0.2	10	20	31	67	12	0.30	160
	Maximum	3432	3432	1456	378	3432	1833	1535	1535	0.76	1456
	Mean	**325**	**338**	**266**	**154**	**389**	**257**	**400**	**331**	**0.60**	**808**
	Std dev	440	444	534	163	479	425	396	364	0.26	917
Shelterbelt carbon (t/km)	Minimum	0.2	3	0.2	8	10	7	61	3	0.19	48
	Maximum	4461	4461	218	284	4461	2016	621	621	0.53	218
	Mean	**265**	**285**	**69**	**120**	**337**	**244**	**251**	**207**	**0.40**	**133**
	Std dev	512	537	103	116	643	445	165	164	0.18	121

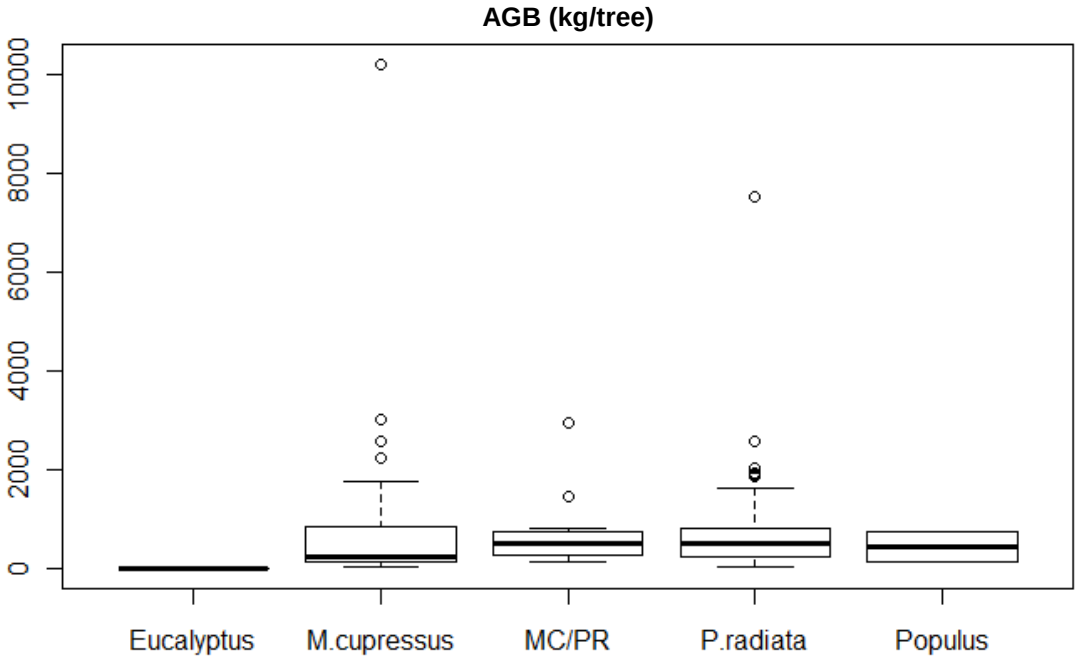


Figure 6.2 Comparing the above ground biomass (AGB) of the average tree by shelterbelt species.

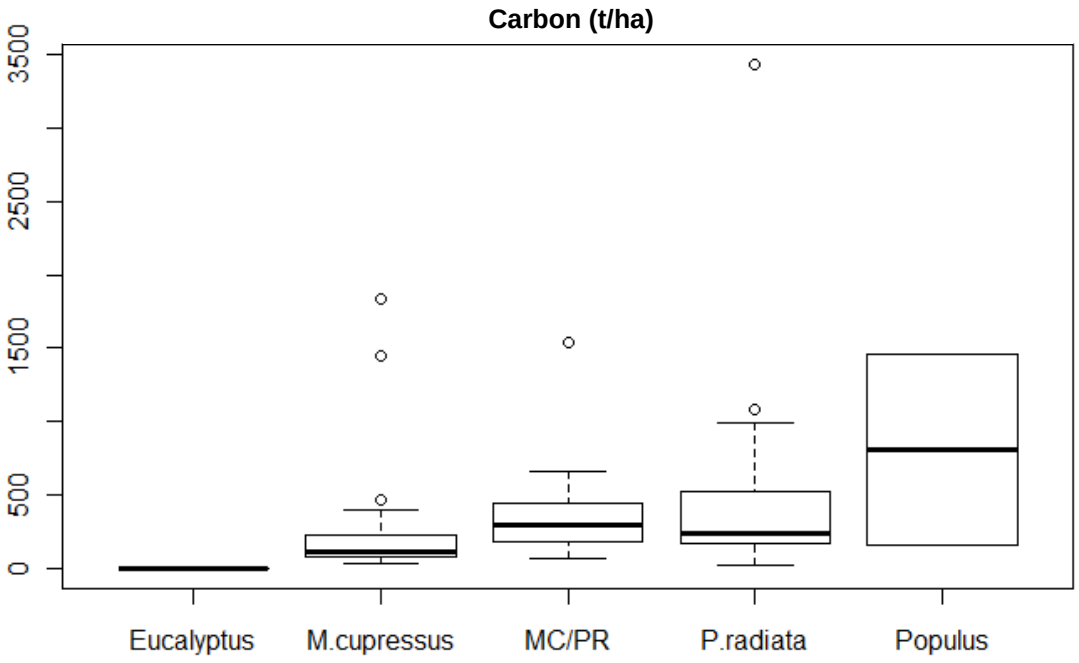


Figure 6.3 Comparing shelterbelt carbon by shelterbelt species.

Table 6.4 Difference in biomass for pruned and unpruned coniferous shelterbelts.

BIOMASS PER TREE		UNPRUNED (U)		PRUNED (P)		DIFFERENCE IN BIOMASS	
		kg/tree	#	kg/tree	#	Actual (U-P)	Percentage
Average tree diameter (DBH)	0 – 12 cm	9	1		0		
	13 – 30 cm	126	8	98	19	28	22 %
	31 – 60 cm	629	34	432	25	197	31 %
	61 – 90 cm	2081	11	2135	3 *		
	91 – 120 cm		0		0		
	> 120 cm	8855	2		0		
Total shelterbelts			56		47	* Sample size too small for comparison	

Table 6.5 Difference in heights for pruned and unpruned coniferous shelterbelts.

AVERAGE HEIGHT		UNPRUNED (U)		PRUNED (P)		DIFFERENCE IN HEIGHT	
		m	#	m	#	Actual (U-P)	Percentage
Average tree diameter (DBH)	0 – 12 cm	3	1		0		
	13 – 30 cm	10.5	8	5.5	19	5.0	48 %
	31 – 60 cm	15.6	34	6.7	25	8.9	57 %
	61 – 90 cm	16.45	11	14.67	3 *		
	91 – 120 cm		0		0		
	> 120 cm	11.0	2		0		
Total shelterbelts			36		22	* Sample size too small for comparison	

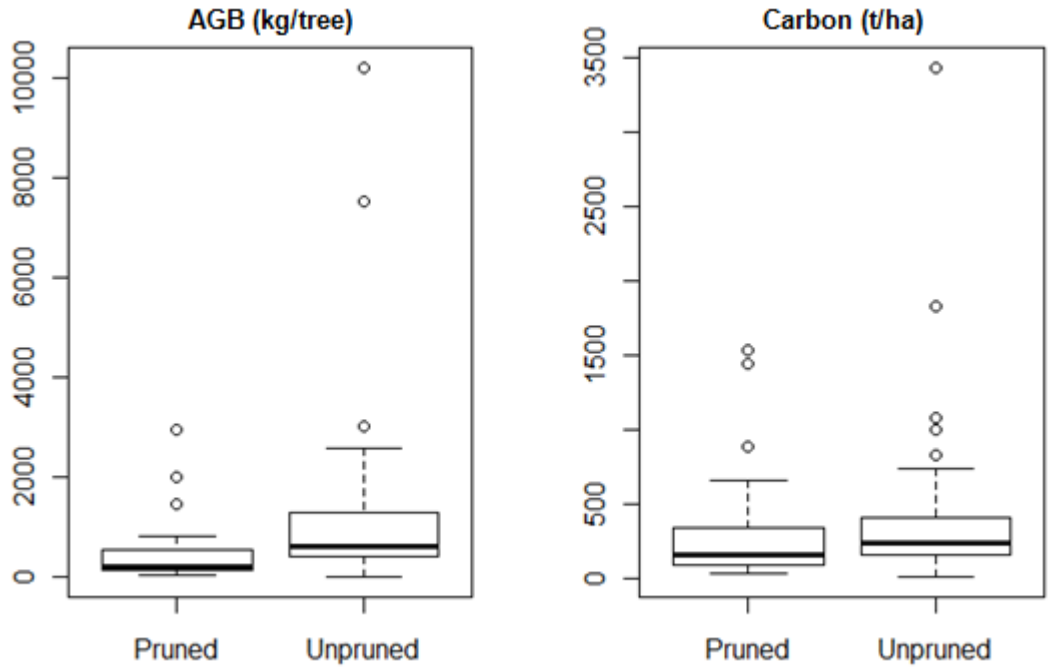


Figure 6.4 Comparing above ground biomass (AGB) and carbon for pruned and unpruned coniferous shelterbelts, as estimated by the field-based method.

Table 6.6 Summary of delineated shelterbelt species by shelterbelt area (ha), proportion of study areas (%), carbon per area of shelterbelt (t/ha), and total shelterbelt carbon (t) for study areas.

RANDOM FOREST PREDICTION OF SHELTERBELT SPECIES		FEATURE ANALYST SHELTEBELTS			
		Hectares of shelterbelt in study areas	% of study areas	Carbon per hectare of shelterbelt	Total shelterbelt carbon for study areas
SHELTERBELT SPECIES	***P.radiata***	115.93 ha	2.42 %	389 t/ha	45097 t
	C.macrocarpa	3.12 ha	0.07 %	257 t/ha	802 t
	Broadleaf	2.10 ha	0.04 %	266 t/ha	559 t
	Mixed (Broadleaf/ Conifer)	1.00 ha	0.02 %	154 t/ha	154 t
TOTAL		**122.19 ha (of 4800 ha)**	**2.55 %**	**381 t/ha**	**46611 t**
Shelterbelt carbon per hectare of land used for agriculture				**= 46611 t / 4800 ha =**	**9.7 t/ha**

6.3.2 Remote sensing-based carbon estimation

Data exploration

PCA 2 indicated that there is a correlation between spectral means and shelterbelt physical characteristics (Figure 6.5). The first two principal component axes explained only 37 and 19 % of spectral and physical variability (Table 6.7). This may be because many of the intercorrelated spectral variables, which were excluded from this PCA, each explain some of the variability. Shelterbelt height and tree diameter were negatively correlated to the means of spectral bands, while tree density was positively correlated to the latter. Physical variables had the highest correlation to the mean of the red band (Appendix D.3). Height was 58, 54 and 53 % correlated to red and near infrared means, and blue variety, respectively. Tree diameter was 52 % correlated to the red mean, while tree density was only 32 and 30 % correlated to the red mean and the green standard deviation, respectively. The PCA loadings (Table 6.8) also suggest that the spectral means are the most important for explaining shelterbelt variability. The results from PCA 2 therefore suggest that spectral means may be useful for modelling shelterbelt carbon.

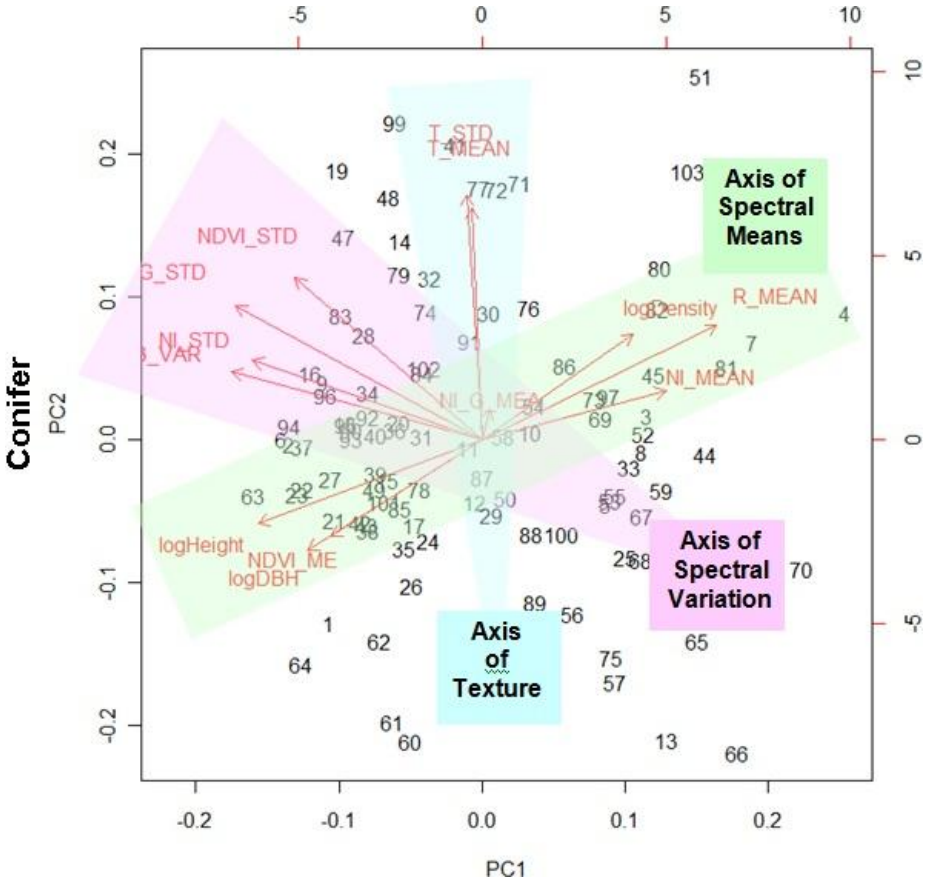


Figure 6.5 Biplot of PC1 and PC2 of image-derived and physical variables for coniferous shelterbelts (PCA 2).

Table 6.7 Proportion of variance explained by principal components.

	PCA 2 IMAGE-DERIVED & PHYSICAL VARIABLES Coniferous spp.		
	PC1	**PC2**	**PC3**
Proportion	37 %	19 %	11 %
TOTAL	37 %	56 %	67 %

Table 6.8 PCA Loadings indicating the importance of variables.

		PCA 2 IMAGE-DERIVED & PHYSICAL VARIABLES Coniferous spp.		
		PC1	**PC2**	**PC3**
R	**Mean**	0.36	0.24	-0.01
G	**Std**	-0.38	0.28	0.14
B	**Variety**	-0.38	0.14	0.09
NI	**Mean**	0.28	0.10	0.52
	Std	-0.35	0.17	0.12
NDVI	**Mean**	-0.23	-0.20	0.50
	Std	-0.29	0.34	-0.07
NI/G	**Mean**	0.01	0.07	0.62
T	**Mean**	-0.02	0.49	-0.13
	Std	-0.02	0.52	-0.05
Physical	$_{log}$**Height**	-0.34	-0.18	-0.16
	$_{log}$**DBH**	-0.27	-0.23	-0.03
	$_{log}$**Density**	0.23	0.22	0.05

Modelling carbon using regression analysis

The best regression model for predicting tree biomass for coniferous shelterbelts explained 56 % of tree biomass variability, and included shelterbelt species, width, density and the red median as significant predictor variables (Table 6.9). Species information was the most significant predictor of tree biomass, followed by the red median, and then shelterbelt width and density. The equation for the linear regression model predicting tree biomass is:

$$\log (\ln AGB_{tree\ S}) = 1.0226 - 0.0024\ R + 0.0982 \log W - 0.0687 \log D + X_S \qquad (20)$$

Where $AGB_{tree\ S}$ = Above ground tree biomass (kg/tree)

R = Red median

W = Width (m)

D = Density (trees/m^2)

Where X_S =

0.0000	for a mixture of other coniferous spp.
0.7341	for *P.radiata*
0.6690	for *C.macrocarpa*
0.7092	for *P.radiata/C.macrocarpa*
0.6755	for *P.radiata*/other coniferous spp.
0.7432	for other coniferous spp.

Table 6.9 Summary of the linear regression model predicting above ground biomass (AGB) for the average tree in a shelterbelt.

		EQUATION COEFFICIENTS			
		Estimate	**Standard Error**	***t* Statistic**	
				p	**Significance**
Intercept	**Mixed Conifer (C/C)**	1.0225645	0.2261757	< 0.001	***
Species	***P.radiata* (PR)**	0.7341349	0.1598404	< 0.001	***
	***C.macrocarpa* (M)**	0.6689583	0.1592788	< 0.001	***
	PR/M	0.7091516	0.1661975	< 0.001	***
	PR/C	0.6755121	0.1816754	< 0.001	***
	Other Conifer (C)	0.7432137	0.1782492	< 0.001	***
Red median		-0.0024096	0.0007848	< 0.01	**
log Width (m)		0.0981689	0.0452711	< 0.1	*
log Density (trees/m^2)		-0.0686610	0.0324409	< 0.1	*
			ADJUSTED R^2		**0.5604**

The best regression model for predicting tree density for coniferous shelterbelts explained 62 % of variability in shelterbelt density, and included shelterbelt width, and the length to width ratio as significant predictor variables (Table 6.10). The equation for the linear regression model predicting tree density is:

$$\log D = -1.3665 - 0.8179 \log W - 0.19894 \log LW \quad (19)$$

Where D = Density (trees/m^2)
W = Width (m)
LW = Length : Width (m)

Table 6.10 Summary of the linear regression model predicting tree density (AGB) for the average tree in a shelterbelt.

	EQUATION COEFFICIENTS			
	Estimate	**Standard Error**	***t* Statistic**	
			p	**Significance**
Intercept	-1.36653	0.25774	< 0.001	***
log Width (m)	-0.81794	0.08442	< 0.001	***
log Length : Width (m)	-0.19894	0.05085	< 0.001	***
		ADJUSTED R^2		**0.6226**

6.4 Discussion

This chapter has shown that shelterbelts represent a significant carbon reservoir on the Canterbury Plains. In this study, above ground carbon stored in shelterbelts ranged from 0.6 t/ha for *Eucalyptus spp.*, to 808 t/ha for *Populus spp.*. Coniferous shelterbelts in this study held 338 t/ha of carbon, while broadleaved shelterbelts held 266 t/ha of carbon. *Pinus radiata* and *Cupressus macrocarpa*, the two most common shelterbelt species, held 389 and 257 t/ha of carbon, respectively. Overall, this study determined that for every hectare of land used for agriculture in Canterbury, shelterbelts sequester 9.7 tonnes of carbon.

Secondly, this chapter has shown that methods described in Chapters 4 (Image-based Shelterbelt Delineation) and 5 (Differentiation of Shelterbelt Species) can be potentially applied for estimating shelterbelt carbon. This chapter presents a method for estimating shelterbelt carbon from variables directly and indirectly derived from satellite imagery (Figure 6.1 Step 8). Shelterbelt tree density can first be estimated from shelterbelt length and width. These measurements can be directly derived from delineated shelterbelt polygons. Secondly, tree biomass is estimated using shelterbelt species, shelterbelt width and density, and the median value of the red band as predictor variables. Lastly, shelterbelt carbon is calculated using tree biomass, tree density and shelterbelt area.

The field-based estimates for shelterbelt carbon (in tonnes per kilometre of shelterbelt) in this study are comparable to those published in the literature. This study estimated that shelterbelt carbon ranged from 69 t/km for broadleaves, to 285 t/km for conifers. In contrast, Kort and Turnock (1999) estimated that carbon stored by shelterbelts in an area of Canada ranged from 11 t/km for shrubs, to 24 to 41 t/km for conifers, to 105 t/km for hardwoods. Whereas Brandle *et al.* (1992) determined that carbon stored by shelterbelts in an area of the USA ranged from 0.68 t/km for shrubs, to 5.41 t/km for hardwoods, to 9.14 t/km for conifers. The estimates for hardwoods and conifers by Brandle *et al.* (1992) and Kort and Turnock (1999) are substantially lower than the estimates for broadleaves (mainly hardwoods) and conifers in this study. This is to be expected due to the differences in shelterbelt age and width among these studies. Shelterbelts sampled by Brandle *et al.* (1992) were 20 years old and had only one row. In contrast, shelterbelts in this study varied in age (as indicated by the DBH ranging from 7 to 150 cm), and had up to 9 rows (mean width 7.6 m). Shelterbelts sampled by Kort and Turnock (1999) also ranged in age, but still had a lower mean DBH (29.5 cm), as well as a narrower mean width (5.3 m) than the shelterbelts sampled in this study (mean DBH 42 cm; mean width 7.6 m).

The linear relationship between canopy reflectance and tree biomass determined in this study is comparable to spectral-biomass relationships determined in other studies. This study found that a linear relationship existed between tree biomass and a mixture of spectral and physical variables (as described above). This relationship explained 56 % of the variability in tree biomass, which is in line with models in the literature that predict tree biomass for forests. According to published studies, the variability in tree biomass explained by remotely-sensed variables varies with the type of model used, study area, timing of image capture, sample size (Thenkabail *et al.*, 2004), which variables are used as predictor variables (Lu, 2005; Thenkabail *et al.*, 2004), and whether the sample is divided into species groups (Zheng *et al.*, 2004). For example, Thenkabail *et al.* (2004) found that spectral variables explained 1 – 73 % of biomass variability, Steininger (2000) found that spectral variables explained 70 % of biomass variability, whereas Phua and Saito (2003) found that spectral variables explained up to 96 % of biomass variability. Lu (2005) found that spectral variables explained 16 – 75 % of biomass variability, whereas a combination of spectral and textural variables explained 50 – 78 % of biomass variability.

The method presented in this chapter for estimating shelterbelt carbon over a large area can be improved in future research by first improving the accuracy of field-based carbon estimates, and secondly by improving the remote sensing-based model. The accuracy of field-based carbon estimates, on which remote sensing-based models are founded, can be improved by (i) making the sample more representative of Canterbury, and (ii) determining allometric relationships specific to shelterbelt species and specific to Canterbury. (i) The sample could be made more representative of Canterbury by including shelterbelts from various districts in Canterbury, as well as including more broadleaved shelterbelts. For example, carbon estimates for *Eucalyptus spp.* were very low compared to other shelterbelt species, probably because all *Eucalyptus spp.* shelterbelts (n = 3) consisted of relatively young trees (DBH of 32 – 35 cm; height of 13 – 14 m). This is in contrast to coniferous shelterbelts which were represented by a wide range of age categories (DBH of 7 – 150 cm; height of 2 – 37 m). (ii) Allometric relationships specific to shelterbelts in Canterbury could be determined by destructive analysis, which involves cutting down and weighing trees. For example, Kort and Turnock (1999) determined allometric biomass equations for different shelterbelt species by cutting down and weighing representative trees. In addition, the effect of pruning on allometric relationships needs to be investigated. This study showed that the "height" parameter in allometric equations accounts for at least some of the effect of pruning.

The remote sensing-based model for predicting tree biomass can be improved by (i) using different vegetation indices which may be better than NDVI at predicting biomass (Foody *et al.*, 2003), (ii) using more sophisticated measures of texture (Lu, 2005), and (iii) using a combination of satellite imagery and LIDAR data (Kalaitzidis & Zianis, 2009). In addition, (iv) indirect biomass estimation techniques could be investigated. For example, tree diameter and height could first be estimated using remotely-sensed variables as predictor variables. Tree biomass could then be estimated from tree diameter and height.

In conclusion, this chapter has shown that shelterbelts are clearly a quantifiable carbon reservoir, which can be utilized for offsetting carbon and carbon trading. Shelterbelts could therefore contribute to the Land Use, Land Use Change, and Forestry Sector (LULUCF) of New Zealand's Greenhouse Gas Inventory. Currently, only major shelterbelts are accounted for in the low producing grassland land use category of this sector (Ministry for the Environment, 2009). This chapter has shown that shelterbelts contribute at least 9.7 t/ha of carbon to the low producing grassland carbon pool, which is currently estimated at 29 t/ha (Ministry for the Environment, 2009).

Chapter 7
Discussion

Shelterbelts are a prominent part of New Zealand's agricultural landscape, particularly of Canterbury (Cameron, 1964; Price, 1993). Shelterbelt hold significant ecological and economic value due to their many beneficial functions, which range from increasing agricultural productivity, to providing wildlife corridors and habitat, to sequestering carbon (Gregory, 1995; Kristensen & Casperson, 2002; McLachlan & Wratten, 2003; Ministry of Forestry, 1992; Stringer, 1977). Despite this, shelterbelts are not particularly well characterized in a spatially explicit manner. This thesis research sought to develop methods that identify and characterize shelterbelts across a landscape.

Results demonstrated that shelterbelts can be successfully delineated and characterized using high spatial resolution satellite imagery. Object-oriented classification of shelterbelts using the specialist software, Feature Analyst (Overwatch, 2010), produced excellent results, with an overall classification accuracy of 92 %. Differentiation of shelterbelts into species type was achieved using classification rules which utilize predictor spectral variables. The Random Forests model for predicting shelterbelt species group (i.e. broadleaved, coniferous, or mixed) explained 91 % of shelterbelt variability within species groups, whereas the Random Forests model for predicting species type (e.g. *P.radiata*, *C.macrocarpa*) of coniferous shelterbelts explained 61 % of shelterbelt variability within coniferous species types. Lastly, shelterbelt carbon was modelled using predictor variables which were directly and indirectly derived from satellite imagery. This provided an example application of shelterbelt delineation and characterization from satellite imagery. The linear regression model for predicting tree biomass explained 56 % of variability in tree biomass, whereas the linear regression model for predicting shelterbelt tree density explained 62 % of variability in tree density. These methods show promise for the automatic mapping and characterization of shelterbelts across large areas.

Methods for delineating and characterizing shelterbelts from satellite imagery can be improved in several ways. First, the use of multitemporal imagery, or imagery captured in autumn rather than summer, is likely to improve species differentiation, and therefore shelterbelt carbon estimates (Holmgren *et al.*, 2008; Key *et al.*, 2001). Second, a combination of satellite imagery and LIDAR data can be used, the latter of which will provide additional information about shelterbelt height and structure (Ke *et al.*, 2010). LIDAR data could not

only increase the accuracy of shelterbelt delineation, but also of shelterbelt species differentiation (Holmgren *et al.*, 2008; Ke *et al.*, 2010) and biomass estimation (Kalaitzidis & Zianis, 2009). LIDAR data will be particularly valuable for differentiating coniferous species with similar spectral signatures, such as *P.radiata* and *C.macrocarpa* (Holmgren & Persson, 2004; Holmgren *et al.*, 2008). Third, further improvements may also be made by using more sophisticated measures of texture for image analysis, and experimenting with vegetation indices other than NDVI (Buddenbaum *et al.*, 2005; Foody *et al.*, 2003; Lu, 2005).

Methods developed on relatively small areas, such as the ones presented here, show promise for application over large areas, because of the potential to automate parts of the methodology (Liknes *et al.*, 2010). However, several challenges will have to be faced before this is possible. First, procedures for correcting QuickBird images captured on different dates for differing atmospheric conditions will have to be developed (Jensen, 2005). Secondly, field data collection will have to cover a larger area, as well as containing a greater number of broadleaved shelterbelts to be more representative of the area to which it is applied. Thirdly, models developed on Canterbury shelterbelts will have to be validated on other areas of New Zealand, if they were to be applied at a national rather than regional scale. This is because most remotely-sensed variables are likely to be sensitive to regional differences (Holmgren & Persson, 2004). Lastly, the accuracy of the model for estimating shelterbelt carbon using remotely-sensed predictor variables needs to be validated. In addition, allometric relationships specific to shelterbelts in Canterbury (and other regions for analysis at a national scale), and that account for the effect of pruning, could be determined by destructive analysis. This will involve cutting down and weighing representative trees from representative shelterbelts (Kort & Turnock, 1999).

Implications for natural resource management

The methods presented in this thesis for semi-automatic delineation and characterization of shelterbelts can be used as natural resource management tools. These methods have several ecological and economic applications, such as modelling the effectiveness of shelterbelts as habitat and wildlife corridors (Lechner *et al.*, 2009), improving the detail of rural habitat maps (Freeman & Buck, 2003), and quantifying the shelterbelt carbon pool across a given landscape (Wiseman *et al.*, 2009). These methods could also be used to assess the need for shelterbelt-related agro-environmental policies, as well as monitoring their success once implemented.

Shelterbelt carbon sequestration has been highlighted as an example application of the methods described in this thesis. Carbon sequestration is an important ecosystem service that

shelterbelts provide, and yet it remains largely unquantified. Shelterbelts can therefore be utilized for offsetting carbon and carbon trading under the Kyoto Protocol. The earning of removal units (RMUs) in carbon trading is of particular interest to this research. The Kyoto Protocol allows countries to earn RMUs for funding carbon sink projects, such as tree-planting projects (Miller, 2009). Planting trees is a prominent strategy to mitigate the effects of greenhouse gases on the environment and global climate (Adams *et al.*, 1993). Thus, New Zealand could earn RMUs if a shelterbelt planting project is implemented. The methods described in this research could be used to estimate the shelterbelt carbon pool at a national level, and to monitor the success of a shelterbelt planting program if instituted. Unfortunately tree planting carbon sink projects have the disadvantage that carbon is released back into the atmosphere when the tree dies, is harvested or burned (Johnson, 2008; Velasquez-Manoff, 2007). However, the duration of shelterbelt carbon offsets can be lengthened by using shelterbelt timber in long term products, such as lumber (Kort & Turnock, 1999).

Shelterbelt planting and management is a current policy topic because (i) agro-environmental policies are becoming increasingly popular around the world as environmental, economic and social challenges caused by globalization increase (Organisation for Economic Co-operation and Development, 2005); (ii) shelterbelt-related policies can maintain or increase connectivity between habitats in large areas of agricultural land, such as the Canterbury Plains; (iii) shelterbelt-related policies can provide incentives for planting native instead of exotic shelterbelts; and (iv) shelterbelt-related policies may hold economic benefits related to ecosystem services. For example, planting trees to sequester carbon is a relatively inexpensive strategy for reducing greenhouse gas emissions, whereas other alternatives are politically and economically unattractive (e.g. carbon taxes) (Adams *et al.*, 1993).

Currently, New Zealand has no legislation in place that specifically protects or encourages the establishment of shelterbelts. Shelterbelts are only indirectly protected under the Resource Management Act 1991 (RMA), which requires that outstanding natural features and landscapes be protected from inappropriate subdivision, use, and development. Case law has determined that shelterbelts are one of the features that make up an outstanding landscape ("Wakatipu Environmental Society Inc v Queenstown-Lakes District Council [2000] NZRMA 59, at 97,"). Methods that delineate and characterize shelterbelts from remote sensing images can therefore be used to assess shelterbelt coverage in terms of land use change, and therefore whether further legislation which specifically protects shelterbelts is needed.

Other countries such as Canada, Denmark and the United Kingdom, already have shelterbelt-related policies and programs (Baudry *et al.*, 2000; Kristensen & Casperson, 2002;

Kulshreshtha *et al.*, 2005). There are three types of policy that can be used to increase or maintain shelterbelt coverage in agricultural land, as well as encourage the planting of native shelterbelts. Firstly, rules or regulations can be used by local or national governments to protect existing shelterbelts, such as classing shelterbelt removal as a controlled activity. For example, legislation in the United Kingdom (Hedgerows Regulations 1997), protect hedgerows by stipulating under what conditions removal is allowed (Baudry *et al.*, 2000). Secondly, local or national governments can provide incentives for planting shelterbelts, such as government subsidies (Adams *et al.*, 1993). Thirdly, seedlings can be supplied at a reduced or no cost to qualifying landowners. For example, the three Prairie Provinces of Canada created the Agriculture and Agri-Food Canada-Prairie Farm Rehabilitation Administration (AAFC-PFRA) Shelterbelt Centre in 1901 (Kulshreshtha *et al.*, 2005). This centre supplies tree and shrub seedlings at no cost to farmers and qualifying agencies, such as municipal and wildlife agencies. This program establishes and maintains shelterbelts on farms and other eligible sites that are exposed to strong winds. It therefore shelters farmsteads, prevents soil erosion and crop damage, and reduces risks along roads. A recent evaluation of this program (Kulshreshtha *et al.*, 2005) showed that it generated substantial environmental benefits, including reducing greenhouse gas accumulation in the atmosphere. This ecosystem service alone was valued at between 56 and 417 million Canadian dollars for the three Prairie Provinces. This Canadian example helps to illustrate the potential value of introducing shelterbelt-related policies and programs in New Zealand.

Summary of recommendations and future research directions

Several recommendations that may improve the accuracy of shelterbelt delineation, shelterbelt species differentiation, and carbon estimation are suggested. First, using a combination of satellite imagery and LIDAR data is likely to improve the accuracy of all the methods described. Second, using images captured in autumn or multitemporal images may improve differentiation between coniferous species. Third, various measures of texture and vegetation indices can be investigated to determine which are most effective for differentiating shelterbelt species and estimating shelterbelt carbon. Fourth, the shelterbelt sample area can be expanded to include the whole of Canterbury. This will make the sample more representative of Canterbury, as well as including more broadleaved shelterbelts. Lastly, shelterbelt carbon estimates can be improved by determining allometric relationships specific to Canterbury shelterbelts by destructive analysis.

Other research directions are also suggested. First, atmospheric correction procedures specific to QuickBird imagery are needed so that methods can be applied over larger areas. Second,

the shelterbelt delineation method needs to be tested on SPOT-5 imagery. This will enable change in shelterbelt landcover resulting from agricultural intensification to be determined (historical SPOT-5 imagery is available). Third, shelterbelt delineation and species differentiation methods can be used to assess the efficiency of shelterbelts as wildlife corridors across the Canterbury Plains.

Summary of contributions to shelterbelt research

This thesis has worked through several technical methods, to identify which methods may be better than others to accurately map and characterize shelterbelts. This thesis has made several contributions to the current body of literature.

First, studies using remote sensing image analysis to delineate shelterbelts, differentiate shelterbelt species, and estimate shelterbelt biomass are uncommon. This thesis therefore fills a gap in current knowledge.

Second, this thesis is the first study of its kind to develop remote sensing-based methods for studying shelterbelts in New Zealand in a spatially explicit manner.

Third, this thesis has confirmed that delineation of narrow vegetation features from satellite imagery is most successful using object-oriented classification with specialist feature extraction software. It has also shown that the performance of Overwatch Feature Analyst is comparable to that of Definiens eCognition software, the latter of which is commonly used in studies extracting linear vegetation features.

Fourth, this thesis is the first study of its kind to attempt the differentiation of *Pinus radiata* and *Cupressus macrocarpa* tree species. It has shown that a classification tree approach which uses spectral variables as predictors produces reasonable results, but that the method has room for improvement.

Fifth, this thesis has shown that shelterbelts represent a significant carbon reservoir on the Canterbury Plains, and that it can be utilized for offsetting carbon and carbon trading under the Kyoto Protocol. This thesis is also the first study of its kind to show that shelterbelt carbon can be modelled using species-specific spectral-biomass relationships, which in turn can be used for estimating carbon across landscapes.

In conclusion, this thesis has made a contribution towards developing methods for studying shelterbelts in a spatially explicit manner. With improvements, the methods presented in this thesis could be used in the future to semi-automatically delineate and characterize shelterbelts across the Canterbury Plains, or even New Zealand.

References

Adams, R. M., Adams, D. M., Callaway, J. M., Chang, C., & McCarl, B. A. (1993). Sequestering carbon on agricultural land: Social cost and impacts on timber markets. *Contemporary Policy Issues, 11*(1), 76-88.

Afifi, A., Clark, V. A., & May, S. (2004). *Computer-Aided Multivariate Analysis* (Fourth ed.). Boca Raton: Chapman & Hall/CRC.

Aksoy, S., Akçay, C. H., & Wassenaar, T. (2010). Automatic mapping of linear woody vegetation features in agricultural landscapes using very high resolution imagery. *Transactions on Geoscience and Remote Sensing, 48*(1), 511-522.

Baudry, J., Bunce, R. G. H., & Burel, F. (2000). Hedgerows: An international perspective on their origin, function and management. *Journal of Environmental Management, 60*(1), 7-22.

Baumgartner, A., Steger, C., Mayer, H., Eckstein, W., & Ebner, H. (1999). Automatic road extraction based on multi-scale, grouping, and context. *Photogrammetric Engineering & Remote Sensing, 65*(7), 777-785.

Beets, P. N. R., K. A., Ford-Robertson, J. B., Gordon, J., & Maclaren, J. P. (1999). Description and validation of C_change: A model for simulating carbon content in managed Pinus radiata stands. *New Zealand Journal of Forestry Science 29*(3), 409-427.

Berkeley Environmental Technology. (2010). BerkeleyImageSeg (Version 1.0rc8). Berkeley

Bi, H. Q., Long, Y. S., Turner, J., Lei, Y. C., Snowdon, P., Li, Y., Harper, R., Zerihun, A., & Ximenes, F. (2010). Additive prediction of aboveground biomass for Pinus radiata (D.Don). *Forest Ecology and Management, 259*(12), 2301-2314.

Blaschke, T. (2010). Object based image analysis for remote sensing. *Journal of Photogrammetry and Remote Sensing 65*, 2-16.

Blaschke, T., Lang, S., Lorup, E., Strobl, J., & Zeil, P. (2000). Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications. In A. Cremers & K. Greve (Eds.), *Environmental Information for Planning, Politics and the Public* (Vol. 2, pp. 555-570). Marburg, Germany: Metropolis-Verlag.

Brandle, J. R., Wardle, T. D., & Bratton, G. F. (1992). Opportunities to increase tree planting in shelterbelts and the potential impacts on carbon storage and conservation. In R. N. Sampson & D. Hair (Eds.), *Forests and Global Change* (Vol. 1: Opportunities for Increasing Forest Cover, pp. 157-176). Washington DC: American Forests.

Brandtberg, T. (2002). Individual tree-based species classification in high spatial resolution aerial images of forests using fuzzy sets. *Fuzzy Sets and Systems, 132*(3), 371-387.

Brazil's National Institute for Space Research (INPE). (2009). SPRING (Version 5.1). São José dos Campos

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32.

Breiman, L. (2002). *Manual on setting up, using, and understanding random forests V3.1*. Retrieved 17 April 2011 from http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf

Breiman, L., & Cutler, A. (2010). *Random Forest: Breiman and Cutler's random forests for classification and regression*. Retrieved 17 April 2011 from http://cran.r-project.org/web/packages/randomForest/

Buddenbaum, H., Schlerf, M., & Hill, J. (2005). Classification of coniferous tree species and age classes using hyperspectral data and geostatistical methods. *International Journal of Remote Sensing, 26*(24), 5453-5465.

Buytaert, W. (2007). Many carbon offset projects have a negative impact on the environment and global warming. In D. A. Miller (Ed.), *Carbon Offsets* (First ed., pp. 105-108). USA: Greenhaven Press.

Câmara, G., Souza, R. C. M., Freitas, U. M., Garrido, J., & Mitsuo, F. (1996). Integrating remote sensing and GIS by object-oriented data modelling. *Computers and Graphics, 20*(3), 395-403.

Cameron, R. J. (1964). Destruction of the indigenous forests for Maori agriculture during the nineteenth century. *New Zealand Journal of Forestry Science, 9*(1), 98-109.

Campbell, J. B. (2007). *Introduction to Remote Sensing* (Fourth ed.). New York: The Guilford Press.

Carroll, J. D., & Green, P. E. (1997). *Mathematical Tools for Applied Multivariate Analysis*. San Diego: Elsevier Science.

Carswell, F. E., Burrows, L. E., & Mason, N. W. H. (2009). Above-ground carbon sequestration by early-successional woody vegetation: a preliminary analysis. *Science for Conservation 297 (NZ Department of Conservation)*, 1-20.

Case, B. S., & Hall, R. J. (2008). Assessing prediction errors of generalized tree biomass and volume equations for the boreal forest region of west-central Canada. *Candian Journal of Forestry Research, 38*, 878-889.

Castillo-Santiago, M. A., Ricker, M., & de Jong, B. H. J. (2010). Estimation of tropical forest structure from SPOT-5 satellite images. *International Journal of Remote Sensing, 31*(10), 2767-2782.

Chatfield, C., & Collins, A. J. (1980). *Introduction to Multivariate Analysis* (First ed.). Boca Raton: CRC Press.

Chatterjee, S., & Hadi, A. S. (2006). *Regression analysis by example* (Fourth ed.). Hoboken: John Wiley & Sons, Inc.

Chavez, P. S. (1996). Image-based atmospheric corrections - revisited and improved. *Photogrammetric Engineering and Remote Sensing, 62*(9), 1025-1036.

Clinton, N. (2010). *BerkeleyImageSeg User's Guide*. Retrieved 18 May 2011 from www.berkenviro.com/berkeleyimgseg/BIS_manual_1_0_20100901.pdf

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37*(1), 35-46.

Coomes, D. A., Allen, R. B., Scott, N. E., Goulding, C., Beets, P. N., & Robertson, K. A. (2002). Designing systems to monitor carbon stocks in forests and shrublands. *Forest Ecology and Management, 164*, 89-108.

Coops, N. C., Waring, R. H., & Landsberg, J. J. (1998). Assessing forest productivity in Australia and New Zealand using a physiologically-based model driven with averaged monthly weather data and satellite derived estimates of canopy photosynthetic capacity. *Forest Ecology and Management, 104*(1-3), 113-127.

Cracknell, A. P. (1998). Synergy in remote sensing - What is a pixel? *International Journal of Remote Sensing, 19*(11), 2025-2047.

Cracknell, A. P., & Hayes, L. (2007). *Introduction to Remote Sensing* (Second ed.). USA: CRC Press.

David, C. A., & Rhyner, V. (1999). An assessment of windbreaks in central Wisconsin. *Agroforestry Systems, 44*, 313-331.

De Brauw, A. (2006). The Kyoto Protocol, market power, and enforcement. *Applied Economics, 38*(18), 2169-2178.

DigitalGlobe Corporate. QuickBird II satellite imagery. Longmont.

Dong, J., Kaufmann, R. K., Myeni, R. B., Tucker, C. J., Kauppi, P. E., Liski, J., Buermann, W., Alexeyev, V., & Hughes, M. K. (2003). Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources, and sinks. *Remote Sensing of Environment, 84*(3), 393-410.

DPI/INPE. (2009). *Georeferenced Information Processing System - SPRING*. Retrieved 18 May 2011 from www.dpi.inpe.br/spring/english/

ERDAS. (2008). *Image Analysis for ArcGIS Desktop Help*. Norcross: ERDAS, Inc.

ERDAS. (2009). Image Analysis Extension for ArcGIS 9.3. Norcross

Erikson, M. (2004). Species classification of individually segmented tree crowns in high-resolution aerial images using radiometric and morphologic image measures. *Remote Sensing of Environment, 91*, 469-477.

ESRI. (2005). ArcPad (Version 7.0). Redlands

ESRI. (2009a). ArcGIS (Version 9.3). Redlands

ESRI. (2009b). ArcGIS 9.3 Desktop Help. Redlands: ESRI

Everitt, B. S. (2005). *An R and S-Plus ® Companion to Multivariate Analysis*. USA: Springer-Verlag London Limited.

Foody, G. M., Boyd, D. S., & Cutler, M. E. J. (2003). Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remotes Sensing of Environment 85*(4), 463-474.

Förster, M., & Kleinschmit, B. (2006). Integration of ancillary information into object-based classification for detection of forest structures and habitats. In S. Lang, T. Blaschke & E. Schöpfer (Eds.), *First international conference on object-based image analysis.* (Vol. XXXVI-4/C42, pp. pp 1-6). Salzburg: ISPRS.

Freeman, C., & Buck, O. (2003). Development of an ecological mapping methodology for urban areas in New Zealand. *Landscape and Urban Planning, 63*(3), 161-173.

Gayon, J. (2000). History of the concept of allometry. *American Zoology, 40*(5), 748-758.

Goetz, S. J., Wright, R. K., Smith, A. J., Zinecker, E., & Schaub, E. (2003). IKONOS imagery for resource management: Tree cover, impervious surfaces, and riparian buffer analysis in the mid-Atlantic region. *Remote Sensing of Environment 88*(1-2), 195-208.

Gregory, N. G. (1995). The role of shelterbelts in protecting livestock: a review. *New Zealand Journal of Agricultural Research, 38*, 423-450.

Gruen, A., & Li, H. (1997). Semi-automatic linear feature extraction by dynamic programming and LSB-snakes. *Photogrammetric Engineering & Remote Sensing, 63*(8), 985-995.

Guyon, I., & Elisseeff, A. (2006). *An Introduction to Feature Extraction* (Vol. 207: Feature Extraction: Foundations and Applications). Netherlands: Springer.

Hájek, F. (2006). *Object analysis of IKONOS XS and pan-sharpened imagery in comparison for purpose of tree species estimation*. Paper presented at the 1st international conference on object-based image analysis (OBIA 2006).

Hájek, F. (2008). Process-based approach to automated classification of forest structures using medium format digital aerial photos and ancillary GIS information. *European Journal of Forest Research, 127*(2), 115-124.

Hay, G. J., Castilla, G., Wulder, M. A., & Ruiz, J. R. (2005). An automated object-based approach for the multiscale image segmentation of forest scenes. *International Journal of Applied Earth Observation and Geoinformation, 7*(4), 339-359.

Hedgerows Regulations, § 5 (1997).

Hengl, T. (2006). Finding the right pixel size. *Computers and Geosciences, 32*(9), 1283-1298.

Holmgren, J., & Persson, A. (2004). Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment, 90*(4), 415-423.

Holmgren, J., Persson, A., & Söderman, U. (2008). Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *International Journal of Remote Sensing, 29*(5), 1537-1552.

Jenkins, J. C., Chojnacky, D. C., Heath, L. S., & Birdsey, R. A. (2003). National-scale biomass estimators for United States tree species. *Forest Science, 49*(1), 12-35.

Jenkins, J. C., Chojnacky, D. C., Heath, L. S., & Birdsey R, A. (2004). Comprehensive database of diameter-based biomass regressions for North American tree species. *General technical report NE-319 (United States Department of Agriculture Forest Service),*

Jensen, J. R. (2005). *Introductory Digital Image Processing* (Third ed.). USA: Pearson Education.

Johansen, K., Coops, N. C., Gergel, S. E., & Stange, Y. (2007). Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification. *Remote Sensing of Environment 4, 110*(1), 29-24.

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (Sixth ed.). Upper Saddle River: Pearson Education, Inc.

Johnson, T. (2008). Are tree-based carbon offsets beneficial? Chapter Overview. In D. A. Miller (Ed.), *Carbon Offsets* (pp. 114-121). USA: Greenhaven Press.

Kalaitzidis, C., & Zianis, D. (2009). Current state of terrestrial and earth observation techniques for biomass estimation in Europe. *Remote Sensing for a Changing Europe: Proceedings of the 28th Symposium of the European Assosiation of Remote Sensing Laboratories*, 109-116.

Katoh, M., Gougeon, F. A., & Leckie, D. G. (2009). Applications of high-resolution airborne data using individual tree crowns in Japanese conifer plantations. *Journal of Forestry Research, 14*, 10-19.

Ke, Y., Quackenbush, L. J., & Im, J. (2010). Synergistic use of Quickbird multispectral imagery and LIDAR data for object-based forest species classification. *Remote Sensing of Environment, 114*, 1141-1154.

Key, T., Warner, T. A., McGraw, J. B., & Fajvan, M. A. (2001). A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood forest. *Remote Sensing of Environment 75*(1), 100-112.

Kort, J., & Turnock, R. (1999). Carbon reservoir and biomass in Canadian prairie shelterbelts. *Agroforestry Systems, 44*(2-3), 175-186.

Kristensen, S. P., & Casperson, O. H. (2002). Analysis of changes in a shelterbelt network landscape in central Jutland, Denmark. *Journal of Environmental Management 66*, 171-183.

Kulshreshtha, S., Knopf, E., Kort, J., & Grimard, J. (2005). The Canadian Shelterbelt Program: Economic Valuation of Benefits. In *Evaluating Agri-Environmental Policies: Design, practice and results* (2005 ed., pp. 347-362): OECD Publishers.

Landcare Research. Retrieved 21 May 2011 from www.landcareresearch.co.nz

Lechner, A. M., Stein, A., Jones, S. D., & 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.

Leckie, D. G., Gougeon, F. A., Walsworth, N., & Paradine, D. (2003). Stand delineation and composition estimation using semi-automated individual tree crown analysis. *Remote Sensing of Environment, 85*, 355-369.

Lennon, M., Mouchot, M. C., Mercier, G., & Hubert-Moy, L. (2000). Segmentation of hedges on CASI hyperspectral images by data fusion from texture, spectral and shape analysis. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium, 2*, 825-827.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest In *R News* (Vol. 2/3, pp. 18-22): R Foundation for Statistical Computing.

Liknes, G. C., Perry, C. H., & Meneguzzo, D. M. (2010). Assessing tree cover in agricultural landscapes using high resolution aerial imagery. *The Journal of Terrestrial Observation, 2*(1), 38-55.

Lillesand, T. M., & Kiefer, R. W. (1994). *Remote Sensing and Image Interpretation* (Third ed.). New York: John Wiley and Sons.

Lu, D. (2005). Above-ground biomass estimation using Landsat TM data in the Brazilian Amazon Basin. *International Journal of Remote Sensing, 26*(12), 2509-2525.

Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. International Journal of Remote Sensing. *27, 7*, 1297-1328.

Luckman, A., Baker, J. R., Kuplich, T. M., Yanasse, C. C. F., & Frery, A. C. (1997). A study of the relationship between radar backscatter and regenerating forest biomass for space borne SAR instrument. *Remote Sensing of Environment, 60*(1), 1-13.

Madgwick, H. A. I. (1994). *Pinus radiata: biomass, form and growth*. Rotorua: Madgwick, H. A. I.

Magcale-Macandog, D. B., Delgado, M. E. M., Ty, E., & Villarin, J. R. T. (2006). A GIS-based model to improve estimation of above-ground biomass of secondary forests in the Philippines. *Journal of Tropical Forest Science, 18*(1), 8-21.

McLachlan, A. R. G., & Wratten, S. D. (2003). Abundance and species richness of field-margin and pasture spiders (Araneae) in Canterbury), New Zealand. *New Zealand Journal of Zoology, 30*, 57-67.

Mickler, R. A., Earnhardt, T. S., & Moore, J. A. (2002). Regional estimation of current and future forest biomass. *Environmental Pollution, 116*, s7-s16.

Miller, D. A. (2009). Introduction. In D. A. Miller (Ed.), *Carbon Offsets* (pp. 16-19). USA: Greenhaven Press.

Miller, J. E., Nelson, S. A. C., & Hess, G. R. (2009). An object extraction approach for impervious surface classification with very-high-resolution imagery. *The Professional Geographer, 61*(2), 250-264.

Ministry for the Environment. *Land Use and Carbon Analysis System (LUCAS)*. Retrieved 18 May 2011 from www.mfe.govt.nz/issues/climate/lucas/

Ministry for the Environment. (2009). *New Zealand's Greenhouse Gas Inventory 1990 - 2007: An Overview*. Retrieved from www.mfe.govt.nz/publications/climate/greenhouse-gas-inventory-2009/index.html

Ministry of Forestry. (1992). *Shelter in Canterbury*. Christchurch: Ministry of Forestry.

Moore, J. R. (2010). Allometric equations to predict the total above-ground biomass of radiata pine trees. *Annals of Forest Science, 67*(8), Article 86.

Nichol, J. E., & Sarker, M. L. R. (2011). Improved biomass estimation using the texture parameters of two high-resolution optical sensors. *IEEE Transactions on Geoscience and Remote Sensing, 49*(3), 930-948.

NZ Defence Force Geospatial Intelligence Oganisation. *KiwImage Project*. Retrieved 18 March 2010 from www.kiwimage.govt.nz

Olofsson, K., Wallerman, J., Holmgren, J., & Olsson, H. (2006). Tree species discrimination using Z/I DMC imagery and template matching of single trees. *Scandinavian Journal of Forest Research., 21*(1), 106-110.

Organisation for Economic Co-operation and Development. (2005). *Evaluating Agri-Environmental Policies: Design, practice and results*. USA: OECD.

Orka, H. O., Næsset, E., & Bollandsås, O. M. (2009). Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sensing of Environment, 113*(6), 1163-1174.

Overwatch. (2010). Feature Analyst (Version 5.0). Missoula: Visual Learning Systems, Inc.

Overwatch Ltd. (2010). *Feature Analyst 5.0 for ArcGIS: Reference Manual*. Missoula: Visual Learning Systems, Inc.

Pankiw, J., Kort, J., & Piwowara, M. (2009). Prairie shelterbelt inventory: Using high resolution imagery and object-based classification. *Proceedings of the 2009 Spatial Knowledge and Information Canada Conference, 1*, 46-47.

Phua M, & Saito, H. (2003). Estimation of biomass of a mountainous tropical forest using Landsat TM data. *Canadian Journal of Remote Sensing, 29*(4), 429-440.

Popescu, S. C., Wynne, R. H., & Nelson, R. F. (2003). Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing, 29*(5), 564-577.

Price, L. R. (1993). Hedges and shelterbelts on Canterbury plains, New Zealand: transformation of an antipodean landscape. *Annals of the Association of American Geographers, 83*(1), 119-140.

Quackenbush, L. J. (2004). A review of techniques for extracting linear features from imagery. *Photogrammetric Engineering & Remote Sensing, 70*(12), 1383-1392.

Richards, J. A., & Jia, X. (2006). *Remote Sensing Digital Image Analysis: An Introduction* (Fourth ed.). Berlin: Springer-Verlag.

Resource Management Act, § 6 (b) (1997).

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings of the 3rd Earth Resource Technology Satellite (ERTS) Symposium, 1*, 48-62.

Schuurman, N. (2004). *GIS: A Short Introduction*. Oxford: Blackwell Publishing Ltd.

Senelwa, K., & Sims, R. E. H. (1998). Tree biomass equations for short rotation eucalypts grown in New Zealand. *Biomass and Bioenergy, 13*(3), 133-140.

Shih, F. Y. (2009). *Image Processing and Mathematical Morphology: Fundamentals and Applications*. Boca Raton: CRC Press.

Specht, A., & West, P. W. (2003). Estimation of biomass and sequestered carbon on farm forest plantations in northern New South Wales, Australia. *Biomass and Bioenergy, 25*, 363-379.

Steininger, M. K. (2000). Satellite estimation of tropical secondary forest above-ground biomass: Data from Brazil and Bolivia. *International Journal of Remote Sensing, 21*(6), 1139-1157.

Stringer, D. J. (1977). *A guide to the establishment of windbreaks*. Timaru NZ: South Canterbury Catchment Board and Regional Water Board Wind Erosion Control Scheme.

Suratno, A., Seielstad, C., & Queen, L. (2009). Tree species identification in mixed coniferous forest using airborne laser scanning. *Journal of Photogrammetry and Remote Sensing 64*, 683-693.

Tansey, K., Chambers, I., Anstee, A., Denniss, A., & Lamb, A. (2009). Object-oriented classification of very high resolution airborne imagery for the extraction of hedgerows and field margin cover in agricultural areas. *Applied Geography*, *29*(2), 145-157.

The R Foundation for Statistical Computing. (2010). R (Version 2.12.0). Available from http://www.r-project.org/

Thenkabail, P. S., Stucky, N., Griscom, B. W., Ashton, M. S., Diels, J., van der Meer, B., & Enclona, E. (2004). Biomass estimations and carbon stock calculations in the oil palm plantations of African derived savannas using IKONOS data. *International Journal of Remote Sensing, 25*(23), 5447-5472.

Thornton, M. W., Atkinson, P. M., & Holland, D. A. (2006). Sub-beeldelement mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping. *International Journal of Remote Sensing, 27*(3), 473-491.

Thornton, M. W., Atkinson, P. M., & Holland, D. A. (2007). A linearised pixel-swapping method for mapping rural linear land cover features from fine spatial resolution remotely sensed imagery. *Computers & Geosciences*, *33*, 1261-1272.

Tomppo, E., Nilsson, M., Rosengren, M. A., P., & Kennedy, P. (2002). Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass. *Remote Sensing of Environment, 82*(1), 156-171.

Vannier, C., & Hubert-Moy, L. (2008). Detection of wooded hedgerows in high resolution satellite images using an object-oriented method. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, *4*, 731-734.

Velasquez-Manoff, M. (2007). The problem with individual carbon offsets now is that they are not properly regulated. In D. A. Miller (Ed.), *Carbon Offsets* (pp. 161-167). USA: Greenhaven Press.

Wakatipu Environmental Society Inc v Queenstown-Lakes District Council [2000] NZRMA 59, at 97.

Wang, X., & Feng, Z. (1995). Atmospheric carbon sequestration through agroforestry in China. *Energy, 20*(2), 117-121.

Waser, L. T., Ginzler, C., Kuechler, M., Baltsavias, E., & Hurni, L. (2011). Semi-automatic classification of tree species in different forest ecosystems by spectral and geometric variables derived from Airborne Digital Sensor (ADS40) and RC30 data. *Remote Sensing of Environment, 115,* 76-85.

Wiseman, G., Kord, J., & Walker, D. (2009). Quantification of shelterbelt characteristics using high-resolution imagery. *Agriculture, Ecosystems and Environment, 131*(1-2), 111-117.

Wu, J., Wang, D., & Bauer, M. E. (2005). Image-based atmospheric correction of Quickbird imagery of Minnesota cropland. *Remote Sensing of Environment, 99*, 315-325.

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., & Ryu, S. (2004). Estimating above-ground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment, 93*, 402-411.

Appendix A
QuickBird Raw Images

A.1 Image specifications

Multispectral and panchromatic images for the three study areas were taken from two out of three image tiles. All images were photographed on 27 February 2008.

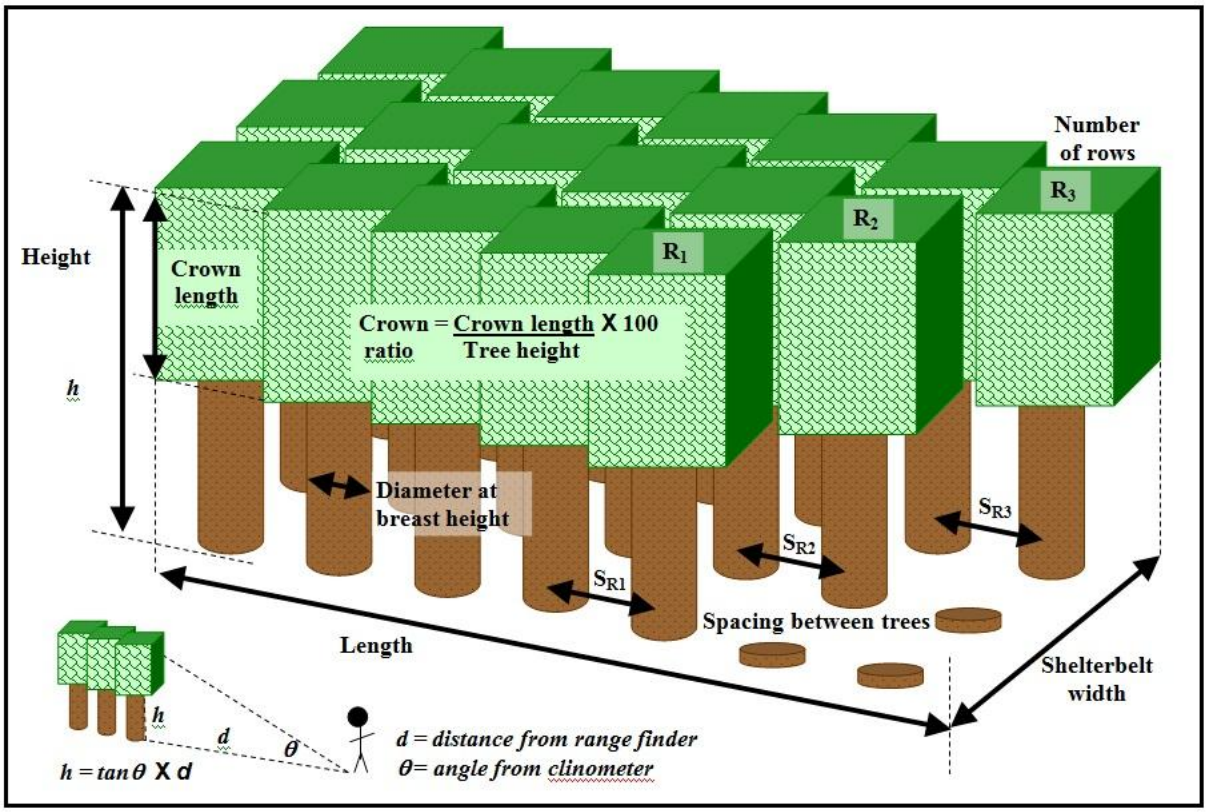
QuickBird II imagery sourced from KiwImage:

Imagery	Resolution		Bands	KiwImage image tiles	
	Spatial	Radiometric			
Panchromatic	0.6 m	16 bit	1 band	▪ bv24_52249656_pan_27feb08_R1C1 ▪ bv24_52249656_pan_27feb08_R2C1	
Multispectral	2.4 m	16 bit	4 bands: RGBNI	▪ bv24_52249656_mul_27feb08_R1C1 ▪ bv24_52249656_mul_27feb08_R2C1	
				Spectral resolution of individual bands	
			1	Blue	450-520 nm
			2	Green	520-600 nm
			3	Red	630-690 nm
			4	Near Infrared	760-900 nm

Appendix B

Data Collection in the Field

B.1 Illustration of Shelterbelt Measurements



B.2 Summary of Shelterbelt and Non-Shelterbelt Observations

	CLASS	# SAMPLES	LENGTH
SHELTERBELT	Single species	107	15.65 km
	Mixed species	26	5.14 km
	TOTAL	**133**	**20.79 km**
OTHER	Chopped down shelterbelt	2	0.37 km
	Shelterbelt perpendicular to road	2	
	Hedge	8	
	Group of trees	11	
	Other vegetation	17	
	Willow trees	4	
	Woodlot block	5	
	TOTAL	**49**	
	TOTAL	**182**	

B.3 Summary of Shelterbelt Observations by Species Type

		CLASS	# SAMPLES	LENGTH
VISIBLE ON IMAGES	Single	*P.radiata*	58	9.55 km
		C.macrocarpa	26	3.59 km
		Other Conifer	3	0.31 km
		Eucalyptus spp.	3	0.42 km
		Populus spp.	2	0.09 km
		Other Broadleaf	1	0.13 km
	Mixed	*P.radiata & C.macrocarpa*	12	2.72 km
		P.radiata & Other Conifer	3	0.63 km
		Other Conifer	1	0.26 km
		P.radiata & C.macrocarpa & Eucalyptus spp.	3	0.40 km
		P.radiata & Other Conifer & *Eucalyptus spp.*	1	0.10 km
		C.macrocarpa & Other Broadleaf	1	0.06 km
		Popupus spp. & Other Broadleaf	1	0.04 km
		Native	1	0.03 km
	Summary	**Conifer**	103	17.06 km
		Broadleaf	8	0.71 km
		Mixed (Conifer & Broadleaf)	5	0.56 km
		TOTAL	**116**	**18.33 km**
NOT VISIBLE ON IMAGE	Single	*P.radiata*	2	0.33 km
		C.macrocarpa	5	0.53 km
		Populus spp.	5	0.67 km
		Native	1	0.03 km
		Other Broadleaf	1	0.02 km
	Mixed	*Eucalyptus spp. & C.macrocarpa*	2	0.65 km
		P.radiata & C.macrocarpa	1	0.24 km
	Summary	**Conifer**	8	1.10 km
		Broadleaf	7	0.72 km
		Mixed (Conifer & Broadleaf)	2	0.65 km
		TOTAL	**17**	**2.46 km**

B.4 Physical Characteristics of All Shelterbelts

Visible on image.

	Width (m)	Length (m)	Rows	Tallest Height (m)	Average DBH (cm)	Density (trees/ 100m^2)	Crown Ratio (%)
Mean	8	158	1.6	12	42	13.69	96
Std Dev	5.5	141	1.1	7.4	21.3	14.98	8.1
Median	6	114	1	11	38	10.21	100
Range	29	733	8	35	143	131.85	40

Not visible on image.

	Width (m)	Length (m)	Rows	Tallest Height (m)	Average DBH (cm)	Density (trees/ 100m^2)	Crown Ratio (%)
Mean	4	145	1.5	5	17	31.57	93
Std Dev	3.5	149	0.8	4.3	14.1	25.23	7.7
Median	3	144	1	5	9	22.37	90
Range	15	513	3	18	54	84.56	20

Difference between visible and not-visible shelterbelts.

	Width (m)	Length (m)	Rows	Tallest Height (m)	Average DBH (cm)	Density (trees/100m^2)
Difference	3 to 5	7 to 19	-0.4 to 1	6 to 8	23 to 27	15.44 to 20.32
	Significant	Significant	Not significant	Significant	Significant	Significant

Where $\bar{x}_1 - \bar{x}_2 = (\bar{x}_1 - \bar{x}_2) - 1.96 \times \sqrt{\left(\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}\right)}$ to $(\bar{x}_1 - \bar{x}_2) + 1.96 \times \sqrt{\left(\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}\right)}$

Where $\bar{x}$ = mean
σ = standard deviation
n = sample size

B.5 Physical Characteristics of Shelterbelt Groups

CONIFEROUS: Visible on image.

	Width (m)	Length (m)	Rows	Tallest Height (m)	Average DBH (cm)	Density (trees/ 100m^2)	Crown Ratio (%)
Mean	8	166	1.6	12	43	13.88	97
Std Dev	5.7	145	1.2	7.4	21.9	15.42	7.0
Median	6	125	1	10	40	10.60	100
Range	29	733	8	35	143	131.85	40

BROADLEAVED: Visible on image.

	Width (m)	Length (m)	Rows	Tallest Height (m)	Average DBH (cm)	Density (trees/ 100m^2)	Crown Ratio (%)
Mean	7	88	1.9	16	29	15.03	81
Std Dev	3.2	67	0.6	7.3	8.7	13.30	6.4
Median	7	61	2	14	30	7.28	80
Range	10	192	2	21	26	35.37	20

Difference between visible coniferous and visible broadleaved shelterbelts.

	Width (m)	Length (m)	Rows	Tallest Height (m)	Average DBH (cm)	Density (trees/100m^2)
Difference	-0.3 to 2	72 to 84	-0.3 to 1	2 to 6	12 to 16	-1.49 to 3.79
	Not significant	Significant	Not significant	Significant	Significant	Not significant

Where $\bar{x}_1 - \bar{x}_2 = (\bar{x}_1 - \bar{x}_2) - 1.96 \times \sqrt{\left(\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}\right)}$ to $(\bar{x}_1 - \bar{x}_2) + 1.96 \times \sqrt{\left(\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}\right)}$

Where $\bar{x}$ = mean
σ = standard deviation
n = sample size

Appendix C

Shelterbelt Delineation

C.1 Per-pixel classification classes

Unsupervised classification:

Landcover class	Original classes	Final class
Shadow	1	1
Shelterbelt	3, 7	2
Other	2, 4, 5, 6, 8-20	3

Supervised classification training shapefile:

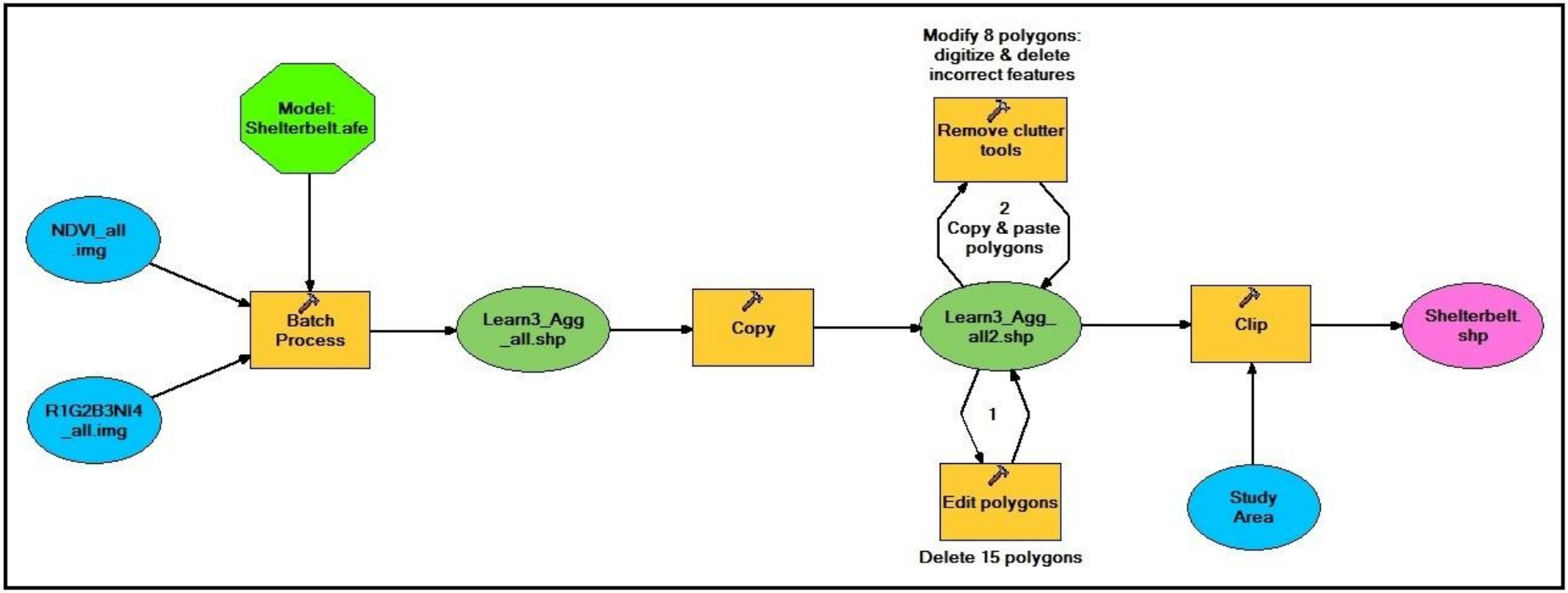
Landcover class	Training sites #	Total area m^2
Shelterbelt	16	14540
Shadow	16	2283
Other	20	459589

C.2 Feature extraction settings

Shelterbelt extraction settings for the Feature Analyst Learner:

	SETTINGS:		
Feature Selector: Narrow Linear Feature (<10 m)	**Input band settings**	Histogram stretch:	Histogram equalize
		Image resolution:	0.6 m/pixel
		Resample factor:	1 x
		Find rotated features:	Yes
	Input representation	Pattern:	Bull's eye 1
		Pattern width:	9 cells
	Masking	Initial learning:	None
		Hierarchical learning:	Clutter removal layers
	Output options	Output format:	Vector
		Score shapes:	Yes
		Post-processing:	Aggregate small regions < 25 m^2
Input Bands	1. R1G2B3NI4.img 2. NDVI.img		
Training Set	**Shelterbelt**	Training sites #:	16
		Total Area:	14540 m^2
Hierarchical learning	1. Remove clutter 2. Remove clutter by shape (invariants setting)		
Post-processing tools	**Aggregate**	Remove shapes	< 50 m^2

C.3 Application of Feature Analyst Model to Other Study Areas



C.4 Batch Processing Summary

Summary of batch processing and post-processing steps.

		File name	# Polygons	Total area m^2	Study area
Batch Prcocessing Tool	**Model**	Shelterbelt.afe			
	Input 1	R1G2B3NI4_all.img			
	Input 2	NDVI_all.img			
	Output	Learn3_Agg_all.shp	1436663	989	
Copy & Edit	**Output**	Learn3_Agg_all2.shp	1422163	973	
Clip to Study Areas	**Input**	Learn3_Agg_all2.shp	1422163	973	Area encompassing all study areas
	Mask 1	BV24.shp	1	16000000	
	Mask 2	BV24_2.shp	1	16000000	
	Mask 3	BV24_3.shp	1	16000000	
	Output 1	Shelterbelts_1.shp	580407	336	Study Area 1
	Output 2	Shelterbelts_2.shp	288016	230	Study Area 2
	Output 3	Shelterbelts_3.shp	356323	321	Study Area 3

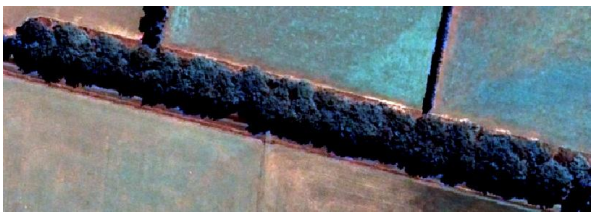
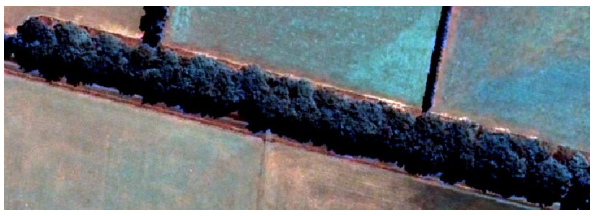
Appendix D

Data Exploration

D.1 Spectral and Textural Information

		INFORMATION EXTRACTED FROM SPECTRAL BANDS, BAND RATIOS AND TEXTURE FOR EACH SHELTERBELT OBJECT.								
		MIN	MAX	RANGE	MEAN	STD	VARIETY	MAJORITY	MINORITY	MEDIAN
RED	Min	1	161	55	70	12	35	1	23	58
	Mean	45	245	199	130	37	168	127	67	130
	Max	123	378	348	224	60	278	251	251	228
GREEN	Min	1	318	98	151	22	44	1	56	134
	Mean	95	399	303	249	58	242	255	99	255
	Max	232	621	556	348	120	513	382	232	363
BLUE	Min	47	243	57	132	13	34	68	48	121
	Mean	80	298	218	194	41	183	202	80	199
	Max	186	417	369	251	77	342	276	197	260
NI	Min	86	602	190	352	33	53	156	86	304
	Mean	263	789	526	559	104	403	577	263	570
	Max	568	1418	1167	868	243	1047	835	568	847
NDVI	Min	0.19	0.61	0.09	0.45	0.02				
	Mean	0.39	0.82	0.43	0.63	0.07				
	Max	0.62	1.00	0.66	0.70	0.10				
NI/G	Min	0.7	2.2	0.8	1.6	0.2				
	Mean	1.2	18.6	17.4	2.5	2.1				
	Max	2.1	920	919	18.2	113				
TEXTURE	Min	0.7	47.5	44.8	17.6	7.6				
	Mean	4.2	85.6	81.4	31.3	14.5				
	Max	13.1	173	170	55.1	24.0				

D.2 Image Interpretation

<table>
<tr>
<td>  
Broadleaf spp.</td>
<td> 
C.macrocarpa
Easily confused with broadleaved spp.</td>
</tr>
<tr>
<td> 
Eucalyptus spp.</td>
<td> 
C.macrocarpa
Easily confused with Eucalyptus spp.</td>
</tr>
<tr>
<td> 
P.radiata</td>
<td> 
C.macrocarpa
Easily confused with P.radiata.</td>
</tr>
</table>

D.3 PCA Correlation Matrices

Image-derived variables: *P.radiata*

		RED (R)			GREEN (G)			BLUE (B)			NEAR INFRARED (NI)			NDVI		NI/G		TEXTURE (T)	
		Mean	Std	Variety	Mean	Std	Variety	Mean	Std	Variety	Mean	Std	Variety	Mean	Std	Mean	Std	Mean	Std
R	Mean	1.00	-0.12	-0.30	0.92	-0.53	-0.53	0.84	-0.66	-0.62	0.52	-0.48	-0.59	-0.76	-0.17	-0.58	-0.59	0.24	0.30
	Std	-0.12	1.00	0.83	-0.12	0.80	0.71	-0.13	0.65	0.63	-0.09	0.72	0.63	0.11	0.65	0.12	0.48	0.16	0.25
	Variety	-0.30	0.83	1.00	-0.30	0.75	0.93	-0.29	0.64	0.86	-0.22	0.69	0.84	0.20	0.57	0.18	0.49	0.01	0.10
G	Mean	0.92	-0.12	-0.30	1.00	-0.50	-0.51	0.97	-0.63	-0.60	0.76	-0.46	-0.59	-0.50	-0.31	-0.37	-0.55	0.25	0.28
	Std	-0.53	0.80	0.75	-0.50	1.00	0.86	-0.50	0.96	0.87	-0.24	0.84	0.81	0.46	0.68	0.47	0.81	0.20	0.31
	Variety	-0.53	0.71	0.93	-0.51	0.86	1.00	-0.50	0.82	0.98	-0.30	0.77	0.94	0.41	0.58	0.37	0.66	0.05	0.13
B	Mean	0.84	-0.13	-0.29	0.97	-0.50	-0.50	1.00	-0.62	-0.57	0.77	-0.45	-0.58	-0.39	-0.35	-0.31	-0.52	0.17	0.21
	Std	-0.66	0.65	0.64	-0.63	0.96	0.82	-0.62	1.00	0.87	-0.33	0.79	0.80	0.54	0.67	0.53	0.89	0.17	0.27
	Variety	-0.62	0.63	0.86	-0.60	0.87	0.98	-0.57	0.87	1.00	-0.33	0.75	0.93	0.48	0.57	0.45	0.74	0.04	0.14
NI	Mean	0.52	-0.09	-0.22	0.76	-0.24	-0.30	0.77	-0.33	-0.33	1.00	-0.33	-0.41	0.15	-0.38	0.31	-0.19	0.29	0.23
	Std	-0.48	0.72	0.69	-0.46	0.84	0.77	-0.45	0.79	0.75	-0.33	1.00	0.84	0.33	0.53	0.23	0.60	0.03	0.30
	Variety	-0.59	0.63	0.84	-0.59	0.81	0.94	-0.58	0.80	0.93	-0.41	0.84	1.00	0.39	0.52	0.31	0.61	-0.03	0.13
NDVI	Mean	-0.76	0.11	0.20	-0.50	0.46	0.41	-0.39	0.54	0.48	0.15	0.33	0.39	1.00	-0.08	0.91	0.56	-0.04	-0.15
	Std	-0.17	0.65	0.57	-0.31	0.68	0.58	-0.35	0.67	0.57	-0.38	0.53	0.52	-0.08	1.00	0.02	0.63	0.26	0.40
NI/G	Mean	-0.58	0.12	0.18	-0.37	0.47	0.37	-0.31	0.53	0.45	0.31	0.23	0.31	0.91	0.02	1.00	0.63	0.14	0.00
	Std	-0.59	0.48	0.49	-0.55	0.81	0.66	-0.52	0.89	0.74	-0.19	0.60	0.61	0.56	0.63	0.63	1.00	0.28	0.35
T	Mean	0.24	0.16	0.01	0.25	0.20	0.05	0.17	0.17	0.04	0.29	0.03	-0.03	-0.04	0.26	0.14	0.28	1.00	0.47
	Std	0.30	0.25	0.10	0.28	0.31	0.13	0.21	0.27	0.14	0.23	0.30	0.13	-0.15	0.40	0.00	0.35	0.47	1.00

Correlation ≥ 0.8

Image-derived variables: *C.macrocarpa*

		RED (R)			GREEN (G)			BLUE (B)			NEAR INFRARED (NI)			NDVI		NI/G		TEXTURE (T)	
		Mean	Std	Variety	Mean	Std	Variety	Mean	Std	Variety	Mean	Std	Variety	Mean	Std	Mean	Std	Mean	Std
R	Mean	1.00	0.08	-0.03	0.97	-0.26	-0.35	0.93	-0.45	-0.43	0.88	-0.35	-0.48	-0.60	-0.18	0.23	0.24	0.01	0.05
	Std	0.08	1.00	0.78	0.01	0.90	0.65	0.02	0.77	0.63	-0.11	0.55	0.39	-0.22	0.78	0.24	0.24	0.55	0.40
	Variety	-0.03	0.78	1.00	-0.08	0.68	0.91	-0.10	0.61	0.87	-0.18	0.40	0.73	-0.11	0.62	0.13	0.13	0.40	0.10
G	Mean	0.97	0.01	-0.08	1.00	-0.30	-0.37	0.98	-0.50	-0.46	0.94	-0.31	-0.46	-0.45	-0.26	0.23	0.24	-0.01	0.11
	Std	-0.26	0.90	0.68	-0.30	1.00	0.71	-0.28	0.92	0.73	-0.38	0.68	0.52	0.00	0.85	0.20	0.20	0.54	0.48
	Variety	-0.35	0.65	0.91	-0.37	0.71	1.00	-0.39	0.72	0.98	-0.45	0.51	0.88	0.05	0.64	0.04	0.03	0.38	0.07
B	Mean	0.93	0.02	-0.10	0.98	-0.28	-0.39	1.00	-0.47	-0.46	0.93	-0.25	-0.46	-0.39	-0.24	0.27	0.28	0.00	0.16
	Std	-0.45	0.77	0.61	-0.50	0.92	0.72	-0.47	1.00	0.78	-0.58	0.66	0.54	0.05	0.80	-0.08	-0.09	0.57	0.46
	Variety	-0.43	0.63	0.87	-0.46	0.73	0.98	-0.46	0.78	1.00	-0.52	0.50	0.87	0.09	0.69	0.04	0.03	0.41	0.10
NI	Mean	0.88	-0.11	-0.18	0.94	-0.38	-0.45	0.93	-0.58	-0.52	1.00	-0.33	-0.47	-0.16	-0.32	0.32	0.32	-0.05	0.10
	Std	-0.35	0.55	0.40	-0.31	0.68	0.51	-0.25	0.66	0.50	-0.33	1.00	0.60	0.18	0.39	-0.01	-0.01	0.31	0.43
	Variety	-0.48	0.39	0.73	-0.46	0.52	0.88	-0.46	0.54	0.87	-0.47	0.60	1.00	0.27	0.38	0.06	0.05	0.06	-0.06
NDVI	Mean	-0.60	-0.22	-0.11	-0.45	0.00	0.05	-0.39	0.05	0.09	-0.16	0.18	0.27	1.00	-0.08	0.10	0.08	-0.11	0.03
	Std	-0.18	0.78	0.62	-0.26	0.85	0.64	-0.24	0.80	0.69	-0.32	0.39	0.38	-0.08	1.00	0.36	0.36	0.54	0.26
NI/G	Mean	0.23	0.24	0.13	0.23	0.20	0.04	0.27	-0.08	0.04	0.32	-0.01	0.06	0.10	0.36	1.00	1.00	-0.05	-0.01
	Std	0.24	0.24	0.13	0.24	0.20	0.03	0.28	-0.09	0.03	0.32	-0.01	0.05	0.08	0.36	1.00	1.00	-0.07	-0.02
T	Mean	0.01	0.55	0.40	-0.01	0.54	0.38	0.00	0.57	0.41	-0.05	0.31	0.06	-0.11	0.54	-0.05	-0.07	1.00	0.66
	Std	0.05	0.40	0.10	0.11	0.48	0.07	0.16	0.46	0.10	0.10	0.43	-0.06	0.03	0.26	-0.01	-0.02	0.66	1.00

Correlation ≥ 0.8

Image-derived and physical variables: Conifers

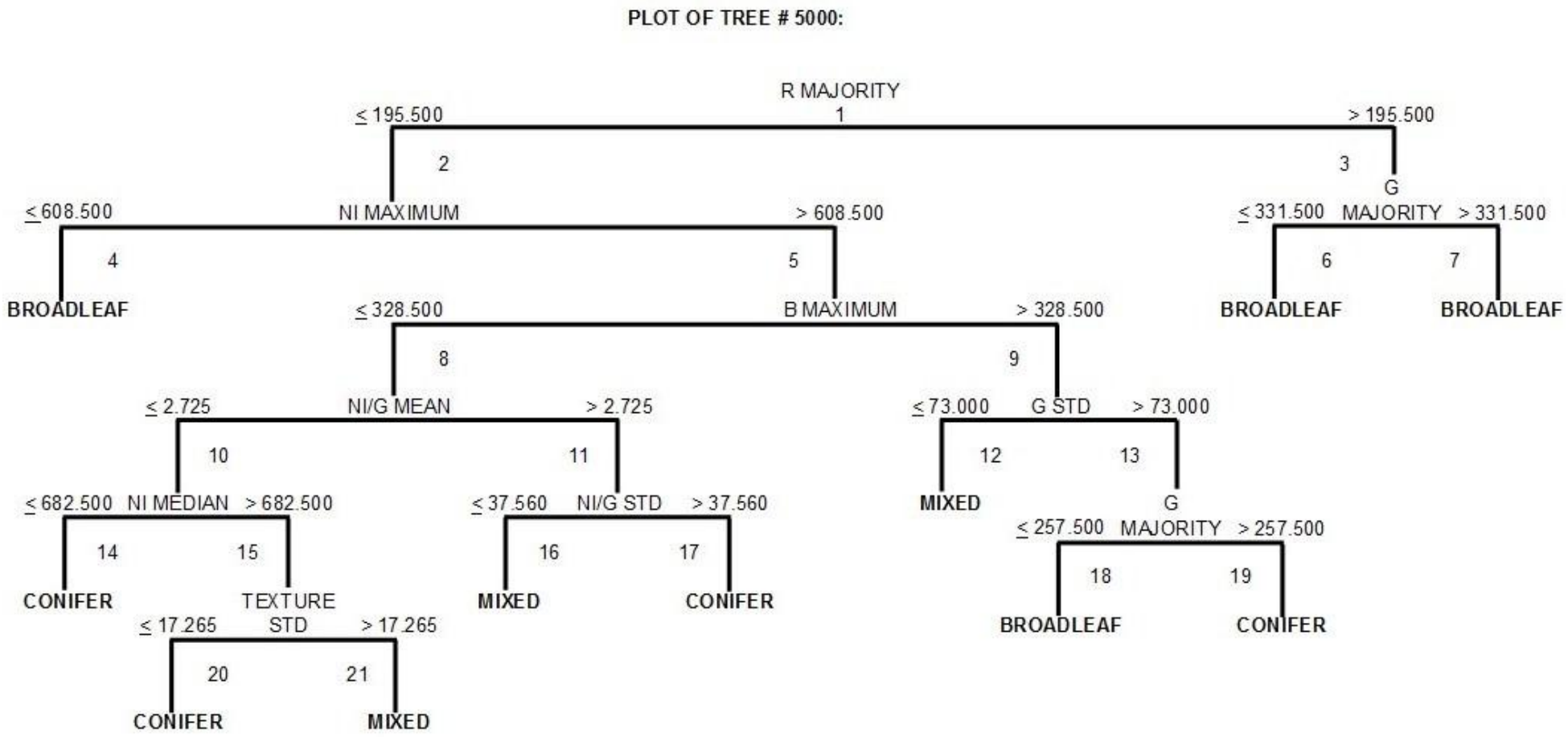
		IMAGE-DERIVED										PHYSICAL (log)		
		RED			NI		NDVI		NI/G	TEXTURE				
		Mean	Std	Variety	Mean	Std	Mean	Std	Mean	Mean	Std	Height	DBH	Density
R	Mean	1.00	-0.42	-0.56	0.66	-0.48	-0.71	-0.23	0.11	0.19	0.22	-0.58	-0.52	0.32
G	Std	-0.42	1.00	0.81	-0.29	0.79	0.34	0.74	0.12	0.31	0.36	0.48	0.28	-0.30
B	Variety	-0.56	0.81	1.00	-0.40	0.67	0.41	0.62	0.02	0.15	0.13	0.53	0.26	-0.27
NI	Mean	0.66	-0.29	-0.40	1.00	-0.32	0.05	-0.40	0.29	0.05	0.10	-0.54	-0.38	0.25
	Std	-0.48	0.79	0.67	-0.32	1.00	0.35	0.48	0.01	0.08	0.29	0.43	0.34	-0.20
NDVI	Mean	-0.71	0.34	0.41	0.05	0.35	1.00	-0.03	0.13	-0.18	-0.18	0.30	0.33	-0.20
	Std	-0.23	0.74	0.62	-0.40	0.48	-0.03	1.00	0.16	0.37	0.35	0.33	0.15	-0.20
NI/G	Mean	0.11	0.12	0.02	0.29	0.01	0.13	0.16	1.00	-0.07	-0.04	-0.12	-0.01	0.02
T	Mean	0.19	0.31	0.15	0.05	0.08	-0.18	0.37	-0.07	1.00	0.58	-0.15	-0.16	0.20
	Std	0.22	0.36	0.13	0.10	0.29	-0.18	0.35	-0.04	0.58	1.00	-0.20	-0.14	0.21
PHYSICAL	Height	-0.58	0.48	0.53	-0.54	0.43	0.30	0.33	-0.12	-0.15	-0.20	1.00	0.47	-0.50
	DBH	-0.52	0.28	0.26	-0.38	0.34	0.33	0.15	-0.01	-0.16	-0.14	0.47	1.00	-0.50
	Density	0.32	-0.30	-0.27	0.25	-0.20	-0.20	-0.20	0.02	0.20	0.21	-0.50	-0.50	1.00

Correlation of physical variables to image derived variables:
Correlation > 0.50

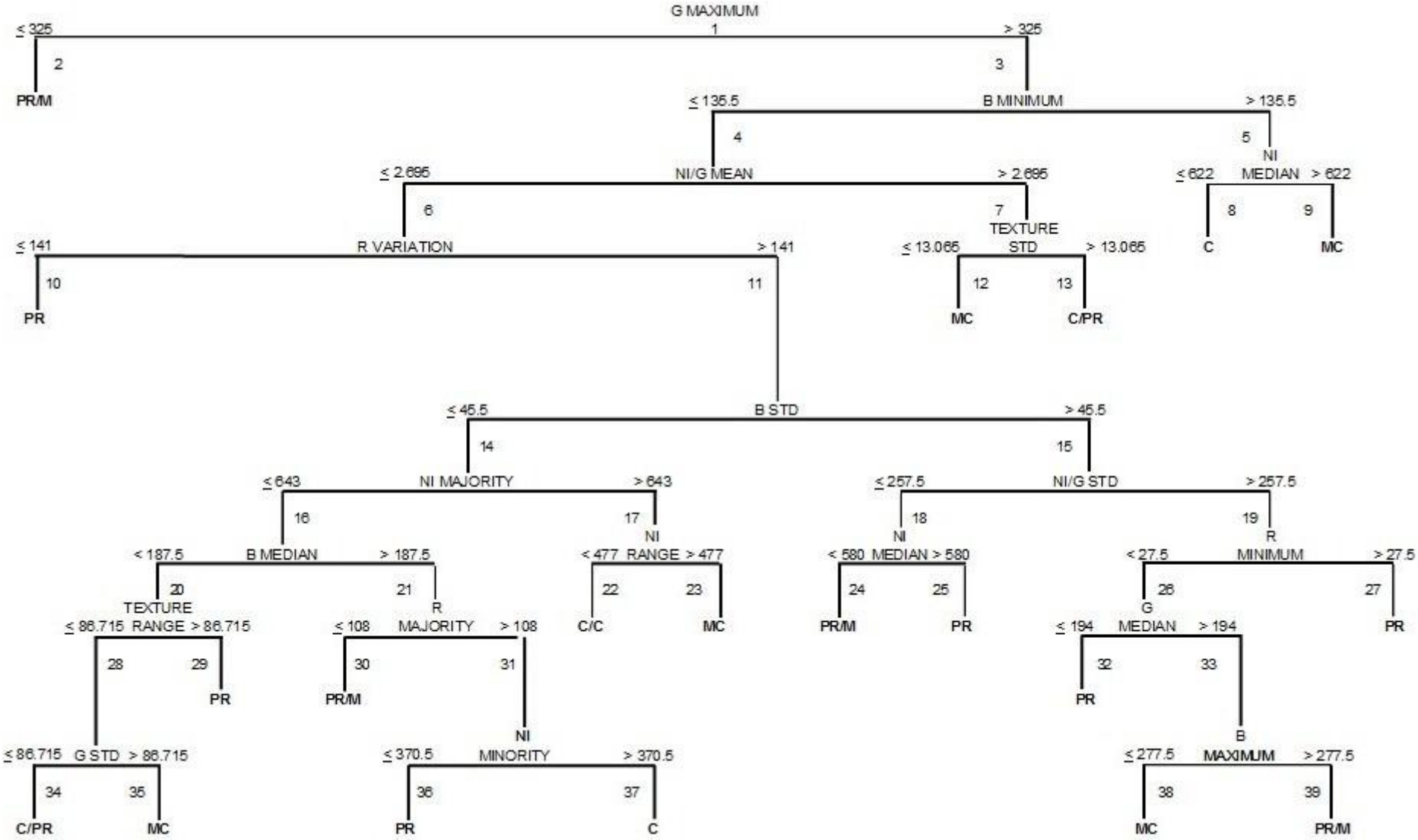
Appendix E

Species Differentiation

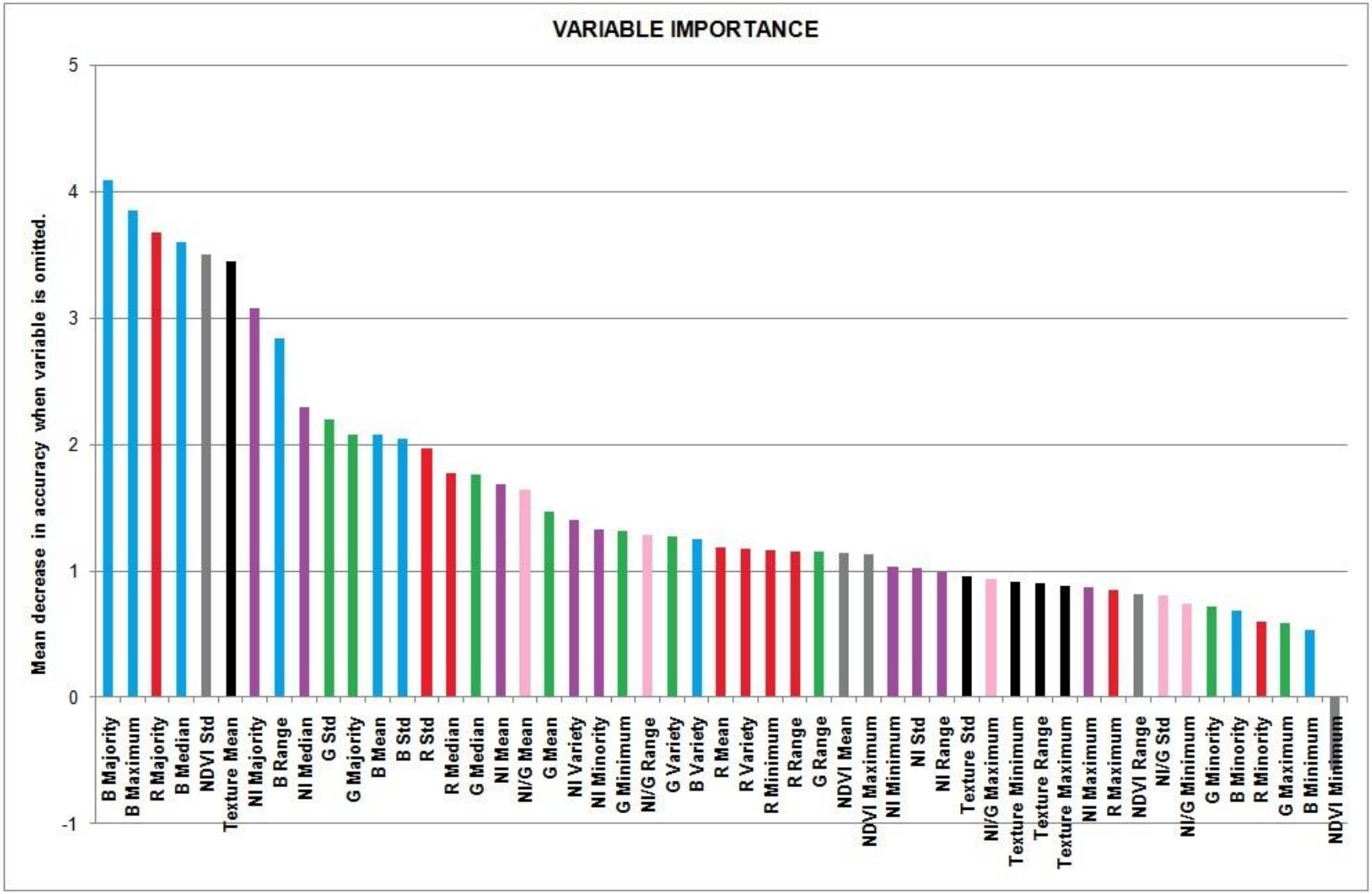
E.1 Random Forest 5000th Tree: Species Groups



E.2 Random Forest 5000th Tree: Coniferous Species



E.3 Importance of Variables for Differentiating Species Groups



E.4 Importance of Variables for Differentiating Coniferous Species