

NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

MULTIVARIATE FOREST MODELLING AND MAPPING USING QUICKBIRD
IMAGERY AND TOPOGRAPHIC DATA IN CHELSEA, QUÉBEC

By

Valerie Torontow
B.A. (Hons.), Carleton University, 2008

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements of the degree of

Master of Science

In

Geography

Department of Geography and Environmental Studies

Carleton University

Ottawa, Ontario

August 2010

© 2010, Valerie Torontow



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-71672-4
Our file *Notre référence*
ISBN: 978-0-494-71672-4

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Abstract

This research determined the capability for modelling and mapping multivariate forest complexity as an indicator of forest biodiversity using remote sensing and topographic data in the municipality of Chelsea, Québec. In 70 field plots, 37 structure and composition variables were measured. Image spectral and spatial variables were derived from Quickbird imagery, and topographic variables were derived from a DEM. Several field based structure/composition indices were developed and then modeled using multiple regression against the geospatial variables. The results were compared to an index derived directly from the geo-spatial data using Redundancy Analysis (RDA). The best models were an index derived additively from a set of ten core field variables and a 4 predictor variable RDA model. These models were then applied over the study area to obtain a map of predicted forest complexity for Chelsea, Québec, which can be used to aid in biodiversity survey planning and conservation efforts.

Acknowledgements

This research was supported by a National Sciences and Engineering Research Council Discovery grant to Dr. Doug King. Field support was provided by the Carleton University Geomatics and Landscape Ecology Research Laboratory through Canada Foundation for Innovation and other funding. Thanks to Nature Chelsea and Carolyn Callaghan for allowing the use of their data, and to the Chelsea landowners for allowing me access to their properties.

My deepest gratitude is given to my supervisor Dr. Doug King for his support and guidance. Doug, thank you for seeing my potential and encouraging me to continue my quest for knowledge. All your time and many valuable suggestions have helped to improve this thesis. Thank you to my committee member Dr. Elyn Humphreys, your support and feedback have been much appreciated. Thank you to Natalie Pressburger, I really value your help with all my applications.

I am indebted to Dr. Jon Pasher for allowing me to extend his research. Jon, your extensive Geomatics knowledge and help have been invaluable. You encouraged and challenged me along the way, and I am very grateful for that. Thank you to Laura Dingle-Robertson for your knowledge and processing help.

The fieldwork for this project would not have been completed on time without all the help I received in the field. Thanks to Chris Czerwinski for sharing your knowledge and field experience with me, as well as braving the insects and rain! Radha Gopalan, your persistence in the field was impressive. I will never forget our trudge through the poison ivy plots in hipwaders! Rebecca Barker, thanks for the help in the field, and especially for being a listening ear.

To the Torontow family (including Grandma), thank you for your love and encouragement and always reminding me I can do anything I put my mind to. Mom and Dad, I'm so lucky to have such special parents. I don't think many other daughters get the opportunity to thank their parents for helping them with fieldwork. Glen and Joel, thanks for the help with hemi photos!

Finally, I would like to express special thanks to Jordan Wynja. Jordan, thank you for supporting me when I encountered difficulties and for celebrating me when I was successful. Thank you for your confidence in me. A special thanks also for helping with the figures in this thesis.

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents	v
List of Tables	vii
List of Figures	ix
List of Appendices	xi
1. Introduction	1
1.2 Objectives	4
1.3 Thesis Structure	5
2. Background and Theory	6
2.1 Biodiversity Indicators	6
2.1.1 Species Based Indicators	6
2.1.2 Forest Structure Based Indicators	9
2.1.2.1 Forest Structure	9
2.1.2.2 Structural Indicators	10
2.2 Structural Complexity Indices	13
2.2.1 Additive Indices	14
2.2.2 Indices Based on the Average Score of Groups of Attributes	16
2.2.3 Indices Based on the Interaction of Attributes	16
2.3 Mapping Biodiversity with Remote Sensing	17
2.3.1 Direct and Indirect Remote Sensing	18
2.3.1.1 Direct Remote Sensing of Species Composition and Species Groups	18
2.3.1.2 Indirect Remote Sensing of Species Composition using Environmental Parameters	20
2.4 Mapping Forest Structure with Remote Sensing	22
2.5 Multivariate Modelling of Forest Structure: Redundancy Analysis	25
3. Methodology	30
3.1 Study Area	30
3.2 Field Plot Creation	33
3.3 Field Variable Selection	36
3.3.1 Tree Size and Distribution	38
3.3.2 Canopy Openness and LAI	39
3.3.3 Ground and Understorey Vegetation	41
3.3.4 Dead Wood	42
3.4 Satellite Imagery Acquisition and Processing	44
3.5 Image Variables	45
3.5.2 Spatial Variables	47

3.5.3 Topographic Variables	49
3.6 Defining Forest Structure and Composition Indices from the Field Data.....	50
3.6.1 Field Based Complexity Index Derived from PCA.....	53
3.6.2 Additive Field Based Complexity Index Derived Subjectively	53
3.7 Complexity Index Derived from Multivariate Relationships between Field and Geo-spatial Data: Redundancy Analysis.....	55
3.8 Model Comparison and Selection	57
3.9 Mapping Complexity.....	58
3.9.1 Model Validation.....	58
4. Results.....	59
4.1 Field Variables	59
4.2 Image Variables.....	64
4.3 Defining Forest Structure and Composition Indices from the Field Data.....	65
4.3.1 Principal Components Analysis	65
4.3.2 Additive Index	68
4.4 Redundancy Analysis Complexity Modelling	74
4.4.1 Model Selection.....	74
4.4.2 Analysis of the Selected Six Variable Model.....	77
4.4.3 Analysis of a Simpler Four Variable Model.....	85
4.5 Model Comparison and Selection	87
4.6 Map and Validation	95
5. Discussion	99
5.1 Significant Findings	99
5.1.1 Model Significance.....	100
5.1.1.1 Field and Image variables	100
5.1.1.2 PCA.....	102
5.1.1.3 Additive index.....	104
5.1.1.4 Redundancy Analysis	105
5.1.2 Applicability to the Municipality of Chelsea	108
5.2 Limitations	109
5.3 Recommendations for further research	112
6. Conclusions.....	115
7. References.....	117
Appendices.....	129

List of Tables

Table 1. Summary of the field data recorded and the field variables derived from them.	37
Table 2. Summary of extracted image variables.....	46
Table 3. Eleven field variables removed due to high correlations with other field variables.	51
Table 4. Core attributes used in additive index.....	54
Table 5. Initial eigenvalues and relative percent variance of the first twelve PCs.	66
Table 6. Factor loadings for PCs 1 - 4.	67
Table 7. Stepwise regression model variables and variance accounted for. Three indices of multivariate forest complexity as described in the text are shown in the first column.	68
Table 8. Eleven core variables identified from an initial suite of 26. Only correlations in which Pearson's $r > 0.60$ were considered.....	69
Table 9. Mean and standard deviation values for the core attributes in mixed forest and deciduous forest communities.....	70
Table 10. Quartile midpoints (12.5, 37.5, 62.5, 87.5) for each of the eleven core attributes and associated score (2.5, 5, 7.5, 10).	72
Table 11. Regression equations for the eleven attributes for scoring on a scale of 0-10..	73
Table 12. Correlation of attribute scores from the linear stretch with the attribute scores from the McElhinny rescaling method. (Highly significant ($p < 0.01$))......	73
Table 13. Stepwise regression model variables and variance accounted for. Two indices of multivariate forest complexity as described in the text are shown in the first column.	74
Table 14. Number of times an image variable type was included in one of the 1000 RDA models, displayed as a percentage.	75
Table 15. Summary statistics for the 1000 models provided through stage two bootstrapping.....	75
Table 16. Bivariate correlations between RDA1 and the individual field variables and bivariate correlations between RDA2 and the individual field variables.	80

Table 17. Biplot scores of RDA1 and RDA2. The scores show the relative contribution of each image variable towards each RDA axis.	81
Table 18. Number of times each of the predictor variables from the 6 variable model was included in the 1000 RDA models, displayed as a percentage.	81
Table 19. Field variables for plot 29.	84
Table 20. Biplot scores of RDA1 and RDA2 for the four variable model. The scores show the relative contribution of each image variable towards each RDA axis.	86
Table 21. Summary of the final seven indices created through three separate methods. .	88
Table 22. Final seven indices created through three separate methods with image variables and variance accounted for. ((-) indicates predictor variable decreased with increasing index value.)	89
Table 23. Number of models in which a particular type of image variable appeared.	90
Table 24. Total explained variance of the 26 variable field dataset by the best indices from each of the three index formulation methods.	92
Table 25. Most common field variables in the complexity indices. The + and – signs demonstrate whether the variables increase or decrease with increasing complexity. ...	101

List of Figures

Figure 1. Comparison between simple and complex forest complexity as described by the USDA forest service (from USDA Forest Service, 2006).....	10
Figure 2. Relations between A) the ordination of a set of Y variables (PCA), B) the ordination of a single y variable against the set of X variables (multiple regression), and C) the ordination of a set of Y variables against the set of X variables (RDA) (Adapted from Legendre and Legendre, 1998).	26
Figure 3. Process map showing the execution of RDA through linear regression and PCA. (Adapted from Legendre and Legendre, 1998).....	28
Figure 4. Map of the municipality of Chelsea, Québec showing expanse of Gatineau Park inside the municipality and the study site for this project in orange (ACRE, 2004).	32
Figure 5. Study area and plot locations displayed over the Quickbird multispectral imagery. Plot location colour corresponds to the qualitative complexity score given in the field.	35
Figure 6. Example plot set up, showing centre and corner stakes at cardinal points.....	36
Figure 7. Example of setup for vegetation sampling, where vegetation height and species was sampled every 50 cm at the orange markings on the rope.....	42
Figure 8. A) Line intercept method used by Pasher and King (2010) B) Adapted line intercept method for sampling coarse woody debris in a plot. (Dominant wind direction: WNW.).....	43
Figure 9. A) Original panchromatic imagery displayed with part of the municipal road network (red) compared to B) improved georeferenced panchromatic imagery.	45
Figure 10. Process map showing the creation, selection, mapping and validation of three separate indices, including two field based methods (PCA and an additive method) and one multivariate statistical method (RDA).	52
Figure 11. Complexity conditions typical of Chelsea, Québec. Complexity of the plots increase from left to right.....	60
Figure 12. Structural conditions typical of an area of relatively low structural complexity (A - mixed (plot 35) and B - deciduous (plot 24))......	61
Figure 13. Structural conditions typical of an area of relatively high structural complexity	

(A – mixed (plot 49) and B – deciduous (plot 48)) within the study area.	62
Figure 14. Example hemispherical photographs taken in a low complexity plot (L) as well as a high complexity plot (H) at both 1 m and 6 m heights.	63
Figure 15. A) Map of shadow and crown, compared to B) panchromatic imagery from which they were derived.	65
Figure 16. The first two principal components (PC1-2) explained 44% of the total variation between sites, mapped by the 11 core attributes. The vectors represent the relative importance of each variable in explaining variation between data points as well as the correlation between variables. (Refer to Table 8 for attribute number.)	71
Figure 17. The triplot for the first and second RDA axes, showing a gradient of increasing forest structure – composition complexity from left to right for RDA1 and a gradient of increasing forest complexity from top to bottom for RDA2.	82
Figure 18. Panchromatic (A), multispectral (B) and NIR imagery (C) used to examine the forest patch of plot 29 (at arrow).	84
Figure 19. A) Comparison of the qualitative complexity score and the additive index score and B) Comparison of the qualitative complexity score and the predicted additive index score. Plots with a low qualitative score received the lowest average additive score, and plots with a high qualitative score had the highest average additive score.	94
Figure 20. Plot frequency by rounded additive index score grouped by the subjective field classification of high, medium and low complexity.	94
Figure 21. Predicted forest complexity mapped across the entire study area using the 10 variable additive index mapped with 5 predictor variables.	96
Figure 22. Predicted forest complexity mapped across the entire study area using the 4 predictor variable RDA index.	97
Figure 23. Map comparison with additive index on the left and RDA index on the right. A) Similar areas of high complexity, and B) Similar areas of low complexity.	98

List of Appendices

Appendix 1: The Formulations for the Five Textures used in this Research (Haralick, 1973).....	129
Appendix 2: Summary Statistics for Field Variables.....	130
Appendix 3: Raw Field Plot Data.....	131
Appendix 4: Comparison between Complexity Score Allocated Qualitatively in the Field and Complexity Score Allocated through RDA.....	139

1. Introduction

Sustainable forest conservation requires continual monitoring, measurement and assessment of forest habitats (Margules & Pressey, 2000). It is not surprising that a key part of forest management involves monitoring and conservation of biodiversity. As defined in the 1992 Convention on Biological Diversity (CBD, 1992), the term biodiversity (biological diversity) represents “the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems.”

Humans require a wide range of natural resources for food, construction materials, medicine and energy, all of which are provided by biodiversity, in order to survive flourish (Leemans, 1999). Biodiversity is also very valuable because it ensures essential ecosystem functions. For example, forests and vegetation have the ability to retain moisture, which helps to prevent erosion. When a hill is clear-cut, there is a higher chance of a landslide occurring, as it lacks roots to stabilize the soil. Agricultural fields and forests provide animals with food and materials for shelter. Ecosystem processes such as soil formation, water recycling and climate control all function through an interaction of biotic and abiotic factors, which benefit humans (Environmental Literacy Council, 2007). Biodiversity can also act as an environmental indicator, alerting researchers when these ecosystem functions may fail (Pykh, 2002).

Biodiversity conservation and management have become a focus of many national governments, local communities and scientific organizations (Redford and Richter, 1999). In 1992, the CBD was formed with the focus to promote the conservation

of biodiversity and the sustainable use of its components (CBD, 1992). The CBD set a goal to cut the rate of loss of biodiversity by 50 percent before 2010. It follows that the causes of biodiversity decline should be identified and evaluated so that useful conservation and monitoring policies can be developed (Spangenberg, 2006). Capturing a representative sample of biodiversity is very difficult, as biodiversity is comprised of many factors. Biodiversity assessment and monitoring are very time consuming if performed manually in the field, even for simple metrics such as species richness, so indicators have been developed as surrogates for biodiversity. They include species or structural indicators that can be used to represent the total biodiversity in an ecosystem and provide repeatable methods for conservation monitoring (Noss, 1990).

In general, a more heterogeneous habitat is able to support a greater number of species (MacArthur and MacArthur, 1961; Newsome and Catling, 1979; Williams et al., 2002; Lassau et al., 2005). Based on this, it should be possible to represent and map the complexity of a forest as an indicator of the biodiversity it can support (Lindenmayer et al., 2000). It is crucial to test this in a forest management setting, as Spies (1998) noted that the diversity of wildlife decreased as the range of forest structures decreased. This was most often seen in controlled forest management for wood products.

By incorporating remotely sensed imagery, more comprehensive and continuous coverage of larger areas may be obtained. In recent years, remote sensing has been evaluated in assessing various aspects of biodiversity in ecosystems. It provides a way to assess biodiversity across different spatial scales (e.g. regional, national, global) and a way to assess change in ecosystem patterns over time (Duelli, 1997).

The focus of this research is directed towards creating a multivariate forest

structure and composition index that can be mapped as an indicator of biodiversity in Chelsea, Québec. This area is a rural municipality that has seen partial development for residential housing and other low intensity land uses. This work is being performed by partnering with Nature Chelsea, an organization with the goal of identifying areas of biodiversity value, and disseminating the information to local landowners. This research was designed to map variations in forest structure and composition to allow identification of potentially high and low biodiversity areas for conservation management actions and municipal planning.

Several methods for derivation and mapping of a multivariate forest structural index have been previously presented (McElhinny et al., 2006; Wunderle et al., 2007; Pasher and King, 2010). They can be categorized into two classes: 1) creating a field based index for modelling and mapping using geospatial data (i.e. the field index is first created independent of the geospatial data); 2) creating an index by direct multivariate statistical modelling of relations between field data and geospatial data (i.e., the field index is defined by relations with the geospatial data) (Pasher and King, 2010). An objective of this research project was to compare these two approaches.

This research follows from that of Pasher and King (2010), who developed a geospatial data-based structural complexity index (approach 2 above) using Redundancy Analysis (RDA). Whereas Pasher and King used 20 cm 3-band multispectral airborne data, based on a scaling analysis in that study, this research used high resolution satellite data (0.6 m panchromatic and 2.4 m multispectral imagery) to cover a larger area. It also integrated compositional diversity in the index, whereas previous studies using geospatial modelling have focused solely on structural indices.

1.2 Objectives

The goal of this research was to model and map multivariate forest structure and composition using high resolution satellite imagery and topographic data as an indicator of biodiversity in Chelsea, Québec. The objectives were:

- 1) Develop indices of forest structure and composition from field measurements using methods previously described in the ecology literature. Through multiple regression, determine how well these indices can be modeled using spectral and spatial information in high resolution satellite imagery as well as topographic information.
- 2) Using a multivariate statistical technique, RDA, derive a model expressing the association between the field and geospatial data.
- 3) Compare the models derived from the regression of pre-determined field-based indices (Objective 1) with the RDA models (Objective 2) to determine which approach is best.
- 4) Produce a map of the best forest model for the municipality of Chelsea, Québec.

Note: This research was designed to include assessment of the relation between mapped (predicted) forest complexity and biodiversity, as biodiversity data were collected in June, 2009 and 2010 by Nature Chelsea. However, the 2009 data were too localized and not geo-referenced well enough and the 2010 data were not compiled before thesis submission. Consequently, the context of this research, forest complexity indices as biodiversity and habitat indicators, was maintained in the thesis but not explicitly tested.

1.3 Thesis Structure

This thesis is structured into seven chapters. Chapter one introduces biodiversity, the use of indicators of biodiversity, and the use of a geomatics-based method for mapping and modelling forest structure complexity as an indicator of biodiversity. It also presents the research objectives, and the general organization of the thesis. Chapter two provides background information for this research obtained from the literature including: an introduction to species and structural indicators, the creation of structural complexity indices, biodiversity and forest structure mapping using remote sensing, and an introduction to RDA as a tool to map forest complexity. Chapter three provides a description of the study site, plot site selection, data and image acquisition, the methods used to process the imagery, including georeferencing, and explains the methods used to create the models, their application, and validation. Chapter four presents the results of the structural complexity modelling and mapping. Chapter five discusses the major findings of the study as well as the limitations and future considerations. Chapter six presents the research conclusions.

2. Background and Theory

2.1 Biodiversity Indicators

The term biodiversity can describe many factors that indicate the state and health of a complex biological system. Its complexity can make it difficult to measure as a whole. Often, the task of collecting information directly in the field is not feasible, as large inventories of biodiversity require extensive resources and time. McElhinny et al. (2005) state that if forest managers and planners need to maintain and conserve biodiversity, a clear and practical method of measuring biodiversity must be first defined. The clear alternative to performing a complete biodiversity survey is to use indicators (Noss, 1990; Dufrene and Legendre, 1997; Smith et al., 2008), which are measurable surrogates of diversity. These biodiversity indicators can help locate areas that represent high or low biodiversity either within forests or between forests for conservation or remedial actions. It is suggested that the ideal indicator should: 1) be sensitive enough to indicate change, 2) be applicable over a large geographical area, 3) provide continuous information, 4) be independent of sample size, and 5) be simple and cost effective to measure (Cook, 1976). Since any one indicator may not fulfill all of the above properties, Noss (1990) suggests a set of indicators be used.

2.1.1 Species Based Indicators

The most commonly used indicators are species based. Generally, indicator species are selected to represent other forms of biodiversity or reflect changes in ecosystem patterns or processes (Lindenmayer et al., 2000). Lawton and Gaston (2001) were able to narrow this down to three particular uses: 1) to signal the biotic or abiotic condition of an ecosystem, 2) to provide evidence of environmental change, and 3) to

indicate the diversity of other neighbouring species. Often, these indicator species are used when it is too expensive or difficult to measure other species that are of interest (Landres et al., 1988). The term indicator species is loosely used and can represent many different things, including ecological indicators, keystone species, umbrella species, flagship species and vulnerable species. Ecological indicators are species whose presence indicates the presence of another set of species generally with the same type of habitat requirements. Keystone species are those which play a large role in maintaining the diversity of plants and animals in a particular ecosystem. They are very important, but as Payton et al. (2002) note, they are sometimes very difficult to identify, and it is unknown if all ecosystems have one. Umbrella species are those which require a very large geographical area. Simberloff (1997) describes that by conserving them, one will automatically have a significant chance of protecting all the other species which fall within that area. Flagship species are generally large charismatic species used in conservation campaigns (NEASPEC, 2007). Conservation of these species can prove to be very expensive. Finally, vulnerable species are those whose populations are threatened due to a variety of factors including low population, low reproductive success or loss of habitat through anthropogenic change (Noss, 1990).

These indicator species help represent the abundance and type of species present, also referred to as composition. Generally, it is the ecological indicators (aquatic macroinvertebrates, fish, birds, and vascular plants), which are most commonly used. Niemi and McDonald (2004) suggest they are used for several reasons, including ease of identification and measurement, interest to the public, and low associated cost. Bird species have been identified as a particularly suitable indicator for forest monitoring and

evaluation in terms of cost (Gardner et al., 2008). They have also been found to respond to habitat structure (Chettri et al., 2005). One particular bird species, the pileated woodpecker (*Dryocopus pileatus*), is often used as an indicator since it can be reliably monitored, and its foraging and nesting activities can help increase the diversity of other avian species (Drever et al., 2008). Relatively easy to spot in the forest, these woodpeckers create cavities in trees for nesting, which are later utilized by many other forest species. They also peel off tree bark while foraging, which exposes food sources for other birds.

Several insects have been used as indicator species. The dragonfly (*Libellulidae spp.*) is gaining new interest as a flagship species. Sahlen and Ekestubbe (2001) explain that it is a good indicator because it attracts the public's attention, due to its likeability, and its presence in wetlands is said to indicate a healthy environment. Additionally, dragonflies are good indicators as they are easy to identify and are taxonomically well understood (compared to other insect species) (Clark and Samways, 1996). Beetles are another common insect used as a bioindicator. For example, Pearson and Cassola (1992) demonstrated that tiger beetle (*Cicindela spp.*) diversity is a good indicator as it can be used to predict bird and butterfly diversity over large areas.

While indicators can be very useful tools, Dudley et al. (2005) warn that they may only be providing information about the presence of available habitat, and very little about its quality or potential for sustaining species. They also warn that a poor choice of indicator species can lead to skewed interpretations of management practices, pointing out that an adaptable species may mask impacts to more sensitive species and areas. Caution must therefore be taken when choosing and interpreting indicator species. In

addition, using indicator species on their own may prove unsuccessful, as strong relations between such indicators and biodiversity are sometimes difficult to establish (Van den Meersschaut, D. and Vandekerckhove, K. 1998). Instead of just monitoring species, Franklin et al. (1981) suggest the importance of monitoring both forest structure (spatial arrangement) and function (ecological processes).

2.1.2 Forest Structure Based Indicators

2.1.2.1 Forest Structure

Forest structure consists of pattern in three dimensions, described in both the horizontal and vertical directions. The horizontal aspect describes patterns of canopy openness, tree size and spacing, and vegetation species in the sense that different species have different structures. The vertical aspect describes the number, spatial density, and arrangement of vegetation layers between the forest floor and the top of the canopy. In ecology literature, forest stand structure is described in two parts: stand structure attributes and stand structure complexity (McElhinny et al., 2005). The stand structure attributes are single attributes which measure variables of abundance (e.g. number of large trees (Acker et al., 1998)), relative abundance (e.g. tree size diversity (Wikstrom and Eriksson, 2000)), richness (e.g. understorey richness (Sullivan et al., 2001)), size variation (e.g. standard deviation of tree height (Zenner, 2000)) and spatial variation (e.g. standard deviation of the number of trees per hectare (Ferreira and Prance 1999)). Stand structural complexity, then, is the multivariate representation of a set of structural attributes found in a given area.

While stand structural complexity can be defined in many ways, the USDA Forest

Service (2006) has provided a succinct definition of structural complexity from a forest management point of view. Complex stands have: 1) more than one tree species, 2) more than one age class, 3) a wide range of tree sizes, 4) abundant and sometimes diverse understory and ground layer plant communities, and 5) abundant numbers of standing dead trees (snags) and dead trees on the ground (Figure 1). They also state that complexity can be expressed as spatial variation or heterogeneity in the above characteristics.

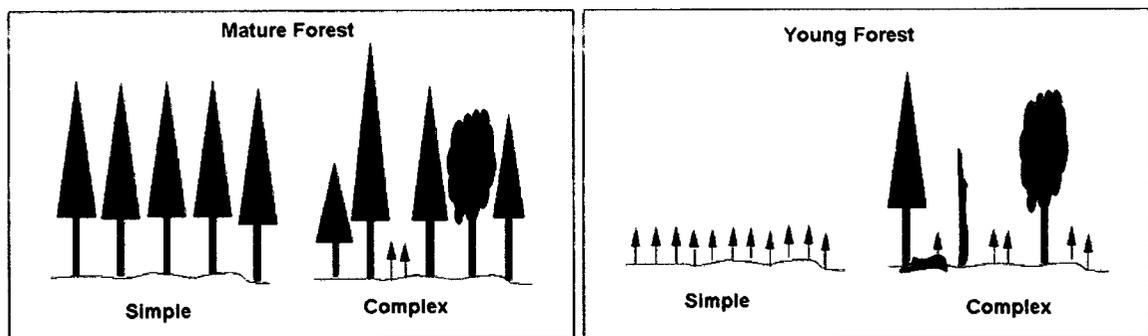


Figure 1. Comparison between simple and complex forest complexity as described by the USDA forest service (from USDA Forest Service, 2006).

This definition of stand structural complexity focuses on the structure and composition of the forest. Composition refers to the type and amount of variation of ecosystem attributes (McElhinny et al., 2005). While these two important components will be the focus of this research, Noss (1990) points out that ecosystem processes such as nutrient cycling and decomposition of coarse woody debris (CWD) are also key factors for maintaining biodiversity.

2.1.2.2 Structural Indicators

The physical structure of a habitat is primarily determined by the plant

communities that are present (Lawton, 1983), i.e. plant species have a significant influence on species distributions and interactions within that environment. In the literature, it is clearly evident that these structural indicators are becoming more commonly used for measuring potential biodiversity (Watson et al., 2001; Ziegler, 2000; Tanabe et al., 2001; Wikstrom and Eriksson, 2000; Acker et al., 1998; Uuttera et al., 2000; Parkes et al., 2003; Bachofen and Zingg, 2001). McElhinny et al. (2005) point out that to be a good indicator of biodiversity, a measure of structural complexity must be based on a suitable group of structural attributes. Some of the important stand elements include: 1) foliage, 2) tree diameter, 3) tree height, 4) stand biomass, 5) understorey vegetation, 6) canopy cover, and 7) dead wood. They state that a more structurally complex forest would be a forest with many overhead gaps of varying sizes and shapes. Gaps would allow sunlight to penetrate to the forest floor and encourage understorey growth of grasses and saplings, leading to a very heterogeneous and biodiverse environment. The presence of dead standing trees (snags) is also an attribute that is associated with increased biodiversity due to the unique ability of such structures to provide niches for a variety of species (Dewalt et al., 2003). Conversely, a non-complex (simple) forest is one where there are relatively few gaps, sunlight is not able to penetrate the forest canopy, and there is little understorey vegetation. There can be exceptions to these definitions. For example, a recently replanted clearcut will receive lots of light but have a low complexity. Also, in some tropical forests, due to climate and site conditions, complexity can be extremely high even with very few canopy gaps and little light penetration to lower levels of the forest.

Smith et al. (2008) developed compositional, structural and functional indicators

of biodiversity for Sitka spruce (*Picea sitchensis*) and ash (*Fraxinus spp.*) plantations in Ireland. They determined that the best overall structural biodiversity indicator was stand stage, which was defined using multiple forest structure variables. Deadwood and distance to forest edge were important structural indicators, though more importantly, canopy cover was found to influence other variables such as cover of lower vegetation layers. Grove (2002) successfully used structural indicators as a surrogate measure for saproxylic insects in Daintree, Australia. These insects proved to be difficult to sample and use as an indicator in their own right, so their known association with deadwood was investigated. It was determined that volume of coarse woody debris was the strongest positive correlate of species richness, while the basal area of larger-diameter trees was found to be a more robust indicator of abundance, incidence, and assemblage composition. The authors suggested the basal area of larger-diameter trees as the final structural indicator for the saproxylic insects due to its greater ease of measurement.

The reasoning behind structural indicators is that a forest, which has various components of structural complexity, is one that should have a variety of resources and therefore be able to support more species. This idea is also known as the habitat heterogeneity hypothesis, which attempts to describe the relation between forest structure and biodiversity at different spatial extents. For example, MacArthur and MacArthur (1961) conducted a study to determine what environmental factors controlled bird species diversity. After testing latitude, plant species diversity, and foliage height diversity, it was determined that the latter accounted for most of the variability in bird species diversity. The heterogeneity hypothesis states that an increase in habitat heterogeneity leads to an increase in species diversity in a landscape. A more heterogeneous landscape

is one with a higher diversity of cover types and a higher complexity in the spatial patterns of those covers types.

In a literature review, Tews et al. (2004) examined over 85 studies which dealt with the relations between habitat heterogeneity and animal species diversity to determine if there were uncertainties or biases in the research. While most studies provided support for this hypothesis, Tews et al. (2004) point out some significant limitations. Many of these studies had been performed in anthropogenically influenced landscapes and had studied vertebrates. As well, the effects of habitat heterogeneity are sometimes very species specific. They note that spatial scale may also influence the effect of habitat heterogeneity from one species to another, though there is a lack of research in this area.

2.2 Structural Complexity Indices

During the past few decades, many indices have been developed for quantifying spatial forest structure (Pommerening, 2002). A structural complexity index is a mathematical formulation in which more than one forest structure attribute has been summarized into a single value or attribute (McElhinny et al. 2006). These indices have been developed in anticipation that they could be used as potential indicators of biodiversity as they are able to group many habitat and structural factors into a single value (Lahde et al., 1999). The general methodology includes selecting the number and type of attributes to be used in the index, selecting the method to mathematically combine the attributes, and deciding whether it is appropriate to weight each attribute by its perceived or expected contribution.

Several methods have been used to create structural indices. These include, but are not limited to: 1) additive indices based on the cumulative score of attributes, 2) indices based on the average score of groups of attributes, and 3) indices based on the interaction of attributes. There is generally little consensus in the literature on which type of index should be used and the appropriate number of attributes, nor does there appear to be a preference of index type. Commonly, indices are created to achieve specific research goals and are very dependent on available data (Watson et al., 2001). The selection of particular attributes and their associated weights are clearly undefined, and generally based on expert opinion (Oliver, 2002; Parkes et al., 2003).

2.2.1 Additive Indices

Additive indices, the most straightforward method of indexing data, involve combining attributes so that a cumulative index is obtained. Here, a subset of attributes are selected, each scored on a point system, and summed together to obtain a total score, as the output index value. This makes it very straightforward to represent the contribution of each attribute and the index is simple to calculate. This method assumes that the attributes selected are independent of each other, and if weights are not used, that they equally contribute to the index and its representation as a habitat or biodiversity indicator. It has been noted by McCarthy et al. (2004) that important stand characteristics may be masked using this method. For example, when several different attributes are combined, a final score is computed. Two plots, with completely different attributes could end up with a similar score. So while they fall within the same structure category, individual differences in plot characteristics are lost.

Newsome and Catling (1979) presented one of the first additive forest structure indices. They included five attributes: soil moisture, and cover of ground vegetation, shrubs, canopy, and debris (litter and coarse woody debris), which were scored visually on a scale of zero to three. Scores were then added together to create their habitat complexity index. This extended work by Barnett et al. (1978) by including canopy cover and soil moisture in the study. It demonstrated that a structural complexity index could be used as an indicator of biodiversity, as it was later correlated with the abundance of ground dwelling mammals by Catling and Burt (1995) and Catling et al. (2000) and to bird species richness by Freudenbeger (1999).

Another additive complexity index is presented by McElhinny et al. (2006) in order to discriminate between different forest types. It used a three-stage analysis to identify a core set of structural/composition attributes from a suite of field variables. Univariate relations were used to identify core attributes based on four criteria: 1) data should be normally distributed, 2) attributes should distinguish between the different vegetation types, 3) attributes should function as a surrogate for other attributes ($r > 0.60$) so not all attributes need to be included, and 4) attributes should be simple to measure. Using these criteria, they identified thirteen core attributes from their suite of original variables, which described stand height, tree diameter at breast height (dbh), canopy cover, understorey cover, ground cover, dead trees, stumps, CWD, and litter. The variables were rescaled to 0-10, by fitting linear regressions to the midquartile points of the attribute distributions, to provide an objective starting point for the index, and then constrained between 0-10. Variables were then added together for each plot to obtain an index, or complexity score out of a total of 130 points. In this thesis, this method was

adapted and the resulting index modelled against geospatial data as described in Chapter 3.6.2.

2.2.2 Indices Based on the Average Score of Groups of Attributes

Rather than adding up attributes to obtain an index value, the average score of groups of attributes is sometimes taken as an index value. This method is demonstrated clearly by the study performed by Koop et al. (1994), where attributes were separated into three groups, which characterized different parts of the ecosystem: 1) forest overstorey, which described tree height, size and canopy layers, 2) light transmission, which described the richness of light demanding species as well as the number of pioneer species, and 3) atmospheric moisture, which described the presence of species that signified high humidity. To create an index for rapid ecological assessment of the Sumatran rainforest, they summed the scores (D) for each group, and compared them to values in control plots (R) in unlogged forests. A relative score (S) was then calculated for each group $S = (D/R) * 100$, and then the three group scores were added together to get a total score.

2.2.3 Indices Based on the Interaction of Attributes

These indices involve more complex, non-linear mathematical calculations. The simplest method is that of multiplication, where attributes are multiplied to achieve a final score. Burgman et al. (2001) have pointed out that this method tends to lead to undesirable results, as it assumes that all variables must be present. If a forest attribute has a score of zero, then the index value becomes zero.

Another common approach for defining structure indices involves the statistical

combination of attributes. One method of doing this involves dimension reduction using principal components analysis (PCA). August (1983) used the first principal component (PC1) of measured field variables (tree size, number of trees, foliage height diversity, and canopy height and density), as a weighted linear index of those variables. This component discriminated between complex and simple habitats based on foliage density and vegetation strata. Another study, performed by Van Pelt and Nadkarni (2004), also used the first PC as a stand indicator. It accounted for 68.5% of the variance of their thirteen variable dataset and showed a gradient of forest complexity based primarily on stand age.

2.3 Mapping Biodiversity with Remote Sensing

Instead of using field based indicators of biodiversity, as described above, some researchers have been using remote sensing data and tools to assess various aspects of biodiversity in ecosystems. Remote sensing provides a way to assess biodiversity across several different spatial scales by helping to determine the extent to which an area contributes to the regional, national or global biodiversity, and also provides a way to assess change in ecosystem pattern over time (Duelli, 1997). These measures are both very important for conservation because they show land cover changes at multiple scales, and they allow researchers to monitor the rate of habitat loss. Quantitative estimates of biodiversity loss can then be derived from the measurements of habitat loss through the use of the species-area relation (Turner et al., 2003).

Many different types of remote sensing data are available. To date, most biodiversity analysis has been conducted using photographs and digital optical imagery,

obtained from satellites and airborne platforms. Recent technological advances have also allowed researchers to incorporate active sensors, such as synthetic aperture radar (SAR) and light detection and ranging (Lidar) data. With the use of remote sensing, phenomena ranging from an ecosystem scale to a local scale can be mapped. Innes and Koch (1998) state that forest mapping at the ecosystem scale can be done efficiently by mapping patches and their arrangement. At an intermediate scale, remote sensing allows for the study of the nature of edges and evaluation of the presence of corridors. At the stand scale, structural attributes of forests can be determined.

2.3.1 Direct and Indirect Remote Sensing

Assessing the landscape using remote sensing data assumes that there are certain key environmental parameters that contain remotely detectable biophysical properties, which drive the distribution and abundance of species across landscapes (Turner et al., 2003). These, in turn, influence how the species occupy habitats. Combining imagery, datasets and ecological models has allowed for greater insight into biodiversity. A combination of these tools provides the potential to offer information about why organisms are found where they are and can be useful in determining locations for future conservation efforts. Previous remote sensing biodiversity studies can be divided generally into two types based on resolution: 1) direct mapping using high resolution imagery, and 2) indirect mapping using coarse resolution imagery.

2.3.1.1 Direct Remote Sensing of Species Composition and Species Groups

Very high spatial and spectral resolution imagery has been used to directly map individual plant species, species groups or ecological communities. High resolution

satellite imagery can be used to remotely detect specific large organisms and communities (Turner et al., 2003), particularly with hyperspectral sensors if spectral signatures exist that are characteristic of specific species or communities.

Airborne multispectral and hyperspectral sensing has been available for several decades, with pixel sizes that range from less than 1 m to approximately 10 m. Recently, numerous satellite systems have been launched that provide very high spatial resolutions. These systems have the ability to detect objects that were only previously attainable from aerial photographs. These resolutions provide the ability to directly identify some larger species, such as those of individual trees, or groups of similar species.

Levin et al. (2007) examined the relation between plant richness, rarity of species, and vegetation indices for forests in Mount Hermon, Israel, using Quickbird imagery. They utilized the Normalized Difference Vegetation Index ($\{NDVI = (NIR - R) / (NIR + R)\}$, where NIR and R are the near infrared and red band reflectance, respectively) because it was found to be strongly correlated to primary production on both local and global scales. All images showed a strong positive correlation between NDVI and both plant species richness and percentage of tree cover. Spatial heterogeneity in NDVI was examined within the defined quadrats and it was found that plant richness was significantly correlated with the standard deviation of NDVI values, this metric being a type of image texture. From this, they concluded that satellite-derived indices could be used to determine richness patterns of landscapes.

Waser et al. (2007) performed a study in the northern Pre-Alps of central Switzerland where they developed regression-based models to predict lichen species richness from Quickbird data. The reflectance of lichen-associated vegetation extracted

from the imagery produced models with predictive power ranging from 0.5 to 0.8 for lichens on soil to lichens on trees, respectively. They warned that linear regression models can be used to predict lichen diversity, but results strongly depend on the sampling design. Remote sensing in this case did not eliminate the need for field studies, but helped determine locations to target for field study.

Hyperspectral data were used by Kokaly et al. (2002) to detect and map invasive species. Using the Hyperion sensor, with 30 m pixels, infestations of leafy spurge (*Euphorbia esula*) were mapped with accuracies of up to 80% for infestations larger than 500 m² in pixels with more than 35% leafy spurge.

2.3.1.2 Indirect Remote Sensing of Species Composition using Environmental Parameters

Coarse resolution data have been used to indirectly map biodiversity using environmental parameters or land cover classes. This approach is possible because, generally, species are restricted to specific habitats (Turner et al., 2003), such as grassland or forest areas that can be easily detected and mapped with remotely sensed data. Species habitat requirements combined with landcover maps derived from coarse satellite imagery have the ability to predict species ranges and patterns of species richness.

One environmental variable, which can be useful for predicting spatial patterns of biodiversity, is primary productivity. The relation between species richness and primary productivity has been explored at numerous scales, though it is difficult to determine their exact relation since it depends on the species and scale being examined (Aava, 2001).

Regardless of the interaction between the two, it is assumed that the available energy in the environment affects the composition of species found (Turner et al., 2003).

Many remote sensing approaches have been used for estimating primary productivity. Multispectral satellite imagery is normally used at spatial resolutions between 4 m and 8 km. Primary productivity estimates are derived from vegetation indices such as NDVI, or field measures such as net primary productivity (NPP) (Turner et al., 2003). NPP is a measurement of plant growth obtained by calculating the quantity of carbon absorbed and stored by vegetation (Waide et al., 1999), sometimes measured by clipping and weighing the vegetation (Ohtsukah et al., 2005).

Bailey et al. (2004) examined, on multiple scales, the relation between maximum productivity or NDVI, heterogeneity of productivity and species richness of birds and butterflies in the Great Basin of western North America. They found positive linear relations between maximum NDVI and the number of functional guilds of birds as well as species richness of neotropical migrant birds. It was also determined that the number of functional guilds of birds and species richness of resident birds was negatively associated with heterogeneity of NDVI. Maximum NDVI was found to be associated with species richness of all butterflies, but no relations were found for NDVI heterogeneity. They therefore concluded that NDVI has the ability to highlight areas of high biodiversity, while heterogeneity of productivity was less valuable.

Environmental variables, which can be useful for predicting spatial patterns of biodiversity, other than primary productivity, can be captured using remote sensors. The environmental variables can include climatic information, which act as limiting factors for species, which can in turn be used to help predict range and location of species

(Turner et al., 2003).

NDVI derived from Advanced Very High Resolution Radiometer (AVHRR) data has been established as a good metric related to interannual climate variability and regional drought conditions. Oindo (2002) hypothesized that in locations with favourable climate and site conditions for flora and fauna, where those environmental conditions remain stable and predictable, the accumulation of high species richness would occur over time. He then studied the relation between large herbivore species richness and AVHRR–NDVI derived climatic-variability indices such as interannual mean NDVI, and the coefficient of variation of $NDVI_{cv}$, at a regional scale in Kenya. Regions with low $NDVI_{cv}$ and high interannual mean NDVI displayed ecoclimatic stability. Through statistical analysis, it was determined that species richness increased with a high interannual mean NDVI and decreased with increasing $NDVI_{cv}$. This, in turn, showed that areas with ecoclimatic stability are those that can contain the maximum number of species.

2.4 Mapping Forest Structure with Remote Sensing

The concept of habitat heterogeneity suggests that if we can detect and map more heterogeneous areas, we can assume that these areas also have higher biodiversity. This means that instead of counting species, one could potentially use remotely sensed data to map cover types, or as in this research, forest structure, to determine the relative biodiversity in the area. Mapping and monitoring forest structure is an essential part in maintaining healthy forests and biodiversity.

In the past, forest structure, was mapped through the use of field studies and

digitized air photos, both of which are very time consuming. Recently developed methods using remote sensing are more cost and time effective. It provides a way to assess biodiversity across spatial scales and to assess change in ecosystem patterns over time. Remotely sensed data can completely cover a large area compared to field methods, which can only sample an area to the extent of available resources and access. While it is agreed upon that remote sensing is an efficient tool used for modelling, detecting and mapping biodiversity indicators, researchers still question the ability of remote sensing to accurately detect individual species, species patterns and parameters which are required to predict species richness and structural levels of an ecosystem (Turner et al. 2003).

Many studies, performed at broad scales (i.e. over large extents using coarse resolution data), use land cover maps created from remotely sensed data to measure composition and configuration attributes such as connectivity or fragmentation of the landscape (Jha et al., 2005). As for vegetation composition and productivity mapping described above, structure studies have used some of the same variables such as NDVI (Waide et al., 1999), rainfall and environmental data (Johnson et al., 1998), and structural indicators. There has been significant success modelling and estimating individual forest structural attributes such as basal area, biomass, crown closure, diameter at breast height (DBH), tree height, vegetation density, and leaf area index (LAI) using spectral and spatial remotely sensed information (Song and Dickinson, 2008). Heterogeneity at the stand scale and landscape scale may also be used as a structural indicator as it helps to indicate both the size and arrangement of stands and trees within stands (Lindenmayer et al., 2000).

Coops and Catling (1997) presented one of the first studies that used remotely sensed imagery to model and map forest structural complexity. Their goal was to determine the benefits of airborne videography to predict fauna habitat complexity in eucalypt forests. The authors performed field based sampling in 29, 0.25 ha plots. At each plot a score was given for the complexity of fauna habitat based on 5 features (canopy cover, shrub cover, ground vegetation cover, the amount of litter, fallen logs and rocks, and a moisture rating). Scores ranged from 0-3, with higher scores representing higher habitat complexity. Local variance was extracted from the NIR band based on the mean standard deviation (of a group of pixels) at increasingly larger window sizes (3x3, 5x5, etc.). This produced an individual local NIR variance graph for each image pixel, from which the maximum local NIR variance was extracted for each plot and regressed against the corresponding habitat complexity scores ($R^2 = 0.75$). The model was then applied across the landscape and validated using 10 independent sites, of which eight were correctly predicted.

A similar study, performed a decade later by Wunderle et al. (2007), focused on spatial variability in a 2000 km² area. It attempted to determine the scale at which spatial variation was significant and to determine what amount of detail was required to capture this variation. Using 44 (2.5 m x 2.5 m) plots, eight structural variables were measured followed by PCA to create a structural complexity index. The structural complexity index, comprised of the first PC (explaining 79% of the total variance), was then modelled using stepwise multiple regression against spectral and textural image data extracted from pan-sharpened SPOT-5 (2.5 m) imagery. The three best predictors were the Normalized Difference Moisture Index ($\{NDMI = (NIR - SWIR) / (NIR + SWIR)\}$),

where NIR and SWIR are the near infrared and shortwave infrared band reflectance, respectively), first-order standard deviation and second-order Grey level co-occurrence matrix (GLCM; see Section 3.5.2). Correlation texture measures accounted for 74% of the structural complexity index variance. The results also demonstrated that the most suitable window size for texture measures was 5x5 pixels for these image data, contradicting the general belief that larger window sizes were best for modelling forest parameters (Wunderle et al., 2007). This approach to modelling forest complexity, where field variables are combined using PCA and the resulting PCs modelled against image variables, was adapted to this research as described in Section 3.6.1.

2.5 Multivariate Modelling of Forest Structure: Redundancy Analysis

Multivariate modelling examines the interaction of multiple attributes simultaneously, instead of modelling individual structural attributes. It is the determination of a relation between a set of predictor and response variables that maximizes the variance accounted for in the set of response variables. A common method to conduct such analysis is RDA. It was used in this research to model forest complexity as described in Section 3.7.

RDA is a multivariate statistical technique one can use to explore linear relations between multiple predictor (image and topographic variables) and response variables (field variables) (Legendre and Legendre, 1998). It is a technique known as constrained ordination, which organizes high dimensionality data along one or more simplified gradient axes (Pielou, 1984). It is also a method of canonical analysis where multiple datasets are simultaneously analyzed. It was originally designed to allow ecologists to

describe the relationships between species composition and environmental descriptors at the same location (Legendre and Legendre, 1998). Unlike the study by Wunderle et al. (2007), which reduced all field variables to a single PC, RDA allows direct examination of the behaviour of more than one dependent variable simultaneously (Figure 2) (Palmer, 1993). PCs are often difficult to interpret, and for varied data types of low correlation, many significant components may be produced, each with little explanation power (Pasher and King, 2010).

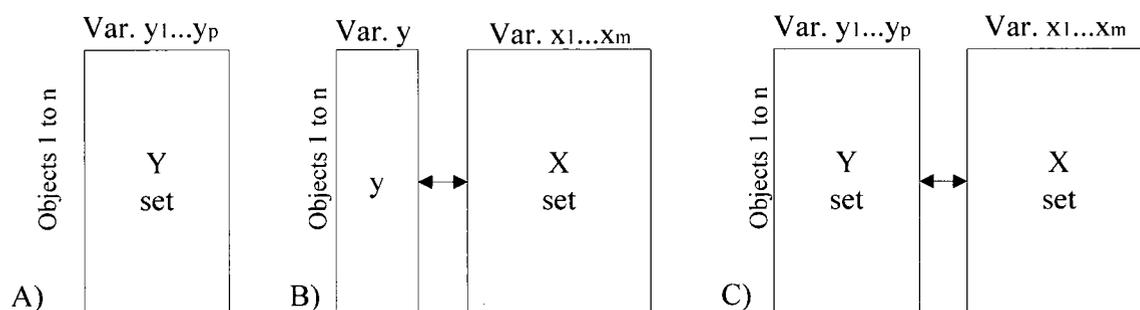


Figure 2. Relations between A) the ordination of a set of Y variables (PCA), B) the ordination of a single y variable against the set of X variables (multiple regression), and C) the ordination of a set of Y variables against the set of X variables (RDA) (Adapted from Legendre and Legendre, 1998).

RDA allows prediction of the set of field variables using the set of predictor variables, which for this thesis were image and topographic variables. The ordination of the field variables during RDA is done in a manner that the resulting vectors are a linear combination of the measured image variables (Legendre and Legendre, 1998). Basically, RDA is an extension of both multiple regression and PCA, and can be broken down into those two main parts (Figure 3). To execute the RDA, each y (response) variable is regressed on the X (predictor) variables. The new fitted/predicted values (\hat{y}) from the multiple regressions are assembled in matrix \hat{Y} and the residuals are placed in matrix

Y_{res} . This means that matrix \hat{Y} contains the portion of Y that is accounted for by X and matrix Y_{res} contains the portion of Y that is independent of X . Then, a single PCA of matrix \hat{Y} is performed using all of the new fitted/predicted field variables to reduce the dimensionality of the dataset (Legendre and Birks, 2010). This produces eigenvalues, a matrix of eigenvectors and a matrix of orthogonal (uncorrelated) components, or canonical ordination axes. Each canonical axis relates to a direction (in Y), which is maximally related to a linear combination of the predictor variables, similar to a PC obtained through PCA. Like a PC, each component or axis represents decreasing portions of the original variance in the set of response variables, and can be interpreted in terms of their relation to the original response and predictor variables (Legendre and Legendre, 1998). The results can be displayed on a triplot, a plot of ‘sample (field plots) scores, species (field) scores, and environmental (image) scores’, displayed in a way that species and samples most alike will be closer together on the plot. The residuals of the regression may also be examined through an ordination (PCA) of matrix Y_{res} .

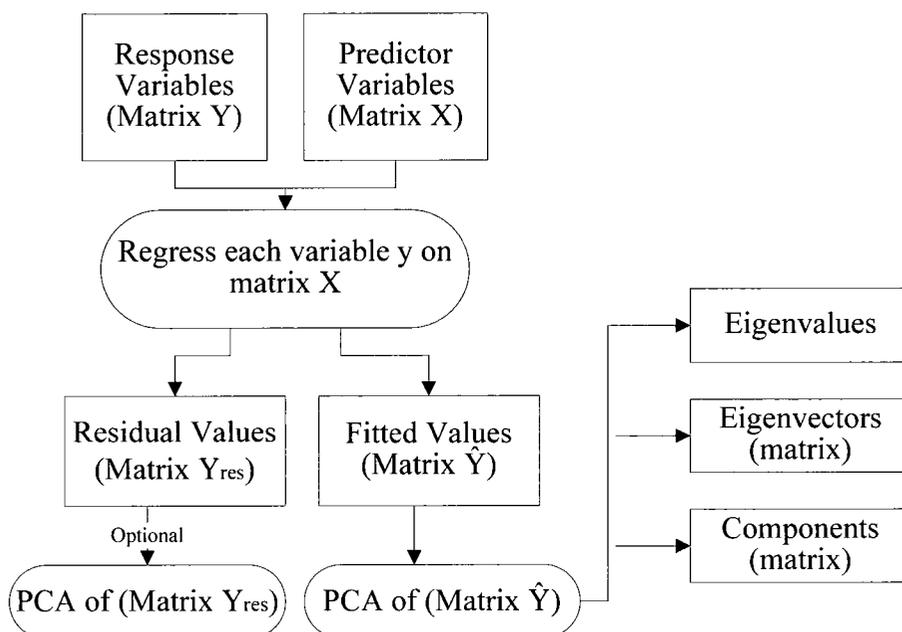


Figure 3. Process map showing the execution of RDA through linear regression and PCA. (Adapted from Legendre and Legendre, 1998).

De la Cueva (2008) performed one of the few studies using RDA to examine the relation between forest attributes and spectral and topographic variables. Structural attributes included canopy crown closure, stand height, # of trees and basal area which were derived from the Spanish National Forest Inventory. Spectral and topographic variables were derived from Landsat Enhanced Thematic Mapper Plus (ETM+) imagery and included spectral reflectance values in six bands as well as four spectral vegetation indices. Elevation, slope, aspect and solar illumination were also derived from a digital elevation model (DEM). Though the RDA results were statistically significant, a low percentage of explained variance was obtained for the first canonical axis (19%) and the combination of the first and second canonical axes (24%). This made it difficult to apply the model in mapping of forest structure. Though the results of this study were not particularly successful, it served as a good starting point for future research.

A Structural Complexity Index (SCI) derived through RDA (Pasher and King, 2010) used high resolution airborne imagery (20 cm pixels) and topographic data. Twenty four field variables and eighty six image variables, representing spectral, spatial and topographic information were used in a two stage bootstrapping procedure which assisted variable selection and provided model validation. Using forward stepwise selection, 500 iterations of the RDA were run, each time randomly choosing 40 of the 50 field plots. Five hundred models were then produced and tested using a second bootstrapping stage. The best model was selected based on the following criteria: 1) fewest number of variables included, 2) the best model fit, and 3) the lowest root mean square error percent (% RMSE). Although models were produced with R^2 up to 0.48 for the first RDA axis, the final model that was selected had a reduced number of predictors that were all easy to extract for mapping. Its first axis explained 35% of the total variance ($R^2 = 0.35$) of the field variables and was interpreted as a general complexity gradient. This model was applied to the airborne imagery to produce a continuous map of predicted complexity. Independent site validation was performed showing that high complexity sites were predicted with 88% accuracy and low complexity sites with 81% accuracy.

The primary objectives for this thesis were to adapt, implement and compare the three types of complexity indices (PCA, additive, and RDA) described in this chapter in order to assess which method was most appropriate for characterizing the forest complexity in Chelsea, Québec, as an indicator of biodiversity.

3. Methodology

3.1 Study Area

This research was conducted in Chelsea, Québec, a 111 km² triangle-shaped municipality located on the Canadian Shield, directly north of Gatineau, Québec. Sixty percent of the municipality of Chelsea is within the southeastern part of Gatineau Park (Figure 4), and is bordered by the Gatineau River on the east (Plan Environnemental de la Municipalité de Chelsea, 1991). Other portions consist of large residential lots and tracts of undeveloped land. The topography in Chelsea consists of a combination of flat regions, small hills and higher steep hills separated by small and generally narrow valleys due to past glacial activity, typical of the southern Canadian Shield. The average elevation is 112.5 m above sea level, and elevation varies from about 100 m near the Gatineau River and Beamish Lake to 230 m in the area of the Precambrian shield west of route 105. This region is the most rugged, as it is composed of rolling hills leading up to a more mountainous area, where several wetlands are found. The northern (Farm Point) and southern (Chelsea and Old Chelsea) extremities of the municipality are characterized respectively by relatively flat terrain and slightly mountainous terrain (Plan Environnemental de la Municipalité de Chelsea, 1991).

The region is one of the most biodiverse areas in the National Capital Region and is composed of deciduous, coniferous and mixed forests, as well as wetlands. Sugar maple (*Acer saccharum* Marsh.) is the most dominant deciduous tree species, but other species including red maple (*Acer rubrum* L.), ironwood (*Ostrya virginiana* (Mill.) K. Koch), white oak (*Quercus alba* L.), red oak (*Quercus rubra* L.), American basswood (*Tilia americana* L.), American beech (*Fagus grandifolia* Ehrh.), white ash (*Fraxinus*

americana L.), paper birch (*Betula papyrifera* Marsh.) and American elm (*Ulmus Americana* L.) are also present. Coniferous species include the eastern white pine (*Pinus strobus* L.), red pine (*Pinus resinosa* Sol.), eastern white cedar (*Thuja occidentalis* L.), white spruce (*Picea glauca* (Moench) Voss), balsam fir (*Abies balsamea* (L.) Mill), and eastern hemlock (*Tsuga canadensis* L.), with the latter being the most dominant species. Scots pine (*Pinus sylvestris* L.) is also found in the area as an introduced species. This composition is in part the result of white pine harvesting, which occurred during the early nineteenth century as land in the surrounding area was converted to farms (Municipality of Chelsea, 2007). As of 2006, Chelsea's population was 6703 and had a population density of 60.1 persons per km², most of it clustered within a few small communities (Municipality of Chelsea, 2007).

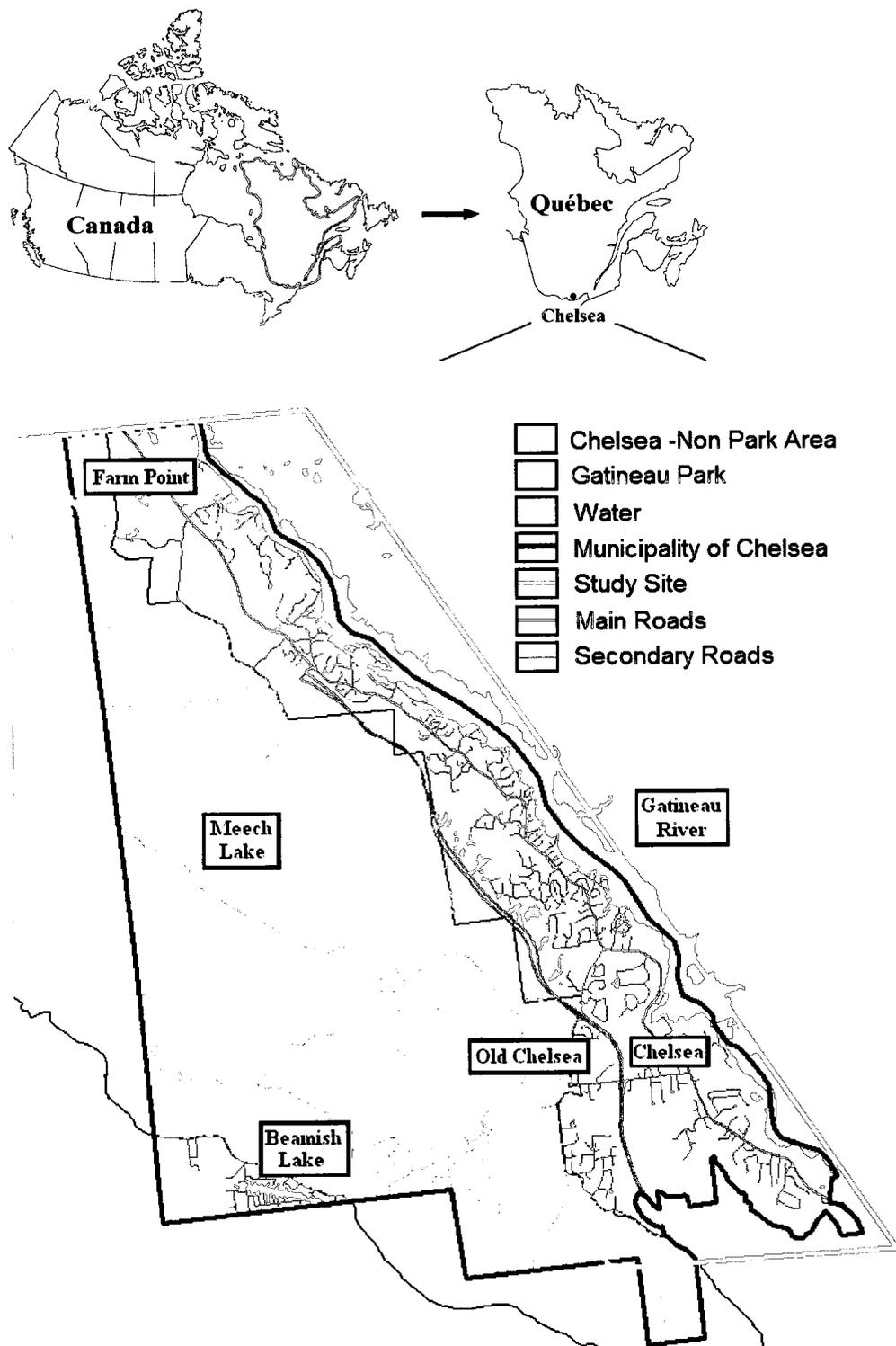


Figure 4. Map of the municipality of Chelsea, Québec showing expanse of Gatineau Park inside the municipality and the study site for this project in orange (ACRE, 2004).

The Chelsea community is well known for environmental programs, and in 1998, was one of the first municipalities to ban the use of pesticides for aesthetic purposes (Municipality of Chelsea, 2007). A current aim of the Chelsea Municipal Plan is to make the community sustainable partially through the conservation of biodiversity in the region. This included support of Nature Chelsea, a two-year project started by Action Chelsea for Respect of the Environment (ACRE). The goal of the project is to identify and begin protecting wildlife habitat in Chelsea and to disseminate the information to citizens, landowners and planners. Its activities include educating the public about environmentally friendly practices, performing Bio-Blitzes to inventory species, and determining existing primary corridors for animal movement. The Bio-Blitz was conducted in June 2009 and 2010 and consisted of surveying and sampling important habitat in Chelsea in order to provide a snapshot of spring biodiversity. The forested land of Chelsea was selected for this research to help meet the goals of Nature Chelsea and because the Bio-Blitz data were expected to be useful for validation against predicted forest complexity.

3.2 Field Plot Creation

The data for this research were collected during the summer of 2009, during leaf-on conditions. An initial survey was qualitatively performed in all the forested areas of Chelsea at over 120 locations, including: the size of overhead gaps, variation in tree size, proportions of deciduous and coniferous trees, relative amounts of snags and CWD, and cover of saplings. From these a relative complexity score (1 (low) to 5 (high)) was given to each potential location to determine its suitability to this study. From these 120

locations, 70 field plots were established (Figure 5). The large number of plots provided a data distribution that tended towards normal and allowed for some plots to be set aside for independent testing of models.

Since forest complexity is considered a continuous gradient, plot locations were selected to represent various conditions based on the following criteria:

1. Representation of a range of forest complexities to maximize the range of the response (y) variables.
2. Representation of a range of elevations, slopes and aspects, to ensure that the full variation in forest attributes that respond to topography was captured.

At each field location, a 20 m x 20 m plot was established and its location recorded using a differential GPS. The size of the plot was selected to be consistent with previous work (Pasher and King, 2010), and based on studies to determine optimal plot size in the similar Gatineau Park forest nearby (Butson and King, 2006). It was important to remain consistent with Pasher and King's work in order to facilitate the comparison of RDA models. Stakes were inserted into the ground at the centre and four corners of the plots (Figure 6), the corners being aligned with the cardinal directions. The coordinates of all five stakes were recorded using a Trimble GeoXT real-time differential GPS (expected horizontal positional accuracy of 0.5-1.0 m (Trimble, 2010)), and post-processed using the Gatineau SOPAC base station. (J. Pasher, *pers. comm.*, 2009).

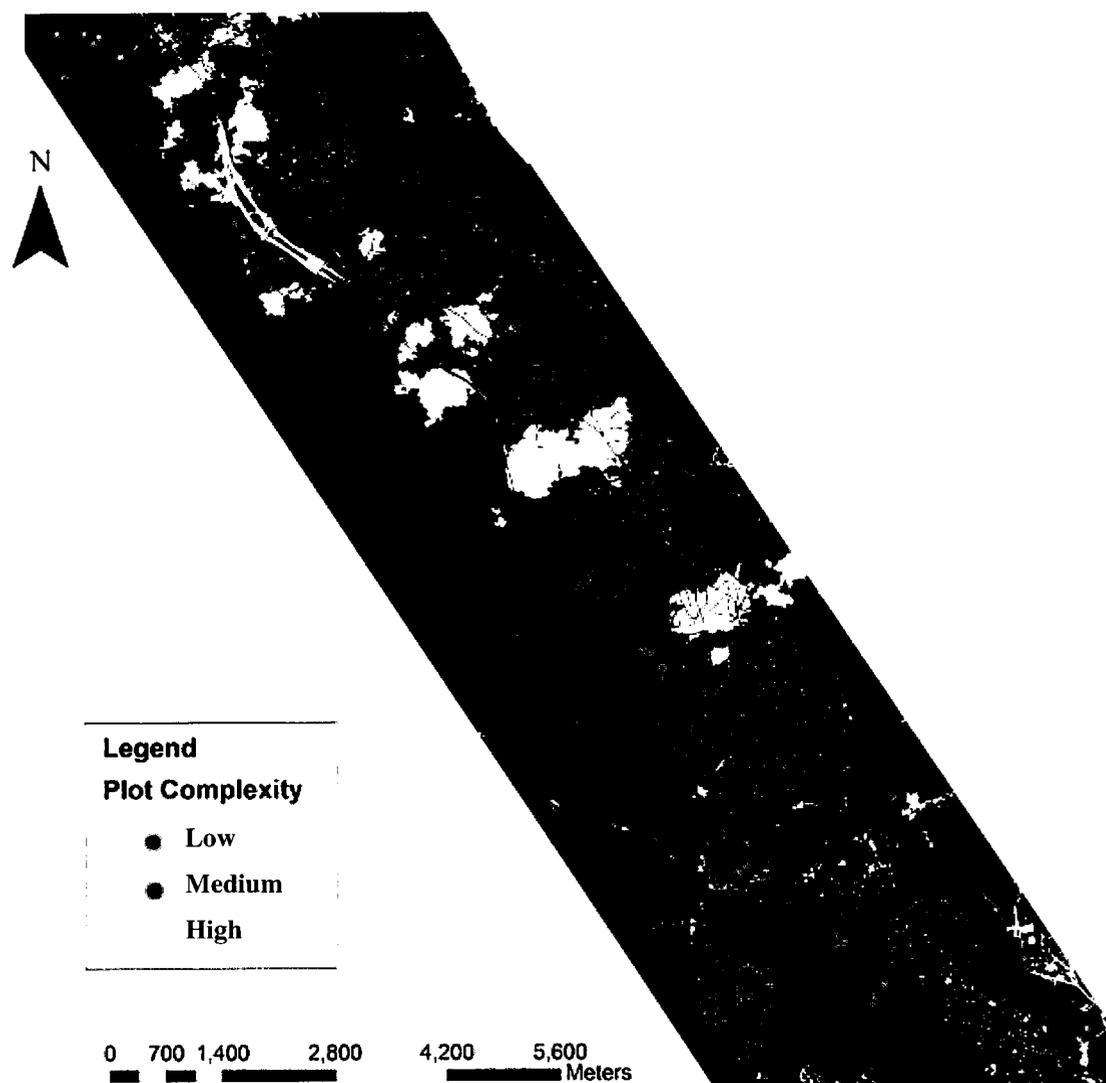


Figure 5. Study area and plot locations displayed over the Quickbird multispectral imagery. Plot location colour corresponds to the qualitative complexity score given in the field.

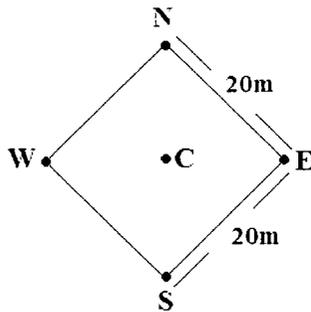


Figure 6. Example plot set up, showing centre and corner stakes at cardinal points.

3.3 Field Variable Selection

Field variables were selected and measured based on: 1) previous multivariate forest structure research (e.g. Cosmopoulos and King, 2004; McElhinney et al., 2006; Wunderle et al., 2007), 2) previous research in the area (e.g. King et al., 2005), 3) variable significance in multivariate models derived in Pasher and King (2010) for a study area in the nearby Gatineau Park, and 4) ease and efficiency of measurement. They included information on: tree size, variability, and distribution; snags and coarse woody debris; ground and understorey vegetation; canopy openness and LAI (the amount of one-sided leaf area per unit area of ground); the number of plant species; and the proportion of coniferous versus deciduous trees. Table 1 provides an overview of all field data recorded, and the variables derived from them.

Table 1. Summary of the field data recorded and the field variables derived from them.

	Field Data Recorded	Calculated field variable
Tree size and distribution	Tree species	# Tree species
	# Trees	# Trees
	Tree DBH and basal area	Mean dbh, std. deviation dbh, live basal area (m^2ha^{-1}), coniferous basal area (m^2ha^{-1}), deciduous basal area (m^2ha^{-1}), % basal area coniferous and deciduous, # large trees (>35 cm dbh)
	Angle to top of tree, Distance to top of tree	Mean tree height, std. deviation tree height
	Nearest Neighbour (NN) distance	Mean NN distance, std. deviation NN distance
Hemi.	Hemispherical photos at 1 m and 6 m	Openness at 1 m (0° - 60°), LAI at 1 m (55° - 60°), openness at 6 m (0° - 60°), LAI at 6 m (55° - 60°), change in openness (%), change in LAI (%)
Vegetation	Litter, Debris, Soil, Rock	% Litter, % debris, % soil, % rock
	Vegetation Types	# Plant species
	Understorey Vegetation	% Vegetation cover (<2 m), % vegetation cover (<2 m) in each height class
Snags	# Snags	# Snags
	Snag DBH and basal area	Mean snag dbh, dead basal area (m^2ha^{-1})
	Angle to top of snag, Distance to top of snag	Mean snag height
CWD	# Pieces CWD	# Pieces CWD, CWD volume
	Condition of CWD	Condition of CWD

3.3.1 Tree Size and Distribution

Tree size and distribution information were collected for all dominant trees (dbh >10 cm, measured at a height of 1.3 m) in each plot, in order to remain consistent with tree measurements performed by Pasher and King (2010). It was anticipated that these trees would form the upper canopy, and would therefore be captured in the imagery. Due to their dominant nature and size, they were also expected to affect the amount of light reaching the ground. Using a clinometer, each tree's height (m) was measured, and its dbh (cm) recorded. The standard deviation of dbh and tree heights were also calculated. The standard deviation of dbh, which represents the variability in tree size, is important as it is often used to represent the diversity of micro habitats (Acker et al., 1998). The standard deviation of tree height is equally important as it describes the vertical stratification of the forest plot. Zenner (2000) showed that an area with more vertical stratification was more likely to contain different tree ages and species, which lead to more microhabitats for species. The number of large trees per plot (>35 cm dbh) was recorded as a measure of habitat availability (Spies and Franklin, 1991), based on the number of trees greater than the 90th size percentile. The horizontal distribution and arrangement of dominant trees was captured using the average nearest neighbour (NN) distance (m) for all trees in the plot (excluding snags) (Spies and Franklin, 1991; Neumann and Starlinger, 2001). Live basal area (m^2ha^{-1}) for each plot was calculated as a relative measure of stand volume and biomass, using each tree's basal area and the plot basal area (Ziegler, 2000). In addition, coniferous basal area (m^2ha^{-1}) and deciduous basal area (m^2ha^{-1}) were also determined.

The number of different tree species was recorded per plot as both a composition

and structural indicator. The Pasher and King (2010) study was conducted in purely deciduous forests whereas the forests of Chelsea include a variety of coniferous species, often in significant groups. Consequently, for this research, the concept of forest complexity was broadened from the structural complexity of Pasher and King and other studies to include composition diversity. It was expected that variability at either the species level (number of plant species) or at a broader level (proportion coniferous vs. deciduous) would be important complexity indicators in terms of provision of a diversity of habitat types and structures. Spies and Franklin (1991) suggest that a variety of tree species (shade-intolerant and tolerant) can produce a multi level canopy, multiple stem sizes, as well as a complex horizontal arrangement of those stems.

3.3.2 Canopy Openness and LAI

Canopy openness and leaf area index (LAI) are two good indirect measures of canopy structure (Frazer et al., 2001) as they are able to characterize the position, size and shape of vegetation, and therefore the amount of light able to penetrate the canopy to the understorey vegetation (Weiss et al., 2004). Digital hemispherical photography has been found to be a reliable, rapid, indirect, non-contact method to record canopy openness and leaf area index (LAI) (Jonckheere et al., 2004). Following the methods described in Pasher and King (2010), hemispherical photographs were taken in each plot using a Nikon Coolpix 8800 (8 megapixels) with an FC-E9 fisheye lens (180° field of view). These were taken under diffuse sky conditions in order to reduce the amount of direct scattering of light off foliage and trunks, as well as to reduce the amount of blooming around the edges of the foliage (Olthof et al., 2003). The exposures were

manually set by adjusting the shutter speed and lens aperture, based on the light conditions in each plot. The camera was mounted on a tripod (with the top of the camera facing north) and photographs were taken at 1 m and 6 m heights in the centre of each plot. One photo was taken instead of several based on tests conducted by Pasher and King (2010) in the Gatineau Park that showed limited spatial variability of LAI within 20 m x 20 m plots. Photos taken halfway between the centre and plot edge would also include canopy outside the plot within the 55-60° view angle used to estimate LAI and openness. Finally, this study strived to have a large number of field plots, which limited the amount of time available to take hemispherical photos. The photos were used to estimate two field variables (LAI and openness) out of the suite of 37 so it was deemed more important to develop the large field dataset than to focus specifically on LAI and openness measurement.

DHP v.4.5.2 and Tracwin v.4.1.1 (Leblanc, 2006) software were used to process the hemispherical photographs. Thresholding of each image was done, for each of the six rings, using DHP to separate sky from canopy pixels. Since Tracwin was created to read TRAC (Tracing Radiation and Architecture of Canopies instrument) photosynthetic photon flux density (PPFD) profiles, the output from DHP is modified to represent azimuth measurements of DHP. The profile was analyzed and the gap size distribution of each segment was used to calculate a clumping index. Using the CLX clumping index, LAI, at both 1 m (includes understorey vegetation) and 6 m (excludes most understorey vegetation) heights, was calculated directly from the imagery at the 55°-60° zenith angle (57.5°). Canopy openness, which represents the relative amount of open sky, was calculated at both heights using gap fraction across zenith angles from 0-60° (Miller,

1967). Using the openness and LAI variables at both heights, the change in openness and the change in LAI from 1 m to 6 m were calculated to determine the change in vegetation cover between the two heights.

3.3.3 Ground and Understorey Vegetation

Overstorey structure has a very large influence on the presence, abundance and structure of understorey vegetation (Berger and Puettmann, 2000). The size and number of gaps in the upper canopy influences the amount of light able to penetrate to the forest floor, thus creating a direct link between the understorey vegetation and the structure of the upper canopy. While direct radiance from the understorey that is propagated to the sensor is very limited, understorey presence and structure may be indirectly related to image shadow fraction or heterogeneity, which respond to overhead canopy openness. It is also important to describe the understorey vegetation as it helps describe the plot structure and complexity (Spies and Franklin, 1991). Understorey vegetation was measured following the line intercept method described by Pasher and King (2010). Vegetation was sampled along both plot diagonals (2 x ~28 m) every 50 cm (Figure 7). This provided 112 vegetation samples per plot. Five height classes were used: < 10 cm; 11 - 50 cm; 51 cm – 1 m; 1 - 2 m; > 2 m. At each 50 cm marker, the presence or absence of vegetation in each of the height classes was recorded as well as its broad vegetation class (moss and lichen; grass; fern; other herbaceous plant; seedling; shrub; sapling). The presence or absence of rock was also recorded. From these measurements, the total percent vegetation cover under 2 m height was recorded for each plot. The percent of vegetation less than 2 m was also calculated for each height class, as well as the

percentage of rock found in each plot. The number of different plant species per plot was recorded as a compositional attribute.



Figure 7. Example of setup for vegetation sampling, where vegetation height and species was sampled every 50 cm at the orange markings on the rope.

3.3.4 Dead Wood

Dead wood, which includes snags and dead fallen logs on the ground or stumps (coarse woody debris, CWD) was measured in this study as a structural attribute. The number of snags (taller than 1.3 m in height) in each plot was counted and their dbh and height were measured. Gibbons and Lindenmayer (2002) state that snags are an important source of tree hollows and show that some hollow dwelling species prefer nesting in the habitat provided by the dead standing trees. The average and standard deviation of snag dbh and snag height were found by Van den Meersschaut and Vandekerckhove (1998) to be indicative of associated hollows and associated diversity of fauna. Dead basal area (m^2ha^{-1}) was calculated for each plot as a relative measure of

volume of dead standing wood, by summing the basal area for all snags.

CWD was considered to be any log on the ground, log leaning against another tree, or stump in the ground, which was larger than 7.5 cm dbh (Stevens, 1997). In each plot, CWD was sampled using a line intercept method, adapted from Pisaric et al. (2008) and Pasher and King (2010). Four N-S transects were placed every four meters resulting in a total transect length of 80 m. This improved on the Pasher and King transects across the plot diagonals (56 m total transect length), which were thought to potentially miss sampling clustered coarse woody debris along the plot sides (Figure 8). The two line intercept methods were tested in eleven plots and it was determined that the 4 transect method captured a better representation of the CWD present as the results were significantly different from the 2 diagonal results (sig. at 0.05). For four of the plots the diagonal transect method missed approximately 50% of the CWD present.

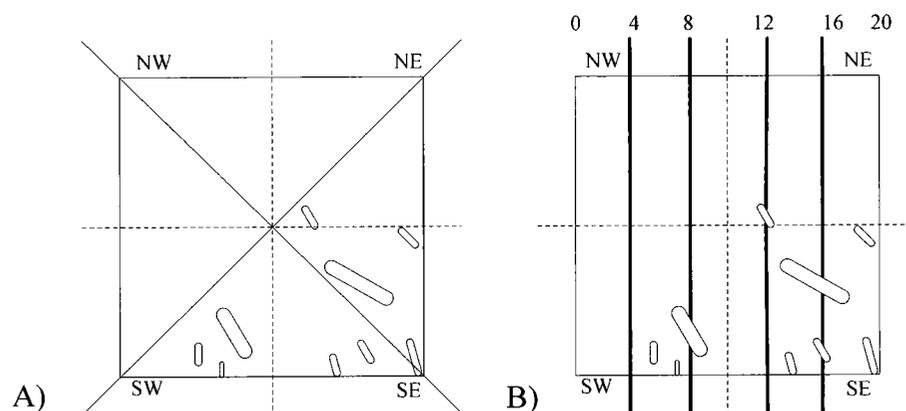


Figure 8. A) Line intercept method used by Pasher and King (2010) B) Adapted line intercept method for sampling coarse woody debris in a plot. (Dominant wind direction: WNW.)

CWD was classified into four decay classes (very solid, solid, rotten, very rotten) as it is considered to be a significant attribute, due to the fact that different vertebrates

find habitat within each decay class (Lindenmayer et al. 1990). The estimated volume of CWD was also determined, using the method described by Van Wagner (1968):

$$V = \frac{\pi \sum d^2}{8L}$$

where V is the volume of CWD per unit area (m^3 / ha), d is the diameter (cm) of the CWD where it intersected one of the four transect lines, and L is the total length (m) of the transect.

3.4 Satellite Imagery Acquisition and Processing

On July 15, 2009, near the middle of the growing season and fieldwork period, one Quickbird scene was acquired for an 18 km x 5 km area covering the study area. It was collected at 2:39 pm with a sun azimuth angle of 225.12° and a sun zenith angle of 30.39° . The data consisted of four multispectral bands (Blue $0.45 - 0.52 \mu\text{m}$, Green $0.52 - 0.60 \mu\text{m}$, Red $0.63 - 0.69 \mu\text{m}$, and NIR $0.76 - 0.90 \mu\text{m}$ (Digitalglobe, 2010)) each with nominal ground pixel size of $2.4 \text{ m} \times 2.4 \text{ m}$, and a panchromatic image ($0.45 - 0.90 \mu\text{m}$) with a $0.6 \text{ m} \times 0.6 \text{ m}$ pixel size. All images were projected to UTM Zone 18T NAD83 datum.

It was noticed that alignment of the acquired imagery with the municipal road database was very poor (Figure 9A). Consequently, additional georeferencing was conducted. The panchromatic image was manually georeferenced to twenty-six 20 cm pixel leaf-off orthophotos (2007) using a thin plate spline model and a 10 m digital elevation model (orthophotos and DEM provided by the Municipality of Chelsea). An RMSE of 3.8 m (X) and 2.7 m (Y) was calculated using a set of independently collected

test points. The multispectral image was then aligned to the panchromatic image, using the thin plate spline model resulting in an RMSE of 3.9 m (X) and 4.4 m (Y) that was also calculated using a different set of independently collected test points. The size of this error was not expected to affect the extraction of image variables from the field plot locations in the imagery, as the error was less than 20% of the plot size. Figure 9B displays the significant improvement after this correction.

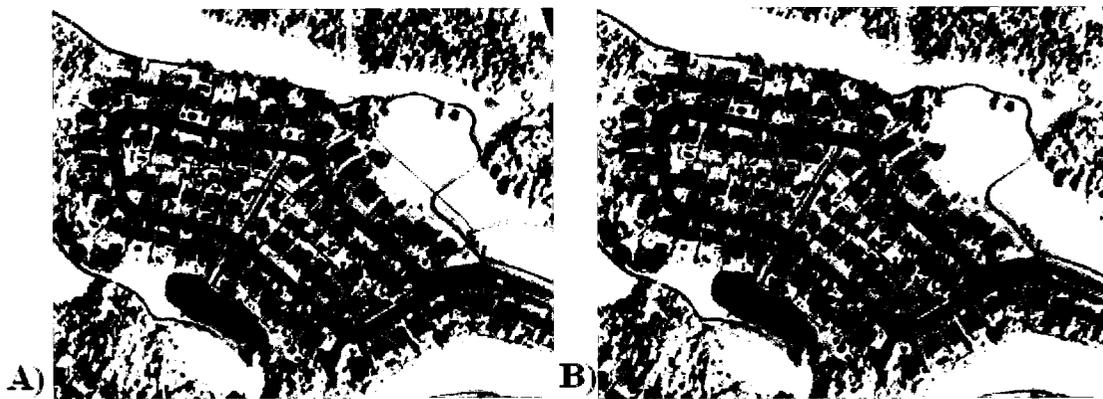


Figure 9. A) Original panchromatic imagery displayed with part of the municipal road network (red) compared to B) improved georeferenced panchromatic imagery.

3.5 Image Variables

Image variables were selected based on the results of a scaling analysis, performed by Pasher and King (2010), which used simulated 1 m pixel imagery. The variables represented 1) spectral information at the pixel level, 2) spatial/textural information from grey-level co-occurrence matrices (GLCM), and 3) topographic information from a digital elevation model. Table 2 provides a summary of the selected variables.

Table 2. Summary of extracted image variables.

Variable Type	Variable Extracted
Spectral	<ol style="list-style-type: none"> 1) % of plot classified as shadow and crown 2) Mean and standard deviation of NIR,R,G, B, pan and NDVI within each plot
Spatial	<ol style="list-style-type: none"> 1) Mean and standard deviation of 5 GLCM texture measures using 3x3 and 9x9 windows and extracted from the NIR band; as above using a 5x5 window and extracted from the panchromatic image
Topographic	<ol style="list-style-type: none"> 1) Elevation, slope, aspect 2) Eastness, northness 3) Solar illumination calculated at the time of image acquisition

3.5.1 Spectral Variables

For each plot, mean image brightness for the blue, green, red, near infrared and panchromatic bands were extracted. It was expected that a more complex plot would have an overall lower image brightness value than a lower complexity plot due to greater presence of shadows from gaps and variable tree height as well as a mix of coniferous and deciduous trees (coniferous trees reflecting less). Mean NDVI was extracted for all plots with the expectation that healthy growing vegetation should have high NDVI values, due to low red reflectance, and high near-infrared reflectance. This is because plant chlorophyll absorbs solar radiation in the red portion of the electromagnetic spectrum, while mesophyll in the plant reflects near infrared (Jackson and Ezra, 1985). Resultant NDVI values range from -1.0 to 1.0. Positive values indicate increasing

amounts of green vegetation, whereas values near zero or negative values indicate non-vegetated surfaces (Jensen, 2007).

Using an ISODATA (Iterative Self Organizing Data Analysis) unsupervised clustering, the panchromatic imagery was classified into sixteen different clusters (Seed and King, 2003; Pasher and King, 2010). The first five clusters were aggregated to create a shadow class, and the remaining eleven clusters represented the vegetation class. Shadow and crown fractions (% of plot covered by shadow and crown) were calculated for each plot. The panchromatic imagery was used as it had a pixel size that was $\frac{1}{4}$ that for the multispectral imagery and, as a result, shadows and crowns could be much better classified.

3.5.2 Spatial Variables

The standard deviation of plot image brightness was extracted from the blue, green, red, NIR, panchromatic and NDVI images as a first order texture measure. Five GLCM texture measures used by Pasher and King, (2010) were extracted for each plot (these had been selected based on cumulative evidence from many previous studies). A GLCM is a two-dimensional histogram of grey levels for pairs of pixels (i,j) which are separated by a fixed spatial relationship (one or more pixels at a given angle (0, 45, 90, 135 degrees)). The GLCM approximates the joint probability distribution of a pair of pixels. This means that each entry within the matrix represents the number of pixel pairs with the given brightness values of the cell entry. After a GLCM is acquired for a sample window, a texture algorithm is then applied in order to calculate a texture measure from the window. Greater image texture occurs when there are higher frequencies of pixel

pairs away from the diagonal (the diagonal being two equal brightnesses), whereas lower image texture occurs when there are higher frequencies close to the diagonal (Haralick, 1973).

The second order texture measures were Homogeneity, Contrast, Entropy, Angular Second Moment (ASM), and Correlation (Haralick et al., 1973), which are considered to be useful measures for distinguishing forest structure (Franklin et al., 2000; Ozdemir et al., 2008; Wunderle et al., 2007). The formulations for the five textures are given in Appendix 1. Contrast is a measure of the amount of local variation in the image. It is high when the local region has a high contrast of grey levels such as at edges of objects (Ozdemir, 2008). Entropy is a measure of statistical uncertainty (or randomness). It is high when the elements of GLCM have relatively equal values (rough texture) and is low when the elements are close to either 0 or 1, or when the image is uniform in the window (fine texture) (Haralick, 1973). ASM is a measure of homogeneity of the image. It is high when the GLCM has few entries of large magnitude, which defines low texture. ASM is low when all entries are almost equal (Haralick, 1973). Homogeneity is a measure of the closeness of the distribution of elements in the image. It is high if the GLCM concentrates along the diagonal (Ozdemir, 2008). Correlation is a measure of grey-tone linear-dependencies in the image. It is higher when grey tones concentrate in either horizontal or vertical directions (Haralick, 1973). These and other GLCM textures are sometimes difficult to interpret individually and are often tested in groups to determine the best ones for a given application.

It was expected that Contrast, Correlation and Entropy, which increase with increasing image texture, would be higher in complex plots, while ASM and

Homogeneity, which decrease with increasing image texture, would be lower. Multiple window sizes were tested, based on their ability to differentiate between high and low plot complexity; 3x3 (7.2 m x 7.2 m) and 9x9 (21.6 m x 21.6 m) moving windows were used to extract the texture measures from the NIR band and a 5x5 (3 m x 3 m) moving window was used to extract the texture measures from the panchromatic image in all directions. The NIR band was selected based on tests performed on all bands to determine which band displayed the largest texture variation between high and low complexity plots. The mean and standard deviation were recorded for each plot, excluding edge pixels, so that only variation within the plot was captured. For textures extracted using the 9x9 window, only the value from the centre pixel in the plot was recorded in order to stay within the plot bounds.

3.5.3 Topographic Variables

Forest structure and composition may be partially determined by topography. Topography has the potential to moderate the amount of incoming light, the type of species which can grow, and the amount of wind entering the canopy (Band, 1993; Chen et al., 2007).

A 10 m resolution DEM was created from a set of 10 m interval contour lines (NCC, 2002) using a spline interpolation. An average elevation was extracted for each plot. Using the DEM and common GIS algorithms, slope and aspect maps were created. Northness (sine of aspect) and eastness (cosine of aspect) values were calculated from the aspect map (Guisan et al., 1999; Lassueur et al., 2006). The mean value for each plot was extracted for the above four variables.

A solar illumination image was also created to ensure residual effects of topographically related brightness variations did not remain in the imagery. It utilized the sun zenith angle (30.39°) and azimuth angle (225.12°) to determine the position of the sun at the time of image acquisition (July 15, 2009; 2:39pm). The solar illumination equation is as follows (Riano et al., 2003):

$$SI = \cos\theta_p \cos\theta_z + \sin\theta_p \sin\theta_z \cos(\varphi_a - \varphi_o)$$

where θ_p is the angle of the slope, θ_z is the solar zenith angle, φ_a is the solar azimuth angle, and φ_o is the aspect angle. It should be noted that this measure provides illumination at the time of image acquisition, and is not necessarily an accurate representation of forest light conditions.

3.6 Defining Forest Structure and Composition Indices from the Field Data

Several methods of creating a field-based index of forest structure and composition diversity were explored. Initially, the field and image variables were checked for normality and transformed when necessary using log or square root transformations. Bivariate Pearson's correlation (r) values were determined and variables were reduced where redundant information was present. A dataset of 37 field variables were reduced to 26, by removing the variable which was more difficult to measure from the study (Table 3), when variable pairs were highly correlated ($r > 0.8$). The first method involved performing a PCA on the field variables to create a field index (e.g. as in Wunderle et al., 2007). The second method followed the steps outlined by McElhinny et al. (2006) to create a simple additive index. The third method used RDA, which does not produce an independent field based index and will be described in Section 3.6.

Figure 10 provides a process map of all three forest complexity indices. Since the forest being examined in this study was structurally different from the Australian woodland and dry sclerophyll forests (McElhinny et al., 2006) and somewhat different from the temperate hardwood forest (Pasher and King, 2010) previously explored, the methods were adapted appropriately.

Table 3. Eleven field variables removed due to high correlations with other field variables.

Type	Variables Removed	r	Correlated Variables
Overstorey	Coniferous basal area (m^2ha^{-1})	0.82	LAI 6m
	Deciduous basal area (m^2ha^{-1}),	-0.82	LAI 6m
	% Basal area deciduous	-0.82	LAI 6m
	Mean tree height	0.80	Std. deviation snag height
	Std. deviation NN distance	0.83	Change in openness
Deadwood	CWD volume	0.86	# pieces CWD
	Condition of CWD	N/A	N/A
	Mean snag height	0.81	Mean tree height
Ground Vegetation	% Litter	-0.85	% Rock
	% Soil	N/A	N/A
	% Debris	0.86	# pieces CWD
Openness and LAI	None	N/A	N/A

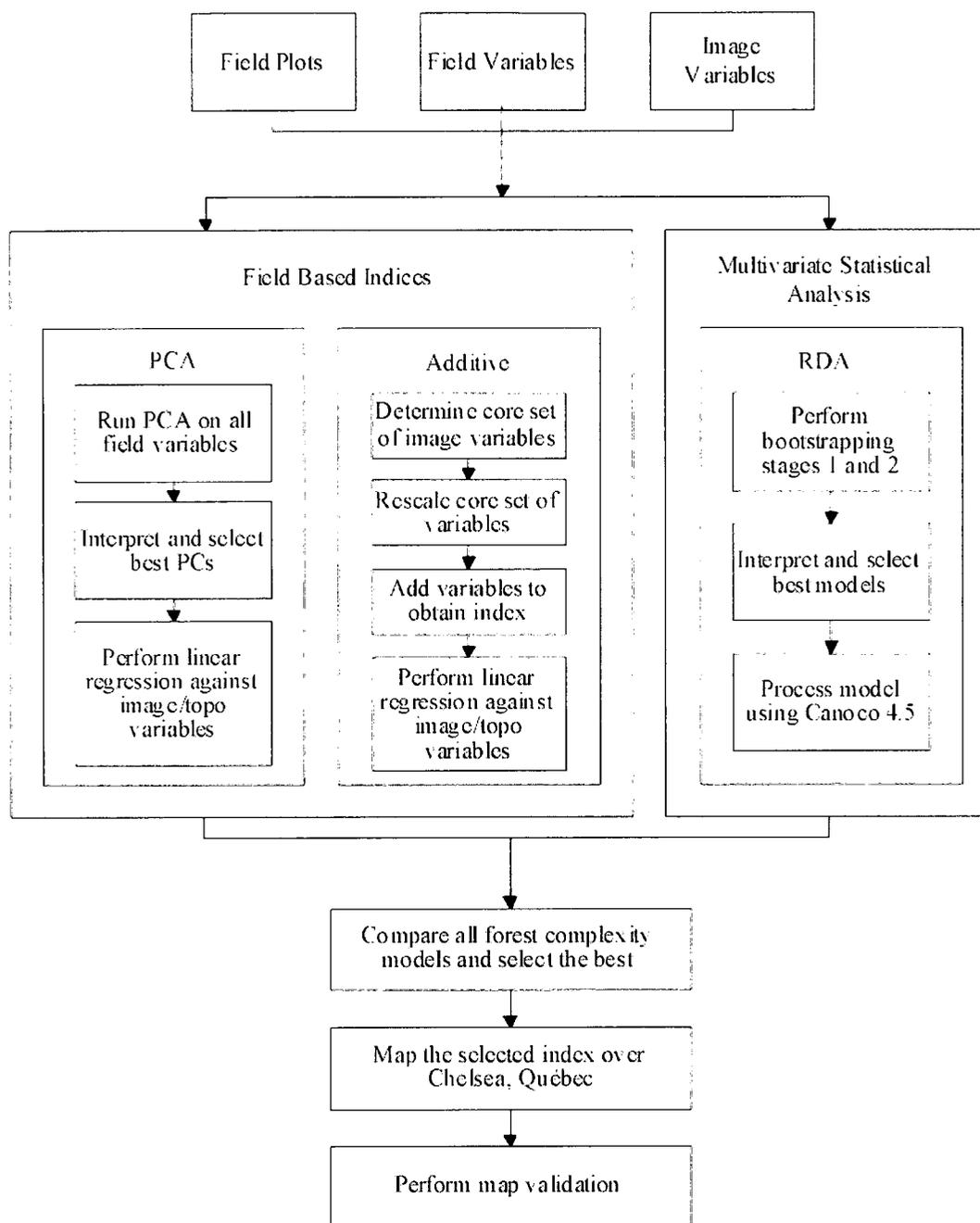


Figure 10. Process map showing the creation, selection, mapping and validation of three separate indices, including two field based methods (PCA and an additive method) and one multivariate statistical method (RDA).

3.6.1 Field Based Complexity Index Derived from PCA

PCA was run on all the field variables with the objective of data reduction and to determine if a complexity gradient could be interpreted in any of the resulting PCs. This reduced the original data set with many (potentially redundant) variables to a new data set with fewer variables that were uncorrelated and retained an equivalent amount of information as the initial group of measures (August, 1983). The PCs were interpreted and then modeled using stepwise regression with the geo-spatial variables as predictors and the PCs as the dependent variables to create a forest complexity index (probability of F-to-Enter 0.05; F-to-Exit = 0.10). A variance inflation factor (VIF) of 5.3 was used as a cut-off threshold (Hair et al., 1998) for all regression models, in order to avoid multicollinearity. Normality and uniformity of the variance of the residuals were checked to ensure that these assumptions were not violated. Models were derived for individual PCs as well as for an additive combination of PCs weighted by their respective proportion of explained variance.

3.6.2 Additive Field Based Complexity Index Derived Subjectively

This method followed and adapted the steps outlined by McElhinny et al. (2006) to create a simple additive index, as described in section 2.2.1. Initially, a three-stage analysis was used to identify a core set of structure/composition attributes from the suite of field variables. The means and standard deviations of the 70 plots were checked to ensure they represented a broad range of forest structure and conditions. In the second stage, univariate relations were used to identify core attributes using three criteria: 1) data should be normally distributed, 2) the attribute should function as a surrogate for other

attributes ($r > 0.60$), and 3) the attribute should be simple to measure. Eleven core attributes (Table 4) were identified from the suite of 26. They were standardized to a mean of zero and a standard deviation of 1 (z-scores), to remove any bias caused by the different scales of measurement for each attribute. PCA was performed to determine if additional redundancy existed between these core variables.

Table 4. Core attributes used in additive index.

Core Attributes	
% Vegetation cover (0 – 10 cm)	Num # of large trees (>35 cm dbh)
% Vegetation cover (50 cm – 1 m)	Std. deviation tree height
# Plant species	Dead basal area (m^2ha^{-1})
# Tree species	# Pieces CWD
Live basal area (m^2ha^{-1})	LAI ≥ 6 m (57.5°)
# Trees	

The core attributes were then combined into an additive index. The variables were rescaled from zero to ten to provide an objective starting point for the index. This was possible as each attribute was measured as a continuous variable, and was generally normally distributed across each plot. Two rescaling methods were explored during this stage. The process described by McElhinny et al., (2006) was compared to a simple linear stretch method. McElhinny et al. used a regression analysis to fit a linear relationship between the attribute score and the quartile midpoints of the attribute distributions. All variables were then manually constrained between 0-10. The linear stretch method initially set the lower bound of the range to zero by subtracting the lowest value from each score. Next a factor was determined by dividing the new maximum into ten. This factor was then applied to each value, constraining the range between 0-10.

Unlike McElhinny's variables, which all increased with increasing complexity, some of the core variables above did not. Live basal area, leaf area index at 6 m (57.5°), and the number of trees decreased with increasing forest complexity. To take this into account, their scores were multiplied by -1, so that they still had the same amount of explanation, but were not positively influencing the index. The rescaled data were then combined to create an additive index. As above, this index was modeled using stepwise regression with the geo-spatial variables as predictors. A second model was also created using only 10 core attributes (excluding leaf area index at 6 m (57.5°)). It was removed as it was still fairly highly correlated with live basal area ($r = 0.625$).

3.7 Complexity Index Derived from Multivariate Relationships between Field and Geo-spatial Data: Redundancy Analysis

The RDA bootstrapping method described in Pasher and King (2010) was used in this research. To create the RDA code, Pasher and King (2010) modified existing code and statistical procedures from both Peres-Neto et al. (2006) and Jones (2008). The process, programmed for Matlab, used two bootstrapping stages, one for variable selection and one for model validation. For the full code, refer to Pasher (2009) Appendix 2.

During the first stage, 1000 iterations of the RDA were run using forward stepwise selection, where 53 (75%) of the 70 plots were randomly selected for each iteration. This created 1000 potentially different models, which were independently analyzed and validated using the second stage of bootstrapping. The second stage provided a means of model validation by providing measures of variance. Here, Pasher

and King (2010) explain that the 1000 models were then run through an RDA, without using stepwise selection, only entering the variables which had been entered during the first stepwise regression stage of the model creation. Again, 53 out of 70 plots were randomly selected for each iteration.

During the bootstrapping validation stage, each RDA model was tested with the seventeen plots that were excluded at that stage, making them independent from the model. Each RDA model was applied to the image variables associated with the seventeen validation plots in order to calculate the predicted first RDA axis, which is simply a linear combination of the field (y) variables as predicted by the geospatial (x) variables. The predicted RDA axis was then compared to the observed axis, which was calculated from the measured field variables in each of the seventeen plots. In order to compare the two axes, the observed field variables needed to be transformed into the equivalent projected space as the modeled RDA axes, using a transformation matrix U, since their projected spaces were not the same (J. Pasher, *pers. comm.*, 2010). Summary statistics were calculated during each iteration, for each of the 1000 models. These included adjusted R^2 , average Akaike information criterion (AIC), the % RMSE, r-validate (correlation) between the predicted and observed, and the significance of the model.

In order to select the most appropriate models, all models were ranked from best to worst adjusted R^2 and % RMSE. The average Akaike information criterion (AIC) (Akaike, 1974) was also examined, as a means of comparing and selecting models. Models can be ranked according to their AIC and models with the lowest value are considered to be the best. While this helped to narrow down the selection, manual

investigation of the models was still required to explore summary statistics and determine which image variables were the most commonly selected in the bootstrapping iterations and therefore most important. CANOCO 4.5 (ter Braak and Šmilauer, 2002), a software package intended for ordination analysis in community ecology, was used to process the final RDA model to obtain the required statistical and visual outputs.

Triplots were created to evaluate the relations between field and geospatial data and to aid interpretation of the RDA axes. They are a type of ordination diagram that plots the field sites (samples), image (environmental) variables and field (species) variables on the first two RDA axes. They help visualize the relations between both sets of variables as well as the position of field sites on the RDA axes, and are a commonly used method for interpretation of RDA results (Pasher and King, 2010, De la Cueva, 2008). Generally, the first RDA 'component' is shown on the x-axis of the triplot, and second RDA 'component' is shown on the y-axis.

3.8 Model Comparison and Selection

The top models which resulted from the above three methods (as described in 3.6.1, 3.6.2 and 3.7) were compared to each other based on: 1) ease of calculation and simplicity of image variables, 2) ease of interpretation of the final index, and 3) total amount of explanation of the field variables (adjusted R^2), the first two criteria being subjectively applied. The best models were selected to map predicted forest complexity over Chelsea, Québec.

3.9 Mapping Complexity

The best models were then applied by extracting the model image variables from the Quickbird imagery and the topographic variables at all pixels. The regression coefficients for the final model were applied to spatial layers representing each of the image and topographic variables that were included in the regression model. A continuous map of predicted complexity was created by multiplying the spatial layers by their respective model coefficients and summing all the results.

3.9.1 Model Validation

The RDA model was validated through a manual bootstrapping process. Initially, 13 plots (75%) were randomly withheld during model creation. The model was applied to the imagery, which resulted in a map of predicted forest complexity. For the additive index, a predicted complexity value was then recorded using the 13 data points which were not included in the modelling procedure. Using three error measures (mean error, mean absolute error and % RMSE), predicted complexity values were compared to actual values. Five iterations of this process were performed and an average value for each error measure was reported in order to provide validation of the continuous gradient of complexity.

4. Results

4.1 Field Variables

A large range of forest structures were recorded when the 70 field plots were established. Of the 70 plots, 15 (21%) were in low complexity areas (as qualitatively assessed in the field), 34 (49%) in medium complexity areas, and 21 (30%) in high complexity areas (see Qualitative Scores in Appendix 4). Figure 11 presents some of the typical forest complexities sampled. Low complexity plots were considered to be those within the following range of conditions: homogeneous closed overstorey canopy, none to little understorey vegetation, and none to little CWD (Figure 12). Medium complexity plots represented the average or most common structural conditions within this mixed temperate forest, with some overstorey variation, moderate amounts of CWD, and some understorey vegetation. Higher complexity plots were those which had more overstorey variation, with larger variation in tree height and larger and variable gap sizes. These plots also had considerably more snags, CWD and understorey vegetation (Figure 13). Hemispherical photographs taken in low complexity plots showed relatively little change in canopy openness at 1 m and 6 m heights, providing evidence of little understorey vegetation and a uniform overstorey. Conversely, hemispherical photographs taken in high complexity plots showed significant change in openness between the 1 m and 6 m heights, highlighting denser understorey and large canopy gaps (Figure 14). For understorey vegetation sampling, the transects were not changed to a four transect method as there was not the same potential for underestimating vegetation, as there was for CWD.

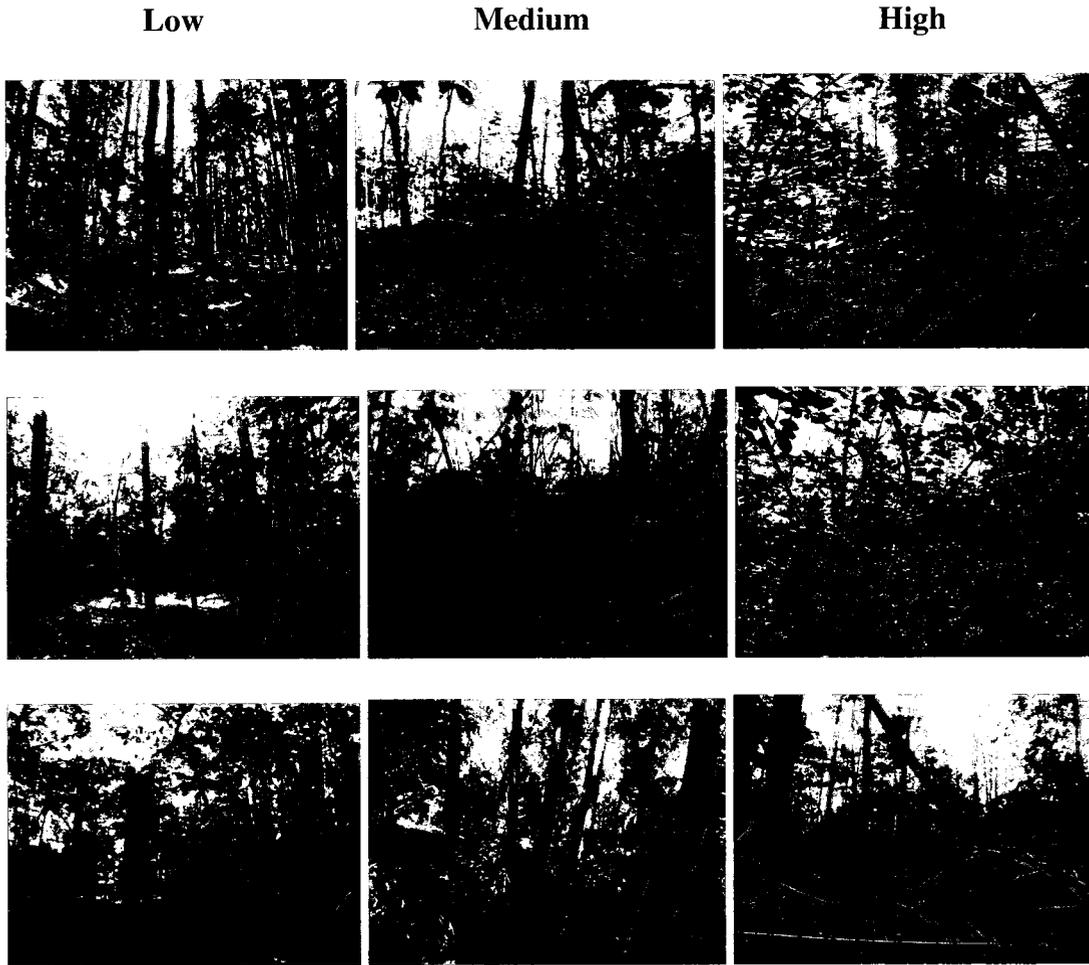


Figure 11. Complexity conditions typical of Chelsea, Québec. Complexity of the plots increase from left to right.

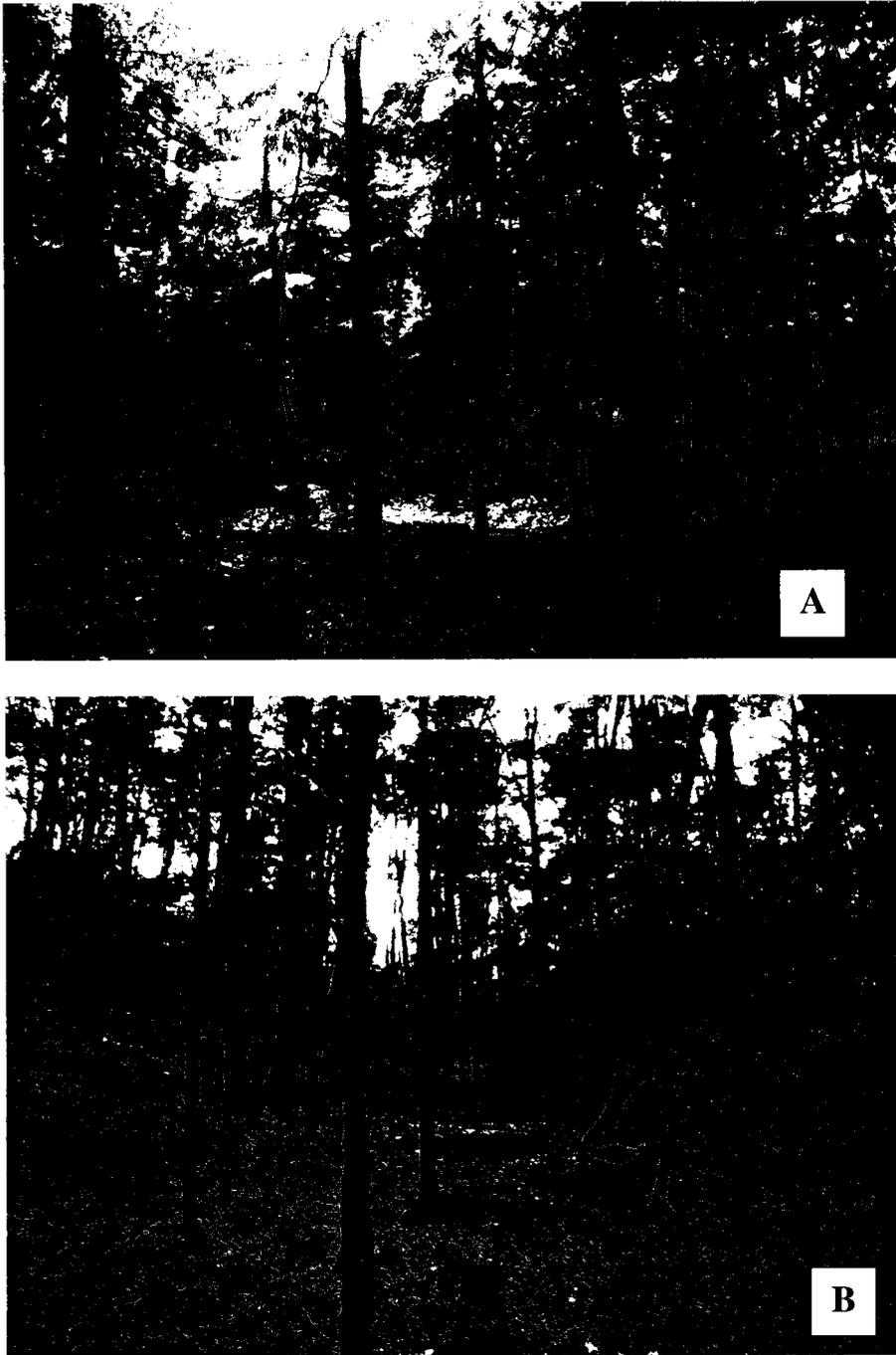


Figure 12. Structural conditions typical of an area of relatively low structural complexity (A - mixed (plot 35) and B - deciduous (plot 24)).



Figure 13. Structural conditions typical of an area of relatively high structural complexity (A – mixed (plot 49) and B – deciduous (plot 48)) within the study area.

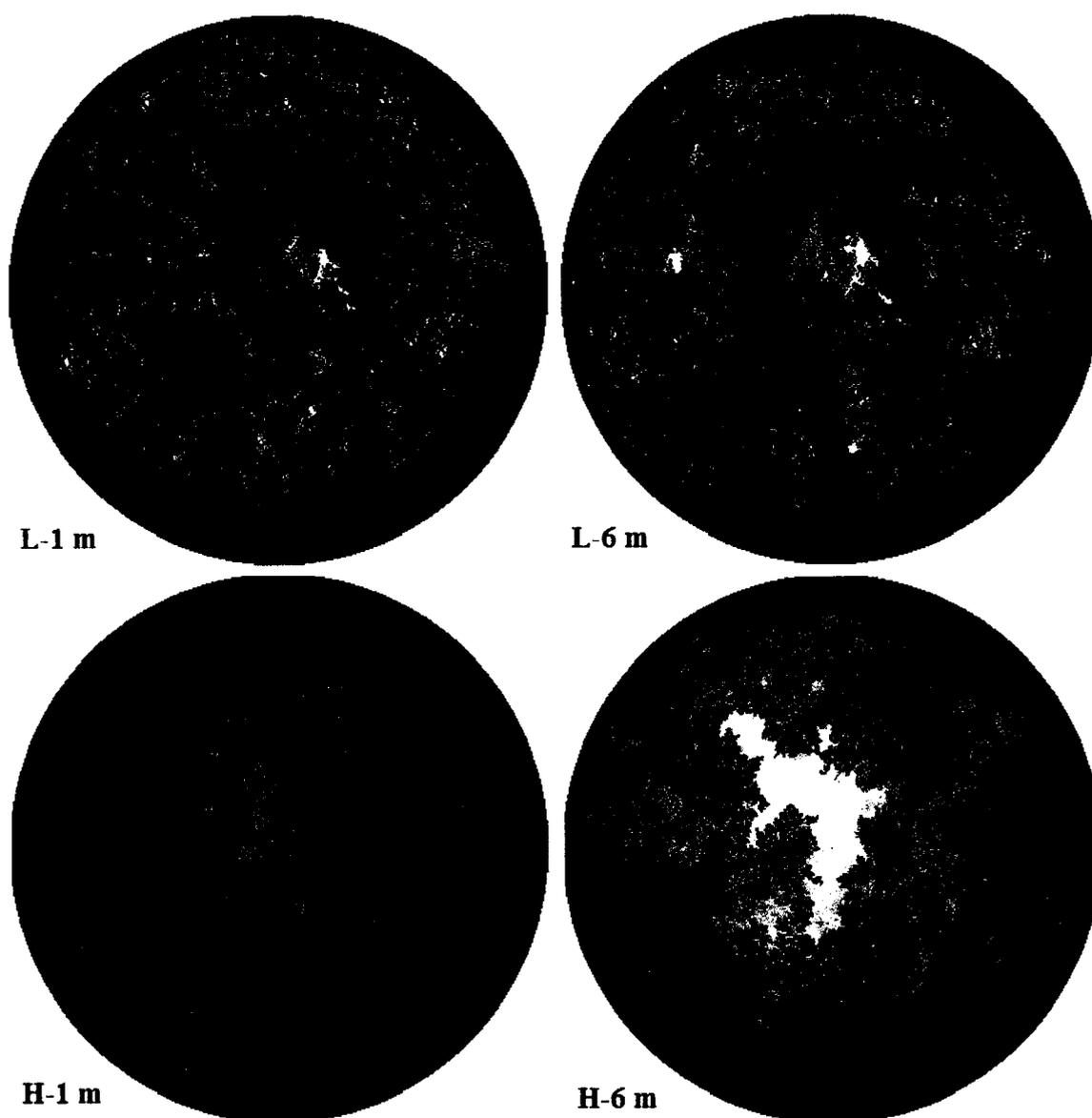


Figure 14. Example hemispherical photographs taken in a low complexity plot (L) as well as a high complexity plot (H) at both 1 m and 6 m heights.

A reduced dataset of 26 field variables, representing various aspects of forest complexity, were selected and calculated for each of the field plots. While a summary of the field data are provided in Appendix 2 and the raw plot data are provided in Appendix 3, a few interesting findings are presented below. Within the 70 plots, a total of 1837 trees with dbh > 10cm were sampled and measured. Plots ranged from having as few as

8 trees to as many as 61, the latter primarily in mixed plots with abundant hemlock. Generally, coniferous trees were less abundant and their dbh was significantly lower than deciduous trees. For example, over all plots, the percentage of coniferous live basal area was 24% compared to 76% for deciduous. Looking at the understory vegetation, the most dominant class was the 1 – 2 m (saplings and shrubs) height class with an average cover of 21%, followed by 10-50 cm (ferns and grass) with 15%, 0-10 cm (mosses and lichens) with 11% and 50 cm – 1 m (seedlings and other herbaceous plants) with 10.5% cover. In terms of relations between the variables, it was found that the nearest neighbour distance was negatively correlated with number of trees in the plot ($r = -0.65$, $p < 0.01$), as expected. LAI at 6 m was positively correlated with the coniferous basal area (%) ($r = 0.792$, $p = 0.01$), meaning that where plots had coniferous trees, LAI was higher, as expected.

4.2 Image Variables

Originally, 68 image variables were extracted from the Quickbird imagery and DEM. For those variable pairs highly correlated ($r > 0.8$) the most difficult to compute was removed from the study, resulting in a reduced dataset of 45 image variables. As the remaining image variables were not highly correlated, each was thought to provide some new or independent information to the analysis.

The crown and shadow fractions, derived through the ISODATA clustering, could not distinguish well between coniferous/mixed forest and shadow, even though with the high resolution of the panchromatic image it was thought that shadows and crowns would be distinguishable. The clustering either: 1) overestimated shadow by including

coniferous trees in the shadow class, or 2) underestimated shadow in deciduous forest but did not include coniferous trees in the shadow class. The output map selected tried to minimize these results (Figure 15). Large shadows and gaps in the canopy were generally well expressed, while small within crown shadows were sometimes missed.

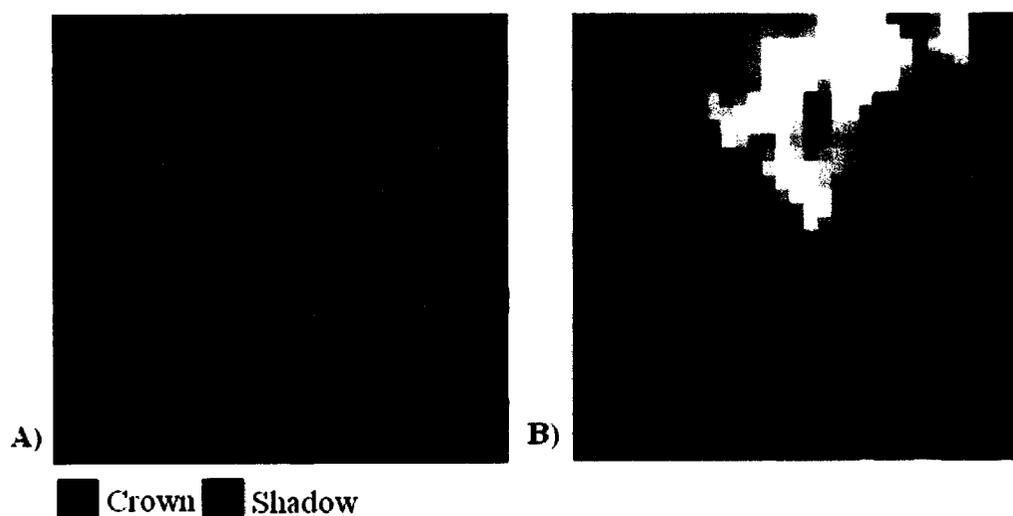


Figure 15. A) Map of shadow and crown, compared to B) panchromatic imagery from which they were derived.

4.3 Defining Forest Structure and Composition Indices from the Field Data

4.3.1 Principal Components Analysis

PCA of the field variables resulted in eight PCs which had eigenvalues greater than 1.0. This common method known as the Kaiser criterion, selects variables with an eigenvalue greater than 1, with the rationale that the others account for less variance than one original variable (Kaiser, 1960). It was used as a simple and objective means to select relevant PCs, although there are several other more complex methods. The resulting eight PCs cumulatively accounted for 75.7% of the variance (Table 5), illustrating the high dimensionality and information content in the field data. PCs 1-4 were the most

interpretable, and together accounted for 55.1% of the variance. Table 6 shows the factor loadings. From their magnitude and sign, the following was interpreted. PC1 increased with increasing openness and decreasing LAI in the canopy, fewer trees and lower basal area, associated with increasing amounts of understorey cover. PC2 increased with increasing tree size and with increasing amount of dead wood. PC3 increased with increasing coniferous basal area, increasing numbers of tree species and smaller tree spacing. PC4 increased with decreasing upper canopy openness and reduced understorey vegetation density. PCs 5-8 were not as easily interpreted. The stepwise regression of each PC against the image variables yielded eight separate models, all significant at $p \leq 0.05$. The models created using PC2 and PC3 had the greatest R^2 values. The regression equations for PC2 and PC3 were as follows:

$$PC2 = 3.277 + 38.957(\text{std NDVI}) + 3.443(\log \text{ mean } 5 \times 5 \text{ pan ASM}) + 0.290(\text{eastness})$$

$$PC3 = 14.608 - 21.684(\text{mean NDVI}) - 1.229(\log \text{ std } 5 \times 5 \text{ pan ASM}) - 0.358(\text{northness})$$

Table 5. Initial eigenvalues and relative percent variance of the first twelve PCs.

PC	Eigenvalue	% of Variance	Cumulative %
1	5.57	21.43	21.43
2	3.88	14.92	36.35
3	2.81	10.82	47.17
4	2.07	7.96	55.13
5	1.71	6.59	61.73
6	1.45	5.59	67.32
7	1.11	4.28	71.59
8	1.07	4.12	75.72
9	.91	3.49	79.22
10	.84	3.22	82.43
11	.78	3.02	85.45
12	.73	2.81	88.26

Table 6. Factor loadings for PCs 1 - 4.

Field Variable	Component			
	1	2	3	4
# Trees	-.669	-.427	.345	-.217
Mean tree dbh (cm)	-.198	.620	-.402	.327
Std. deviation tree dbh (cm)	-.204	.606	-.239	.225
Live basal area (m ² ha ⁻¹)	-.830	.275	-.051	.044
# Large trees	-.392	.602	-.295	.201
Std. deviation tree height (m)	-.084	.311	-.053	-.090
Mean NN distance (m)	.433	.318	-.593	.085
# Snags	.259	.576	.491	-.115
Mean snag dbh (cm)	.299	.563	-.171	-.129
Dead basal area (m ² ha ⁻¹)	.320	.770	.147	-.190
# Pieces CWD	.329	.386	.441	.045
Amount of rock (%)	.388	-.082	-.238	-.239
% Vegetation cover (<2 m)	.788	-.044	.311	.182
% Vegetation cover (0 – 10 cm)	-.022	-.212	.431	-.060
% Vegetation cover (10 – 50 cm)	.456	-.186	.263	.134
% Vegetation cover (50 cm – 1 m)	.726	-.055	.229	.303
% Vegetation cover (1 – 2 m)	.689	.269	-.055	.015
# Tree species	-.088	.164	.562	-.093
# Plant species	.534	.025	-.006	.007
Live coniferous basal area (%)	-.406	.504	.566	-.023
LAI ≥ 1 m (57.5°)	-.181	.351	.255	-.686
1 m openness (0 - 60°) (%)	.300	-.023	.419	.541
LAI ≥ 6 m (57.5°)	-.657	.436	.320	.014
6 m openness (0 - 60°) (%)	.604	.275	.161	-.057
Change LAI (55 - 60°) (%)	-.355	.189	.184	.677
Change Openness (0 - 60°) (%)	.395	.220	-.205	-.549

Table 7 shows the model variables and associated R² values. It also provides this information for the regression of the index created by weighted addition of PCs 1-8. In each model, NDVI or its standard deviation were the main variables entered, accounting for most of the variance in the model. Minor variables accounting for less than 8% of the

variance included Texture (GLCM ASM) and topographic aspect (northness or eastness). PC2 and PC3 produced stronger models individually than all eight PCs together in an index, demonstrating that the geo-spatial data are more related to some aspects of forest structure and composition than others. In PCs 5-8, the proportion of exposed rock, number of plant species at heights less than 2 m, and per cent ground vegetation cover (0-10 cm) are examples of variables with high factor loadings. These cannot be directly sensed through the image data but may be weakly related to the topographic data.

Table 7. Stepwise regression model variables and variance accounted for. Three indices of multivariate forest complexity as described in the text are shown in the first column.

Field Index	Predictor Variables	R ²	Adj. R ²
PC2	Std. Dev. NDVI; 5x5 Pan ASM; Eastness	0.41	0.38
PC3	Mean NDVI; 5x5 Pan ASM; Northness	0.54	0.52
Weighted Sum of PC 1-8	Std. Dev. NDVI; Northness; Eastness	0.36	0.33

4.3.2 Additive Index

The additive index adapted from McElhinny et al. (2006) was created by reducing the 26 field variables down to 11 core attributes: % vegetation cover (0 cm – 10 cm); % vegetation cover (50 cm – 1 m); LAI \geq 6 m (57.5°); # of plant species; # of tree species; live basal area; # of trees; # of large trees (>35 cm dbh); std. deviation tree height; dead basal area; and # of pieces of coarse woody debris. As a group, these variables best represent complexity as they are able to capture information on the horizontal and vertical aspects of the forest, as well as understory vegetation, dead wood and

composition. As well as representing important complexity attributes, these variables were also simple to calculate. Table 8 shows the correlations between the 11 core variables and the other attributes, which were used as criteria for selecting the core variables.

Table 8. Eleven core variables identified from an initial suite of 26. Only correlations in which Pearson's $r > 0.60$ (McElhinney et al. 2006) were considered.

#	Core attribute	Correlates (>0.60)
1	% Vegetation cover (0 – 10 cm)	None
2	% Vegetation cover (50 cm – 1 m)	% Vegetation cover (<2 m) (0.824)
3	# Plant species	None
4	# Tree species	None
5	Live basal area (m^2ha^{-1})	LAI \geq 6 m (57.5°) (0.625)
6	LAI \geq 6 m (57.5°)	Live basal area (0.625) % Basal area coniferous (0.792) % Basal area deciduous (-0.792)
7	# Trees	Mean NN distance (-0.645)
8	Dead basal area (m^2ha^{-1})	# Snags (0.756) Mean snag dbh (0.779)
9	# Large trees (>35 cm dbh)	Mean tree dbh (0.689)
10	Std. deviation tree height	None
11	# Pieces CWD	% Debris (0.766) Volume CWD (0.780)

Results of the preliminary analysis indicated that the means (Table 9) and standard deviations of complexity attributes of the 70 plots represented a broad range of forest structure and conditions within each community (mixed and deciduous). Both types of plots, mixed and deciduous, were combined in the same data set in this study in order to introduce the composition component to the complexity index by including coniferous trees. Simple t-tests showed no significant difference between the two groups.

Table 9. Mean and standard deviation values for the core attributes in mixed forest and deciduous forest communities.

Complexity Attributes	Mixed forest		Deciduous forest	
	Mean	Std. deviation	Mean	Std. deviation
% Vegetation cover (0 – 10 cm)	11.58	7.82	10.03	6.41
% Vegetation cover (50 cm – 1 m)	0.78	0.42	1	0.37
# Plant species	4.82	1.26	5.06	1.10
# Tree species	5.43	1.73	3.51	1.81
Live basal area (m ² ha ⁻¹)	32.55	13.73	24.01	7.94
LAI _{≥ 6 m} (57.5°)	7.42	2.32	4.88	1.11
# Trees	28.17	11.79	23.51	8.34
Dead basal area (m ² ha ⁻¹)	0.73	0.38	0.46	0.34
# Large trees (>35 cm dbh)	3.02	2.10	1.69	1.81
Std. deviation tree height	3.83	1.17	3.29	0.90
# Pieces CWD	15.09	9.99	13.13	9.35

PCA of the standardized core attributes resulted in four PCs which had eigenvalues greater than 1.0. The four PCs cumulatively accounted for 68% of the variance, and the first two components explained 44% of the variation between study sites. The vectors, representing the size and direction of the 11 attributes defined by the first four PCs are shown in Figure 16. Vectors, representing the core attributes, were used to demonstrate the relative importance of each attribute in explaining variation between sites, and the degree of correlation between attributes. The angle between vectors shows the correlation between variables (smaller angle being more correlated) and the length of the vectors demonstrates the importance of each variable in explaining variation between data points. The best arrangement of the vectors would be if attributes were of equal importance in explaining variation between sites (same length) and correlations between variables were minimized (larger angles) (McElhinny et al., 2006). All vectors were similar in length with the exception of attribute 4 (number of tree

species) and attribute 10 (standard deviation of tree height (m)), which demonstrated that the attributes generally had a similar weighting in mapping the location of study sites in the 2D space. Relations between the selected field variables were not considered to be too strong to justify reducing the number of core attributes further. The eleven core attributes were therefore accepted as an optimum set of complexity attributes.

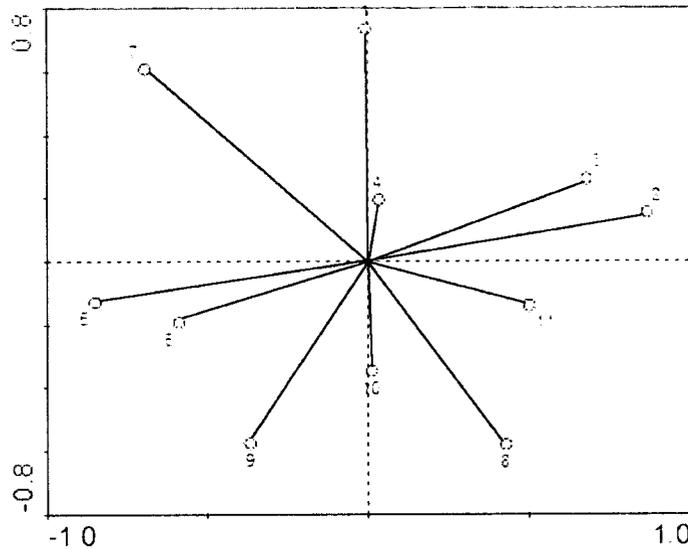


Figure 16. The first two principal components (PC1-2) explained 44% of the total variation between sites, mapped by the 11 core attributes. The vectors represent the relative importance of each variable in explaining variation between data points as well as the correlation between variables. (Refer to Table 8 for attribute number.)

The eleven attributes were rescaled using two different methods. The first method, following the steps outlined by McElhinny et al. (2006), resulted in the generation of eleven regression equations, which fit a linear relation between the attribute score (0-10) and the quartile midpoints of the attribute distributions. Scores of 2.5, 5, 7.5, and 10 were allocated respectively to the quartile midpoints 12.5, 37.5, 62.5 and 87.5 (Table 10). The maximum possible score was 10, set by the 87.5 percentile. To prevent

extreme values from distorting the scaling of the scores, the data were manually constrained between 0 and 10. The sets of equations used are shown in Table 11.

Table 10. Quartile midpoints (12.5, 37.5, 62.5, 87.5) for each of the eleven core attributes and associated score (2.5, 5, 7.5, 10).

Core attribute	Quartile mid points			
	12.5 (2.5)	37.5 (5)	62.5 (7.5)	87.5 (10)
% Vegetation cover (0 – 10 cm)	3.57	7.70	12.83	18.75
% Vegetation cover (50 cm – 1 m)	0.27	0.74	1.08	1.30
# Plant species	4	5	5	6
# Tree species	2	4	5	7
Live basal area (m ² ha ⁻¹)	16.27	24.46	30.02	46.48
LAI \geq 6 m (57.5°)	4.37	5.34	6.68	9.04
# Trees	15.00	22.00	28.38	37.25
Dead basal area (m ² ha ⁻¹)	0.17	0.43	0.74	1.19
# Large trees (>35 cm dbh)	0.0	1	3	5
Std. deviation tree height	2.56	3.08	3.70	5.02
# Pieces coarse woody debris	4	9.63	15	24

The second method consisted of a simple linear stretch of the attribute scores. Both the linear stretched attribute scores and the McElhinny method scores, on a scale from 0-10, were highly correlated with the original attribute data, indicating there was only a little information lost in the rescaling process. As well, the linear stretch scores and the McElhinny scores were highly correlated (Table 12) and significant ($p \leq 0.01$). The linear stretch was selected as the rescaling method to be used in the additive index, as it was found to be equivalent to the McElhinny et al. (2006) method, and avoided data loss when an attribute had a score greater than 10. While McElhinny et al. (2006) argue that high values may distort the scaling of a score, it was important to retain those high values as they represented important information on extremely high or low complexity plots.

Table 11. Regression equations for the eleven attributes for scoring on a scale of 0-10.

Core attribute	Regression equations	R ²
% Vegetation cover (0 – 10 cm)	$y = 0.4962x + 0.9335$	0.9924
% Vegetation cover (50 cm – 1 m)	$y = 7.2103x + 0.1592$	0.9767
# Plant species	$y = 3.75x - 12.5$	0.9
# Tree species	$y = 1.5385x - 0.6731$	0.9846
Live basal area (m ² ha ⁻¹)	$y = 0.2478x - 1.0346$	0.9416
LAI \geq 6 m (57.5°)	$y = 1.6152x - 3.9862$	0.9646
# Trees	$y = 0.3448x - 2.5431$	0.9931
Dead basal area (m ² ha ⁻¹)	$y = 7.2813x + 1.6343$	0.9829
# Large trees (>35 cm dbh)	$y = 1.4407x + 3.0085$	0.9797
Std. deviation tree height	$y = 2.9921x - 4.4798$	0.9404
# Pieces coarse woody debris	$y = 0.3783x + 1.2369$	0.9837

Table 12. Correlation of attribute scores from the linear stretch with the attribute scores from the McElhinny rescaling method. (Highly significant (p<0.01)).

Core attribute	Pearson's r
% Vegetation cover (0 – 10 cm)	0.912
% Vegetation cover (50 cm – 1 m)	0.979
# Plant species	0.954
# Tree species	0.977
Live basal area (m ² ha ⁻¹)	0.963
LAI \geq 6 m (57.5°)	0.942
# Trees	0.921
Dead basal area (m ² ha ⁻¹)	0.990
# Large trees (>35 cm dbh)	0.971
Std. deviation tree height	0.956
# Pieces coarse woody debris	0.916

The rescaled attributes were then added up to create the index. Each attribute accounted for either +10 or -10 (LAI; live basal area; number of trees) points and was combined together to create a total index value of 50 points. An index value for the ten variable model (excluding LAI \geq 6 m) was also calculated, as LAI was still correlated ($r = 0.625$) with live basal area. Summaries of the stepwise regression models for both the ten and eleven core attributes are given in Table 13. They included similar predictor variables to the PC models but with additional minor variables. All regression models did not suffer from multicollinearity, having a VIF at each model step significantly lower than the cut-off threshold of 5.3 (Hair et al., 1998).

Table 13. Stepwise regression model variables and variance accounted for. Two indices of multivariate forest complexity as described in the text are shown in the first column.

Field Index	Equations	R ²	Adj. R ²
11 variable Field Index (adapted from McElhinny et al. (2006))	-16.025 + 2.549(std. deviation blue) + 73.452(mean 5x5 PAN Correlation) + 5.308(mean slope) + 183.413 (std. deviation NDVI)	0.44	0.41
10 variable Field Index (adapted from McElhinny et al. (2006))	-58.637 + 351.809(std. deviation NDVI) + 2.204(eastness) + 123.367 (std. deviation 5x5 pan Homogeneity) + 28.441(mean 3x3 nir Entropy) + 4.082 (mean slope).	0.53	0.49

4.4 Redundancy Analysis Complexity Modelling

4.4.1 Model Selection

The 1000 models produced during the first stage of the RDA bootstrapping consisted of different combinations of image variables (Table 14). Spectral information

was included in 98% of the models, with the mean plot NDVI included in 94% of all models. Topographic information was included in 27% of all models with the main variable, log mean slope, included in 24% of all models. First order spatial measures were also included in 27% of all models, with the standard deviation of the blue band being the most often included (22% of all models). GLCM textures and shadow and crown fractions were included in 23% and 16% of the models, respectively. During the second stage, models were examined and validated. Models had adjusted R^2 values that ranged between 0.186 and 0.367 (mean = 0.233) and contained between one and twelve variables (Table 15). The % RMSE values were fairly high and ranged between 41% and 54%, while the AIC ranged between 546.13 and 556.08. All models were highly significant at $p < 0.001$.

Table 14. Number of times an image variable type was included in one of the 1000 RDA models, displayed as a percentage.

Type of variable	% of all models
Spectral	98%
Topographic	27%
1st order spatial	27%
Texture	23%
Crown and shadow fractions	16%

Table 15. Summary statistics for the 1000 models provided through stage two bootstrapping.

Model	Adjusted R^2	R validate	Number of variables	% RMSE	AIC
Minimum	0.19	0.61	1	40.89	546.13
Maximum	0.37	0.78	12	54.64	556.08
Average	0.23	0.75	2.283	48.05	548.29

The highest ranked model, ranked by highest adjusted R^2 and lowest % RMSE, had an adjusted R^2 value of 0.37 and was comprised of ten variables. This model included all five types of variables: spectral, textural (1st and 2nd order), radiometric fractions and topographic. Though it was the best model, it was not selected as the final model due to the fact that it included too many predictor variables. This decision was based on a common rule of thumb, the 10:1 ratio, (Babyak, 2004) which recommends a maximum of 1 predictor variable for every 10 observations (field plots). With 10 variables entered, and 53 plots, the ratio for these models was 5.3:1, which was considered too low. Using the 10:1 ratio, only models with seven or fewer variables were evaluated and compared.

The best model with seven variables (3rd best overall; adjusted $R^2 = 0.325$) was examined but it was not selected for further analysis as the VIF for two of the variables, GLCM ASM and Entropy was very high (75 and 78, respectively) and these two variables were highly correlated ($r = -0.992$). As a result, any model which incorporated both GLCM ASM and Entropy textures was not analyzed further. Models which incorporated either crown or shadow fraction and average NDVI were examined with caution, as both fraction types were highly correlated with NDVI ($r = 0.882$ and $r = -0.771$, respectively). However, the VIFs produced for these variables were significantly lower (VIF = 2-3), and still within the acceptable range of <5.3 (Hair et al. 1998).

Unfortunately, the above two conditions occurred very frequently in the top ranked models making it difficult to select an appropriate model. To avoid this issue, the seven variable model (above) was examined and manipulated by removing Entropy from the suite of image variables. ASM was not removed because it appeared in significantly

more models (95 vs. 33) than Entropy. An RDA was the re-run using these six variables, resulting in a model with an adjusted R^2 of 0.32 (from the first two axes), which was a higher adjusted R^2 than any of the six variable models produced through the bootstrapping process. The model's first RDA axis (RDA1) had an adjusted R^2 of 0.229 ($p = 0.001$) and the second RDA axis (RDA2) had an adjusted R^2 of 0.091 ($p = 0.009$) and was found to be significant, by testing RDA1 as a covariable. The axes were linearly weighted by their explained variance and additively combined to produce the forest complexity index.

4.4.2 Analysis of the Selected Six Variable Model

By examining the sign of the significant correlations in Table 16 and the triplot of Figure 17 for RDA1 and RDA2, it is possible to describe them in terms of a gradient of complexity. For RDA1, it can be seen that an increase in coniferous basal area, amount of deadwood (snags and CWD), number of tree species and LAI at 1 and 6 m as well as a decrease in the amount of exposed rock corresponded to an increase in the index. The increase in LAI at 6 m was not expected as Pasher and King (2010) found that it decreased with increasing forest complexity as the canopy opens. However, the fact that the proportion of coniferous trees and an increased number of species increases the complexity as represented by RDA1 shows that RDA1 is more of a composition—structure complexity gradient than the structure only gradient in Pasher and King (2010) where conifers were not present. Thus, increased proportions of coniferous trees are associated with greater LAI whether at 1 m or 6 m as many of them were below the upper canopy but taller than 1 m. Fifteen of the 24 field variables were found to be

uncorrelated with RDA1. For RDA2, it can be seen that a decrease in understorey vegetation, variation in tree height, the amount of exposed rock, and CWD, accompanied by an increase in number of trees and standing live basal area corresponded to an increase in the index. It has no significant relation with composition but represents increased overstorey trees and volume, which is generally accompanied by reduced understorey. Plots 2, 7, 26, 50 and 51 are the farthest out on this axis and all have very low variation in dbh and very little understorey vegetation, demonstrating that these are indeed lower complexity plots. This may also be illustrated by the lack of inclusion of LAI as a significant variable even though live basal area was significant. Thus, increasing RDA2 represents decreasing forest complexity. Twelve of the 24 variables were found to be uncorrelated with RDA2.

After examining the biplot scores of the image variables included in the model (Table 17), it was determined from RDA1 that forest complexity increased as: 1) mean NDVI decreased, 2) percent shadow increased, 3) mean 9x9 NIR band GLCM Correlation increased, 4) mean 3x3 NIR band GLCM ASM increased, 5) the variation in 5x5 Panchromatic GLCM Homogeneity increased, and 6) slope decreased. This shows (for the first three predictor variables) that as the canopy becomes more open, with more shadow and more image texture, forest complexity increases. It must be noted that such a relation would only hold for mature canopies with an overstorey and an openness range similar to that of this study (1 - 57%). If openness continues to increase, there would be reduced overstorey, spatial heterogeneity, image texture, and shadow fraction. The three other predictor variables in the model were very weakly correlated with RDA1 but all had higher contributions towards the second axis, RDA2. Both slope and GLCM

Homogeneity variation decreased with decreasing forest complexity (increasing RDA2) and mean GLCM ASM increased with decreasing complexity. This demonstrates that as the terrain becomes steeper, forest complexity increases, image texture increases (GLCM ASM is lower), and texture variation increases (increased variation in GLCM Homogeneity). ASM is a measure of homogeneity of the plot, and therefore it was expected to decrease as complexity increased. Slope changed direction of effect from the weak relation with RDA1 (lower slopes associated with greater complexity) to the strong relation with RDA2 (higher slopes associated with greater complexity). The reason for this is currently not known. The number of times each of the six image variables were included in other models is expressed in Table 18 as a percentage of all models.

Table 16. Bivariate correlations between RDA1 and the individual field variables and bivariate correlations between RDA2 and the individual field variables.

	RDA1		RDA2	
Field Variables	Pearson r	p	Pearson r	p
Live coniferous basal area (%)	0.76	< 0.001	0.06	0.604
# Snags	0.47	< 0.001	-0.19	0.120
LAI \geq 6 m (57.5°)	0.46	< 0.001	0.21	0.084
LAI \geq 1 m (57.5°)	0.41	< 0.001	0.20	0.099
# Tree species	0.41	< 0.001	-0.19	0.121
# Pieces of CWD	0.32	0.006	-0.26	0.030
Dead basal area (m ² ha ⁻¹)	0.30	0.011	-0.24	0.460
% Vegetation cover (0 – 10cm)	0.27	0.023	0.02	0.898
6 m openness (0 - 60°) (%)	0.26	0.027	-0.06	0.606
Amount of rock (%)	-0.23	0.052	-0.26	0.029
1 m openness (0 - 60°) (%)	0.19	0.110	-0.08	0.504
Mean NN distance (m)	-0.18	0.130	-0.10	0.402
Live basal area (m ² ha ⁻¹)	0.16	0.190	0.30	0.012
% Vegetation cover (>2 m)	0.14	0.230	-0.45	< 0.001
% Vegetation cover (10 – 50 cm)	0.11	0.370	-0.22	0.068
# Large trees	0.09	0.470	0.01	0.933
# Trees	0.09	0.470	0.31	0.008
Std. deviation tree height (m)	0.06	0.600	-0.23	0.052
Mean snag dbh (cm)	0.06	0.629	-0.19	0.106
Mean tree dbh (cm)	0.04	0.720	0.04	0.754
# Plant species	-0.02	0.847	-0.18	0.132
Std. deviation tree dbh (cm)	-0.01	0.950	-0.28	0.019
% Vegetation cover (50 cm – 1 m)	0.00	0.980	-0.39	< 0.001
% Vegetation cover (1 – 2 m)	0.00	0.990	-0.40	< 0.001

Table 17. Biplot scores of RDA1 and RDA2. The scores show the relative contribution of each image variable towards each RDA axis.

Image Variable	Biplot Score	
	RDA1	RDA2
Mean NDVI	-0.98	0.13
% Shadow in the plot	0.78	-0.02
Mean GLCM ASM (NIR band; 3x3 window)	0.17	0.41
Mean GLCM Correlation (NIR band; 9x9 window)	0.35	-0.08
Variation in GLCM Homogeneity (Pan imagery; 5x5 window)	0.16	-0.34
Slope	-0.14	-0.59

Table 18. Number of times each of the predictor variables from the 6 variable model was included in the 1000 RDA models, displayed as a percentage.

Type of variable	% of all models
Mean NDVI	98%
% Shadow	10%
Mean GLCM Correlation (9x9 NIR)	4%
Mean GLCM ASM (3x3 NIR)	10%
Std. deviation GLCM Homogeneity (5x5 pan)	3%
Slope	24%

Through the RDA analysis in CANOCO, a triplot was generated to visually explore the relations between the image variables, field variables, and RDA axes (Figure 17). The red arrows on the plot represent the six image variables and their relation to each axis, determined by their length and direction of the arrows. The blue arrows represent the field variables, and their relations to the image variables can be determined based on their direction. Correlations between a pair of variables are higher if they point in similar directions, while variables which are at large angles to each other or perpendicular are uncorrelated. The field plots are also displayed, which allows them to be assessed in terms of their structure and composition.

The triplot allowed the plots to be visually examined to determine if they were real combinations of the conditions interpreted for RDA 1 and 2. For example, plots 62 and 65 (high complexity) are found on the opposite sides of both axes from plots 27 and 37 (low complexity). Plots 62 and 65 had significantly larger amounts of snags and CWD, resulting in a higher complexity. As well, the percent of shadow in these two plots is much larger than in plots 27 and 37. In plots 27 and 56, the sites are rocky and composed of 100% deciduous trees, thus placing them opposite from plot 9, which had no rock and was composed of 50% coniferous trees. The percent of shadow and the GLCM Correlation were also high in plot 9 putting it opposite plots 27 and 37.

It was noted that plot 29 appeared to be an outlier, as it was located at the far top right of the triplot. The statistics for this plot were examined and it was determined that it had the highest ASM value (very low texture) of all the plots and a low variation in Homogeneity within the plot. A visual assessment of this plot (Figure 18) determined that it was not an outlier but an extreme case of low complexity in Chelsea on RDA2. On RDA1, this plot appears to be high complexity due perhaps to the very high percent of coniferous basal area. It should be noted that a very high percentage of coniferous basal area and few tree species represents low complexity as plots become very uniform in nature. Thus, this approach of including coniferous trees in the complexity index is more suitable to natural forests of mixed composition and structure and not to forests of very uniform structure and coniferous composition such as this plantation. The panchromatic imagery in Figure 18 A) clearly shows a uniform patch of coniferous trees, which was a red pine plantation. Figure 18 B) shows the red, green and blue bands in natural colour and Figure 18 C) shows the NIR band. The ASM texture was extracted from the NIR

band, which again has very low texture in this uniform patch of coniferous trees. The attributes for plot 29, as related to RDA1 and 2 are listed in Table 19.

Table 19. Field variables for plot 29.

% Coniferous	87
# Snags	6
# Pieces coarse woody debris	5
Std. deviation tree height	4.7
# Trees	35
LAI \geq 6 m (57.5°)	13
LAI \geq 1 m (57.5°)	12
Live basal area	65
# Tree species	3
% Rock	0
% Vegetation cover (< 2 m)	15

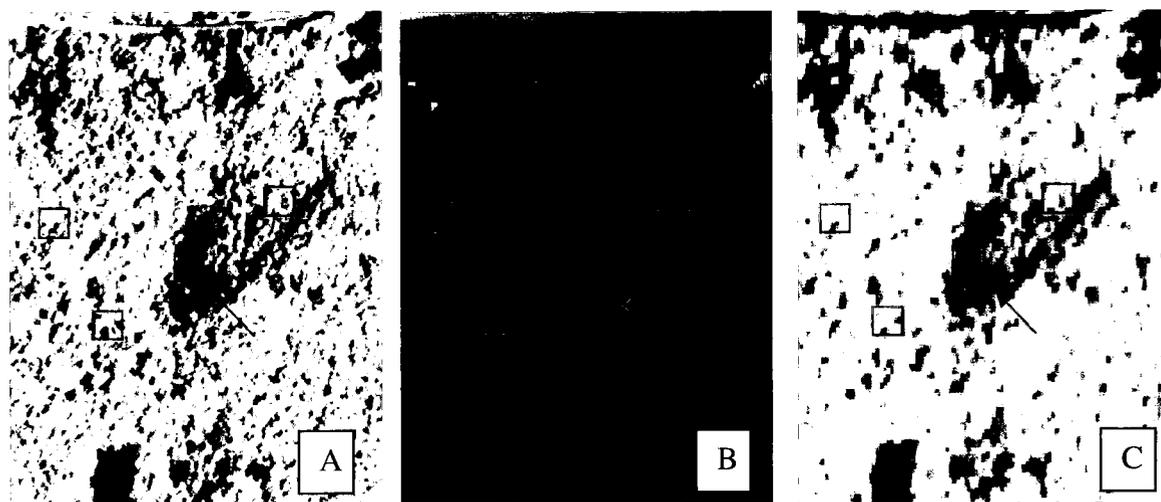


Figure 18. Panchromatic (A), multispectral (B) and NIR imagery (C) used to examine the forest patch of plot 29 (at arrow).

The complexity index (represented by RDA 1 and 2) was created by linearly combining the regression/canonical coefficients for the standardized image variables and

is presented below. These were obtained through the output provided by CANOCO 4.5 (ter Braak and Šmilauer, 2002).

$$\begin{aligned} &\text{Equation of first RDA axis (eigenvalue = 0.229)} \\ \text{Index} &= -0.88(\text{mean NDVI}) + 0.10(\% \text{ shadow}) + 0.13(\text{mean GLCM Correlation}) + \quad (1) \\ &0.16(\text{mean GLCM ASM}) - 0.13(\text{std. deviation GLCM Homogeneity}) - 0.02(\text{slope}) \end{aligned}$$

$$\begin{aligned} &\text{Equation of second RDA axis (eigenvalue = 0.091)} \\ \text{Index} &= 0.51(\text{mean NDVI}) + 0.35(\% \text{ shadow}) + 0.31(\text{mean GLCM Correlation}) + \quad (2) \\ &0.75(\text{mean GLCM ASM}) - 0.68(\text{std. deviation GLCM Homogeneity}) - 0.73(\text{slope}) \end{aligned}$$

4.4.3 Analysis of a Simpler Four Variable Model

While a six variable model may have produced interesting results, a model with fewer and simpler variables is often desired for practical use and implementation in a forest management setting. Some variables are difficult to compute and are not always ideal to use. For example, it was difficult to accurately produce crown and shadow fractions due to the problems previously discussed concerning inclusion of coniferous species in shadow fractions. By examining the best four variable models, it was possible to select a model which was comparable to the six variable model, but which did not include shadow fractions.

This model's first axis (RDA1) had an adjusted R^2 of 0.225 ($p = 0.001$) and the second axis (RDA2) had an adjusted R^2 of 0.077 ($p = 0.004$) and was found to be significant, again by testing RDA1 as a covariate. The four variables in this model included: 1) mean NDVI, 2) variation in the blue band, 3) average 3x3 NIR band GLCM ASM, and 4) slope. Variables 2 (increasing variation) and 3 (increasing homogeneity) both represent spatial variation, and are showing contrasting effects. Both variables are minor on RDA1 in relation to mean NDVI (Table 20), and the reasons why opposite effects are being shown or why the blue band variation was selected (in 22% of 4 variable

models) are not understood at this time. The variables in this model were identical to those in the six variable model, with the exception of variation in the blue band. The explanation difference between the two models for RDA1 was only 0.4%, while the difference between RDA2 was 1.4%, for a total difference of 1.8% of the data variance.

The biplot scores for this model (Table 20) and the bivariate correlations between RDA1 and the individual field variables were examined and it was determined that forest complexity increased as: 1) mean NDVI decreased, 2) variation in the blue band increased, 3) average 3x3 NIR band GLCM ASM increased, and 4) slope decreased. Like the six variable model, mean NDVI made the greatest contribution (-0.99) to RDA1 and both slope (-0.63) and variation in the blue band (-0.63) made the greatest contributions to RDA2. Again for RDA1, complexity increased with an increase in coniferous area, amount of deadwood (snags and CWD), number of tree species and LAI at 1 and 6 m as well as a decrease in the amount of exposed rock.

Table 20. Biplot scores of RDA1 and RDA2 for the four variable model. The scores show the relative contribution of each image variable towards each RDA axis.

Image Variable	Biplot Score	
	RDA1	RDA2
Mean NDVI	-0.99	0.04
Variation in blue band	0.33	-0.63
Average GLCM ASM (NIR band; 3x3 window)	0.16	0.48
Slope	-0.11	-0.63

Again, the model equations were created through the linear combination of the regression/canonical coefficients, which were obtained through the output provided by CANOCO 4.5 (ter Braak and Šmilauer, 2002).

$$\text{Equation only using the first RDA axis (eigenvalue = 0.225) = -1.0(\text{mean NDVI}) - 0.03(\text{variation in blue band}) + 0.09(\text{average GLCM ASM}) + 0.03(\text{slope}) \quad (3)$$

$$\text{Equation using only second RDA axis (eigenvalue = 0.077) = -0.04(\text{mean NDVI}) - 0.58(\text{variation in blue band}) + 0.46(\text{average GLCM ASM}) - 0.67(\text{slope}) \quad (4)$$

4.5 Model Comparison and Selection

The seven models selected from the three approaches are listed in Table 21 as a summary of the outcome of the modelling tests. They were compared based on three features: 1) simplicity of image variables entered into the model, 2) ease of interpretation of the final index, and 3) total amount of explanation of the field variables (adjusted R^2). Table 22 presents the image variables that were entered into the models. Since it is difficult to compare the models based on individual image variables, they will be compared by the type of image variable entered into the model (Table 23). Topographic variables were the most common, and were entered into every single model. They were followed closely by GLCM textures, entered into six of the seven models, and first order textures, entered into five of the seven models. Of the GLCM textures, ASM was entered the most often, appearing in 4 of the 7 models. Spectral variables and fractions were the least common type of variable to be entered, although either mean NDVI or std. deviation of NDVI were in all of the models ($r = -0.76$). The calculation of image fractions was a subjective process as users are required to decide which groups should be allocated to the shadow or crown class, based on ISODATA clustering. For this reason, the calculation of fractions by different users would most likely yield different results, and is therefore not easily repeatable. As well, this variable was one of the most difficult to calculate, as previously discussed. Since fractions were the least used in model creation and one of the more difficult variables to calculate, any model, like the 6 predictor variable RDA

model, which used fractions was considered to be less desirable. Any models which had all three of the more common variables (topographic and 1st and 2nd order textures) were considered to be more desirable as they incorporated variables that were more straightforward to calculate and variables which were more often used for explaining variance in the field data. These models included PC2, the 11 and 10 variable additive indices and the 4-predictor variable RDA index.

Table 21. Summary of the final seven indices created through three separate methods.

	Index	Model Summary
PCA	PC2	Increased with increasing tree size and with increasing amount of dead wood.
	PC3	Increased with increasing coniferous basal area, increasing numbers of tree species and smaller tree spacing.
	Weighted Sum of PC 1-8	Un-interpretable
Additive	11 variable Field Index (adapted from McElhinny et al. (2006))	Increased with increasing understorey vegetation cover, # plant species, # tree species, # of large trees, std. deviation tree height, dead basal area and # pieces coarse woody debris, and decreasing LAI ≥ 6 m (57.5°), live basal area and # trees.
	10 variable Field Index (adapted from McElhinny et al. (2006))	Increased with increasing understorey vegetation cover, # plant species, # tree species, # of large trees, std. deviation tree height, dead basal area, and # pieces coarse woody debris and decreasing live basal area and # trees.
RDA	4 and 6 variable RDA1	Increased with increasing in coniferous basal area, amount of deadwood (snags and CWD), number of tree species and LAI at 1 and 6 m and decreasing amount of exposed rock.
	4 and 6 variable RDA2	Increased (ie. decreasing complexity) with decreasing understorey vegetation, variation in tree height, the amount of exposed rock, and CWD, and increasing number of trees and standing live basal area.

Table 22. Final seven indices created through three separate methods with image variables and variance accounted for. ((-) indicates predictor variable decreased with increasing index value.)

Index	Predictor Variables	Adj.R²	# Variables
PC2	Std. Dev. NDVI; 5x5 Pan ASM; Eastness	0.38	3
PC3	(-) Mean NDVI; (-) 5x5 Pan ASM; (-) Northness	0.52	3
Weighted Sum of PC 1-8	Std. Dev. NDVI; (-)Northness; Eastness	0.33	3
11 variable Field Index (adapted from McElhinny et al. (2006))	Std. Dev. Blue; Mean 5x5 Pan Correlation; Mean slope; Std. Dev. NDVI	0.41	4
10 variable Field Index (adapted from McElhinny et al. (2006))	Std. Dev. NDVI; Eastness; Std. Dev. 5x5 Pan Homogeneity; Mean 3x3 NIR Entropy; Mean slope.	0.49	5
6 var RDA1	(-) Mean NDVI, Shadow fraction, Mean 3x3 NIR ASM; Mean 9x9 NIR Correlation; Std. Dev. 5x5 Pan Homogeneity; Mean slope	0.32	6
4 var RDA1	Std. Dev. Blue; (-) Mean NDVI, (-) Mean 3x3 NIR ASM; (-) Mean slope	0.30	4

Table 23. Number of models in which a particular type of image variable appeared.

	Spectral	1st order texture	2nd order texture	Crown/ Shadow Fractions	Topographic
PC2		X	X		X
PC3	X		X		X
PC1-8 weight		X			X
11 var additive		X	X		X
10 var additive		X	X		X
6 var RDA	X		X	X	X
4 var RDA	X	X	X		X
Sum	3	5	6	1	7

The second factor on which the models were compared was ease of interpretation of the final index. Models that had fewer predictor variables were considered to be easier to interpret. As well, simple models are more desirable for practical use and implementation. The number of predictor variables in each model (Table 22) varied between 3 and 6. Though this factor was important, it was given less consideration than the other two factors. Each PC model contained only 3 variables, though each also accounted for the smallest amount of total variance explained (refer to Table 24).

The final factor for comparison was the total amount of explanation of the field variable variance (adjusted R^2). Looking only at the adjusted R^2 value in Table 22, it would appear that PC3 was the highest ranked model, followed by the 10 and 11 variable additive indices, PC2, the weighted sum of PC1-8, the 6-predictor variable RDA index and the 4-variable RDA index. This is not necessarily the case, as each PC originally

only accounted for a portion of the variance in the 26 variable field dataset. By examining column two in Table 24, it can be seen that PC2 only accounted for 14.92% of the original variance in the field data, PC3 for 10.82% and PC1-8 for 75.72%. The adjusted R^2 value, derived from linear regression therefore describes how much variance in the PC the image variables were able to explain. For example, the image variables were able to explain 38% of the variance in PC2, which originally only explained 14.92% of the field data. This example demonstrates that the PC2 model only explained 5.67% (i.e., $14.92\% \times 0.38$) of the total field data variance. It is possible that this may represent all of the variance that is related to complexity, but other PCs such as PC3 and PC4 also represented gradients that were interpretable as aspects of complexity. By examining column four (Table 24), it is clear the PC models are not really as strong as first suspected, even when the first eight PCs were combined in a single model. The same holds true for the additive indices as they both only incorporate either 10 or 11 field variables out of the suite of 26 that were not inter-correlated ($r \leq 0.8$). For the RDA models, all field 26 variables were incorporated into the modelling process, therefore 100% of the variance is included. If this research had only collected the 10 core variables, then this index would have had the highest explained variance at 49%. The RDA models therefore are better able to explain the variance in the whole field dataset than the PC or additive models. Due to the nature of the RDA analysis, the adjusted R^2 value reports the total variance explained for each of these methods because each index used all the field variables input into the model.

Table 24. Total explained variance of the 26 variable field dataset by the best indices from each of the three index formulation methods.

Index	% variance of whole dataset	Adj.R ²	Total variance explained by the model
PC2	14.92%	0.38	5.67%
PC3	10.82%	0.52	5.63%
Weighted Sum PC 1-8	75.72%	0.33	24.99%
11 variable Additive Index	42.5%	0.41	17.43%
10 variable Additive Index	38.5%	0.49	18.87%
6 predictor variable RDA1,2		0.32	32%
4 predictor variable RDA1,2		0.30	30%

After examining all the models, the 10 variable additive index and the 4 predictor variable RDA index were selected as the best models. Both contained the three most common image variable types (topographic and 1st and 2nd order texture measures) and they did not contain fractions which were subjective to derive. The additive index model contained 5 image variables and the RDA model 4 image variables, which met the requirements of the 10:1 ratio rule (Babyak, 2004). Additionally, of all the models, the 4 predictor variable RDA index had the second highest explained variance at 30%. While the 10 variable additive index did not explain the most variance of the original 26 variables, it was a simple and straightforward index to create. Of the PC indices, only that for PC2 contained all three image variable types, though it was not able to explain much variance. Again, the 6-predictor variable model was not selected because it contained fractions, it included the highest number of image variables, and it did not

contain the three most common variables. The 11 variable additive index was the most comparable, in that it contained the three variable types, and only included 4 predictor variables, however it was only able to explain 17.43% of the total variance.

The 10 variable additive index was then compared to the original subjective complexity score assigned in the field. This produced a score of low, medium or high complexity based on qualitative analysis of the plots in the field. Figure 19 shows that the 10 variable field index calculated from the field data corresponds well with the average scores for the three subjective complexity classes. This demonstrates that the additive index was very comparable to the qualitative score given in the field. The line chart in Figure 20 demonstrates that plots with low and high complexity were generally distinguishable from each other, and from the medium complexity plots. There was, however, some overlap between the high and medium, and between the medium and low classes. Three plots that were classified as high, and two plots which were classified as low (qualitatively) both ended up with medium index scores. While these plots appeared to be a particular complexity in the field, the higher complexity plots actually had fewer tree species, large trees and less variation in tree height than the visually classified low plots. These attributes can be difficult to accurately assess qualitatively in the field and could have caused the confusion seen in Figure 20. The 4 predictor variable RDA index was also compared to the original subjective complexity score assigned in the field. A table comparing the results is given in Appendix 4.

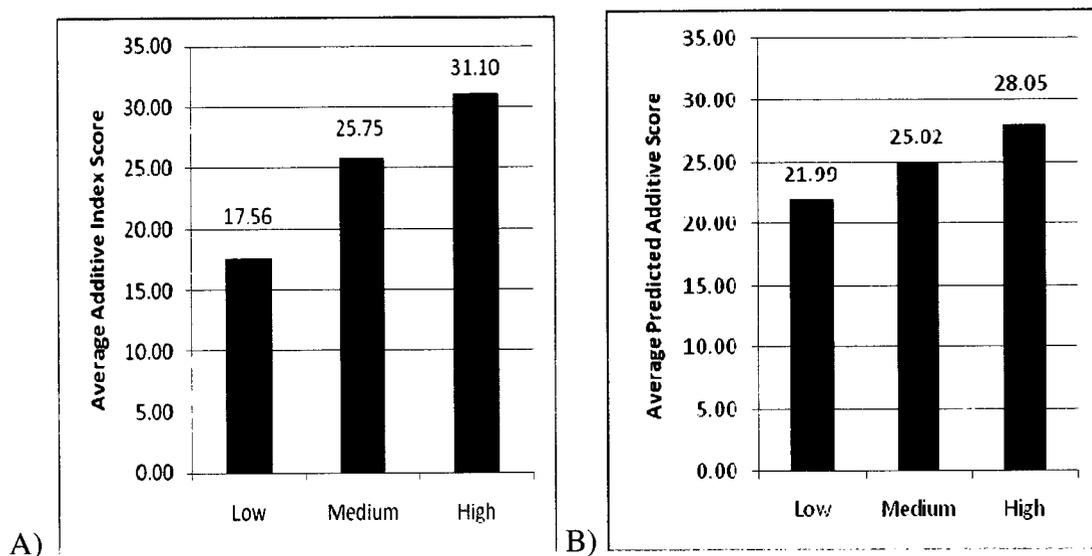


Figure 19. A) Comparison of the qualitative complexity score and the additive index score and B) Comparison of the qualitative complexity score and the predicted additive index score. Plots with a low qualitative score received the lowest average additive score, and plots with a high qualitative score had the highest average additive score.

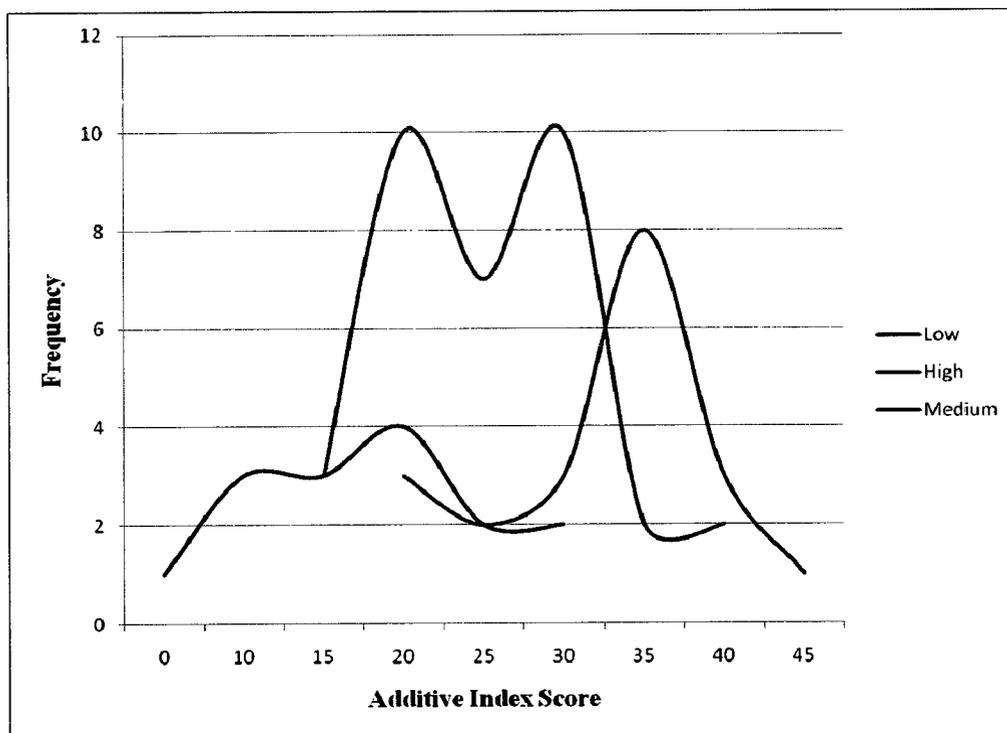


Figure 20. Plot frequency by rounded additive index score grouped by the subjective field classification of high, medium and low complexity.

4.6 Map and Validation

The best two indices, 10 variable additive and four predictor variable RDA, were mapped across Chelsea. The maps were standardized to the same scale, using a linear stretch, for comparison. The map of predicted forest complexity (Figure 21), created with the additive index, reveals several patterns about the complexity of the forest in Chelsea. First, it appears that much of the area is composed of low to medium complexity forest. This was expected since 70 percent of the field plots surveyed were either low or medium complexity and because it was a lot more difficult to locate areas of high complexity. Secondly, areas of high complexity or very low complexity appear in small patches across the area. Again, this was expected, as areas of high complexity are often smaller and more discrete entities, sometimes resulting from local disturbances such as wind. Many areas of high complexity were apparent on steeper slopes, such as those along Chelsea Creek (bottom right). The 10 variable additive model was validated through manual bootstrapping. The validation analysis provided an average mean error of -0.05 and an average mean absolute error of 6.74. The distribution of % RMSE ranged between 22.37% to 49.97%, with an average RMSE of 33.08%.

The map of predicted forest complexity (Figure 22), created with the four variable RDA index, revealed similar patterns to the complexity map created with the additive index. This model was validated automatically through the bootstrapping process during stage two. Overall, both maps look fairly similar. Figure 23 shows two examples of locations where the predicted complexity of the maps were fairly similar. Figure 23 A) shows that both maps were able to map the high complexity of plots 59 and 60.

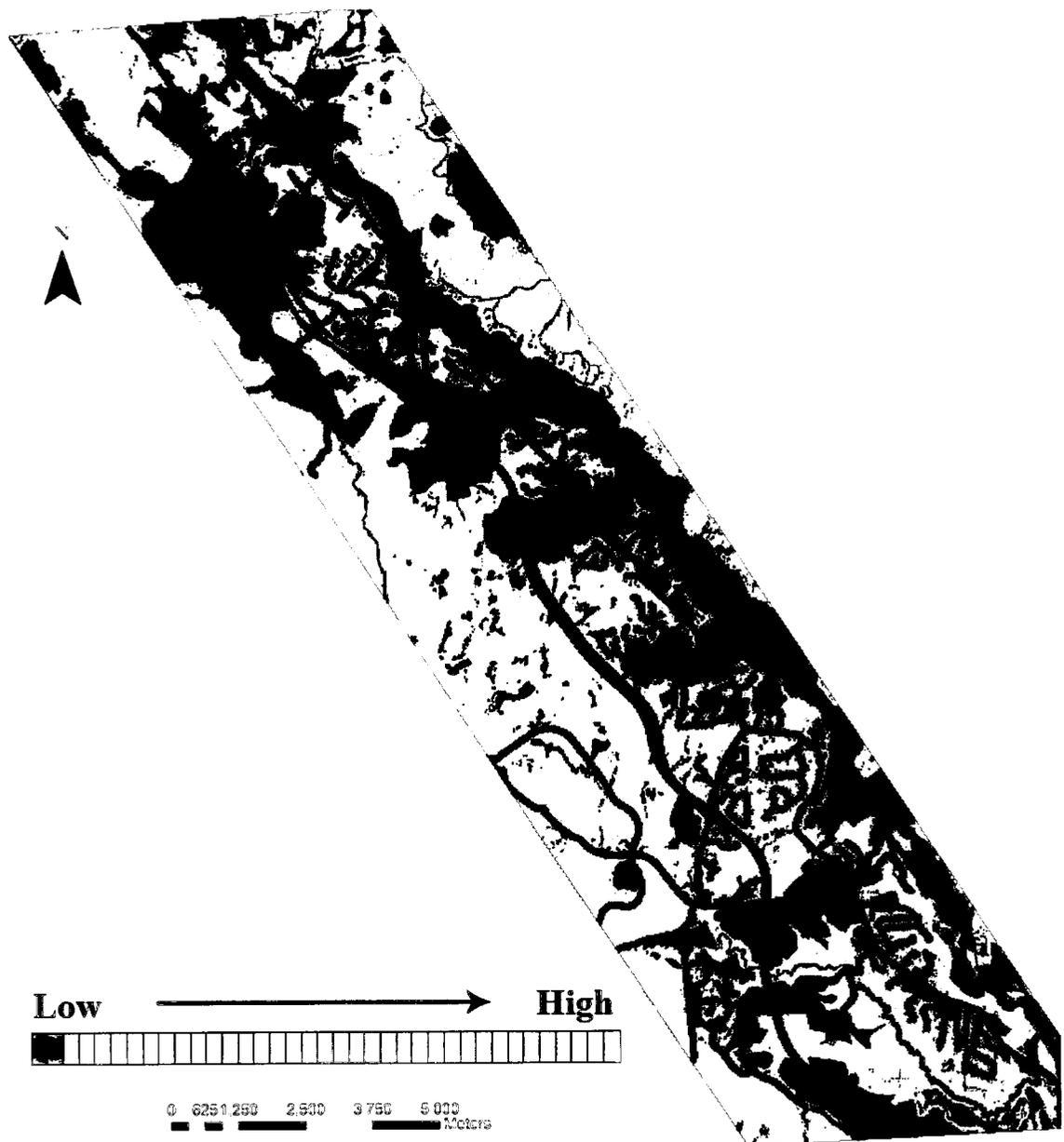


Figure 21. Predicted forest complexity mapped across the entire study area using the 10 variable additive index mapped with 5 predictor variables.

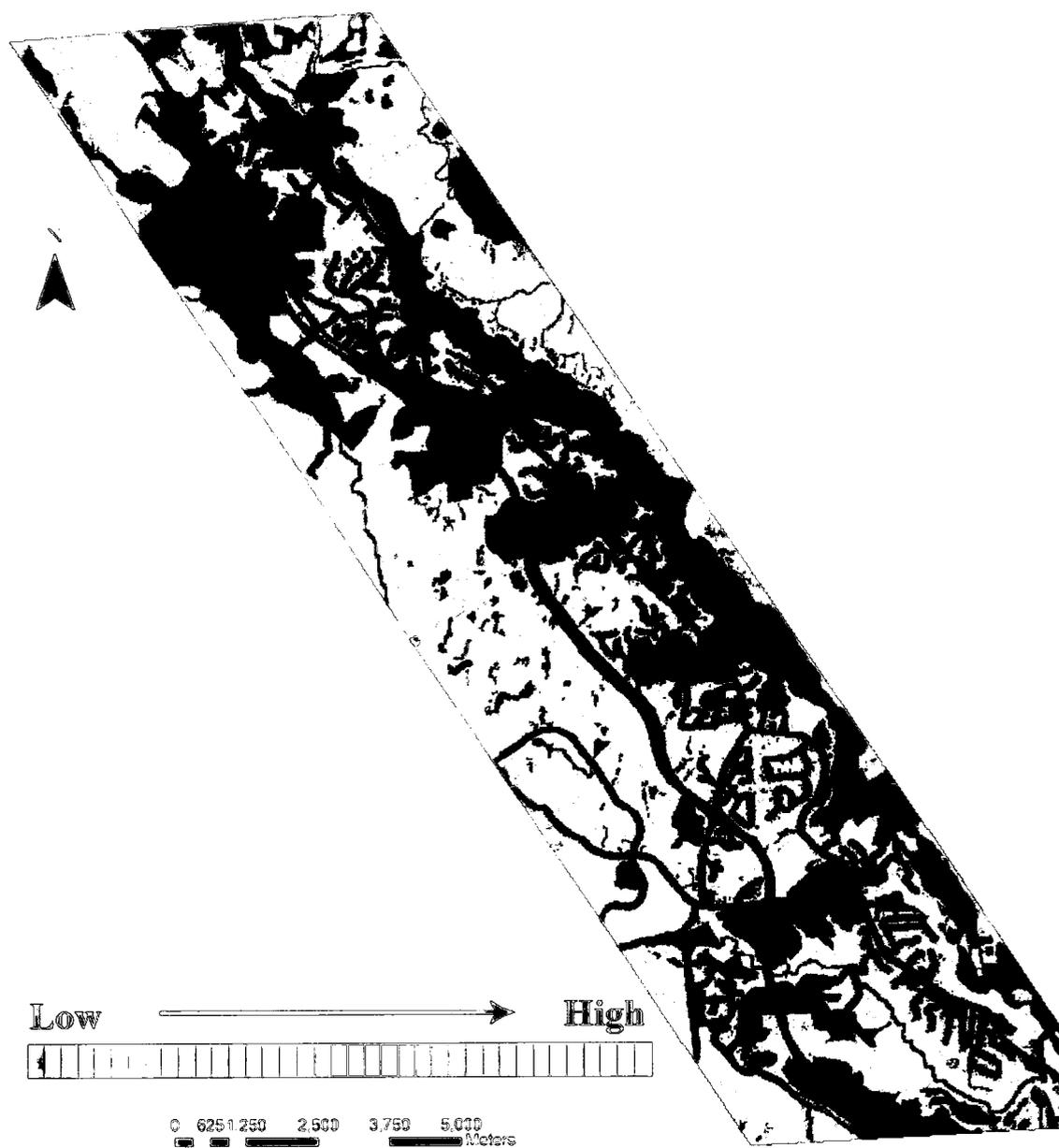
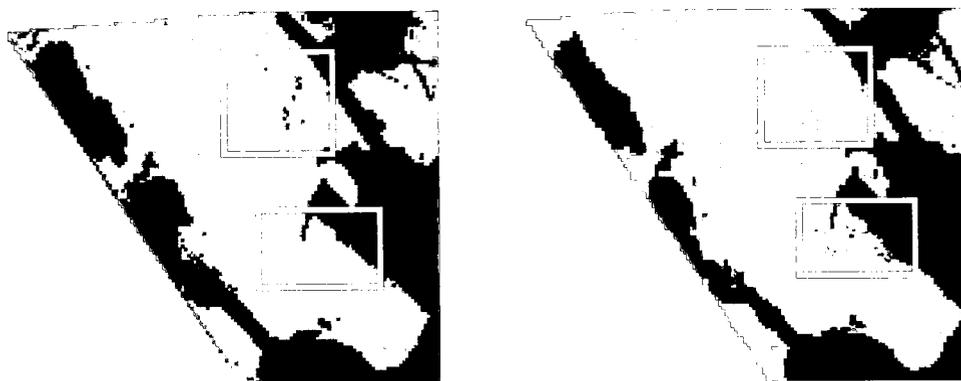


Figure 22. Predicted forest complexity mapped across the entire study area using the 4 predictor variable RDA index.



A)



B)

Figure 23. Map comparison with additive index on the left and RDA index on the right.
A) Similar areas of high complexity, and B) Similar areas of low complexity.

5. Discussion

A primary objective of this study was to determine if it is more advantageous to first define a field index and then model it with geospatial data, or if an index should be defined based on multivariate statistical analysis of the relations between the geospatial and field data. There appeared to be no best approach to modeling and mapping forest complexity in Chelsea Quebec. Both the additive and RDA methods had their own advantages and disadvantages. While the RDA method was best able to explain the variance in the field data, the additive method was simpler and easier to implement. Figure 23 demonstrated that both models produced similar results and that extreme high and low complexity areas were clearly indicated with both methods. Once all models were created, the best models of forest complexity were then selected to map predicted forest complexity across Chelsea, Québec. This objective is important as the selection of the optimal model helped to create a map which can be used by the municipality of Chelsea to determine possible target areas for biodiversity assessment. This study also contributes to research involving the use of remote sensing to map forest complexity.

5.1 Significant Findings

Prior to this study, there had been no attempt to compare field based forest complexity indices (e.g. McElhinney et al., 2006; Wunderle et al., 2007) to those created through multivariate statistical analysis (e.g. Pasher and King, 2010; de la Cueva, 2008). Researchers used one method or the other to map forest complexity, without recognizing if a particular method was more appropriate. This study attempted to determine if one method was more appropriate than the others for mapping forest complexity in Chelsea.

Generally, all models examined were composed of similar types of predictor variables, and model explained variance of the whole set of field 26 variables ranged between 5.67% - 32%. Of the 7 models created, the 10 variable additive index (adapted from McElhinney et al., 2006) and the 4 predictor RDA index, selected to map complexity, were able to explain 18.87% and 30% of the variance in the field data respectively. Though this value is not especially high, the result was still considered successful as the goal of the study was to be able to map a gradient of complexity across the area, and not individual complexity attributes. The unexplained variance could be accounted for by several factors including human error (introduced during field measurement and georeferencing), GPS positional accuracy error, or simply the unsystematic nature of the environment. As well, most radiance leaving the canopy is from the upper canopy and radiance from sub canopy elements (understorey, ground vegetation, CWD, rock, etc.) is limited. All models produced were able to distinguish between the different forest complexity classes identified in the field.

5.1.1 Model Significance

The following section discusses the significant findings for each model type and compares the results to previous work in the literature.

5.1.1.1 Field and Image variables

While all field variables that were not highly correlated (26 of the original 37; $r \leq 0.8$) were utilized in the RDA analysis, only a subset were utilized in the additive indices or were highly correlated to each PC index. Table 25 summarizes the most common field variables and the ones that were most highly correlated with the PCs, additive indices and first RDA axis. The six most common variables included both composition and structural

data. The variables, including how they varied with increasing complexity are: 1) number of tree species; increased, 2) dead basal area; increased, 3) number of large trees; increased, 4) number of snags; increased, 5) percent coniferous basal area; increased, and 6) LAI >6 m; both increased and decreased depending on the model. Variables 1-5 were all expected to increase with increasing complexity, and variable 6 was expected to decrease with increasing complexity based on increased overstory gaps as found in Pasher and King (2010). However, that study was conducted in a purely deciduous forest. The variable LAI effect found here was probably due to mixed effects of structure and composition related to the presence of conifer trees in the overstory which have much higher LAI than deciduous trees. Based on Pasher and King (2010), when the additive index was formulated, LAI >6 m was reversed to represent an expected decrease with increasing complexity. However, the RDA modeling demonstrated that greater LAI is related to higher complexity in the forests studied in Chelsea, and therefore it may not have been appropriate to reverse LAI in the additive index.

Table 25. Most common field variables in the indices. The + and – signs demonstrate whether the variables increase or decrease with increasing complexity.

	PC2	PC3	11 add.	10 add.	4 RDA	6 RDA
% Vegetation cover (0 – 10cm)			+	+		
% Vegetation cover (50 cm – 1 m)			+	+		
# Plant species			+	+		
# Tree species		+	+	+	+	
Live basal area			-	-		
LAI > 6 m (57.5°)			-		+	+
# Trees			-	-		
Dead basal area	+		+	+		
# Large trees (>35 cm)	+		+	+		
Std. deviation tree height			+	+		
# Pieces coarse woody debris			+	+		
Mean tree dbh	+					
Mean snag dbh	+					
# Snags	+				+	+
% Coniferous		+			+	+
Mean NN distance		-				

Unlike Bailey et al. (2004) and Oindo (2002) who found no relations between NDVI heterogeneity and species richness at the landscape scale, this study found that NDVI heterogeneity increased with increasing forest complexity and was therefore expected to be associated with species richness. At a landscape scale, greater NDVI heterogeneity represents lower biodiversity since it implies that there are features on the landscape other than vegetation (water, roads, buildings, etc.). However, within forest patches, greater NDVI variability indicates that there is structural variation due to variable gap sizes and tree heights which is associated with higher biodiversity. The opposite holds true for mean NDVI. At a landscape scale, mean NDVI is expected to be higher and represent higher complexity, whereas at the forest patch scale, mean NDVI was lower in higher complexity plots, again to due variable gap sizes, shadow and variability in tree heights.

The forests studied in this research have been subject to disturbance through patchy ice storm damage, particularly during the ice storm of 1998. This creates distinct areas, some of which are highly damaged and some with little damage. This is thought to lead to higher biodiversity in the area by increasing the amount of CWD and dead wood, as well as creating new gaps in the canopy.

5.1.1.2 PCA

The field based index, created through PCA analysis was used to reduce the dimensionality of the set of field variables to a single variable. Similar methods were previously applied in various studies (Cohen & Spies, 1992; Hansen et al., 2001; Danson & Curran, 1993). A recent study performed by Estes et al. (2010) compared both an additive method (modified from McElhinny et al., 2006) and a PCA method for mapping

microhabitat features in Kenya's Aberdare Mountains. They created both types of indices using thirteen microhabitat scale vegetation structure attributes, which described the mountain bongo antelope habitat. Field variables were collected from 127 circular field plots (0.04 hectare). They regressed the index values against variables derived from ASTER optical imagery, a Shuttle Radar Topography Mission (SRTM) DEM and rainfall data. The accuracy of the predicted structural complexity score was then assessed. PC1, which accounted for 28% of the total variance produced the best model with an R^2 value of 0.51, which means that the image variables accounted for a total of 14.28% of the variance in the whole field dataset. This was better than either PC2 or PC3 used in this study (5.6% of explained variance), though not as good as the weighted sum of PC1-8 (25% of explained variance). Their second best model was the additive index (summarizing 8 variables), with an R^2 value of 0.46. The R^2 value of 0.51 (adjusted $R^2 = 0.49$) achieved in this study with 10 core variables is very comparable, and even slightly better (R^2 values not compared to the whole dataset). These results are partly dependent on the variance of the original dataset, and it is unknown whether Estes et al. had a lower overall variance.

Wunderle et al. (2007) also created a PCA-based index by incorporating mainly spatial information, and one spectral index, but did not use any topographic variables. They investigated three simple forest types including recently harvested, regenerating and old growth forests and were able to produce a three variable model, which had an adjusted R^2 of 0.74. This model, which resulted from a single component, was able to capture a large amount of variance, probably due to the simple nature of the forests studied. Due to the complex nature of the forest in the current study, eight PCs were

significant representing 75.7% of the total field data variance. PC2 and PC3 were used to derive models and each accounted for 14.9% and 10.8% of the total field data variance, respectively, as their models had adjusted R^2 values of 0.38 and 0.52. This demonstrates that the PCA method may not be robust enough to capture the multidimensional nature of the forest studied here. One major disadvantage of this method is that it is sometimes very challenging to interpret PCs, particularly when many are significant.

5.1.1.3 Additive index

McElhinny et al. (2006) did not apply their additive index to remotely sensed imagery. The objective of their study was to create a user-friendly spreadsheet which would easily allow landowners to assess the condition of the vegetation on their properties. While this helped to rank one location relative to another, continuous coverage of the area was never obtained. This thesis research demonstrated that it was possible to successfully map an additive index, like the one created in McElhinny et al. (2006) as a continuous gradient of forest complexity across an area, using satellite imagery.

During the index creation a multiple stage analysis was used in both methods to identify a core set of structural/composition attributes from the suite of field variables. This involved examining the correlation between all field variables, reducing variables where redundancy occurred, selecting variables which were simple to measure in the field, rescaling the field variables, and adding the variables together for each plot to obtain an index. McElhinny et al. (2006) claim that their novel methodology provided an objective basis for identifying, scoring and combining structural attributes into an index. While their methods may have reduced some of the ambiguity in attribute combination, it

was still difficult to determine how many field variables should be selected as the core set, which ones were simpler to measure, and what weights to apply to a variable if necessary. It should therefore be noted that one limitation of this method was that it was quite subjective, despite the authors' statements on being objective, as it depends predominantly on user interpretation of the field variables measured and selected. This being said, McElhinny et al. (2006) were able to avoid subjective scoring of the field plots, unlike Coops and Catling (1997) who subjectively scored the complexity in their field plots from 0-3. One advantage of this method is that it provides the user considerable control over variable selection, which can be desirable in many situations.

5.1.1.4 Redundancy Analysis

Though the RDA work was based on research performed by Pasher and King (2010), several significant changes and improvements were made to the study. One main difference was the environment in which the two studies were performed. The Pasher and King study was performed in Gatineau Park, in a purely deciduous forest, which had seen minimal human disturbance due to its location within the park. The current study was situated within the municipality of Chelsea and as such, it was expected that human disturbance would have some influence on the forest complexity in the area. Forests in this area were modified due to walking trails, logging on private land and even through the collection of dead wood. In addition, the forest in Chelsea was a mixedwood temperate forest and included several species of coniferous trees. This influences both the structure component of this study as well as the composition. Through their research, Pasher and King (2010) developed an index of forest structure. In addition to

incorporating forest structure variables, this study also included three measures of forest species composition: 1) number of different tree species, 2) number of different plant species found in the ground and understorey vegetation, and 3) percent of basal area of coniferous trees. These measures added an additional level of depth to the complexity index.

One important improvement from the Pasher and King (2010) study was the inclusion of a larger number of plots. It had been suggested (in peer review) they use a reduced set of variables and a larger number of plots. This study accounted for both those suggestions by including 70 plots and reducing the number of image variables to 45 from the 86 used by Pasher and King (2010).

The study tested two different aspects of scale based on results of Pasher and King (2010): size of area and image resolution. The size of the study area used by Pasher and King (2010) was relatively small at approximately 400 m x 4.5 km. The current study area was significantly larger at approximately 18 km x 5 km. Since a model was successfully produced and mapped across a larger area, it demonstrates that the methods tested in Pasher and King (2010) are scalable. The second area of scalability was in image resolution. It is not always feasible to acquire or use high resolution (20 cm) aerial imagery to map and model forest complexity as was done by Pasher and King (2010). It is costly to acquire, difficult to cover a large area, and time consuming to process, mosaic and correct. Instead of using aerial imagery, this study used coarser resolution satellite imagery (60 cm panchromatic and 2.4 m multispectral) to determine if the methods were still applicable at a coarser resolution. While the model explanation for this study was

lower than that found by Pasher and King (2010), the model was still significant and was able to predict forest complexity.

Similar to the Pasher and King (2010) research, this study used spectral, spatial and topographic data to derive a forest complexity index using RDA. Using only the first RDA axis, Pasher and King were able to obtain a model, which presented an explanation of 35% of the original variance in the field data. This is significantly higher than the 23% explanation of the first RDA (6 variable model) axis in this study. This could be due to the higher resolution imagery used by Pasher and King, giving the ability to extract within crown and shadow measures (4 of the 5 image variables included in their model). While the first axis was the only one significant in their study, both the first and second axes were significant ($p = 0.01$) in this study, adding another 9% explanation. This resulted in a total explanation of 30%, making it very comparable to the Pasher and King (2010) model. Image variables found to be significant in their model included within crown variance GLCM Homogeneity, within crown mean GLCM Correlation, within shadow mean NDVI, within shadow variance of the green band, and elevation. Both studies included spectral, 2nd order textures and topographic variables. Pasher and King (2010) also performed a scaling analysis and produced models using rescaled 1 m imagery. The average adjusted R^2 value for these models was 0.25 for the first axis. This is very comparable to the four predictor variable RDA model presented here. Similar to the results presented here, the most common variables entered into the Pasher and King (2010) models were topographic, subpixel fractions and full plot 1st order textures.

The de la Cueva (2008) study incorporated only spectral and topographic information as it was not feasible to incorporate spatial information with the Landsat

imagery and plot size used. By combining both the first and second axes, they presented a model, which explained 24% of the original variance in the field data. This is comparable in explanation to the first axis (4 variable RDA) extracted in this research. This demonstrates that the incorporation of spatial variables in forest complexity modelling can have a significant effect on the model produced.

In addition to producing the highest adjusted R^2 values, there are many advantages to using RDA models. First, it avoids subjective selection, scoring or weighting of variables, which is often introduced in field-based indices. Additionally, it allows for the inclusion of many field and image variables, when compared with field based indices, which only examine one dependent variable at a time. RDA eliminates the sometimes difficult interpretation of components in PCA by presenting the results in a triplot, which lets the user visually examine the relation between the RDA axes, image variables and field variables.

5.1.2 Applicability to the Municipality of Chelsea

The map of predicted forest complexity will be used by the municipality of Chelsea to determine possible target areas for biodiversity assessment and incorporation into a municipal plan. It is important to be able to locate both areas of high and low complexity for conservation and remediation efforts, respectively. At a local scale, within the municipality of Chelsea, the forest complexity map can aid in the location of areas of high biodiversity within forest patches. These higher complexity areas can be further investigated, perhaps in the next Bio-Blitz to determine if they do indeed provide more habitat and higher biodiversity than surrounding areas. The map may also help the

municipality assess the condition of the forests as well as help them make appropriate management decisions based on the complexity found in a particular area.

It is also important to locate areas of low biodiversity, which may require remediation efforts to improve their biodiversity status. Residents of Chelsea may remove CWD and snags for firewood or simply while cleaning their properties. These activities decrease the composition and complexity of the forest as well as remove essential habitat used by smaller organisms (Takahashi et al., 2000). By using the complexity map to locate areas of low complexity, the municipality of Chelsea may be able to investigate these areas to determine if any of the above activities are taking place that could be reduced or stopped. The municipality may also be able to educate the landowners about the effects of these activities. If remote sensing has the ability to detect these areas of low complexity, perhaps remediation and education methods may become more common. While greater stand complexity may result in greater habitat types within that stand, there will always be species that are best suited to low complexity stands and as a result, heterogeneity among stands within a forest is important for the preservation of biodiversity and the ecological integrity of the forest itself.

5.2 Limitations

There were some methodological limitations to this research. The main limitation was the large extent of cloud cover/shadow captured in the Quickbird imagery. After most of the field plots had been established, the Quickbird imagery was received. Fourteen of the 70 sites were obscured by cloud cover or cloud shadow. This resulted in prolonging the field season to go back and set up more field plots. If the imagery had been acquired before the field season had commenced, it would have been possible to

establish sites in known locations with no cloud cover. Since only one image was acquired, it was not possible to correct for the cloud cover or substitute in a section of another photo. The other limitation of having a large amount of cloud/shadow was that a continuous map of forest complexity could not be created. A map was still created, but large portions of forest were masked out as they were covered by cloud. The municipality of Chelsea will therefore not receive a complete forest complexity map.

Another potential limitation to this study was GPS positional accuracy errors. GPS devices have limited signal capabilities under the canopy and scattering of the signal may cause some positional errors (Pirti, 2008). The expected horizontal accuracy was 0.5-1 m (Trimble, 2010), but this could vary and be larger in some locations. This may have affected the validation of the models, which were created during this research. Georeferencing errors may also affect the accuracy of the final model. Problems may arise in higher complexity areas as they are less uniform, and are found in smaller patches than the low complexity areas. A small shift in the imagery may therefore have resulted in the extraction of image data, which was not associated with a high complexity site. Further exploration should be done to assess what effect these small errors had on the final indices.

Statistical models are limited in application to data of a similar nature, and geographically this usually means they are not spatially extendable over large distances. Areas in Chelsea, which are mapped as high complexity are not comparable to areas of high complexity in a tropical rain forest. It is very important to note that forest complexity is a relative, not an absolute concept. For example, consider two separate forests, in different locations. They may each contain similar attributes but may be

considered to have high complexity in one forest, and medium complexity in the other, based solely on the range of other attributes found within their respective forests. The structural and compositional attributes of each forest or patch are inherently influenced by the characteristics of that particular environment. Environmental characteristics may include the type of soil, topography, drainage, altitude, and latitude.

Since complexity is a relative concept, a new set of reference sites need to be established if the method is to be applied in a different type of forest. It would not make sense to map a boreal forest using an index formulation from structure data in Australian forests. If it is not possible to obtain new reference data, the application of the indices shown in this study should be limited to communities, which are similar in structure to the forest in Chelsea, Québec. While the index itself may not be transferable to other forest types, the methods certainly are, as was demonstrated by the transference of both the additive index and RDA methods. Both the additive index, created by McElhinny et al. (2006) for Australian forests and the RDA method created originally by Pasher and King (2010) for temperate hardwood forests were effectively used to model forest complexity in temperate mixedwood forests in Chelsea, Québec.

As scale becomes coarser, there is less spatial variation so the data types must become simpler and mostly spectral. This reduces the accuracy with which complexity can be modelled and mapped. At scales coarser than that of this study (e.g. Landsat) only the broadest of gradients of complexity could be mapped and the minimum mapping unit (about 1ha) may be too large to capture small areas of very high complexity (in forest management conditions, large areas of low complexity would be abundant). Though the results of this study were comparable to those from Pasher and King (2010), their results

showed that it was more beneficial to have higher resolution imagery as more types of image information could be extracted (e.g. within shadow brightness and brightness variation, possibly indicating understorey radiance). In order to obtain the spatial coverage necessary to map areas equal to or larger than that of this study, satellite imagery would be required. This is because it would be very time consuming and costly to acquire, mosaic, and correct hundreds or thousands of aerial photographs. That being said, no satellite imagery currently available to the public has a multispectral resolution as small as 20 cm (the resolution used in Pasher and King).

Lastly, it proved very difficult to compare the results from the three methods of creating a forest complexity index. Each method had its own advantages and disadvantages and model comparison was not a straightforward task. Each model accounted for different amounts of total variance of the field data, and it was essential to examine which field variables were used in the indices. It is difficult to compare R^2 values between indices and between studies as they are all dependent on the variance in the field datasets.

5.3 Recommendations for further research

This study did not compare the complexity map to biodiversity data to determine if predicted forest structure and composition diversity corresponded to field-assessed biodiversity. Based on a review of the literature, this study assumed that a more complex area should provide more available habitat, and therefore have a higher biodiversity. The reverse was also assumed with low complexity areas providing less habitat for fewer species. While some biodiversity data, collected during the June 2009 Bio-Blitz, was available through Nature Chelsesa, it was limited in two ways. The first was that the

sample design was opportunistic and scientists concentrated on collecting species data in areas where there was already suspected high biodiversity. This led to a lack of data for low biodiversity areas. When a species was located, the area (1-4) that it was found within was recorded. The exact spatial location was not always recorded with a GPS and thus data were not considered to be well georeferenced. Future studies should consider assessing the relation between the predicted complexity and biodiversity in the areas of the 2009 and 2010 Bio-Blitzes.

Further investigations could be done to test the additive index approach adapted from McElhinny et al (2006). Instead of using the 10 core variables that were selected, it may be useful to try a smaller subset of core variables comprised only of those manifested in the overstorey/trees. These variables could include # of tree species, live basal area, # trees, # of large trees, and std. deviation of tree height. These overstorey variables are ones which are most likely to manifest themselves in the imagery, and may be able to produce models with higher explained variance, albeit of a reduced dataset.

Other sources of GIS data could be explored to help improve the models or aid in model validation. While the following may not necessarily manifest itself in the imagery, they may be important in describing complexity patterns. Information about soil composition may provide an idea of the species of plants or trees that could grow in particular locations, which result in a particular complexity. Human impact may also be assessed through trail maps or landuse change maps. Lidar data could be incorporated to model individual forest attributes that could be combined in an index for biodiversity prediction.

Goetz et al. (2006) studied the use of Lidar to estimate habitat metrics of canopy

structural diversity as predictors of bird species richness in the temperate forests of Maryland. They also wanted to compare the Lidar-derived predictors with those created from traditional remote sensors. The results of their study showed that canopy vertical distribution information was the strongest predictor of species richness. Canopy vertical distribution was also best predicted when it was stratified into groups that were dominated by forest, scrub, suburban and wetland species. Lidar metrics of habitat heterogeneity were found to have performed better than traditional optical remote sensor variables. Ultimately, this study determined that the best models could only explain 45% of variation in bird species richness. Goetz et al. (2006) concluded that when trying to predict species richness, factors other than habitat heterogeneity should be considered. Factors to consider include a combination of the temporal changes in environmental conditions across a given site, the responses of organisms to those environmental conditions, and intra and inter specific competition interactions between the organisms. The results of this demonstrated that Lidar could prove to be very useful when combined with other datasets or multispectral information, which contain other habitat information.

In a more recent study, Anderson et al. (2008) came to similar conclusions. They modelled individual forest structural parameters using Lidar and airborne hyperspectral data, as well as an integration of both datasets. They determined that, for six structural parameters, Lidar was only capable of explaining 26% of the variation, whereas the hyperspectral imagery could explain 32%. When combined, the Lidar and hyperspectral data were able to explain 42% of the variation in the structural parameters, confirming the benefits of combining both types of data.

6. Conclusions

Two different methods of creating a field based multivariate forest structure and composition index were explored in this work followed by modelling and mapping of the indices using Quickbird imagery and topographic data within the municipality of Chelsea, Québec. In addition, these were compared to indices derived from relations between the field and geospatial data using RDA methods to determine if it is better to first define a field index and then model it with geo-spatial data, or if an index should be defined based on multivariate statistical analysis of the relations between the geo-spatial and field data. It was determined that the RDA method was best able to explain the variance in the field data, though the additive method was simpler and easier to implement. In 70 field plots, 41 structure and composition variables were measured, including information on leaf area index (LAI) and canopy openness from hemispherical photographs (1 m and 6 m heights), tree size and spacing, ground and understorey vegetation, snags, coarse woody debris, number of plant species and the proportion of coniferous and deciduous trees. Multispectral imagery, panchromatic imagery and a DEM, were used to derive image spectral, spatial and topographic variables, respectively.

Initially, the two field based structure/composition indices were independently developed and then modeled using multiple regression with the geospatial variables as predictors. The results were then compared with the RDA approach, which directly produced a multivariate index from the geo-spatial data. The 10 variable additive index (Adjusted $R^2 = 0.18$) and the 4 predictor variable RDA index was selected as the best model (Adjusted $R^2 = 0.30$) and was applied to the imagery to map complexity. The indices were constructed using either four or five image variables, which represented 1st

and 2nd order image texture, and topographic information. The map will be used by the municipality of Chelsea to determine possible target areas for biodiversity assessment.

This study was performed to help advance research in the field of forest complexity mapping using remotely sensed data, with the goal of developing structural indicators that can be mapped over areas of local management (e.g. rural municipalities, conservation areas, parks, etc.), are spatially continuous, and which facilitate monitoring efforts. These are three advantages of a geospatial data approach over field based sampling which covers small areas, yields discreet point based samples, and is costly and time consuming. In addition, this study makes a contribution towards furthering research using RDA, a multivariate statistical technique that has only recently been applied with geospatial data to map forest complexity. Overall, this research was successful in creating and comparing multiple indices of forest complexity, many of which had the potential to be mapped across Chelsea, Québec. The final indices selected proved capable of mapping a gradient of forest complexity over the study area.

7. References

Aava, B. 2001. Can resource use be the link between productivity and species richness in mammals? *Biodiversity and Conservation*, Vol. 10, pp. 2011–2022.

Acker, S.A., T.E. Sabin, L.M. Ganio, and W.A. McKee. 1998. Development of old-growth structure and timber volume growth trends in maturing Douglas-fir stands. *Forest Ecology Management*, Vol. 104, pp. 265–280.

ACRE (Action Chelsea for Respect of the Environment). 2004. (Available from <http://acrechelsea.qc.ca/>; accessed on January 29, 2009).

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Vol. 19, pp. 716–723.

Anderson, J.E., L.C. Plourde, M.E. Martin, B.H. Braswell, M-L. Smith, R. O. Dubayah, M.A. Hofton and J.B. Blair. 2008. Integrating waveform Lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, Vol. 112, pp. 1856-1870.

August, P.V. 1983. The role of habitat complexity and heterogeneity in structuring tropical mammal communities. *Ecology*, Vol. 64, pp. 1495-1507.

Babyak, M.A. 2004. What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, Vol. 66, pp. 411-421.

Bachofen, H., and A. Zingg. 2001. Effectiveness of structure improvement thinning on stand structure in subalpine Norway spruce (*Picea abies* (L.) Karst.) stands. *Forest Ecology Management*, Vol. 145, pp. 137–149.

Bailey, S.A., M.C. Horner-Devine, G. Luck, L.A. Moore, K.M. Carney, S. Anderson, C. Betrus, and E. Fleishman. 2004. Primary productivity and species richness: relationships among functional guilds, residency groups and vagility classes at multiple spatial scales. *Ecography*, Vol. 27, pp. 207- 217.

Band, L.E. 1993. Effect of land surface representation on forest water and carbon budgets. *Journal of Hydrology*, Vol. 150, pp. 749-772.

Barnett, J.L., R.A. How, and W.F. Humphreys. 1978. The use of habitat components by small mammals in eastern Australia. *Australian Journal of Ecology*, Vol.3, pp. 227-285.

Berger, A.L., and K.J. Puettmann. 2000. Overstorey composition and stand structure influence herbaceous plant diversity in the mixed Aspen forest of northern Minnesota. *The American Midland Naturalist*, Vol. 143, pp. 111–125.

- Burgman, M.A., D.R. Breininger, B.W. Duncan, and S. Ferson. 2001. Setting reliability bounds on habitat suitability indices. *Ecological Applications*, Vol. 11, pp. 70-78.
- Butson, C. and D. King. 2006. Lacunarity analysis to determine optimum extents for sample-based spatial information extraction from high-resolution forest imagery. *International Journal of Remote Sensing*, Vol. 27, pp. 105-120.
- Catling P.C., and R.J. Burt. 1995. Studies of the ground-dwelling mammals of eucalypt forests in south-eastern New South Wales: the effect of habitat variables on distribution and abundance. *Wildlife Research*, Vol. 22, pp. 271-288.
- Catling, P.C., R.J. Burt, and R.I. Forrester. 2000. Models of the distribution and abundance of ground-dwelling mammals in the eucalypt forests of north-eastern New South Wales in relation to habitat variables. *Wildlife Research*, Vol. 27, pp. 639-654.
- Chen, X.F., J.M. Chen, S.Q. An, and W.M. Ju. 2007. Effects of topography on simulated net primary productivity at landscape scale. *Journal of Environmental Management*, Vol. 85, pp. 585-596.
- Chettri, N., D. Chandra Deb, E. Sharma, and R. Jackson. 2005. The relationship between bird communities and habitat: A study along a trekking corridor in the Sikkim Himalaya. *Mountain Research and Development*, Vol. 25, pp. 235-243.
- Clark, T.E., and M.J. Samways. 1996. Dragonflies (Odonata) as indicators of biotope quality in the Kruger National Park, South Africa. *Journal of Applied Ecology*, Vol. 33, pp. 1001-1012.
- Cohen, W.B., and T.A. Spies. 1992. Estimating structural attributes of Douglas-fir/Western hemlock forest stands from Landsat and SPOT imagery. *Remote Sensing of Environment*, Vol. 41, pp. 1-17.
- Convention on Biological Diversity, 1992. (Available from <http://www.cbd.int/>; accessed on February 6, 2009).
- Cook, S. E. K. 1976. Quest for an index of community structure sensitive to water pollution. *Environmental Pollution*, Vol. 11, pp. 269- 288.
- Coops, N. and P. Catling. 1997. Predicting the complexity of habitat in forests from airborne videography for wildlife management. *International Journal of Remote Sensing*, Vol. 18, pp. 2677-2682.
- Cosmopoulos, P. and D.J. King. 2004. Temporal analysis of forest structural condition at an acid mine site using multispectral digital camera imagery. *International Journal of Remote Sensing*, Vol. 25, pp. 2259-2275.

- Danson, F. M., and P.J. Curran. 1993. Factors affecting the remotely sensed response of coniferous forest plantations. *Remote Sensing of Environment*, Vol. 43, pp. 55–65.
- de la Cueva, A.V. 2008. Structural attributes of three forest types in central Spain and Landsat ETM plus information evaluated with redundancy analysis. *International Journal of Remote Sensing*, Vol. 29, pp. 5657-5676.
- Dewalt, S.J., S.K. Maliakal, and J.S. Denslow. 2003. Changes in vegetation structure and composition along a tropical forest chronosequence: implications for wildlife. *Forest Ecology Management*, Vol. 182, pp. 139–151.
- Digital globe. Quickbird imagery products: product guide. 2007. (Available from <http://www.digitalglobe.com/>; accessed on March 26, 2010).
- Drever, M.C., K.E.H. Aitken, A.R. Norris and K. Martin. 2008. Woodpeckers as reliable indicators of bird richness, forest health and harvest. *Biological Conservation*, Vol. 141, pp. 624-634.
- Dudley, N., D. Baldock, R. Nasi, and S. Stolton. 2005. Measuring Biodiversity and sustainable management in forests and agricultural landscapes. *Philosophical Transactions: Biological Sciences*, Vol. 360, pp. 457-470.
- Duelli, P., 1997. Biodiversity evaluation in agricultural landscapes: An approach at two different scales. *Ecosystems and Environment*, Vol. 62, pp. 81-91.
- Dufrene, M. and P. Legendre. 1997. Species Assemblages and Indicator Species: The Need for a Flexible Asymmetrical Approach. *Ecological Monographs*, Vol. 67, pp. 345-366.
- Environmental Literacy Council, 2007. Value of Biodiversity. (Available from <http://www.enviroliteracy.org>; accessed on April 7, 2008).
- Estes, L.D., P.R. Reillo, A.G. Mwangi, G.S. Okin, and H.H. Shugart. 2010. Remote Sensing of structural complexity indices for habitat and species distribution modeling. *Remote Sensing of Environment*, Vol. 114, pp.792-804.
- Ferreira, L.V., and G.V. Prance. 1999. Ecosystem recovery in terra firme forests after cutting and burning: a comparison on species richness, floristic composition and forest structure in the Jau National Park, Amazonia. *Botanical Journal of the Linnean Society*, Vol. 130, pp. 97–110.
- Franklin, J.F., K. Cromack, W. Denison, A. McKee, C. Maser, J. Sedell, F. Swanson, and G. Juday. 1981. Ecological characteristics of old-growth Douglas-fir forests. USDA Forest Service. General Technical Report PNW-118. Pacific North-west Forest and Range Experiment Station, Portland, Oregon. pp.48.

- Franklin, S.E., R.J. Hall, L.M. Moskal, A.J. Maudie, and M.B. Lavigne. 2000. Incorporating texture into classification of forest species composition from airborne multispectral images. *International Journal of Remote Sensing*, Vol. 21, pp. 61-79.
- Frazer G.W., R.A. Fournier, J.A. Trofymow, and R.J. Hall. 2001. A comparison of digital and film fisheye photography for analysis of forest canopy structure and gap light transmission. *Agricultural and Forest Meteorology*, Vol. 109, pp. 249–263.
- Gardner, T.A., J. Barlow, I.S. Araujo, et al. 2008. The cost-effectiveness of biodiversity surveys in tropical forests. *Ecology Letters*, Vol. 11, pp. 139–150.
- Gibbons, P., and D. Lindenmayer. 2002. *Tree Hollows and Wildlife Conservation in Australia*. CSIRO, Collingwood VIC, pp. 211.
- Goetz, S., D. Steinberg, R. Dubayah, and B. Blair. 2007. Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest. *Remote Sensing of Environment*, Vol. 108, pp. 254–263.
- Grove, S.J. 2002. Tree basal area and dead wood as surrogate indicators of saproxylic insect faunal integrity: a case study from the Australian lowland tropics. *Ecological Indicators*, Vol. 1, pp. 171-188.
- Guisan, A., S.B. Weiss, and A.D. Weiss. 1999. GLM versus CCA spatial modelling of plant species distribution. *Plant Ecology*, Vol. 143, pp. 107-122.
- Hair, J. F., R. E. Anderson, R.L. Tatham, and W.C. Black. 1998. *Multivariate Data Analysis (Fourth Edition)*. Prentice Hall, Upper Saddle River, NJ.
- Hansen, M. J., S.E. Franklin, C. Woudsma, and M. Peterson. 2001. Forest structure classification in the North Columbia mountains using the Landsat TM Tasseled Cap wetness component. *Canadian Journal of Remote Sensing*, Vol. 27, pp. 20–32.
- Haralick, R.M., K. Shanmugan, and I. Dinstein. 1973. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, Vol. 3, pp. 610-621.
- Innes, J.L., and B. Koch. 1998. Forest biodiversity and its assessment by remote sensing. *Global Ecology and Biogeography Letters*, Vol. 7, pp. 397-419.
- Jackson, T.J., and C.E. Ezra. 1985. Spectral response of cotton to suddenly induced water stress, *International Journal of Biometeorology*, Vol. 6, pp. 177-185.
- Jensen, J. R. 2007. *Remote Sensing of the Environment: An Earth Resource Perspective*, Prentice Hall, Upper Saddle River, NJ, USA, pp. 592.

- Jha, C.S., L. Goparaju, A. Tripathi, B. Gharai, A.S. Radhubanshi, and J.S. Singh. 2005. Forest fragmentation and its impact on species diversity: an analysis using remote sensing and GIS. *Biodiversity and Conservation*, Vol. 14, pp. 1681–1698.
- Johnson, D.D.P., S.I. Hay, and D.J. Rogers. 1998. Contemporary environmental correlates of endemic bird areas derived from meteorological satellite sensors. *Proceedings of the Royal Society of London Series B-Biological Sciences*, Vol. 265, pp. 951–959.
- Jonckheere, I., S. Fleck, K. Nackaerts, B. Muys, P. Coppin, M. Weiss, and F. Baret. 2004. Review of methods for in situ leaf area index determination Part I. Theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology*, Vol. 121, pp. 19-35.
- Jones, D. FATHOM: A Matlab toolbox for ecological and oceanographic data analysis. 2008. (Available from <http://www.rsmas.miami.edu/personal/djones/matlab/matlab.html>; accessed on March 26, 2010).
- Kaiser, H.F. 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, Vol. 20, pp. 141-151.
- King, D.J., Olthof, I., Pellikka, P.K.E., Seed, E.D. and Butson, C. 2005. Modelling and mapping forest ice storm damage using remote sensing and environmental data. *Natural Hazards, Special Issue on Remote Sensing*. 35(3): 321-342.
- Kokaly, R.F., R.R. Root, K. Brown, G.L. Anderson, and S. Hager. 2002. Mapping the invasive species leafy spurge (*Euphorbia esula*) in Theodore Roosevelt National Park using field measurements of vegetation spectra and imaging spectroscopy data. *Geological Society of America bulletin*, Vol. 34, pp. 552.
- Koop, H., H.D. Rijkesen., and J. Wind. 1994. Tools to diagnose forest integrity: an appraisal method substantiated by Silvi-Star assessment of diversity and forest structure. In: *Measuring and Monitoring biodiversity in tropical and temperate forest*, ed. By T.J.B. Boyle and B. Boontawee. CIFOR, Chaing Mai, Thailand.
- Lahde, E., O. Laiho, Y. Norokorpi, and T. Saksa. 1999. Stand structure as the basis of diversity index. *Forest Ecology and Management*, Vol. 115, pp. 213-220.
- Landres, P. B., J. Verner, and J. W. Thomas. 1988. Ecological uses of vertebrate indicator species: a critique. *Conservation Biology*, Vol. 2, pp. 316-327.
- Lassau, S.A., D.F. Hochuli, G. Cassis, and C.A.M. Reid. 2005. Effects of habitat complexity on forest beetle diversity: Do functional groups respond consistently. *Diversity and Distributions*, Vol. 11, pp. 73-82.

- Lassueur, T., S. Joost and C.F. Randin. 2006. Very high resolution digital elevation models. Do they improve models of plant species distribution? *Ecological Modelling*, Vol. 198, pp. 139-153.
- Lawton, J.H. (1983) Plant architecture and the diversity of phytophagous insects. *Annual Review of Entomology*, Vol. 28, pp. 23–39.
- Lawton, J.H., and K.J. Gaston. 2001. Indicator species. In: Levin, S.A. (Ed.), *Encyclopedia of Biodiversity*, vol. 3. Academic Press, New York, pp. 437–450.
- Leblanc, S.G. 2006. Digital hemispherical photography manual v. 1.3b. Canadian Centre for Remote Sensing, Natural Resources Canada, Ottawa.
- Leemans, R. 1999. Modelling for species and habitats: new opportunities for problem solving. *The Science of the Total Environment*, Vol. 240, pp.: 51-73.
- Legendre, P., and H. J. B. Birks. 2010. From classical to canonical ordination. Chapter 7 in: *Tracking Environmental Change using Lake Sediments, Volume 5: Data handling and numerical techniques*. H. J. B. Birks, A. F. Lotter, S. Juggins and J. P. Smol [eds.]. Springer, Dordrecht (in press).
- Legendre, P. and L. Legendre. 1998. *Numerical ecology*. 2nd English edition. Elsevier Science BV, Amsterdam. Pp. 870.
- Levin, N., A. Shmida, O. Levanoni, H. Tamari, and S. Kark. 2007. Predicting mountain plant richness and rarity from space using satellite-derived vegetation indices. *Diversity and Distributions*, Vol. 13, pp. 692-703.
- Lindenmayer, D.B., R.B. Cunningham, M.T. Tanton, A.P. Smith, H.A. Nix. 1990. The conservation of arboreal marsupials in the montane ash forests of the central highlands of Victoria, southeast Australia: I. Factors influencing the occupancy of trees with hollows. *Conservation Biology*, Vol. 5, pp. 111–131.
- Lindenmayer, D.B., C.R. Margules, and D.B. Botkin, 2000. Indicators of Biodiversity for Ecologically Sustainable Forest Management. *Conservation Biology*, Vol. 14, pp. 941-950.
- MacArthur, R.H. and J.W. MacArthur. 1961. On bird species diversity. *Ecology*, Vol. 3, pp. 594-598.
- Margules C.R. and R.L. Pressey. 2000. Systematic conservation planning. *Nature*, Vol. 405, pp. 243-253.

McCarthy, M.A., K.M. Parris, R. van der Ree, M.J. McDonnell, M.A. Burgman, N.S.G. Williams, N. McLean, M.J. Harper, R. Meyer, A. Hahs, and T. Coates. 2007. The habitat hectares approach to vegetation assessment: An evaluation and suggestions for improvement. *Ecological Management and Restoration*, Vol. 5, pp. 24-27.

McElhinny, C., P. Gibbons, C. Brack and J. Baulhus. 2005. Forest and woodland stand structural complexity: Its definition and measurement. *Forest Ecology and Management*, Vol. 218, pp. 1-24.

McElhinny, C., P. Gibbons, and C. Brack. 2006. An objective and quantitative methodology for constructing and index of stand structural complexity. *Forest Ecology and Management*, Vol. 235, pp. 54-71.

Miller, J.B. 1967. A formula for average foliage density. *Australian Journal of Botany*, Vol. 15, pp. 141-144.

Municipality of Chelsea, 2007. Chelsea: Demographic profile. (Available from http://www.chelsea.ca/reenseignements_generaux/profil_e.php; accessed on January 28, 2009).

National Capital Commission (NCC). 2002. 1:25,000 Scale Topographic Mapping - Contours.shp [computer file]. NCC, Ottawa, Ontario.

NEASPEC. Saving the flagship species of north east asia. 2007. (Available from http://www.neaspec.org/publication/Publication_SavingNatureConservation.pdf; accessed on February 19, 2010).

Niemi G.J. and M.E. McDonald. 2004. Application of ecological indicators. *Annual Review of Ecology, Evolution, and Systematics*, Vol. 35, pp. 89-111.

Neumann, M. and F. Starlinger. 2001. The significance of different indices for stand structure and diversity in forests. *Forest Ecology and Management*, Vol. 145, pp. 91-106.

Newsome, A.E. and P.C Catling. 1979. Habitat preferences of mammals inhabiting heathlands of warm temperate coastal, montane and alpine regions of southeastern Australia. In *Ecosystems of the World*. Vol. 9A. Heathlands and Related Shrublands of the World. Edited by R.L. Specht. Elsevier. Amsterdam, Netherlands.

Noss, R.F. 1990. Indicators for monitoring biodiversity: A hierarchical approach. *Conservation Biology*, Vol. 4, pp. 355-364.

Ohtsuka, T., T. Akiyama, Y. Hashimoto, M. Inatomi, T. Sakai, S. Jia, W. Mo, S. Tsuda, and H. Koizumi. 2005. Biometric based estimates of net primary production (NPP) in a cool-temperate deciduous forest stand beneath a flux tower. *Agricultural and Forest Meteorology*, Vol. 134, pp. 27-38.

- Oliver, I. 2002. An expert panel approach to the assessment of vegetation condition within the context of biodiversity conservation. *Ecological Indicators*, Vol. 2, pp. 223-237.
- Olthof, I., D.J. King, and R.A. Lautenschlager. 2003. Overstorey and understorey leaf area index as indicators of forest response to ice storm damage. *Ecological Indicators*, Vol. 3, pp. 49-64.
- Oindo, B.O. 2002. Patterns of herbivore species richness in Kenya and current ecoclimatic stability. *Biodiversity and Conservation*, Vol. 11, pp. 1205–1221.
- Ozdemir, I., D.A. Norton, U.Y. Ozkan, A. Mert., and O. Senturk. 2008. Estimation of tree size diversity using object oriented texture analysis and aster imagery. *Sensors*, Vol. 8, pp. 4709-4724.
- Palmer, M. W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology*, Vol. 74, pp. 2215-30.
- Pasher, J. 2009. Forest structural complexity in a temperate hardwood forest: A geomatics approach to modelling and mapping indicators of habitat and biodiversity. PhD Thesis, Department of Geography and Environmental Studies, Carleton University, Ottawa, Ontario. Appendix 2.
- Pasher, J., and D.J. King. 2010. Multivariate forest structure modelling and mapping using high resolution airborne imagery and topographic information. *Remote Sensing of Environment*, Vol. 114, pp. 1718-1732.
- Parkes, D., G. Newell, and D. Cheal. 2003. Assessing the quality of native vegetation: The 'habitat hectares' approach. *Ecological Mangement and Restoration*, Vol. 4, pp. S29-S38.
- Payton, I.J., M. Fenner, and W.G. Lee. 2002. Keystone species: the concept and its relevance for conservation management in New Zealand. *Science for conservation*, Vol. 203, pp. 1-29.
- Pearson, D. L., and F. Cassola. 1992. World-wide species richness pat-terns of tiger beetles (Coleoptera:Cicindelidae): indicator taxon for biodiversity and conservation studies. *Conservation Biology*, Vol. 6, pp. 376-391.
- Peres-Neto, P.R., P. Legendre, S. Dray and D. Borcard. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, Vol. 87, pp. 2614–2625.
- Pielou, E. C. 1984. *The Interpretation of Ecological Data: A Primer on Classification and Ordination*. John Wiley and Sons, New York, NY.

Pirti, A. 2008. Accuracy analysis of GPS positioning near the forest environment. *Croatian Journal of Forest Engineering*, Vol.29, pp. 189 – 199.

Pisarcic, M.F.J., D.J. King, A.J.M. MacIntosh and R. Bemrose. 2008. Impact of the 1998 ice storm on the health and growth of sugar maple (*Acer saccharum* Marsh) dominated forests in Gatineau Park, Quebec. *Journal of the Torrey Botanical Society*, Vol. 135, pp. 530-539.

Plan Environnemental de la Municipalité de Chelsea. 1991. (Available from http://www.chelsea.ca/environnement/plan_e.php; accessed on January 19, 2009).

Pommerening, A. 2002. Approaches to quantifying forest structure. *Forestry*, Vol. 75, pp. 305-324.

Pykh, Y.A. 2002. Lyapunov functions as a measure of biodiversity: theoretical background. *Ecological Indicators*, Vol. 2, pp. 123-133.

Redford, K.H., and B.D. Richter. 1999. Conservation of Biodiversity in a world of use. *Conservation biology*, Vol. 13, pp. 1246- 1256.

Riano, D., E. Chuvieco, J. Salas, and I. Aquado. 2003. Assessment of different topographic corrections in Landsat-TM data for mapping vegetation types. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, pp. 1056-1061.

Sahlen G., and K. Ekestubbe. 2001. Identification of dragonflies (Odonata) as indicators of general species richness in boreal forest lakes. *Biodiversity and conservation*, Vol. 10, pp. 673-960.

Seed, E.D. and D.J. King. 2003. Shadow brightness and shadow fraction relations with effective LAI: Importance of canopy closure and view angle in mixedwood boreal forest. *Canadian Journal of Remote Sensing, Special Issue on Measurement and use of leaf area index in monitoring vegetated ecosystems*, Vol. 29, pp.324-335.

Simberloff, D. 1997. Flagships, umbrellas and keystones: is single-species management passé in the landscape era? *Biological conservation*, Vol. 83, pp. 247-257.

Smith G. F., T. Gittings, M. Wilson, L. French, A. Oxbrough, S. O'Donoghue, J. O'Halloran, D.L. Kelly, F.J. G. Mitchell, T. Kelly, S. Iremonger, A. McKee, and P. Giller. 2008. Identifying practical indicators of biodiversity for stand-level management of plantation forests. *Biodiversity Conservation*, Vol. 17, pp. 991–1015.

Song, C. and M.B. Dickinson. 2008. Extracting forest canopy structure from spatial information of high resolution optical imagery: tree crown size versus leaf area index. *International Journal of Remote Sensing*, Vol. 29, pp. 5605–5622.

- Spangenberg, J.H., 2006. Biodiversity pressure and the driving forces behind ecological economics. *Ecological Economics*, Vol. 61, pp. 146-158.
- Spies, T.A. 1998. Forest structure: A key to the ecosystem. *Northwest Science*, Vol. 72, pp. 34-39.
- Spies, T.A. and J.F. Franklin. 1991. The structure of natural young, mature, and oldgrowth Douglas-Fir forests in Oregon and Washington. In: Aubry, K.B., M.H. Brookes, J.K Agee, R.G. Anthony, and J.F. Franklin (Eds.), *Wildlife and vegetation of unmanaged Douglas-Fir forests*. USDA Forest Service, Portland, Oregon, pp. 91-109.
- Stevens, V. 1997. The ecological role of coarse woody debris: An overview of the ecological importance of CWD in B.C. forests. Research Branch, British Columbia Ministry of Forestry. Victoria, B.C. Working Paper 30/1997.
- Sullivan, T.P., D.S. Sullivan, and P.M.F. Lindgren. 2001. Stand structure and small mammals in young Lodgepole Pine forest: 10 years after thinning. *Ecological Applications*, Vol. 11, pp. 1151-1173.
- Takahashi, M., Y. Sakai, R. Ootomo and M. Shiozaki. 2000. Establishment of tree seedlings and water-soluble nutrients in coarse woody debris in an old-growth *Picea-Abies* forest in Hokkaido, northern Japan. *Canadian Journal of Forest Research*, Vol. 30, pp. 1148-1155.
- Tanabe, S., Toda, M.J., Vinokurova, A.V., 2001. Tree shape, forest structure and diversity of drosophilid community: comparison between boreal and temperate birch forests. *Ecological Research*, Vol. 16, pp. 369-385.
- ter Braak, C.J.F. and P. Šmilauer. 2002. CANOCO reference manual and CanoDraw for Windows. user's guide: Software for Canonical Community Ordination (v. 4.5). Microcomputer Power, Ithaca, NY. 500 pp.
- Tews, J., U. Brose, V. Grimm, K. Tielborger, M.C. Wichmann, M. Schwager, and F. Jeltsch. 2004. Animal species diversity driven by habitat heterogeneity/diversity: The importance of keystone structures. *Journal of Biogeography*, Vol. 31, pp. 79-92.
- Trimble. 2010. GeoXT handheld datasheet. (Available from http://trl.trimble.com/docushare/dsweb/Get/Document-414895/022501-164H_GeoXT_DS_0510_MGIS_LR.pdf; accessed on June 20, 2009).
- Turner, W., S. Spector, N. Gardiner, M. Fladeland, E. Sterling, and M. Steininger. 2003. Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution*, Vol. 18, pp. 306-314.

- USDA Forest Service. 2006. Forest management guides. (Available from http://www.ncrs.fs.fed.us/fmg/nfm/fm101/silv/p4_ecology.html; accessed on March 2, 2009)
- Uutera, J., T. Tokola, and M. Maltamo. 2000. Differences in the structure of primary and managed forests in East Kalimantan Indonesia. *Forest Ecology Management*, Vol. 129, pp. 63–74.
- Van den Meersschaut, D. and K. Vandekerkhove. 1998. Development of a stand scale forest biodiversity index based on the State Forest Inventory. In *Integrated Tools for Natural Resources Inventories in the 21st Century*. Eds. M. Hansen and T. Burk, USDA, Boise, Idaho, USA, pp. 340-349.
- Van Pelt, R. and N.M. Nadkarni. 2004. Development of canopy structure in *Pseudotsuga menziesii* forests in the southern Washington Cascades. *Forest Science*, Vol. 50, pp. 326-341.
- Van Wagner, C.E. 1968. The line intersect method in forest fuel sampling. *Forest Science*, Vol. 14. pp 20-26.
- Waide, R.B., M.R. Willig, C.F. Steiner, G. Mittelbach, L. Gough, I. Dodson, G.P. Juday and R. Parmenter. 1999. The relationship between productivity and species richness. *Annual Review of ecology and systematics*, Vol. 30, pp. 257–300.
- Waser, L.T., M. Kuechler, M. Schwarz, E. Ivits, S. Stofer, and C. Scheidegger. 2007. Prediction of lichen diversity in an UNESCO biosphere reserve – correlation of high-resolution remote sensing data with field samples. *Environmental Modeling and Assessment*, Vol. 12, pp. 315–328.
- Watson, J., D. Freudenberger, and D. Paull. 2001. An assessment of the focal-species approach for conserving birds in variegated landscapes in southeastern Australia. *Conservation Biology*, Vol. 15, pp. 1364–1373.
- Weiss, M., F. Baret, G.J. Smith, I. Jonckheere, and P. Coppin. 2004. Review of methods for insitu leaf area index (LAI) determination Part II. Estimation of LAI, errors and sampling. *Agricultural and Forest Meteorology*, Vol. 121, pp. 37-53.
- Wikstrom, P., and L.O. Eriksson. 2000. Solving the stand management problem under biodiversity-related considerations. *Forest Ecology Management*, Vol. 126, pp. 361–376.
- Williams, S.E., H. Marsh, and J. Winter. 2002. Spatial scale, species diversity, and habitat structure: Small mammals in Australian tropical rain forest. *Ecology*, Vol. 83, pp. 1317-1329.

Wunderle, A.L., S.E. Franklin, and X.G. Guo. 2007. Regenerating boreal forest structure estimation using SPOT-5 pansharpened imagery. *International Journal of Remote Sensing*, Vol. 28, pp.4351-4364.

Zenner, E.K. 2000. Do residual trees increase structural complexity in Pacific Northwest coniferous forests? *Ecological Applications*, Vol. 10, pp. 800-810.

Ziegler, S.S., 2000. A comparison of structural characteristics between oldgrowth and postfire second growth hemlock-hardwood forests in Adirondack Park, New York, USA. *Global Ecology and Biogeography*, Vol. 9, pp. 373–389.

Appendices

Appendix 1: The formulations for the five textures used in this research (Haralick, 1973).

$$\text{Angular Second Moment} = \sum_i \sum_j \{p(i,j)\}^2$$

$$\text{Contrast} = \sum_{i \neq j} P_{i,j} (i-j)^2$$

$$\text{Correlation} = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

$$\text{Entropy} = -\sum_i \sum_j p(i,j) \log(p(i,j))$$

$$\text{Homogeneity} = \sum_i \sum_j \frac{1}{1 + (i-j)^2} p(i,j)$$

Where i is the row number, j is the column number, and μ_x , μ_y , σ_x and σ_y are the means and standard deviations of the marginal distributions associated with $P(i,j)$, and R is a normalizing constant.

Appendix 2: Summary Statistics for Field Variables

Calculated field variable	Minimum	Maximum	Average	Std. deviation
# tree species	1.00	9.00	4.64	1.99
# Trees	8.00	61.00	26.24	10.68
Mean dbh	13.10	33.83	21.90	4.46
Std. deviation dbh	4.37	25.88	9.90	3.71
Live basal area (m ² ha ⁻¹)	5.20	65.84	29.02	12.36
% basal area coniferous	0.00	97.72	23.77	29.18
# Large trees (>35cm dbh)	0.00	8.00	2.47	2.08
Std. deviation tree height	1.90	6.85	3.61	1.09
Average NN distance	0.90	3.39	1.86	0.55
# Snags	0.00	14.00	4.47	3.30
Mean snag dbh	0.00	57.51	19.52	9.44
Dead basal area (m ² ha ⁻¹)	0.00	30.04	5.30	6.55
# Pieces CWD	0.00	48.00	14.29	9.71
% Rock	0.00	50.89	6.58	10.12
# Vegetation species	1.00	7.00	4.93	1.20
% Vegetation cover (<2 m)	4.46	100.00	57.63	25.20
% Vegetation cover (0-10 cm)	0.00	38.39	10.94	7.26
% Vegetation cover (10-50 cm)	0.00	60.71	14.89	10.77
% Vegetation cover (50 cm-1 m)	0.00	95.54	10.56	13.62
% Vegetation cover (1-2 m)	0.00	58.93	21.03	15.75
Openness at 1 m (0°-60°) (%)	0.70	24.70	4.23	3.91
Openness at 6 m (0°-60°) (%)	1.00	56.60	12.15	12.31
LAI at 1 m (57.5°)	3.41	13.72	7.10	2.30
LAI at 6 m (57.5°)	3.85	17.28	8.81	2.57

Appendix 3. Raw Field Plot Data.

Plot num	# of Trees	Avg Tree dbh (cm)	Stdev Tree dbh (cm)	Live Basal Area (m ² /ha)	# of Large Trees	Stdev Tree Height (m)	Avg NN Distance (m)
1	27	20.91414	9.323322	27.62618	3	3.643214	2.307407
2	31	19.55039	7.07516	26.21368	0	4.339729	1.722581
3	34	16.55211	10.7779	25.81698	1	3.747575	1.567647
4	31	18.04721	9.789156	25.69105	1	3.966031	1.791935
5	18	22.06949	12.2037	20.42758	4	4.283272	3.041667
6	17	26.17631	9.422812	25.66101	4	2.582998	2.591176
7	28	25.07827	10.07343	39.95634	4	3.257421	1.785714
8	26	27.88884	12.24698	47.06939	2	3.933036	1.265385
9	20	20.32409	10.51122	20.34303	3	3.913197	2.405
10	29	26.97951	11.41221	48.60762	6	2.788875	1.524138
11	31	24.52312	8.994586	31.9932	5	3.819828	1.608958
12	45	20.83161	8.300106	44.29511	3	2.596855	1.803333
13	11	33.59616	14.01385	28.23435	4	2.91711	3.159091
14	30	21.66629	11.89742	35.71167	5	2.828083	1.541667
15	16	29.18504	9.649021	29.50123	7	3.384031	2.1125
16	20	26.24465	8.013158	29.44393	4	2.841429	2.605
17	61	20.41358	11.58993	65.73632	3	1.900358	1.043443
18	16	27.81233	7.859848	26.12057	2	1.978108	2.51875
19	12	27.90517	12.36713	21.65109	2	2.719337	0.9
20	25	19.70975	6.445232	21.0268	0	4.493086	1.802
21	19	18.78028	8.959291	15.99491	1	3.276444	2.489474
22	33	20.84447	5.969966	30.3925	0	3.671515	1.869697
23	31	21.86789	9.734875	33.56486	2	2.883825	1.63871
24	41	19.57218	9.055097	37.27835	2	4.614715	1.386585
25	32	22.37122	9.395486	36.81879	3	2.671052	2.25625
26	54	16.63464	5.41151	32.38691	1	2.298754	1.003519
27	27	25.8892	15.19755	47.32404	6	3.499892	1.566296
28	27	26.5494	14.71215	48.41823	5	3.851912	1.32037
29	35	30.49409	5.385212	65.84037	5	4.790964	1.901429
30	16	22.42095	6.662715	17.10024	0	2.175297	2.06875
31	28	22.65684	14.84209	39.90044	6	3.748246	1.626786
32	22	17.17427	9.507003	16.468	3	2.400344	1.715909
33	14	25.62395	11.26693	21.28921	1	3.134765	2.678571
34	30	19.60789	7.007107	25.44297	1	3.001516	2.071667
35	35	24.24612	12.49317	50.81988	4	3.235436	1.562857
36	28	15.7336	5.0918	14.98408	0	3.036669	1.514286

37	31	20.69014	7.973435	29.80163	1	3.160882	1.698387
38	22	21.77529	11.90831	26.32966	6	5.006059	1.952273
39	39	17.83352	6.522384	27.5281	0	4.706669	1.192308
40	22	19.43137	9.259677	19.84567	2	3.224601	1.465909
41	37	18.10925	7.836137	28.16552	0	3.083444	1.267568
42	17	18.78028	7.806708	13.68756	1	2.864073	1.891176
43	58	18.50039	8.141164	46.39596	3	2.699762	1.250862
44	26	16.56436	5.43964	15.45975	0	2.286231	1.751923
45	27	17.50704	7.834861	19.38253	1	2.938999	1.623214
46	24	18.85986	7.284179	19.15792	0	2.990996	1.539583
47	22	25.68182	14.47369	37.12875	8	4.381888	1.625
48	30	17.37972	6.453193	20.16378	0	2.200255	1.173667
49	22	20.8493	10.64351	23.44855	4	3.068776	2.113636
50	23	22.26785	12.39236	29.02695	2	3.793069	1.736957
51	14	33.83179	10.60607	34.33497	7	2.744694	3.196429
52	26	19.8454	10.05897	25.07273	3	2.368108	1.680769
53	18	14.23553	4.796202	7.930113	0	2.678135	1.936111
54	16	22.32148	13.60485	21.10439	1	3.50653	1.878125
55	13	19.68624	6.2765	10.82057	0	3.1036	2.257692
56	16	32.8456	13.14339	38.98052	5	3.596462	1.9625
57	15	22.17559	11.93414	18.39855	2	6.182617	3.386667
58	18	22.84758	25.87633	40.79986	2	4.904969	2.333333
59	22	25.40692	10.75484	32.6535	4	6.045026	2
60	8	23.43557	11.3614	10.4014	1	3.584997	2.14375
61	21	25.69216	23.47741	48.86287	4	6.275298	2.652381
62	40	19.45669	6.633667	33.10212	1	3.324229	0.97625
63	30	19.23653	11.0325	28.72813	2	5.119566	1.618333
64	14	13.09618	4.374892	5.203185	0	3.222021	2.796429
65	15	25.40113	8.934736	21.1977	4	5.198643	2.24
66	35	20.45368	11.53906	37.63924	2	6.853932	1.464286
67	16	21.92593	10.86685	19.75694	2	5.305994	1.561765
68	26	16.95612	5.763823	16.30845	0	3.626665	1.732692
69	45	16.62992	6.724061	28.34178	0	5.562394	0.909091
70	29	17.2985	8.48516	20.99736	2	4.766133	1.718966

Plot num	# of Snags	Avg Snag dbh (cm)	Dead Basal Area (m ² /ha)	# of Pieces of CWD	% Rock
1	2	23.23662	2.198333	9	5.357143
2	2	14.96056	0.880527	8	7.142857
3	0	0	0	0	4.464286
4	2	19.09859	1.433989	6	1.785714
5	10	31.03521	20.764	23	5.357143
6	6	23.87324	7.888135	13	0
7	2	16.2338	1.067136	10	0
8	6	20.37183	4.996283	21	0
9	8	33.58169	20.02015	22	0
10	0	0	0	15	0
11	6	32.41456	18.5406	17	0
12	4	14.56268	1.741756	21	2.678571
13	2	18.30282	1.535252	3	0.892857
14	8	15.91549	4.200905	24	0
15	3	21.43287	3.05618	32	0.892857
16	2	10.98169	0.473686	9	6.25
17	4	15.59718	2.061061	11	0
18	3	22.81221	3.162615	22	2.678571
19	8	24.50986	10.2317	23	0.892857
20	3	13.89953	1.186702	11	0
21	0	0	0	14	5.357143
22	1	10.82254	0.229979	2	4.464286
23	4	13.84648	1.53943	9	0.892857
24	6	16.0216	3.253135	16	1.785714
25	4	22.44085	3.976496	9	5.357143
26	2	12.09578	0.574949	11	16.07143
27	4	23.23662	5.786489	12	0
28	3	23.55493	4.431282	5	3.571429
29	6	17.24179	3.716078	5	0
30	1	20.05352	0.789609	21	25
31	1	44.56338	3.899305	15	24.10714
32	7	17.73441	4.629032	15	8.928571
33	6	26.84413	10.92243	4	8.035714
34	1	14.00563	0.385156	10	39.28571
35	4	25.30564	5.252523	7	3.571429
36	9	17.36557	5.618382	12	0
37	3	15.59718	1.462041	12	18.75
38	2	25.46479	2.815458	14	1.785714

39	5	12.85972	1.751504	9	0
40	6	19.2047	4.615106	6	0
41	10	13.49634	3.696382	42	0
42	1	21.96338	0.947173	24	0
43	0	0	0	4	0
44	1	19.09859	0.716199	4	0
45	3	14.96056	1.425037	4	0
46	1	14.00563	0.385156	4	0
47	4	20.92888	3.954015	5	0
48	1	10.18592	0.203719	27	0
49	11	24.71242	14.64348	38	0
50	5	24.31888	7.584547	10	0
51	1	27.37465	1.471391	9	0.892857
52	4	14.64225	1.811983	1	3.571429
53	3	57.50799	30.04096	12	13.39286
54	3	15.80939	1.651435	24	10.71429
55	10	30.36676	22.99436	21	24.10714
56	3	25.57089	4.558805	15	8.035714
57	6	17.7723	4.994294	20	35.71429
58	4	42.33521	18.41308	9	14.28571
59	11	20.16927	9.786261	29	0
60	14	26.6156	23.20783	14	6.25
61	7	18.50745	5.597691	12	4.464286
62	12	12.2284	3.631725	23	12.5
63	4	22.36127	5.06293	13	2.678571
64	3	10.39812	0.638611	5	21.42857
65	12	23.95282	14.66795	32	19.64286
66	2	11.61831	0.532573	10	8.035714
67	7	24.19155	8.347298	48	3.571429
68	7	19.50785	5.854329	17	4.464286
69	4	13.28944	1.419466	7	10.71429
70	3	17.61315	1.989839	9	50.89286

Plot num	% Understorey Cover (<2m)	% 0-10cm	% 10-50cm	% 50cm-1m	% 1-2m
1	76.78571	9.821429	13.39286	11.60714	41.96429
2	57.14286	25.89286	14.28571	10.71429	6.25
3	34.82143	16.07143	16.07143	0.892857	1.785714
4	10.71429	2.678571	3.571429	1.785714	2.678571
5	32.14286	3.571429	12.5	3.571429	12.5
6	72.32143	13.39286	8.035714	12.5	38.39286
7	22.32143	0	0.892857	5.357143	16.07143
8	53.57143	18.75	12.5	8.928571	13.39286
9	56.25	14.28571	7.142857	10.71429	24.10714
10	54.46429	16.07143	21.42857	5.357143	11.60714
11	40.17857	8.928571	16.07143	4.464286	10.71429
12	8.035714	5.357143	0	0	2.678571
13	78.57143	7.142857	18.75	13.39286	39.28571
14	79.46429	38.39286	15.17857	1.785714	24.10714
15	73.21429	5.357143	24.10714	15.17857	28.57143
16	60.71429	5.357143	12.5	12.5	30.35714
17	4.464286	0.892857	0.892857	2.678571	0
18	34.82143	1.785714	10.71429	3.571429	18.75
19	64.28571	6.25	10.71429	16.96429	30.35714
20	100	0	0	95.53571	4.464286
21	52.67857	14.28571	18.75	2.678571	16.96429
22	36.60714	8.928571	16.07143	2.678571	8.928571
23	74.10714	8.928571	18.75	14.28571	32.14286
24	42.85714	8.928571	15.17857	7.142857	11.60714
25	78.57143	10.71429	9.821429	9.821429	48.21429
26	66.07143	33.92857	17.85714	4.464286	9.821429
27	21.42857	3.571429	8.035714	3.571429	6.25
28	53.57143	14.28571	16.07143	5.357143	17.85714
29	15.17857	6.25	4.464286	0	4.464286
30	73.21429	7.142857	14.28571	11.60714	40.17857
31	43.75	18.75	12.5	2.678571	9.821429
32	58.92857	15.17857	10.71429	4.464286	28.57143
33	80.35714	8.035714	19.64286	16.96429	35.71429
34	68.75	15.17857	14.28571	6.25	33.03571
35	22.32143	6.25	0.892857	0.892857	0
36	91.07143	4.464286	19.64286	56.25	10.71429
37	50.89286	12.5	16.96429	7.142857	14.28571
38	33.03571	3.571429	8.928571	2.678571	17.85714

39	19.64286	14.28571	5.357143	0	0
40	71.42857	15.17857	21.42857	8.928571	25.89286
41	73.21429	23.21429	22.32143	16.07143	11.60714
42	52.67857	8.928571	25.89286	13.39286	4.464286
43	29.46429	14.28571	6.25	0.892857	8.035714
44	99.10714	20.53571	60.71429	8.035714	9.821429
45	31.25	8.035714	7.142857	4.464286	11.60714
46	79.46429	9.821429	34.82143	26.78571	8.035714
47	23.21429	11.60714	2.678571	0	8.928571
48	86.60714	13.39286	26.78571	21.42857	25
49	81.25	13.39286	22.32143	17.85714	27.67857
50	16.96429	5.357143	3.571429	0.892857	7.142857
51	28.57143	15.17857	12.5	0.892857	0
52	32.14286	7.142857	4.464286	2.678571	17.85714
53	78.57143	9.821429	16.96429	19.64286	32.14286
54	94.64286	8.928571	14.28571	23.21429	48.21429
55	82.14286	7.142857	12.5	15.17857	47.32143
56	73.21429	3.571429	4.464286	6.25	58.92857
57	78.57143	5.357143	12.5	15.17857	45.53571
58	74.10714	16.96429	30.35714	12.5	14.28571
59	89.28571	11.60714	44.64286	26.78571	6.25
60	86.60714	9.821429	18.75	5.357143	52.67857
61	33.03571	5.357143	8.035714	1.785714	17.85714
62	69.64286	24.10714	16.07143	6.25	23.21429
63	68.75	7.142857	13.39286	12.5	35.71429
64	85.71429	1.785714	25	20.53571	38.39286
65	78.57143	4.464286	10.71429	15.17857	48.21429
66	34.82143	16.96429	8.035714	4.464286	5.357143
67	74.10714	8.928571	12.5	8.035714	44.64286
68	90.17857	13.39286	11.60714	18.75	46.42857
69	67.85714	10.71429	49.10714	3.571429	4.464286
70	71.42857	18.75	15.17857	5.357143	32.14286

Plot num	# of different tree species	# of plant species	% basal area coniferous	LAI 1 m (57.5)	Openness 1 m (0-60)	LAI 6 m (57.5)	Openness 6 m (0-60)
1	2	4	0	4.3	6.4	4.5	5.4
2	2	6	0	4.06	5.3	4.66	4.3
3	6	5	0	6.59	1.2	4.5	4.8
4	3	4	0	5.18	1.7	5.37	3.8
5	4	4	0	6.56	1.9	4.33	6
6	4	6	48.10057	8.55	5.6	5.58	9.6
7	5	3	26.39538	7.06	2.4	6.69	5.5
8	7	5	70.33424	7.31	4.8	7.75	3.7
9	7	6	50.0709	10.85	8.3	8.5	21.7
10	4	5	67.96724	11	3.5	10.26	16.6
11	3	5	22.1932	7.97	2.5	5.28	11.5
12	5	2	55.65392	7.75	3.9	7.91	5.4
13	2	5	0	6.03	3.6	4.67	9.1
14	5	5	43.22084	9.16	2.6	7.72	28.7
15	5	5	8.403186	7.49	2.7	5.73	10.2
16	1	6	0	4.9	7.4	4.38	11
17	5	2	53.52212	9.2	4	8.92	2.8
18	2	5	0	6.09	2.6	5.02	13.9
19	2	7	0	6.56	8.7	5.93	14.5
20	2	2	0	3.41	15.4	3.2	14.1
21	3	6	0	6.83	2.9	4	12
22	1	6	0	4.89	2.6	5.25	3.6
23	4	3	0	5.9	2.2	4.42	5.3
24	2	4	0	5.42	3.3	5.78	5.1
25	4	5	0	5.95	2.6	5.07	5.6
26	4	6	64.30458	9.52	14.3	7.19	15.1
27	9	5	1.248129	4.5	4	5.57	4.2
28	5	5	69.87131	4.71	3.8	10.29	7.4
29	3	4	87.34911	12.63	0.7	13.8	9.4
30	3	6	0	6.35	1.5	5.83	3.9
31	6	5	8.768405	6.05	1.4	5.56	2.4
32	5	6	0	5.02	2.2	4.1	26.8
33	1	5	0	6.57	0.8	6.24	9.6
34	4	5	0	6.35	3.9	6.68	4.3
35	4	2	73.85613	11.06	1.2	10.38	2
36	7	6	34.82036	6.48	9	5.02	20.4
37	4	6	9.414616	7.33	1.5	5.58	5.6
38	4	5	8.279749	6.59	2.2	5.65	4.9

39	5	1	2.224455	6.92	1.3	5.44	7.9
40	4	4	22.25553	5.25	7.1	7.26	8
41	6	4	0	7.07	6.6	5.06	21.8
42	6	6	0	5.13	4.4	5.62	6.5
43	9	5	72.37566	9.79	3.2	7.87	7
44	4	5	0	3.93	5.9	4.71	5.6
45	6	5	2.669691	6.97	7.6	4.68	5.1
46	6	4	0	5.54	2.1	5.39	3
47	4	5	29.09355	7.45	1.9	7.27	5.4
48	8	6	0	6.39	2.1	4.89	21.5
49	5	5	83.23675	3.61	24.7	7.89	40.8
50	3	5	57.32908	13.72	3.5	11.92	2.9
51	2	5	0	5.41	3.3	5.53	3.2
52	7	6	56.92579	6.24	2	9.86	15.6
53	5	6	8.040441	10.05	1.5	4.17	33.3
54	6	6	74.93826	4.58	8.8	8.6	25.2
55	5	6	21.53705	11.03	8.9	4.14	33.9
56	4	6	0	7.87	1.6	6.68	9.4
57	3	6	0	7.07	1.9	2.07	55.6
58	6	6	3.359144	5.3	1.8	5.29	24.2
59	5	5	97.71528	12.85	2.7	11.27	7.5
60	8	5	58.04946	9.94	9.3	4.51	43.8
61	9	4	55.86232	5.71	1.7	11.44	6.1
62	8	5	51.71676	8.34	1.1	8.13	1
63	3	5	42.49773	10.52	0.7	8.02	1.1
64	5	6	0	7.24	5.1	2.05	56.6
65	6	4	51.2121	4.55	8.6	7.44	15.5
66	5	6	10.24393	5.17	3.2	6.73	1
67	8	6	35.99573	9.71	3.3	4.81	24.8
68	5	5	0	8.11	2.5	5.61	2.4
69	7	4	22.79361	6	1.6	8.1	1.4
70	3	6	0	7.62	1.8	6.08	3.5

Appendix 4: Comparison between Complexity Score Allocated Qualitatively in the Field and Complexity Score Allocated through RDA.

Plot Number	Qualitative Complexity	RDA Complexity	Plot Number	Qualitative Complexity	RDA Complexity	Plot Number	Qualitative Complexity	RDA Complexity
1	M	M	26	M	M	51	L	M
2	M	L	27	M	L	52	M	L
3	L	L	28	M	H	53	H	M
4	M	L	29	L	M	54	H	H
5	H	M	30	H	L	55	H	H
6	H	H	31	M	M	56	H	M
7	L	L	32	M	L	57	H	L
8	M	H	33	H	M	58	M	H
9	M	M	34	H	M	59	M	H
10	M	M	35	L	M	60	H	H
11	M	M	36	M	M	61	L	M
12	L	M	37	M	L	62	M	H
13	H	M	38	M	M	63	M	H
14	H	H	39	H	L	64	M	M
15	H	M	40	M	M	65	H	H
16	M	L	41	H	L	66	L	M
17	L	M	42	M	L	67	H	H
18	M	L	43	M	M	68	M	M
19	M	M	44	M	M	69	L	M
20	M	M	45	L	M	70	M	M
21	M	M	46	M	M			
22	L	L	47	L	L			
23	M	L	48	H	M			
24	L	L	49	H	H			
25	H	L	50	L	L			