

Winter 2012

A multi-temporal image analysis of habitat modification in the Coastal Watershed, NH

Meghan Graham MacLean
University of New Hampshire, Durham

Follow this and additional works at: <https://scholars.unh.edu/dissertation>

Recommended Citation

MacLean, Meghan Graham, "A multi-temporal image analysis of habitat modification in the Coastal Watershed, NH" (2012). *Doctoral Dissertations*. 698.
<https://scholars.unh.edu/dissertation/698>

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.

A MULTI-TEMPORAL IMAGE ANALYSIS OF HABITAT MODIFICATION IN THE
COASTAL WATERSHED, NH

BY

MEGHAN GRAHAM MACLEAN
B.S., Clarkson University, 2006
M.S., University of New Hampshire, 2008

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy
in
Natural Resources and Environmental Studies

December, 2012

UMI Number: 3537820

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.

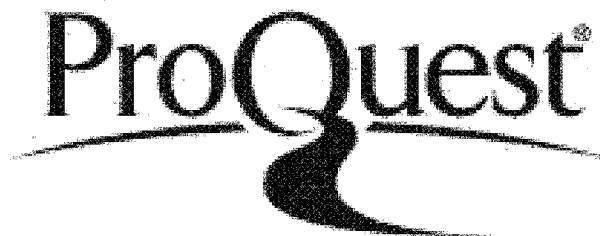


UMI 3537820

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code.



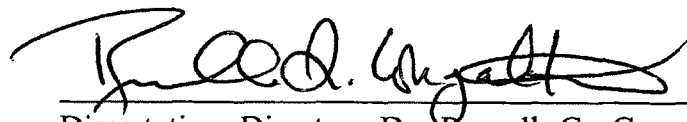
ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ALL RIGHTS RESERVED

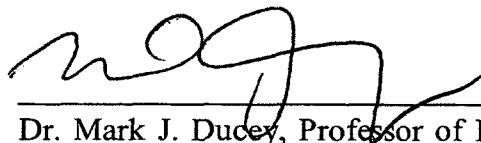
© 2012

Meghan Graham MacLean

This dissertation has been examined and approved.



Dissertation Director, Dr. Russell G. Congalton,
Professor of Remote Sensing and Geographic
Information Systems




Dr. Mark J. Ducey, Professor of Forest Biometrics
and Management



Dr. Joel Hartter, Roland H. O'Neal Assistant
Professor of Geography



Dr. Thomas D. Lee, Associate Professor of Forest
Ecology



Dr. Mary E. Martin, Research Assistant Professor of
Forest Ecosystem Analysis and Remote Sensing

September 6, 2012

Date

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Russ Congalton, for all of the support he gave me during my exploration of the worlds of science and teaching and through the process of defining myself within those worlds. Also, thank you to my committee: Dr. Mark Ducey, Dr. Joel Hartter, Dr. Tom Lee, and Dr. Mary Martin, for lending me their expertise, as well as all of their encouragement throughout this process.

I would also like to thank the many people that helped me in the field as well as in the lab, gathering both data and ideas: Mickey Campbell, Dan Maynard, Alexis Rudko, and Paul Sokoloff. Thank you to Emma Congalton, for being a great help as a NHView intern. Thank you to all of the BASAL students, both past and present, I really could not have done my work without such a wonderful group of students.

I can't thank my friends and family enough for making my life so joyous as I completed my dissertation. The positive outlook of my friends and the constant love from my family has made my time as a PhD student that much more fun and rewarding. As always, a special thank you to my mom for reading and editing my entire dissertation, to my dad for being my cheering section during my defense, to my sisters for pushing me to explore, and of course to my husband Rich, for being everything from a field assistant to editor, to constant companion and fun seeker.

Thank you to MS-33 and MS-66 NH Agriculture Experiment Station Grants (Congalton), NHView, and the Graduate School Summer TA Fellowships (2009-2011) for making my research possible.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iv
TABLE OF CONTENTS.....	v
LIST OF TABLES.....	viii
LIST OF FIGURES	x
ABSTRACT.....	xii
I. INTRODUCTION.....	1
II. BACKGROUND AND LITERATURE REVIEW.....	8
Object-based Image Analysis.....	8
Sampling Techniques for Classification and Accuracy Assessment.....	10
Multi-temporal Image Analysis	15
Atmospheric Correction	18
Forest Fragmentation/Modification Metrics	21
Invasive Species Mapping.....	31
Prediction Mapping with Presence-only Data.....	32
III. REQUIREMENTS FOR LABELING FOREST POLYGONS IN AN OBJECT- BASED IMAGE ANALYSIS CLASSIFICATION.....	36
Abstract	36
Introduction	37
Methods.....	45
Study Site.....	45
Object-Based Image Segmentation	46
Sampling.....	48
Bootstrap Calculations.....	49
Minimum Sample Requirement for Classification.....	52
Results	55
Conclusions.....	60
IV. APPLICABILITY OF MULTI-DATE LAND COVER MAPPING USING LANDSAT 5TM IMAGERY IN THE NORTHEASTERN US	65
Abstract	65
Introduction.....	66

Methods.....	68
Image Selection and Processing.....	72
Image Segmentation.....	74
Reference Data Collection.....	75
Classification.....	76
Accuracy Assessment.....	78
Results and Discussion.....	80
Single-Date and Multi-Date Maps.....	80
Comparison of the Multi-Date and Single-Date Accuracies.....	84
Conclusions.....	89
V. POLYFRAG: A VECTOR-BASED PROGRAM FOR COMPUTING LANDSCAPE METRICS.....	91
Abstract.....	91
Introduction.....	92
Software Uses.....	103
Conclusions and Future Directions.....	106
VI. A REVIEW OF USING FRAGMENTATION PROGRAMS TO IDENTIFY POSSIBLE INVASIVE SPECIES LOCATIONS.....	107
Abstract.....	107
Introduction.....	108
FRAGSTATS.....	111
CLEAR Shape Metrics Tool.....	113
Patch Analyst.....	113
PolyFrag.....	114
Methods.....	116
Land Cover Map and Invasive Species Data.....	116
Fragmentation Map and Predictive Map Creation.....	121
FRAGSTATS.....	121
CLEAR Landscape Fragmentation Tool.....	124
CLEAR Shape Metrics Tool.....	125
Patch Analyst.....	126
PolyFrag.....	127
Results and Discussion.....	128
Quantitative Assessment of the Fragmentation Programs.....	134
Qualitative Assessment of the Fragmentation Programs.....	136
Conclusions and Future Work.....	139
VII. OVERALL CONCLUSIONS.....	142
LITERATURE CITED.....	146

APPENDICES	158
APPENDIX A. NEW HAMPSHIRE LAND COVER CLASSIFICATION SCHEME	159
APPENDIX B. BOOTSTRAP CODE USED IN CHAPTER III.....	161
APPENDIX C. EXAMPLE MANUAL CHECK OF THE COMPUTATIONS COMPLETED WITHIN POLYFRAG.....	163

LIST OF TABLES

Table 1. A limited selection of the metrics produced by FRAGSTATS	26
Table 2. Current guidelines for sampling in mixed hardwood forests in North America	43
Table 3. The minimum number of prism samples necessary to meet the conditions each of the three thresholds.....	56
Table 4. Variables used in the best predictive model of SE of percent coniferous	59
Table 5. Classification system used to map the Coastal Watershed, NH.	70
Table 6. Parameters used during segmentation.	75
Table 7. A comparison of the traditional error matrix (<i>a</i>) and the area-based error matrix (<i>b</i>).....	79
Table 8. The 2010 multi-date traditional error matrix	82
Table 9. The 2010 multi-date area-based error matrix	83
Table 10. The traditional and area-based overall accuracies of the two maps created for each mapping year and the differences in the accuracies	85
Table 11. Explanatory variables used in estimating the difference between the single-date and multi-date accuracies	88
Table 12. List of landscape metrics available in PolyFrag.....	98
Table 13. Each of the land cover classes as placed in one of three categories: Fragmented classes; Fragmenting classes; and Background classes	119
Table 14. The list of invasive species found within the TNC properties.....	121
Table 15. Edge width used between fragmented and fragmenting classes.....	122
Table 16. The significant predictors of the presence of woody invasive species as determined by FRAGSTATS.....	124
Table 17. The significant predictors of the presence of woody invasive species as determined by LFT	125

Table 18. The significant predictors of the presence of woody invasive species as determined by Shape Metrics.	126
Table 19. The significant predictors of the presence of woody invasive species as determined by PA	127
Table 20. The significant predictors of the presence of woody invasive species as determined by PolyFrag.....	128
Table 21. The accuracies of the predictive maps.....	135
Table 22. Computations done by hand to verify PolyFrag.....	163
Table 23. Extra computations needed to produce diversity indices.	165

LIST OF FIGURES

Figure 1. An example of an output from LFT	29
Figure 2(a). A three-by-three reference unit, as recommended for pixel-based classification and 2(b). Segments as reference units for an OBIA classification	41
Figure 3. The Coastal Watershed of New Hampshire	46
Figure 4. Example of forested segments produced using ERDAS Imagine	48
Figure 5. One example of standard error (SE) of percent coniferous basal area versus the number of prism samples used to make the estimate of percent coniferous basal area....	54
Figure 6. The minimum number of prism samples necessary when the change in standard error (SE) for one additional prism sample is $\leq 1\%$	57
Figure 7. The standard error (SE) of the final percent coniferous (when $n=N$) for each polygon plotted against the final percent coniferous	58
Figure 8. The Coastal Watershed study area in New Hampshire	69
Figure 9. The trend of mean NDVI values computed using Landsat 5TM data in the Coastal Watershed for the 1991 growing season.....	71
Figure 10. Distribution of Landsat 5TM images for each map year with map year along the y-axis and date in that year along the x-axis.....	73
Figure 11. The hierarchical classification system used to classify each of the images	78
Figure 12(a). The single-date map created for 2010 and 12(b). The multi-date map created for 2010	81
Figure 13. The single-date and multi-date overall accuracies computed using the traditional error matrix for each year.....	84
Figure 14. The multi-date overall accuracies computed using the traditional error matrix approach and the area-based error matrix approach	86
Figure 15. ArcMAP 10 tool window for PolyFrag.....	97
Figure 16. Example of the Output Fragmented Land Cover shapefile from PolyFrag ..	101

Figure 17. Example attribute table outputs for (a) the Output Fragmented Land Cover shapefile and (b) the Output Patch Metrics shapefile	103
Figure 18. Example of the Output Fragmented Land Cover shapefile from PolyFrag ..	105
Figure 19. The 2010 vector land cover map used to study the Coastal Watershed of New Hampshire	117
Figure 20. The 2010 land cover map clipped to the extent of the sampled TNC properties with known invasive species overlaid.....	120
Figure 21. The predicted probability map created by FRAGSTATS	130
Figure 22. The predicted probability map created by LFT	131
Figure 23. The predicted probability map created by Shape Metrics	132
Figure 24. The predicted probability map created by PA.....	133
Figure 25. The predicted probability map created by PolyFrag	134
Figure 26(a). The predicted probability of the presence of woody invasive species throughout the Coastal Watershed as predicted using FRAGSTATS and 26(b). The predicted probability of the presence of woody invasive species as predicted using PolyFrag.....	138

ABSTRACT

A MULTI-TEMPORAL IMAGE ANALYSIS OF HABITAT MODIFICATION IN THE COASTAL WATERSHED, NH

By

Meghan Graham MacLean

University of New Hampshire, December, 2012

Habitat modification has become a progressively important concern as human populations increase and urbanization continues to replace natural environments with anthropogenic landscapes. Habitat modification concerns both the loss and fragmentation of environments, and these actions can have profound effects on ecosystem function, including increasing the potential of invasion by exotic species in vulnerable landscapes. The Coastal Watershed of New Hampshire (NH) has seen a 52% growth in population over the last 30 years which has led to marked urbanization and land use change. However, little has been done to study current land cover types, levels of fragmentation, and how fragmentation might be affecting the spread of woody invasive species. This research investigated new ways of using remote sensing techniques, such as object-based image analysis (OBIA) and multi-temporal image analysis, to create accurate land cover maps and corresponding fragmentation metrics. These products were

then used to determine if habitats of interest in the Coastal Watershed were potentially more susceptible to invasion by woody invasive species.

To map the Coastal Watershed, new sampling protocols were designed and implemented for labeling forest types on Landsat 5 Thematic Mapper (TM) imagery. In classification, an OBIA approach, coupled with the multi-temporal analysis, performed better than creating maps using a single Landsat 5TM image. A new fragmentation program, PolyFrag was also created to compute fragmentation metrics from the vector land cover maps generated by the OBIA approach. Finally, The Nature Conservancy (TNC) woody invasive species data were used along with the PolyFrag fragmentation maps to create a predicted probability map of the presence of woody invasive species. When compared to other programs, PolyFrag performed equally well to the more prevalent FRAGSTATS program in creating a predictive model from fragmentation metrics. However, the advantage of PolyFrag over FRAGSTATS is that it creates a fragmentation map in addition to the patch, class, and landscape metrics. Interestingly, both predictive models indicated that woody invasive species were less likely to be found in deciduous forests than in either coniferous or mixed forests. The maps and methods designed in this research are useful for fragmentation and invasive species management.

CHAPTER I

INTRODUCTION

As the human population of New England has grown considerably in the last 25 years, southeastern New Hampshire (NH) has become increasingly susceptible to issues associated with urban development and sprawl (TNC, 2010). According to The Nature Conservancy (TNC), the Coastal Watershed in southeastern New Hampshire contains some of the most valuable habitat in the state; however, it is also one of the regions with the highest population growth rates. Many of the habitats that are particularly critical to the region are at high risk of suffering irreversible losses and woody invasive species have become a threat to many of the natural plant communities (PREP, 2010). The Coastal Watershed of NH encompasses approximately 10% of the total area of the state, as well as one of the National Estuarine Research Reserves (NERR), Great Bay. The unique features of the watershed have also made it a popular place to live, leading to a 52% growth in population from 1980 to 2010 (USCB, 2012). Therefore, mapping the changing land cover within the watershed from the mid 1980's to present is of critical importance, as it can indicate how different habitats are shifting due to human expansion pressures (Vitousek, 1994; Xiuwan, 2002). Through the synergistic blend of remote sensing technologies and geospatial analyses, the observation of habitat change within the

watershed may be conducted more effectively and efficiently than in previous analyses of land use change (Xiuwan, 2002).

Urbanization in the Coastal Watershed has increased the amount of impervious surface in the area and modified many of the crucial habitats, especially forests. Forest modification includes both the loss and the fragmentation of these critical habitats. While the effects of habitat loss and fragmentation can be different, it can be difficult to study these processes separately in natural systems (Wiens, 2008). The modification of important habitats, through loss and fragmentation, can have negative effects on ecosystem function, which can in turn affect the vegetation structure, wildlife, water quality, and other ecosystem metrics of the watershed (Moran, 1984; With 2002; Fahrig, 2003; Turner, 2005; Johnson *et al.*, 2006; Fischer and Lindenmayer, 2007; Brown and Boutin, 2009). However, the extent to which fragmentation alone impacts biodiversity is still somewhat unknown (Fahrig, 2003). Therefore, it is becoming increasingly important to study the specific impacts of fragmentation on different species using the most appropriate tools for the habitat of interest.

Previous studies have indicated that forest fragmentation can increase the potential for invasive species to establish in modified areas, especially along forest edges (e.g. Moran, 1984; Brothers and Spingarn, 1992; With, 2002; Johnson *et al.*, 2006; Brown and Boutin, 2009). Therefore, accurate measures of land cover change and fragmentation are of critical importance for conserving and protecting habitats at risk of invasion. Many country-wide efforts to quantify land cover change currently exist (Homer *et al.*, 2007; C-CAP, n.d.), but there are few quantitative measures of landscape fragmentation available for the Coastal Watershed of NH. The lack of information about the Coastal Watershed

may hinder future efforts at conservation, since due to limited time and budgets, the most effective conservation efforts are generally focused and well defined in scope. However, in order to study landscape fragmentation in this area more closely, new land cover maps were needed as well as the appropriate fragmentation analysis software to compute fragmentation metrics.

Due to the need for timely and accurate creation of land cover maps, remote sensing has become essential to the process of detecting landscape modification (Foody, 2002; Congalton and Green 2009). Images captured using remote sensing are one of the preferred ways to create maps because the imagery can easily be used to create consistent, and spatially continuous, land cover maps (Foody, 2002). Currently, the most complete set of consistent remotely sensed imagery of the Coastal Watershed for the last 25 years is Landsat 5 Thematic Mapper (TM) satellite image data. Fortunately, the United States Geological Survey (USGS) has recently changed its policy regarding Landsat satellite imagery. Instead of selling each scene individually, all images are now free for anyone to download. Therefore, not only is Landsat an ideal source for quantifying landscape fragmentation in this region, but the free availability of the imagery makes the methods laid out by this research valuable to anyone wishing to study land cover change and landscape fragmentation.

In this study, quantifying current levels of forest fragmentation was of particular concern for analyzing the progression and effects of fragmentation in the Coastal Watershed. However, to study forest fragmentation, a current land cover map identifying major forest types (i.e. coniferous; mixed; and deciduous forest) was necessary. Prior to this research, such a current map did not exist and therefore had to be created. However,

forest classes in the Coastal Watershed are difficult to label, even on the ground, since these habitat types are often quite variable and complex (Justice *et al.*, 2002). Labeling these habitats on remotely sensed imagery, especially moderate resolution imagery such as Landsat 5TM (30 m pixels), is even more difficult since several tree species can be found within a single Landsat pixel. Therefore, creating accurate land cover maps of forested areas in the Coastal Watershed using Landsat 5TM can be quite challenging. Consequently, new mapping techniques were explored to try to improve the accuracy of mapping forest types in the Coastal Watershed.

Traditionally, most land cover maps created from Landsat imagery were classified using a pixel-based approach, where each pixel is classified individually. However, recent advances in image processing have introduced an object-based image analysis (OBIA) technique that mimics the way humans interpret images (Warner *et al.*, 1998; Blaschke and Strobl, 2001). The OBIA technique groups pixels with similar spectral characteristics into segments or polygons which can then be classified as a whole, instead of pixel by pixel. Each of the segments has its own characteristics, such as: size; shape; and texture; that can be used to help classify the pixels within the segments. The added information gained from using an OBIA technique may allow for the use of more specific land cover types when classifying Landsat imagery as compared to the traditional pixel-based approach (Lu and Weng, 2007). Since, as noted earlier, classifying types of forested polygons in the Coastal Watershed using Landsat 5TM images can be problematic, the new OBIA approach was chosen for this investigation. An OBIA approach coupled with multi-temporal image analysis techniques (e.g. Conese and Maselli, 1991; Wolter *et al.*, 1995; Justice *et al.*, 2002; Lu and Weng, 2007; Duveiller *et*

al., 2008) was also assessed to determine whether these methods could be used to improve the accuracy of distinguishing coniferous, deciduous, and mixed forest segments, thereby making the analysis of fragmentation in the Coastal Watershed more meaningful.

Finally, when identifying fragmentation in a landscape, it is important to compute the fragmentation metrics that are most meaningful for the study. There are currently many software programs that compute different combinations of metrics (e.g. Riitters *et al.*, 2002; Parent *et al.*, 2007; Vogt *et al.*, 2007; MacLean and Congalton, 2012c; McGarigal *et al.*, 2012). However, the most commonly used program is FRAGSTATS, a freely available program that computes many landscape fragmentation metrics on raster datasets (McGarigal *et al.*, 2012). FRAGSTATS was not deemed the most appropriate program to compute fragmentation metrics for this study for two important reasons. First, FRAGSTATS does not easily create a spatial output of the fragmentation metrics, so spatial analyses using these metrics are challenging. Also, because the land cover maps for this study were vector shapefiles, they were not compatible with FRAGSTATS unless they were converted to raster files prior to use (McGarigal *et al.*, 2012), which can impact the accuracy of land cover maps (Congalton, 1997). Although there are a few programs that will compute landscape fragmentation metrics using vector datasets, these programs are not nearly as well reviewed and do not compute the number of metrics that FRAGSTATS does. Therefore, a new fragmentation program was designed to compute similar metrics to FRAGSTATS, create a spatial output of fragmentation, and be compatible with vector shapefiles.

In summary, the focus of my dissertation was to create land cover maps of the Coastal Watershed, assess forest fragmentation, and estimate the probability of invasion by exotic species at different locations throughout the watershed. Throughout the process, many related issues involving the mapping of fragmentation and invasion were addressed. The specific objectives were to:

- 1. Determine the appropriate number of samples required to label reference samples to be used as both training and accuracy data in an OBIA approach.**
- 2. Investigate whether using a multi-temporal analysis improves the accuracy and efficiency of using an OBIA approach to create land cover maps from Landsat 5TM imagery.**
- 3. Create a new fragmentation program (PolyFrag) that can be used within ArcGIS (esri®) to investigate the extent of fragmentation of forested land cover at different scales and specificity using vector land cover maps.**
- 4. Compare the new PolyFrag program to the more traditional FRAGSTATS, Landscape Fragmentation Tool, Shape Metrics tool, and Patch Analyst programs regarding ease of use and effectiveness in creating fragmentation metrics and predicting whether the current locations of woody invasive species are correlated with areas of forest fragmentation.**

The research was accomplished using both new data collected in the field, and already existing data analyzed in the lab. Field data were collected in the Coastal Watershed and used to classify segmented Landsat 5TM images. New land cover maps were created for every three years from 1986 to 2010, meaning there were nine mapping years. Two maps were created for each mapping year, one using more traditional

classification methods and another using multi-date classification. The accuracies of the two methods for classification were compared over the nine mapping years. The 2010 map with the higher accuracy was used to analyze forest fragmentation within the Coastal Watershed using the new fragmentation program, PolyFrag. The resultant fragmentation map was then compared to observed locations of woody invasive species to assess whether the metrics computed by PolyFrag could be used to predict the presence of these invasive species. The results from PolyFrag were compared to the results created using several other fragmentation programs to determine the usefulness of PolyFrag.

The results of this work are valuable to the Coastal Watershed community, as well as to the students of landscape ecology, in several ways. First, a better program for mapping forest fragmentation, PolyFrag, was produced. Second, a map of the probability of invasive species presence in the Coastal Watershed was created. Third, and perhaps most importantly, the documented methods used in this study can be used for remotely monitoring the effects of forest modification on this landscape, and others, for years to come. These new methods and maps can be used to help inform the decisions of New Hampshire's law and policy makers as human development continues to influence the area. In the future, the methods developed in this study should be tested and applied in other areas of the world to help address the growing concern of loss and degradation of critical habitats.

CHAPTER II

BACKGROUND AND LITERATURE REVIEW

Many studies have used satellite imagery to monitor land cover change or study landscape fragmentation (e.g. Conese and Maselli, 1991; Wolter *et al.*, 1995; Du *et al.*, 2002; Paolini *et al.*, 2006; Schroeder *et al.*, 2006; Duveiller *et al.*, 2008). To understand the previous work that is the basis of this study, four major bodies of knowledge must be reviewed. They are: (1) object-based image analysis (OBIA) techniques, including the sampling methods used for the classification and accuracy assessment of maps created using an OBIA technique; (2) multi-temporal Landsat image analysis; (3) forest fragmentation/modification metrics; and (4) mapping and predicting invasive species locations.

Object-based Image Analysis

Most current land cover maps are created using computer-based land cover classification techniques with remotely sensed images (McGarigal and Cushman, 2002; Xiuwan, 2002; Jensen, 2005; Turner, 2005). In computer-based land cover classifications, there are generally two ways to analyze an image for classification: the traditional pixel-based approach; and the newer object-based image analysis (OBIA) approach (Blaschke and Strobl, 2001; Jensen, 2005; Congalton and Green, 2009). Pixel-based approaches

classify the pixels of an image individually without accounting for the context of the pixels. In contrast, the OBIA approach groups contiguous pixels with similar properties into segments or polygons. The resulting segments represent areas of similar spectral response that can be classified as a whole, rather than pixel by pixel (Baatz *et al.*, 2001; Desclée *et al.*, 2006). In a well performed segmentation, the pixels within a single segment should all have the same land cover type. Using an OBIA approach increases the number of attributes that can be used to identify the pixels within each segment, including segment shape and texture (Baatz *et al.*, 2001; Desclée *et al.*, 2006; Lu and Weng, 2007).

This innovative process mimics how a human interprets an image, and is considered an improvement over the traditional pixel-based approaches (Warner *et al.*, 1998; Blaschke and Strobl, 2001; Desclée *et al.*, 2006; Congalton and Green, 2009). The resulting groups of pixels, or segments, reduce the ‘salt and pepper’ effect often found on land cover maps created using pixel-based classification approaches. Therefore, the OBIA approach creates maps that are more visually pleasing as well as potentially more accurate. The segments are also more easily translated to management units than individually classified pixels, so maps with segments are more useful to land management groups.

Although there are many advantages to using the OBIA approach over the pixel-based approach, there are some caveats when dealing with segments rather than individual pixels. For instance, the added processing time required for grouping pixels based on similarities can be cumbersome for extremely large datasets. Also, the methods used to create the segments can also highly influence the success of the subsequent

classification of the segments, making OBIA approaches more complex than pixel-based approaches (Blaschke, T, 2010). In OBIA classification approaches, the image is broken into segments of similar unlabeled pixels so that there is less spectral variation within each segment than between the segments, based on chosen input parameters such as maximum variability or minimum segment size (Batz *et al.*, 2001; Desclée *et al.*, 2006). However, land cover types are naturally heterogeneous. Therefore, the segmentation of the image may or may not always place conterminous pixels of the same land cover type into the same segment. Additionally, it is likely that the pixels within each of the segments will have slightly different spectral properties (Blaschke and Strobl, 2001). The heterogeneity of the pixels within each of the segments can make the process of classifying the segments more complex than classifying individual pixels. However, in many instances, the added information regarding segment properties that can be used in OBIA classification processes outweigh the increased complexities of using an OBIA approach.

Sampling Techniques for Classification and Accuracy Assessment

When generating a land cover map from remotely sensed data, reference units are needed for both training and validation (Congalton *et al.*, 1983; Congalton, 1991; Gopal and Woodcock, 1994; Foody, 2002; Congalton and Green, 2009). Training data are used to guide the classification of the image and validation data are used to assess the accuracy of the map. The reference units are usually either collected through photo-interpretation or ground visits (Congalton and Green, 2009). When performing a classification, the training and validation data are assumed correct, so that any discrepancies between the

land cover map and the validation data are assumed to be errors in the map, rather than in the validation data (Congalton, 1991; Gopal and Woodcock, 1994; Stehman, 1995; Foody, 2002; Congalton and Green, 2009). Therefore, the accuracy of the reference data is of the utmost importance when creating a land cover map.

Generally, attaining acceptable thematic accuracies of ground collected reference data is straightforward when using a pixel-based approach, especially when the image being classified is of medium to high spatial resolution (Stehman and Czaplewski, 1998). In these images, the reference data units are generally defined as squares of at least 3x3 pixels in size within an area of a single land cover type. Since the reference units cover a relatively small area and should contain only one land cover type, the variability of the land contained within the reference unit should be small. Therefore, since the variability of the land in the reference unit is small, the reference unit can often be easily classified using a single sample observation within the unit. However, as the pixels get larger or more variability is captured within a single pixel, it may be more difficult to accurately label a reference unit using a single observation (Stehman and Czaplewski, 1998). In forest classifications this is especially important since, in the case of eastern US forests for example, stands can be highly variable and depending on the level of detail desired in the classification, finding pixels of 'pure' forest classes may be difficult for a low resolution image.

In an OBIA approach, the pixels are grouped so that within-segment variances are less than between-segment variances, with the thresholds for both minimum size and maximum variability of the segments defined by the analyst creating the map (Blaschke and Strobl, 2001). Therefore, the segments are generally not all the same size and are

dependent on the properties of the image (Desclée *et al.*, 2006; Congalton and Green, 2009; Blaschke, 2010). In order to validate maps created using an OBIA approach, the reference units should be identical to the segments (i.e. polygons) used in classification, rather than pixels, so that the units are directly comparable to the map segments (Congalton and Green, 2009; Radoux *et al.*, 2011). However, with an effective OBIA approach, the average segment usually contains considerably more pixels than a 3x3 pixel square, and the polygons range in size from the minimum mapping unit (mmu) specified to much larger (Desclée *et al.*, 2006; Drăguț and Blaschke, 2006; Blaschke, 2010; Radoux *et al.*, 2011). Since most of the reference units are larger than the 9 pixel squares recommended in the pixel-based approach, there is a wider variety of pixels within each reference unit, making it more difficult to label the polygons (Stehman and Czaplewski, 1998; Congalton and Green, 2009). Currently, there is not a recommended sampling method for determining the map class of polygon reference units in remote sensing (Stehman and Czaplewski, 1998; Jensen, 2005). However, since the larger reference units are generally more variable, a single sample within the reference unit may not be sufficient to label that unit in many land cover types (Stehman and Czaplewski, 1998). As part of the work of this dissertation, as discussed in Chapter III, the appropriate sampling strategy for collecting OBIA approach reference data given the specific study area and objectives of this research was determined.

Once a method for sampling is chosen, reference data are collected, and a land cover map is created, it is then necessary to test the accuracy of the map. When using polygons as validation units, the statistics used to determine the accuracy of the map are different than those used in a pixel-based approach, where the size of each reference unit is the

same (Radoux *et al.*, 2011). In the traditional pixel-based approach, overall accuracy is estimated using:

$$\hat{\pi} = \frac{\sum_{i=1}^n C_i}{n} \quad (1)$$

where $\hat{\pi}$ is estimated overall accuracy, C_i is equal to 1 or 0 if the validation data unit i is correctly classified on the map or not, and n is the number of validation data units collected. Currently, many researchers use the same equation to calculate accuracy when polygons are used as part of an OBIA approach (Radoux *et al.*, 2011). However, this equation does not account for the variability in polygon sizes in the accuracy assessment.

The actual accuracy of the map should be computed using:

$$\pi = \frac{\sum_{i=1}^N C_i S_i}{\sum_{i=1}^N S_i} \quad (2)$$

where N is the total number of segments in the image, and S_i is the area of a single unit i . However, the accuracies for all of the polygons within a map are usually not known, so two alternative estimates of overall accuracy have been proposed. The first equation just replaces N with n :

$$\hat{\pi} = \frac{\sum_{i=1}^n C_i S_i}{\sum_{i=1}^n S_i} \quad (3)$$

which effectively weights the pixel-based accuracy assessment by the size of the validation polygons (Radoux *et al.*, 2011). However, Radoux *et al.* (2011) propose another estimate of overall accuracy which incorporates the size of the remainder of the polygons not used as validation polygons. Radoux *et al.* (2011) note that in most mapping exercises the size, S_i , of all of the polygons in the study area are known, but the accuracy, C_i , is not. They propose that the information gained from knowing the S_i of the

remainder of the unsampled polygons can reduce the variance of the estimate of overall accuracy. Their estimate of accuracy is:

$$\hat{\pi} = \frac{1}{S_T} (\sum_{i=1}^n C_i S_i + \hat{p} \sum_{i=n+1}^N S_i) \quad (4)$$

where S_T is the total area of the map and \hat{p} is the estimate of the probability of an object being classified correctly. As long as C_i is independent of S_i , \hat{p} can be estimated using:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n C_i \quad (5)$$

Radoux *et al.* (2011) found that when using this estimate of accuracy, fewer polygons were needed as validation data to achieve the same accuracy and variance estimates as compared to the units needed in a pixel-based approach.

Since the accuracy of maps created using an OBIA approach must be calculated while taking into account the area of the reference units, an error matrix that incorporates area into each cell is appropriate for reporting thematic accuracy in conjunction with the traditional error matrix (Congalton *et al.*, 1983). The new polygon OBIA error matrix would be set up similarly to the traditional error matrix, but instead of each reference unit having the same weight, the individual cells would reflect the total area of the reference units that fell into that cell. These new methods are discussed in Chapter IV.

As with the pixel-based approach, the new method of accuracy assessment for maps created using OBIA still assumes that the reference polygons are 100% correct (Radoux *et al.*, 2011). However, the accuracy of the reference units can be affected by the positional and thematic accuracy of the sampling method used to decide the label of the reference polygons. As discussed above, the variability within a polygon, or segment, often makes it difficult to label a polygon with a single observation. Therefore, the number of necessary observations for each reference polygon should be determined so

that the reference polygon labels can be as close to 100% accurate as possible. With the reference labels as accurate as possible, the accuracy assessment of the map created using the OBIA approach should reflect the accuracy of the map, rather than the accuracy of the reference data.

Multi-temporal Image Analysis

There have been many studies that have used multi-temporal image analysis to either perform a change detection, or improve the accuracy of a land cover classification for a single date (e.g. Conese and Maselli, 1991; Lunetta *et al.*, 1993; Wolter *et al.*, 1995; Du *et al.*, 2002; Lu *et al.*, 2002; Paolini *et al.*, 2006; Schroeder *et al.*, 2006; Duveiller *et al.*, 2008). One of the earliest programs designed specifically to look at land cover change over time was the Landsat Pathfinder program that used Landsat images from 1973, 1986, and 1992 (± 1 year) to identify areas of land cover change throughout the conterminous United States (Lunetta *et al.*, 1993).

More recent studies have applied the knowledge gained from land cover change detection studies to classification processes by using multiple images to create a single land cover map. In these studies, the information contained in Landsat images taken at different times throughout the growing season was used to improve the accuracy of the creation of a single land cover map (e.g. Conese and Maselli, 1991; Wolter *et al.*, 1995; Justice *et al.*, 2002; Lu and Weng, 2007; Duveiller *et al.*, 2008). Phenological changes in vegetation types observed throughout the growing season can be useful in distinguishing land cover types that would otherwise be very difficult to determine using a single date of imagery (Lu and Weng, 2007). For instance, Justice *et al.* (2002) found that they were

better able to separate different forest cover types when completing the 2001 NH Land Cover Dataset by using Landsat imagery from different dates throughout the growing season. However, many of the images used in that study were also from different years, presenting further complication. In general, this new application of multi-temporal image analysis pushes the boundaries of land cover classification techniques toward greater accuracy, and allows for the separation of certain land cover types during classification that otherwise would remain indistinct.

Several issues arose in the early application of multi-temporal image analysis to Landsat imagery. The two primary issues with multi-temporal image analysis are registration errors and radiometric errors (Lunetta *et al.*, 1991; Lunetta *et al.*, 1993; Jensen, 2005; Congalton and Green, 2009). Registration errors are introduced when images are not correctly georeferenced to the ground or to each other, meaning the image, or parts of it, have not been given the correct x, y locations (Lunetta *et al.*, 1991; Lunetta *et al.*, 1993; Jensen, 2005; Congalton and Green, 2009). Radiometric error occurs when images used in the multi-temporal analysis have different radiometric properties, usually due to either sensor differences or changes in the environment between image acquisition dates (Hall *et al.*, 1991; Moran *et al.*, 1992; Lunetta *et al.*, 1993; Dwyer *et al.*, 1996; Lunetta *et al.*, 1998; Song *et al.*, 2001; Jensen, 2005; Paolini *et al.*, 2006).

Currently, the Landsat data of the US that are now freely available to the public from USGS have been processed with the Standard Terrain Correction process (USGS, n.d.). The images have gone through terrain and geometric correction so that the images are all in the same format and displayed using the WGS84 UTM map projection and coordinate system. In addition, all geometric correction has been completed using cubic

convolution. Therefore, all of the Landsat images disseminated by USGS should be registered correctly, and images of the same area should overlay properly, as well as be visually appealing. Unfortunately, the accuracy of the standard correction done by USGS will depend on the accuracy of the ground control points and the Digital Elevation Model (DEM) used in the terrain correction, and some of the precision of the raw data is lost during convolution (USGS, n.d.). If the same ground control points and DEM were used for all images in the time series, the images should, at a minimum, be comparable to each other. However, a check for geometric error should always be completed prior to a multi-date analysis regardless of whether the same ground control points and DEM were used for all images in the time series. With the standard correction, only the radiometric properties associated with either sensor degradation/error and terrain have been corrected for these images, leaving many other sources of radiometric error in the Landsat images.

Radiometric error is usually defined as occurring when the radiance recorded by the sensor is not an accurate representation of the radiance leaving the surface of the object of interest (Hall *et al.*, 1991; Jensen, 2005; Paolini *et al.*, 2006; Schroeder *et al.*, 2006). As opposed to reflectance, which is the light that bounces off of an object in any direction, radiance is the “radiant intensity per unit of projected source area in a specified direction” (Jensen, 2005 p. 193). In other words, radiance can be described as the amount of light leaving an object in a certain direction as observed at a specific location away from the object (such as by an orbiting optical sensor). Therefore, satellite sensors record the amount of light radiating in the direction of the sensor per unit surface area observed by the sensor. The digital numbers (DN) logged by the sensor represent radiance values recorded as the light enters the optical sensor.

Radiometric errors can be caused by factors both internal and external to the satellite that change how the satellite records radiance values. Most internal errors, caused by factors such as: random bad pixels; line-start/stop problems; or striping; are typically corrected by USGS prior to dissemination. The most prevalent external issue for satellite remote sensing is the effect of the Earth's atmosphere on the transmission of light from the surface of the Earth to the satellite sensor (Hall *et al.*, 1991; Lunetta *et al.*, 1991; Dwyer *et al.*, 1996; Jensen, 2005; Paolini *et al.*, 2006). When light is radiated off of the surface of the Earth and passes through the atmosphere to the sensor, the properties of the atmosphere can scatter and/or absorb the light so that the light reaching the sensor is different than the light that was radiated off the surface in that direction (Song *et al.*, 2001; Lu *et al.*, 2002; Jensen, 2005; Paolini *et al.*, 2006). The issues caused by the changing atmosphere are generally addressed using atmospheric correction.

Atmospheric Correction

When completing a multi-temporal image analysis, atmospheric correction can typically be accomplished using either absolute or relative methods. Generally the first step of any atmospheric correction uses information about the sensor to convert the DN values to at-satellite radiance, and then relates these converted radiance values to either scaled surface reflectance values or other radiance values (Markham and Barker, 1986; Schroeder *et al.*, 2006). Absolute atmospheric corrections relate the at-satellite radiance values to scaled surface reflectance values for the same locations (Song *et al.*, 2001; Lu *et al.*, 2002; Paolini *et al.*, 2006; Schroeder *et al.*, 2006). Relative atmospheric corrections relate the

radiance values of one image to the values of another image of the same location (Song *et al.*, 2001; Lu *et al.*, 2002; Paolini *et al.*, 2006; Schroeder *et al.*, 2006).

Currently, the most common form of relative image-to-image atmospheric correction technique is regression analysis (Lu *et al.*, 2002; Jensen, 2005; Paolini *et al.*, 2006). In regression analysis, pseudo-invariant features (PIFs), or areas that are assumed to be constant between two images, are chosen for the two images that are being relatively corrected. The DN or radiance values of the PIFs are compared on a bispectral plot and a regression line is defined to relate the values of the pixels from one image to the values of the pixels from the other image (Jensen, 2005). One image is chosen as the base image and the second image is atmospherically corrected to match the conditions of the base image using the modeled relationship. While relative atmospheric corrections are easier to accomplish than absolute corrections, the resulting radiance values of the corrected images do not have any relation to the surface reflectance values of the same locations.

Unfortunately, absolute correction techniques can be very time consuming, are more processing intensive than relative correction techniques, and may require *in situ* atmospheric data to accurately relate at-satellite radiance values to surface reflectance (Moran *et al.*, 1992; Lu *et al.*, 2002). However, in a multi-temporal image analysis, using absolute atmospheric correction can be very advantageous since it allows any corrected images to be compared to each other, as well as surface reflectance values for different land cover types (Jensen, 2005). There are two general types of absolute atmospheric correction techniques used for Landsat imagery: image-based and physically-based models (Lu *et al.*, 2002). Image-based models use only information that can be attained from the image to perform atmospheric correction, while physically-based models rely on

in situ data about the atmosphere to correct for atmospheric effects. There are several different models that fall within each of these categories.

The most common physically-based models used to correct Landsat images are the Second Simulation of the Satellite Signal in the Solar Spectrum (6S) model and the Moderate Resolution Atmospheric Radiance and Transmittance model (MODTRAN) (Lu *et al.*, 2002; Jensen, 2005; Kotchenova *et al.*, 2006). These physically-based models use known properties of gasses within the atmosphere, as well as data about the atmosphere collected at the same time as the imagery, to model the absorption and scattering of light under the given conditions (Vermote *et al.*, 1997; Kotchenova *et al.*, 2006). While this form of atmospheric correction generally yields the best results, *in situ* data for historical images or remote locations can be difficult to obtain, in which case image-based atmospheric correction techniques may be necessary.

The most common forms of image-based models are the Dark-Object Subtraction (DOS) model, and modifications of the DOS, such as the Cosine of the Solar Zenith Angle (COST or DOS2) method (Song *et al.*, 2001; Schroeder *et al.*, 2006). The DOS method assumes that the darkest objects on the image should actually be black and therefore have near zero percent reflectance values (Moran *et al.*, 1992; Chavez, Jr., 1996; Jensen, 2005). Therefore, any radiance values recorded with values greater than 1% for the dark objects are attributed to atmospheric scattering and are removed from the image (Chavez, Jr., 1996; Song *et al.*, 2001; Jensen, 2005). The COST method uses the same technique as the DOS method, but includes an approximation of atmospheric transmittance loss when converting at-satellite radiance to surface reflectance (Song *et al.*, 2001; Schroeder *et al.*, 2006). These methods generally produce consistent results,

do not need any *in situ* data, and can be applicable for imagery with or without atmospheric data.

One method that has become quite prevalent in multi-temporal image analyses combines absolute and relative correction techniques (Schroeder *et al.*, 2006). The so-called “absolute-normalization” method corrects one base image using an absolute correction method so that the values on the base image represent surface reflectance values, and then a relative correction method is used to correct the remainder of the time-series images to the conditions of the base image (Schroeder *et al.*, 2006). The method has shown great promise in studies using multi-temporal image analysis (Schroeder *et al.*, 2006) because it reduces both the processing time and the need for ancillary data for each of the images in the time series, while still producing images with values comparable to surface reflectance values.

Forest Fragmentation/Modification Metrics

When discussing changing landscape or forest conditions, the terms forest fragmentation or modification are traditionally used interchangeably (Haila, 2002; McGarigal and Cushman, 2002; Fahrig, 2003; Fischer and Lindenmayer, 2007; Wiens, 2008). In many cases, these terms are used to mean either the combined effects of both forest loss and the breaking apart of forests, or just the breaking apart of forests independent of forest loss (McGarigal and Cushman, 2002; Fahrig, 2003). Forest fragmentation/modification is often studied to determine the effects of these landscape changes on the biodiversity of the remaining forest fragments (Blake and Karr, 1987; Andr n, 1994; McGarigal and Cushman, 2002; Fahrig, 2003; Prugh *et al.*, 2008; McGarigal *et al.*, 2012). These

differing definitions of forest fragmentation/modification have appeared throughout the years because there are a number of dominant theories regarding the relationship between landscape modification and biodiversity, and each of these theories have different underlying assumptions about the state of fragments within a landscape (Andrén, 1994; Haila, 2002). Since the terms 'forest fragmentation' and 'forest modification' may be used slightly contrarily in different studies, this research will use the following definitions: *forest loss* will refer to the reduction of forest; *forest fragmentation* will refer to the breaking apart or the change in configuration of the forest, independent of forest loss; and *forest modification* will refer to the combined effects of forest loss and fragmentation.

Fahrig (2003) analyzed several studies on habitat modification and attempted to separate out the effects of habitat loss and fragmentation. In this study, it was incredibly difficult to separate the effects, but when it was possible, there was a distinct negative effect of habitat loss on species biodiversity within forests, but fragmentation was as likely to have a positive effect as a negative one. In general, the species found in smaller patches were usually a selection of the species found in the larger patches (Blake and Karr, 1987; Flather and Sauer, 1996; Rosenblatt *et al.*, 1999; Boulinier *et al.*, 2001; Damschen *et al.*, 2008; Brown and Boutin, 2009). The sensitivity of species to habitat loss can often be correlated with their dispersal ability, although most studies found that animal species with high dispersal capabilities were the most sensitive to habitat loss than others (Blake and Karr, 1987; Flather and Sauer, 1996; Gibbs, 1998; Boulinier *et al.*, 2001). In vegetation studies, some species showed more sensitivity to habitat loss (Brown and Boutin, 2009) and species with high dispersal capability were generally

found in all forest patches while species with lower dispersal capabilities were only found in the patches connected to the source of seeds (Damschen *et al.*, 2008). Therefore, the effect of habitat loss on species richness and composition can be very dependent on a species dispersal ability, so it is hard to generalize how much of a reduction in species richness there will be for patches in response to habitat loss.

The positive or negative effect of habitat fragmentation on biodiversity is different than what would have been predicted by the theory of island biogeography, but it is not completely unexpected. The metapopulation concept predicted that if habitat amount remained the same, a few smaller patches close together may provide habitat for more species than one large patch if the amount of total habitat remained the same and the species were able to disperse between all patches (Levins, 1969, 1970; Pulliam, 1988). Again, the dispersal capabilities of the species becomes a very important factor in determining if habitat fragmentation will have a positive or negative effect on species richness and composition of patches. Factors such as the type of land cover fragmenting the landscape and the scale at which the landscape is fragmented can have significant effects on dispersal capabilities of species, so even determining the dispersal capabilities of a species within a landscape may not be straight forward (Moran, 1984; With, 2002; Damschen *et al.*, 2008). Since so much depends on individual species dispersal capabilities, it is very difficult to generalize how habitat modification will affect species richness and composition within specific fragments. Most landscape ecology literature shows that habitat modification will in general have a negative effect on biodiversity since habitat loss has such a negative effect on biodiversity and habitat fragmentation

often does not have as significant effect on biodiversity, in either the positive or negative direction (Fahrig, 2003).

The reaction of woody invasive species to habitat modification may be even more complicated, since often it is reliant on how native species are affected (Moran, 1984; With, 2002; Brown and Boutin, 2009). If we assume that native species are negatively affected by forest modification, it may mean that some resources become available at the edge of the patches where woody invasive species may be able to establish, provided they are able to get there (Brothers and Spingarn, 1992; With, 2002). Again, this process is very dependent on the dispersal ability of the invasive species, competition between species, as well as the intervening habitat type (With, 2002; Johnson *et al.*, 2006; Prugh *et al.*, 2008). Corridors have often been proposed as a mechanism to connect patches to allow between patch movement of species with low dispersal capabilities (Levey *et al.*, 2005; Proches *et al.*, 2005). However, because woody invasive species are usually fairly efficient at moving along forest edges, these corridors may increase invasive species spread (Proches *et al.*, 2005). Since the direct effects of forest fragmentation are relatively unknown, especially on invasive species, it has become increasingly important to study the interaction between forest fragmentation and invasive species, particularly woody invasive species, spread.

Several software programs have been designed to analyze the amount of forest fragmentation occurring in the landscape using spatial data such as satellite image derived land cover maps (Riitters *et al.*, 2002; Parent *et al.*, 2007; MacLean and Congalton, 2010; McGarigal *et al.*, 2012;). Some of the more popular programs are: (1) FRAGSTATS; (2) Patch Analyst (PA); (3) the Landscape Fragmentation Tool (LFT);

and (4) Shape Metrics. FRAGSTATS, developed at the University of Massachusetts, Amherst, is a strictly statistical program used to assess the fragmentation of a landscape (McGarigal *et al.*, 2012). The program provides excellent quantitative measures but does not produce visual results. The Landscape Fragmentation tool (LFT) from the Center for Land Use Education and Research (CLEAR) at the University of Connecticut, is a newer program, built upon older work done by Riitters *et al.* (2002), that produces a visual output of forest fragmentation and is written in Python so it can be run in ArcGIS with a graphical user interface (GUI) (CLEAR, 2009). However, LFT does not compute any fragmentation metrics. Both FRAGSTATS and LFT can only use raster land cover maps for processing. Patch Analyst (PA) and Shape Metrics both are able to use vector datasets in processing, but they are far more limited in their ability to compute fragmentation metrics than FRAGSTATS, and therefore are rarely found in the literature.

FRAGSTATS produces measures of fragmentation at three different landscape scales: (1) patch; (2) class; and (3) landscape. Several measures of fragmentation are produced for each scale, including, but not limited to the list in Table 1. When quantifying the amount of fragmentation of different forest types, the measures at the class level usually produce the most useful information. For instance, the total amount of class area is available for each individual forest type. FRAGSTATS will also compute the amount of edge and core habitat within each forest type. Edge habitat is defined as the area along the border between the two different land cover types. Generally, a group of land cover classes is chosen as the land cover types of interest, or fragmented landscape (e.g. forest) and the other group of land cover classes are defined as the

fragmenting landscape (e.g. development) (McGarigal *et al.*, 2012). Some areas can be considered neither, or background land cover types, such as open water.

Table 1. A limited selection of the metrics produced by FRAGSTATS (McGarigal *et al.*, 2012).

Scale	Metric
Patch	Patch Area Patch Perimeter Core Area Number of Core Areas Proximity Index (Isolation)
Class	Total (Class) Area Percentage of Landscape Number of Patches Total Edge Total Core Area Core Area Percentage of Landscape
Landscape	Total Area Number of Patches Patch Density Total Edge Total Core Area

In FRAGSTATS, edge areas are only delineated along the boundary of the fragmented land cover types when they are bordered by a fragmenting land cover type. The distance that the edge extends into the land cover type of interest is defined by the user to delineate areas that are thought to be suffering from effects from the bordering fragmenting land cover type. In the latest version of FRAGSTATS (v. 4.0, released in March 2012) different edge widths can be defined for different fragmented/fragmenting land cover type interactions. This ability to model different effects between land cover types is very important in landscape ecology because, for instance, the effect of a roadway on a forested area may have more far reaching effects than an agricultural field,

depending on what traits of forest fragmentation are being assessed. The areas within the patch that are not included in the edge habitat are often referred to as “core” areas and are thought to suffer from fewer effects from the surrounding land cover types (McGarigal *et al.*, 2012). The two largest drawbacks to using the FRAGSTATS program are its inability to calculate metrics for vector datasets and its limited spatial output. Vector files must be converted to raster before being used in FRAGSTATS, which is often not recommended, depending on the methods used to create the vector dataset (Congalton, 1997). Vector to raster conversion is generally not a suitable option when trying to compute fragmentation metrics because the choice of pixel size can have profound effects on the accuracy of the resulting raster land cover map (Congalton, 1997). However, if the land cover map is in raster format, the latest version of FRAGSTATS will output a raster file with pixels labeled with the patch number it was placed in for analysis. With some manipulating of the data, the output fragmentation metrics can be tied to these patches for use in further spatial analysis, but the tying of the fragmentation metrics to the spatial data is not intuitive and remains problematic.

LFT is a complementary program to FRAGSTATS, in that it produces a raster map of fragmentation, but no landscape metrics. LFT uses a raster land cover map recoded to three categories: forest; non-forest; and other; where forest is the land cover type being fragmented, non-forest is fragmenting the forest, and other is background. The program then takes the input data, along with a user defined edge width (which is limited to a single value), and produces a map of forest fragmentation (Figure 1). The seven categories of the output map are: non-forest, patch, edge, perforated, small core (<250 acres), medium core (250-200 acres), and large core (>500 acres). Patch, edge, and core

areas are the same as those produced with FRAGSTATS, but LFT adds the category of 'perforated', defined as an edge area around a small section of non-forest completely encased by core area, and patch, which is an area of forest not large enough to have any core habitat (CLEAR, 2009). Although the visual output produced by LFT is quite useful and can be utilized in further spatial analyses, it is limited in how it defines forest, non-forest, and other, since it will only accept these three land cover categories (MacLean and Congalton, 2010). Without doing further analysis, there is no way of determining how different forest land cover types are being affected by forest fragmentation, whereas in FRAGSTATS, each land cover type of interest can be assessed separately with different edge widths.

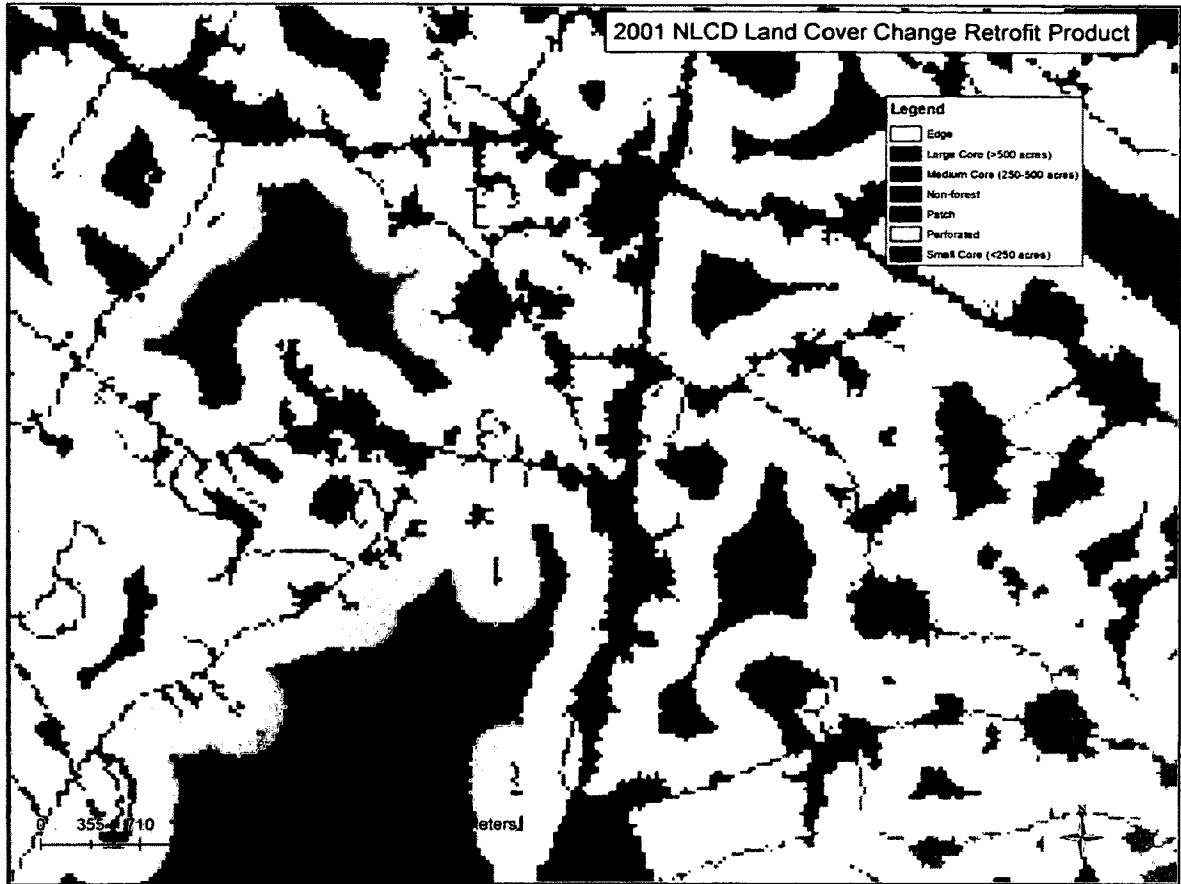


Figure 1. An example of an output from LFT using the 2001 National Land Cover Database (NLCD) Land Cover map as input. Open water was considered background in this analysis (shown in white).

Shape Metrics and Patch Analyst (PA) use similar measures of fragmentation as those used in FRAGSTATS to produce a map of forest fragmentation for a given area, but these two programs differ from those described above in that they compute these metrics using vector shapefiles (CLEAR, 2009; Rempel *et al.*, 2012). PA is most similar to FRAGSTATS, in that it computes a small subsection of the metrics produced by FRAGSTATS, but the program will also create a vector shapefile with an associated attribute table detailing the patch metrics for the landscape. Another function of PA will

also create a shapefile of core areas, but is only able to use a single edge width in its creation of core areas. Unlike FRAGSTATS, PA runs within the ArcGIS (esri®) framework, and therefore can be easier to use for those who are familiar with ArcGIS. Similarly, Shape Metrics also runs within ArcGIS. However, this tool only computes landscape metrics that have historically been difficult to compute for polygons (CLEAR, 2009), such as shape cohesion or spin, which makes these metrics also less common and therefore less comparable to metrics in the current literature. Unfortunately, because the metrics computed by Shape Metrics are more difficult computationally than many of the other widely used landscape metrics, the program also takes considerably longer to run than the other three presented here.

Since the most common landscape fragmentation programs that are able to work with vector data, Shape Metrics and PA are so limited, a new program was written in the course of this research. The new program, PolyFrag, is introduced and tested in Chapters V and VI, respectively. The advent of this program will help researchers that would like to use an OBIA approach to classification create landscape fragmentation metrics that are similar to those produced by FRAGSTATS do so with the flexibility of defining different edge widths, and without having to first convert their data to raster format. Fortunately, the conversion from raster to vector is generally risk free, so raster land cover maps, converted into vector format, can also be used in PolyFrag. The program has the added benefit of running within the ArcGIS (esri®) platform (ArcGIS 10 or higher) as a new tool, so it has a very user-friendly interface for those that are familiar with ArcGIS. In addition, the tool also intuitively creates spatial maps of landscape fragmentation, so spatial analyses with these metrics are quite easy.

Invasive Species Mapping

In general, woody invasive species are best suited for disturbed landscapes that allow for the establishment of the species when there is high resource availability, such as light and/or soil nutrients (e.g. Moran, 1984; With 2002; Fahrig, 2003; Turner, 2005; Johnson *et al.*, 2006; Fischer and Lindenmayer, 2007; Brown and Boutin, 2009). Currently in New Hampshire, most woody invasive species are limited to forest patch edges and old fields (e.g. agriculture or other cleared areas like clear cuts) that have been allowed to regenerate and are transitioning into forested land cover types (Johnson *et al.*, 2006). These areas provide hospitable habitat for invasive species where there is greater availability of the resources they require and the landscape is not constantly being disturbed (Moran, 1984; Brothers and Spingarn, 1992; Johnson *et al.*, 2006). Therefore, native vegetation loss, land use change, and fragmentation may increase the potential for invasion of a landscape by increasing the number of disturbed sites and total available edge area of the remaining forest patches.

A landscape that is constantly kept clear of vegetation may not allow the woody invasive species to establish, thereby limiting its movements and possibly allowing for some containment of the species. However, if the disturbed landscape is open to vegetation, the invasive species may flourish within the disturbed landscape, out growing and perhaps outcompeting native species (With, 2002; Johnson *et al.*, 2006; Brown and Boutin, 2009). The dispersal capabilities of the invasive species will in part determine whether the species can spread beyond the fragments of the landscape (With, 2002). In some cases, if the woody invasive species is a poor disperser, the species may be limited to certain areas. However, if the forest is modified to create long stretches of edge

habitat, the woody invasive species may have no problem dispersing along edges. Therefore, the configuration of forest patches and the amount of edge habitat present are important to either promoting or limiting invasive spread (With, 2002). In 1984, Moran found that there were more introduced species in forest edge habitat that abutted residential land cover types than those that bordered either agriculture or road, indicating that the dispersal of introduced species can be enhanced by human activity. Therefore, the type of edge can also play a role in determining the potential of invasion of certain forest fragments.

Prediction Mapping with Presence-only Data

Unfortunately, mapping invasive potential in a landscape can be quite difficult. When creating a map of potential invasion, a predictive model must be created. Generally, the predictive model uses known information about the landscape, such as current land cover maps, maps of fragmentation, and known locations of invasive species to determine what characteristics are significant in predicting invasive species presence (Zaniewski *et al.*, 2002; Anderson *et al.*, 2003; Brotons *et al.*, 2004; Elith *et al.*, 2006; VanDerWal *et al.*, 2009; Barbet-Massin *et al.*, 2012). The accuracy of these models is highly influenced by the accuracy of the land cover map used in modeling, as well as what fragmentation program is used to determine fragmentation metrics. However, even more important to the accuracy of the model is the invasive species data used to indicate known locations of presence and absence of invasive species.

When mapping something rare, such as invasive species, it is quite uncommon to find data that records both presence and absence (Zaniewski *et al.*, 2002; Elith *et al.*, 2006).

Most datasets include only information on presence, and many of the methods used to record presence may not have followed any form of statistically valid sampling protocol. So, unless a great deal of time and money is expended to gather new data with a statistically sound sampling method that records both presence and absence, modeling potential invasion is done using less than ideal presence-only data. There are generally two ways of using presence-only data to create predictive models: (1) through the use of iterative models that can use presence-only data; or (2) by creating pseudo-absence data by assuming most locations of presence were recorded. Both methods have advantages and disadvantages in individual scenarios.

Since presence-only data has become so prevalent, especially in the case of historical data, several models have been designed to create predictive maps with presence-only datasets. Some of these models include: Bioclimatic Envelope Model (BIOCLIM); DOMAIN; and Ecological Niche Factor Analysis (ENFA) (Elith *et al.*, 2006; Brotons *et al.*, 2004). BIOCLIM uses climatic data and presence-only data to create a species profile for a specific study (Busby, 1991). The species are profiled across a number of environmental variables (n), creating an 'environmental envelope' in n-dimensional space of all possible min and max values for each environmental variable. The 'environmental envelope' can be used to model species' presence, and a predictive map can be created by comparing the environmental variables at an unknown location to the 'environmental envelope' that was produced using the known presence locations. If the variables at the unknown location fall within the 'environmental envelope' created in BIOCLIM, the location can be predicted as a location of possible presence. DOMAIN works similarly to BIOCLIM, but DOMAIN uses the Euclidean distance (in n-dimensional space) rather

than an envelope to predict whether a location should be predicted as presence or absence (Carpenter *et al.*, 1993). Both BIOCLIM and DOMAIN can be implemented in DIVA-GIS (Hijmans *et al.*, 2001). Another model, ENFA, also compares unknown areas on a map to known presence location in n-dimensional space, but instead of using envelopes or distances, ENFA compares the distributions of the known presence location (i.e. species distribution) along each environmental factor with the distribution of all of the cells in the image (i.e. global distribution). Factors that best predict presence are chosen when the marginality (the difference between the global mean and the species mean) is the largest (Hirzel *et al.*, 2002).

If the assumption can be made that nearly all presence data were recorded in the study area, pseudo-absence data can be created from within the study area. Most studies have found that using pseudo-absence data and more typical logistic regression techniques are actually more accurate than the presence-only methods (Zaniewski *et al.*, 2002; Brotons *et al.*, 2004; Elith *et al.*, 2006; Barbet-Massin *et al.*, 2012). When creating pseudo-absence data, locations are chosen from within the sampled area to represent areas 'absent' of the invasive species. As long as most areas of presence were recorded during sampling, these pseudo-absence locations should effectively represent areas absent of the species (Zaniewski *et al.*, 2002; Barbet-Massin *et al.*, 2012). These locations can either be chosen by random methods, or by using some form of weighting to attempt to match any bias in the presence data. If the presence data were sampled using a known bias (e.g. only along roadways), the pseudo-absence data should be sampled in the same way (VanDerWal *et al.*, 2009; Barbet-Massin *et al.*, 2012). However, if the presence data were sampled randomly, or if the bias is unknown, the pseudo-absence locations should

be chosen at random (Barbet-Massin *et al.*, 2012). When sampling at random, some studies recommend setting a minimum distance between pseudo-absence locations and presence locations to minimize false positives and spatial autocorrelation (Barbet-Massin *et al.*, 2012). In either sampling method, as long as there is a sufficient number of pseudo-absence locations (i.e. equal to or larger than the number of presence locations, preferably over 1000 samples), any false-negatives should be inconsequential as compared to the number of actual absence locations (Barbet-Massin *et al.*, 2012).

With presence-only data, it is important to choose a modeling technique wisely. Each prediction mapping endeavor will require a different strategy depending on what data are available. The method used for modeling should be dependent on whether the assumption that the presence data represents all known locations of presence within the study area can be met. If that assumption cannot be met, the presence-only modeling techniques should be used. However, if it is assumed that nearly all locations of presence were recorded, pseudo-absence data should be created, since the regression modeling techniques produce more accurate models (Zaniewski *et al.*, 2002; Brotons *et al.*, 2004; Elith *et al.*, 2006; Barbet-Massin *et al.*, 2012).

CHAPTER III

REQUIREMENTS FOR LABELING FOREST POLYGONS IN AN OBJECT-BASED IMAGE ANALYSIS CLASSIFICATION

Abstract

The ability to spatially quantify changes in the landscape and create land cover maps is one of the most powerful uses of remote sensing. Recent advances in Object-Based Image Analysis (OBIA) have also improved classification techniques for developing land cover maps. However, when using an OBIA technique, collecting ground data to label reference units may not be straight forward, since these segments generally contain a variable number of pixels as well as a variety of pixel values, which may reflect variation in land cover composition. Accurate classification of reference units can be particularly difficult in forested land cover types, since these classes can be quite variable on the ground. This study evaluates how many prism sample locations are needed to attain an acceptable level of accuracy within forested reference units in Southeastern New Hampshire. Typical forest inventory guidelines suggest at least ten prism samples per stand, depending on the stand area and stand type. However, because OBIA segments group pixels based on the variance of the pixels, fewer prism samples may be necessary in a segment to properly estimate the stand composition. A bootstrapping statistical technique was used to find the necessary number of prism samples to limit the variance

associated with estimating the species composition of a segment. Allowing for the lowest acceptable variance, a maximum of only six prism samples was necessary to label forested reference units. All polygons needed at least two prism samples for classification.

Introduction

Currently, land cover and land use change are some of the most important factors for quantifying global ecological change and predicting future change to our environments (Vitousek, 1994; Xiuwan, 2002). Land cover change is indicative of changes in ecosystem goods and services, such as water quality, nutrient cycling, and overall biodiversity (e.g. Binkley and Brown, 1993; Vitousek, 1994; Xiuwan, 2002; Foody, 2002). Due to the need for timely and accurate creation of land cover maps, remote sensing has become inherent to the process of detecting land cover change. Traditionally, most land cover maps were created by classifying images using a pixel-based approach, where each pixel is classified individually. However, recent advances in image processing have introduced an Object-Based Image Analysis (OBIA) approach that mimics the way humans interpret images (Warner *et al.*, 1998; Blaschke and Strobl, 2001).

When using the OBIA approach, pixels with similar spectral characteristics are grouped into segments and the segments are then classified as a whole, instead of pixel by pixel. The size of the segments is generally determined by the variability of the spectral characteristics of the pixels in the segment: the more variable the pixels on an image, the smaller the segment; the less variable, the larger the segment. Once created,

the segments have their own characteristics, such as size, shape, texture, and a variety of zonal statistics, that can be used to help classify those segments. Other advantages of using an OBIA approach may include a less 'noisy' land cover map and groupings of pixels that are more representative of management units (Robertson and King, 2011). Therefore, maps created using OBIA can be more understandable and useful for land managers and owners than the maps created using a pixel-based approach. The added information gained and usefulness of the maps created using an OBIA approach have made it a preferred method for land cover classification (Warner *et al.*, 1998; Blaschke and Strobl, 2001; Desclée *et al.*, 2006; Congalton and Green, 2009).

When classifying an image to use as a land cover map, reference sample units are needed to use as both training and validation data. Training data are used to guide the classification of the image and validation data are used to assess the accuracy of the resultant map. Reference sample units are usually either collected through photo-interpretation or ground reconnaissance (Congalton and Green, 2009). The accuracy and interpretability of the classification is fully dependent on the accuracy of both the training data and the validation data. The accuracy of the training data will influence the success of the classification, and the validation data are assumed to be 100% correct in an accuracy assessment, so that any discrepancies between the land cover map and the validation data are assumed to be errors on the map (Congalton, 1991; Gopal and Woodcock, 1994; Stehman, 1995; Foody, 2002; Congalton and Green, 2009). Therefore, the sampling approach used to collect the reference data can highly influence the success of the land cover classification. In very broad classes or relatively homogeneous landscapes, photo-interpretation of reference data may be sufficient for accurate

collection of reference data. However, in highly variable landscapes, ground visits may be necessary to ensure the accuracy of the reference data.

In the more traditional pixel-based classification approach, a small group of pixels (a three-by-three cluster or larger) within a homogeneous land cover type is the recommended size for a reference sample unit (Congalton and Green, 2009). When the imagery is of medium to high spatial resolution and pixels are relatively small, the area covered by the reference unit is also quite small and generally covers only a small amount of variability in the landscape (Figure 2a). Therefore, a single observation taken on the ground within that reference sample unit may be sufficient for accurately labeling that group of pixels. However, if the pixels of the image are large, or the area covered by the reference unit is larger and/or more variable, a single ground sample observation will often not be adequate for labeling the reference unit (Congalton and Biging, 1992). When using an OBIA approach, the reference units should be segments (i.e. polygons), rather than a small square of pixels, so that the units are directly comparable to the map segments (Congalton and Green, 2009; Radoux *et al.*, 2011). With an effective OBIA approach, the average segment usually contains substantially more pixels than a three-by-three pixel square, and the polygons range in size from the minimum mapping unit (mmu) to the maximum allotted spectral variability, which, in homogeneous segments can produce very large segments (Desclée *et al.*, 2006; Drăguț and Blaschke, 2006; Blaschke, 2010; Radoux *et al.*, 2011). Therefore, each reference unit in an OBIA approach encompasses more variation in the landscape than in a pixel-based approach, even with relatively high spatial resolution imagery. With greater variability in the reference units, more than a single sample observation may be necessary to accurately

label each unit (Figure 2*b*). Potentially high landscape variability within reference units combined with insufficient sampling would lead to inaccurate reference data, which would in turn make it increasingly more difficult to design and implement a classification scheme. A poorly designed classification scheme and inaccurate reference data would cause the accuracy of the resulting land cover map to be quite low (Foody, 2002). Thus, when larger, more variable reference units are used, it is imperative to determine how many sample observations are needed to accurately label a reference unit.

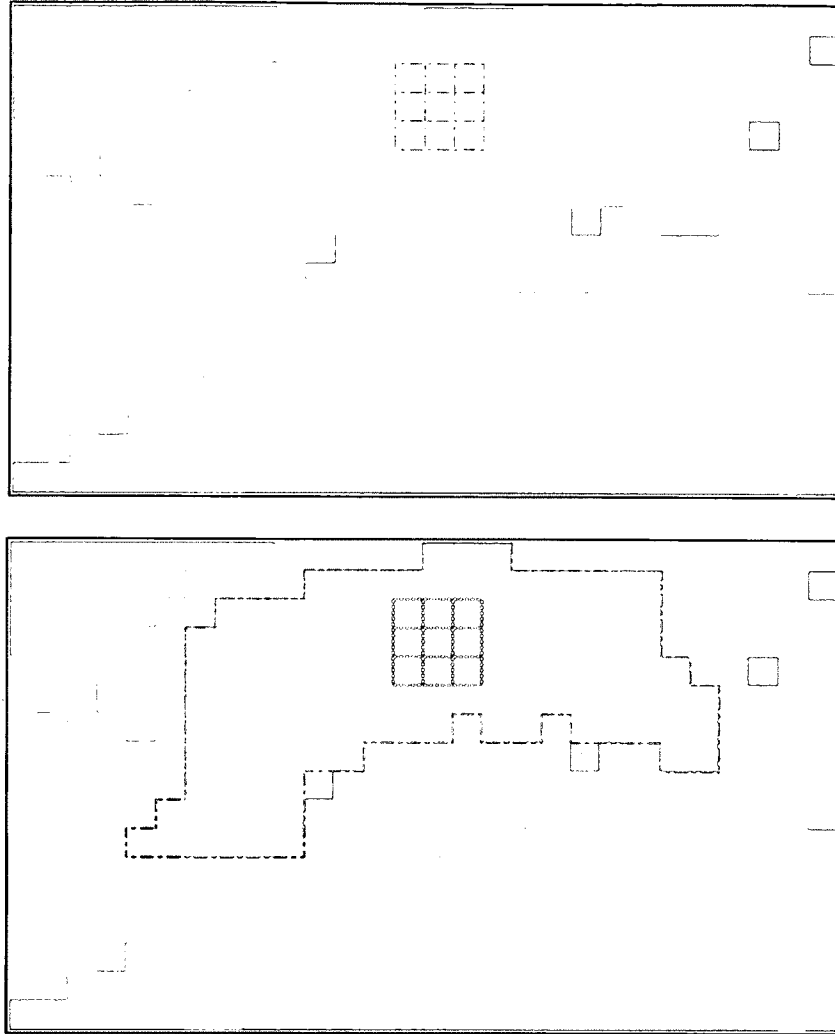


Figure 2(a) (above). A three-by-three reference unit (dashed black box), as recommended for pixel-based classification, does not encompass a large amount of landscape variability and a single observation within the reference unit is sufficient for labeling the unit. The example raster dataset was generated to be a clear representation of landscape variability on a medium resolution image, such as Landsat 5 TM, which has 30 m pixels. **2(b) (below).** In an OBIA classification, segments are used as reference units (shown by the black dashed polygon) and a single observation in the segment does not accurately assess the majority of the reference unit.

In many landscapes, forested habitats provide substantial values and are the subject of intensive mapping efforts, especially for natural resource, human-environment, or wildlife studies (e.g. Congalton *et al.*, 1993; Wolter *et al.*, 1995; Warner *et al.*, 1998; Foody, 2002; Justice *et al.*, 2002; Riitters *et al.*, 2002; Xiuwan, 2002; With, 2002; Johnson *et al.*, 2006; Duveiller *et al.*, 2008). However, forest stands can be quite variable in comparison to other land cover types (Justice *et al.*, 2002). Therefore, more ground visits are usually necessary to accurately label reference data for forested land cover types (Squires and Wistendahl, 1975; Held and Wistendahl, 1978). In particular, differentiating between coniferous, deciduous, and mixed forest land cover types can be particularly challenging in the northeastern United States, since forest composition changes continuously (Justice *et al.*, 2002).

In most projects, sampling efforts are limited by time and money. Accordingly, ways of reducing the quantity and/or increasing the efficiency of sampling, while still attaining accurate results, are always desirable. One recommended method for quickly sampling forests for composition is through prism sampling (i.e. horizontal point sampling or Bitterlich sampling). Prism sampling is a quick and efficient method of quantifying tree basal area using a variable radius plot, wherein the probability of sampling a tree is proportional to its size (Bitterlich, 1947; Squires and Wistendahl, 1975; Held and Wistendahl, 1978; Mitchell *et al.*, 1995; Husch *et al.*, 2003). Basal area is defined as the cross sectional area of a tree, inside the bark, at breast height (1.3 meters above the ground), and the total basal area per tree species can be determined for each prism plot (Bitterlich, 1947). Prism sampling does not require any plot set-up and only trees that are large enough, or close enough, to be counted when using a prism with a given Basal Area

Factor (BAF) are included in the sample for any one particular location. Different BAFs are chosen based on a general understanding of the density and size of the trees in the forest stand that is going to be sampled (Mitchell *et al.*, 1995; Husch *et al.*, 2003). However, the number of prism samples necessary to accurately label a polygon created from an OBIA approach has not been assessed in the literature.

Previous prism sampling studies, focused primarily on traditional timber inventory objectives, have suggested that ten or more prism samples are necessary to quantify stand structure and composition, and the number is dependent on the size and type of stand (Held and Wistendahl, 1978; Mitchell *et al.*, 1995; Husch *et al.*, 2003). Current guidelines for mixed hardwood forests, modified from the standard forest inventory text by Husch *et al.* (2003), are as follows (Table 2):

Table 2. Current guidelines for sampling in mixed hardwood forests in North America, modified from Husch *et al.* (2003). Original values from Husch *et al.* (2003) were given in acres, as shown in parentheses.

Area of Stand (ha)	Number of Prism Samples Required
<4.05 (<10 ac)	10
4.05-16.19 (11-40 ac)	2.47 per ha (1 per ac)
16.19-32.37 (41-80 ac)	20 + 1.235*(area in ha) (20 + 0.5*(area in ac))
32.37-80.94 (81-200 ac)	40 + 0.6175*(area in ha) (40 + 0.25*(area in ac))
>80.94 (>200 ac)	Use equation (1)

If the area is greater than 80.94 ha, the following equation is used:

$$n = \frac{t^2(CV^2)}{E^2} \quad (1)$$

where n = the number of required prism samples

t = Student's t -value

CV = coefficient of variation (in %) of the target variable in the stand

E = allowable error of the estimate of the target variable (in %), which can be calculated using:

$$E = \frac{t * SE_{\bar{x}}}{\bar{x}} * 100 \quad (2)$$

where \bar{x} = estimated mean of the target variable

$SE_{\bar{x}}$ = standard error of the mean

In most remote sensing studies, few if any quantitative measurements are taken to label reference units (Congalton and Biging, 1992). However, quantitative measurements for determining forest composition, such as through prism sampling, are important since it ensures that labeling is objective and accurate for each reference unit, especially in areas where forest composition is quite variable (Congalton and Biging, 1992). But, for most remote sensing studies, which involve the collection of hundreds of different forested reference units (Foody, 2002; Congalton and Green, 2009), the collection of ten prism samples per reference unit may not be feasible, depending on the available resources for completing ground surveys. Moreover, the guidance exemplified by Husch *et al.* (2003) focuses on accuracy for a single continuous variable (such as timber volume per unit area), not accuracy of cover type classification. However, since reference sample units are generally assumed to be 100% correct in remote sensing studies, labeling these polygons correctly and efficiently is incredibly important. In previous studies, a maximum allowable error (E) of between 4% to 10% (for a 95% confidence level) has been deemed acceptable for labeling reference sample units generated using remote sensing techniques (Anderson *et al.*, 1976; Fitzpatrick-Lins, 1981), but ultimately

allowable error should be determined by the needs of each individual study based on available resources and the purpose of the classification.

Therefore, this research aims to determine whether the ten prism sample minimum is necessary in forested polygons created through the segmentation of a Landsat 5 TM image. Since the segmentation of the image limits the amount of pixel variability within a polygon, we hypothesize that the segmentation also limits the amount of variability of tree composition within the polygon. Limiting the variability of the trees within the polygon would effectively delineate stands with more uniform composition and less variability than implied by traditional inventory guidelines (Husch *et al.*, 2003).

Methods

Study Site

The study was performed in the Coastal Watershed of New Hampshire (Figure 3). The Coastal Watershed is approximately 61% forested and is dominated by hemlock-hardwood-pine forest stands. These stands are generally mixed and contain a variety of species including: *Pinus strobus* (white pine); *Tsuga canadensis* (Eastern hemlock); *Fagus grandifolia* (American beech); *Quercus* spp. (oak species); as well as some *Acer* spp. (maple species) and *Betula* spp. (birch species). Classification can be difficult in these forests because they are generally quite variable in composition over short distances (Justice *et al.*, 2002). For this study, we focused on the ability to separate and classify coniferous, deciduous, and mixed forest types, using class definitions derived from a previous study of the area (Justice *et al.*, 2002). Coniferous forest was defined as a forested polygon with more than 65% coniferous basal area per unit area; deciduous

forest was less than 25% coniferous basal area per unit area; and mixed forest was between 25% and 65% coniferous basal area per unit area.

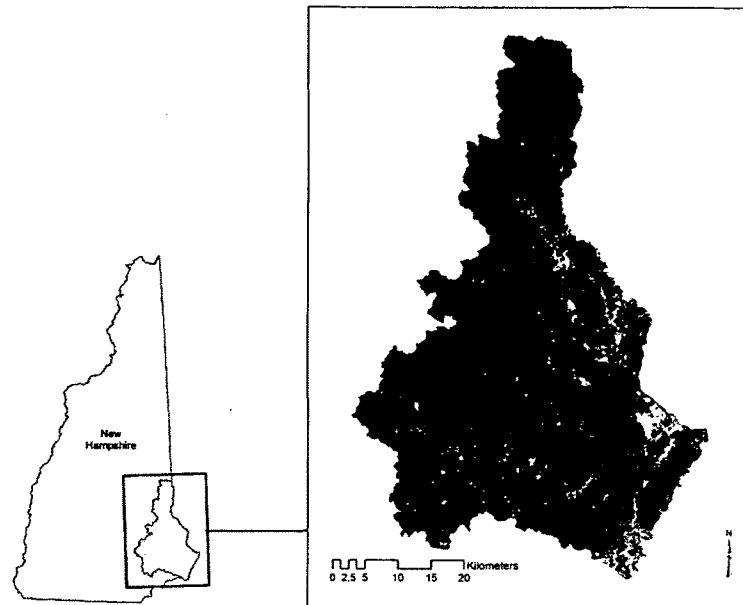


Figure 3. The Coastal Watershed of New Hampshire. The image is the base image for the study: a Landsat 5 TM image from 30 August 2010.

Object-Based Image Segmentation

A cloudless Landsat 5 TM image from 30 August 2010 was selected for use in this study. The Landsat image was from path 12 and row 30, and all bands, except for the thermal band (band 6), were used in the analysis, all with 30 m pixels. The image was clipped to the extent of the Coastal Watershed in New Hampshire and all six remaining bands were corrected for atmospheric effects using the cosine of the solar zenith angle (COST) method (Chavez, 1996). A normalized difference vegetation index (NDVI; Rouse *et al.*, 1974) band and the first three tasseled cap bands (brightness, greenness, and wetness; Kauth and Thomas, 1976) were also calculated and added to the six-banded Landsat

image (all except the thermal band). A vector layer delineating forest and non-forest areas of the watershed was created using the 2001 NH Land Cover Dataset (Justice *et al.*, 2002). The separation of the forested areas from the non-forested areas allowed the segmentation to be completed using only the reflectance values of the forested areas, rather than the entire image. Since the inclusion of the non-forested areas increases the variance of reflectance values to be grouped, the segmentation could not delineate different forest stands as efficiently using the entire image (Drăguț and Blaschke, 2006). The benefits of first delineating forest from non-forest using the NH Land Cover Dataset far outweighed the possibility of including small areas of non-forest, or missing small areas of forest for this project, especially since all of the study sites were chosen from segments within the forest delineation.

Once all forested areas were delineated, the segmentation of the forested areas of the image was completed using ERDAS Imagine Image Segmentation software (ERDAS, Inc.) with a minimum segment size of nine pixels, a minimum value difference of 0.02, and a variance factor of 2.50 (Figure 4). The minimum value difference determines how different the spectral values of each segment must be to be considered a separate polygon; a low number creates more segments, while a larger number creates fewer segments. The variance factor determines how important variation in pixel values within a segment is for expanding a segment; a small value restricts the amount of variation allowed in a single segment, while a larger number allows for more. These numbers are unique to each image.

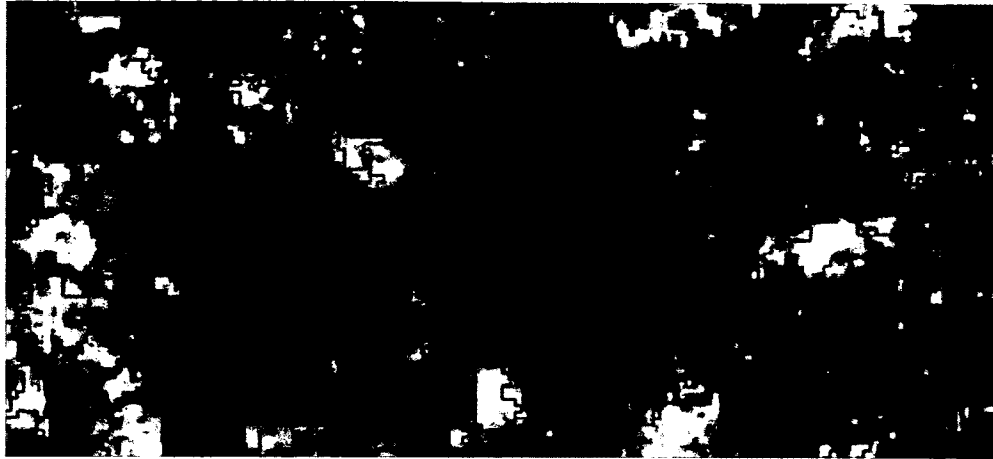


Figure 4. Example of forested segments (in black) produced using ERDAS Imagine. The image is a Landsat 5 TM image of Durham, NH.

Sampling

In order to determine how many prism samples were necessary to accurately label different forested polygons in the Coastal Watershed, several locations within the watershed were extensively sampled and analyzed. The properties involved in the study were either owned by the University of New Hampshire (UNH) and managed by the UNH Office of Woodlands and Natural Areas, or located in Pawtuckaway State Park. All of the locations were sampled using the UNH Office of Woodlands and Natural Areas protocol using a prism with a BAF $4.59 \text{ m}^2/\text{ha}$ ($20 \text{ ft}^2/\text{acre}$), which is the recommended BAF for operational inventory in this region (Wiant *et al.*, 1984; Ducey, 2001). The protocol dictates that prism samples are to be systematically located throughout the stands so that there is one sample per hectare. Therefore, polygons delineated through the segmentation process were chosen so that a minimum of ten prism samples could be placed in each polygon (i.e. 10 ha or larger in size). Ten polygons of each forest type (resulting in 30 total polygons) were chosen and sampled for this study. The locations of

the prism samples were determined by walking North-South transects systematically spaced approximately 100 m apart through each of the polygons, and placing a sample every 100 m along each transect. Following standard techniques, each tree determined to be “in” using the prism was identified by species and tallied at each location (Husch *et al.*, 2003).

Bootstrap Calculations

Each of the 30 polygons analyzed in this study contain a different number of total collected prism samples ranging from ten to 46. Each prism sample also contains a different number of total trees. All polygons were treated as independent units, with tree totals for each polygon produced by summing tree counts for each species at all prism samples within that unit. In this study, since classification was based on percent coniferous basal area, the total number of coniferous trees at each prism location was summed to produce the total coniferous basal area at each location. The same was done for the deciduous trees. Combining the species into two groups allowed for the relatively easy calculation of the estimate of percent coniferous by basal area. The totals could then be summed for a “stand” or polygon, as in a traditional forest stand inventory.

For each polygon, a bootstrap estimate (Efron and Tibshirani, 1993) of the percent coniferous trees within the polygon was generated in the R statistical software package, along with the standard deviation (SD) of that estimate. Using a bootstrap estimator, instead of calculating the SD of percent coniferous within a polygon using the variability of percent coniferous in each of the prism samples, ensures that no assumptions are made about the distribution of the population of coniferous trees in each prism sample, but that

instead the assumption is that the prism samples are independent. A bootstrap estimate was completed to estimate total percent coniferous for each polygon so that the SD of the bootstrap estimate for each possible sample size (n) was actually representative of the standard error (SE) of the mean of all estimates of total percent coniferous. The bootstrap process computes an estimate of total percent coniferous for each bootstrap run and then averages those estimates to come up with a mean estimate of total percent coniferous. The standard error of the mean represents the range of all means possible given all possible combinations of n prism samples. For instance, if six prism samples are chosen randomly (with replacement) from the 20 possible prism samples in a particular polygon, the estimate of total percent coniferous will depend on which six prism samples are chosen. Therefore, the bootstrap estimate was necessary to produce all possible estimates of the total percent coniferous given n samples and illustrate how variable that estimate is within a given polygon. If the SD of the percent coniferous in the polygon was computed on a sample by sample basis (basically how variable percent coniferous is from one prism sample to another) using a single selection of n prism samples, the SD of percent coniferous would be highly dependent on which samples were chosen and not a true reflection of how variable the estimate of total percent coniferous is when selecting only a few prism samples.

The bootstrap estimate of the percent coniferous basal area within a polygon was calculated by first summing the tree counts (x) for each specified group of trees (coniferous or deciduous) (b) over the number of selected prism samples (n) for any individual bootstrap run (m). The prism samples were randomly selected with replacement from the total number of prism samples (N) within the polygon. These

values were then divided by the total tree count for that run (m), resulting in the estimate of how much total basal area each group represents (in %). The estimate was calculated using:

$$\widehat{D}_b^m = \frac{\sum_{i=1}^n x_{ib}}{\sum_{j=1}^B \sum_{i=1}^n x_{ij}} * 100 \quad (3)$$

where \widehat{D}_b^m = the estimate of the percent basal area of each group

n = the number of prism samples used to create the estimate

x_{ib} = the tree count of one group (b) at one prism sample location (i)

m = the bootstrap run number

Note that the estimate in equation (3) is not the simple mean of the percent coniferous on a sample-by-sample basis. The estimation process was repeated 400 times ($M=400$). The average of percent basal area for the 400 estimates was calculated on a per species basis, using:

$$\widehat{D}_b = \frac{1}{M} \sum_{m=1}^M \widehat{D}_b^m \quad (4)$$

where \widehat{D}_b^m = the estimate of the average percent basal area of each group calculated using equation (3)

M = the number of times the estimates are calculated in equation (3) ($M=400$)

The SD of the estimates of percent basal area was calculated for each group using:

$$\widehat{SD}_b = \sqrt{\frac{1}{M-1} \sum_{m=1}^M (\widehat{D}_b^m - \widehat{D}_b)^2} \quad (5)$$

For a given sample size n , \widehat{SD}_b represents the SE of the mean of the bootstrap estimates of percent coniferous if the inventory were conducted with that sample size. These steps were repeated for n in 2:N, so that the estimates of \widehat{D}_b and \widehat{SD}_b for each group were calculated for all possible numbers of prism samples. The \widehat{SD}_b was then used to

determine the variability in the prediction of the percent basal area for each group using n prism samples.

Minimum Sample Requirement for Classification

The classification system for this study utilizes the percent coniferous in each polygon to label the polygon as deciduous, coniferous, or mixed. Therefore, the certainty with which a classification can be made is based on the variability of percent basal area of coniferous trees within the polygon. The accuracy of an estimate of the true percent basal area of coniferous tree species depends on the SD of the percent basal area of coniferous tree species, and also on the sampling intensity (number of samples). There is some natural variability in percent coniferous basal area and basic considerations from sampling theory predict a declining marginal return in accuracy for each additional sample (Thompson, 2002). Our objective was to determine at what point that declining return meant that additional sampling effort would not be lead to substantial increases in accuracy.

Three thresholds were used to determine when additional prism samples did not result in a substantially better estimation of percent coniferous. These thresholds were used to find the minimum number of prism samples needed before the effort required for additional samples was greater than the reduction in the SE of the estimated percent coniferous. Since sampling for different projects can entail different costs, the three thresholds presented in this work represent three different sampling costs. To calculate the relationship between the reduction of SE and the number of additional prism samples, the SE of the percent coniferous was plotted against the number prism samples (n) used

to calculate that SE, and a power curve was fit to the relationship (Figure 5). The three thresholds for sampling were designed so that when the change in SE over the change in n is: 1) less than 1% per sample ($\frac{dSE}{dn} \leq 1\%$); 2) less than 2% per sample ($\frac{dSE}{dn} \leq 2\%$); and 3) less than 4% per sample ($\frac{dSE}{dn} \leq 4\%$). The first threshold resulted in the most conservative estimate of minimum samples needed or for when the cost of sampling is low. The third threshold resulted in the least conservative estimate, which may be useful when the marginal cost of additional samples is high. The most conservative estimate should result in a higher accuracy of reference data labeling since the precision with which percent coniferous is estimated is relatively high, while the least conservative should have lower accuracy.

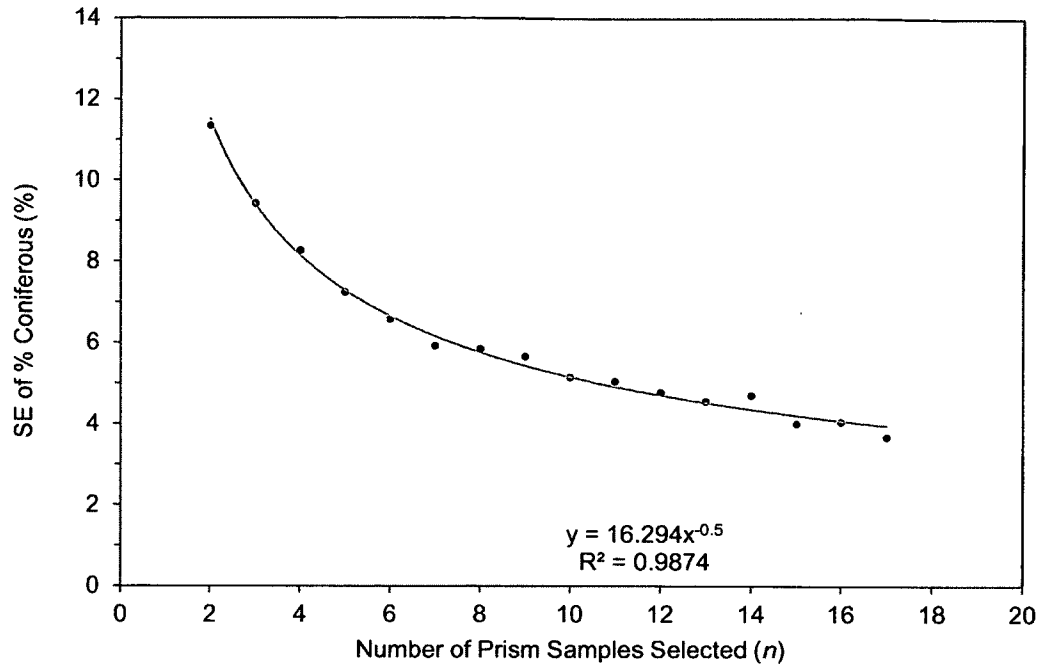


Figure 5. One example of standard error (SE) of percent coniferous basal area versus the number of prism samples used to make the estimate of percent coniferous basal area in one particular polygon. The fitted power curve is shown as the grey line.

When labeling reference data units, not only is the SE of percent coniferous important in labeling the unit correctly, but how close the estimate of the mean is to the boundary value between land cover types can also determine whether or not a polygon is labeled correctly. For instance, using this study's classification scheme the difference in labeling a polygon estimated at $80\% \pm 6\%$ coniferous when $n=3$ and labeling the same polygon estimated at $80\% \pm 2\%$ coniferous when $n=4$ is inconsequential, and the more cost effective choice would be to use only three samples. However, if the polygon was estimated at $70\% \pm 6\%$ when $n=3$ and $70\% \pm 2\%$ at $n=4$, the difference could have an impact on the labeling of the polygon depending on the three samples actually chosen in

the $n=3$ scenario. Assuming that all percent coniferous values were equally probable, the maximum possible error in classification, due purely to the missed opportunity of an additional prism sample, was computed as follows for each of the three thresholds:

$$E_{\max} = \frac{dSE}{dn} * 2(B) \quad (6)$$

where B = the number of boundaries between classes (e.g. 2 for this study)

Using equation (6), E_{\max} for $\frac{dSE}{dn} \leq 1\%$ was 4%, meaning a maximum of 4% more of the polygons could be mislabeled solely by not adding an additional sample. Similarly, E_{\max} for $\frac{dSE}{dn} \leq 2\%$ was 8% and E_{\max} for $\frac{dSE}{dn} \leq 4\%$ was 16%. These values only represent the error associated with not taking another prism sample and do not reflect any other error associated with the sampling process. However, the maximum errors in classification for both the 1% and 2% thresholds fall within the generally accepted allowable errors (between 4% and 10%), while the 4% threshold represents a more extreme case, where each additional sample is very costly and accuracy must be sacrificed due to resource limitations.

Results

The minimum number of prism samples necessary to meet each of the thresholds was calculated for each of the 30 sampled polygons and the results were compared in order to determine the appropriate guideline for sampling polygons created using an OBIA approach (Table 3). In the most conservative case ($\frac{dSE}{dn} \leq 1\%$) the largest number of prism samples needed to meet the threshold was eight, while the minimum was three samples. For the less conservative thresholds, the number of samples needed was much

lower, with two being the lowest number of samples needed. For all cases there was a significant positive linear relationship between SE of the estimate of total percent coniferous when $n=N$ and the number of prism samples needed to meet the thresholds ($p<0.05$). The overall average minimum number of prism samples needed in the sampled polygons for the three thresholds were six, four, and three, from most conservative to least conservative. When averaged by forest type, the differences in number of prism samples needed are negligible. However, as seen in Figure 6, as the stand composition becomes less mixed, the number of prism samples necessary decreases. This observation follows the same general trend as the SE of the final percent coniferous in each stand (Figure 7). Both the number of prism samples and the SE of the final percent coniferous attain a maximum when the stand is between 25% and 65% coniferous (i.e., is a mixed stand). In the non-mixed stands the number of prism samples needed and the SE of final percent coniferous decreases as the final percent coniferous decreases from 25% and increases from 65%, especially past 75%.

Table 3. The minimum number of prism samples necessary to meet the conditions each of the three thresholds. The table summarizes the results from the 30 sampled polygons.

	Number of Prism Samples when $dSE/dn \leq 1\%$	Number of Prism Samples when $dSE/dn \leq 2\%$	Number of Prism Samples when $dSE/dn \leq 4\%$
Maximum	8	5	3
Minimum	3	2	2
Deciduous Average	5	4	3
Mixed Average	6	4	3
Coniferous Average	6	4	3
Overall Average	6	4	3

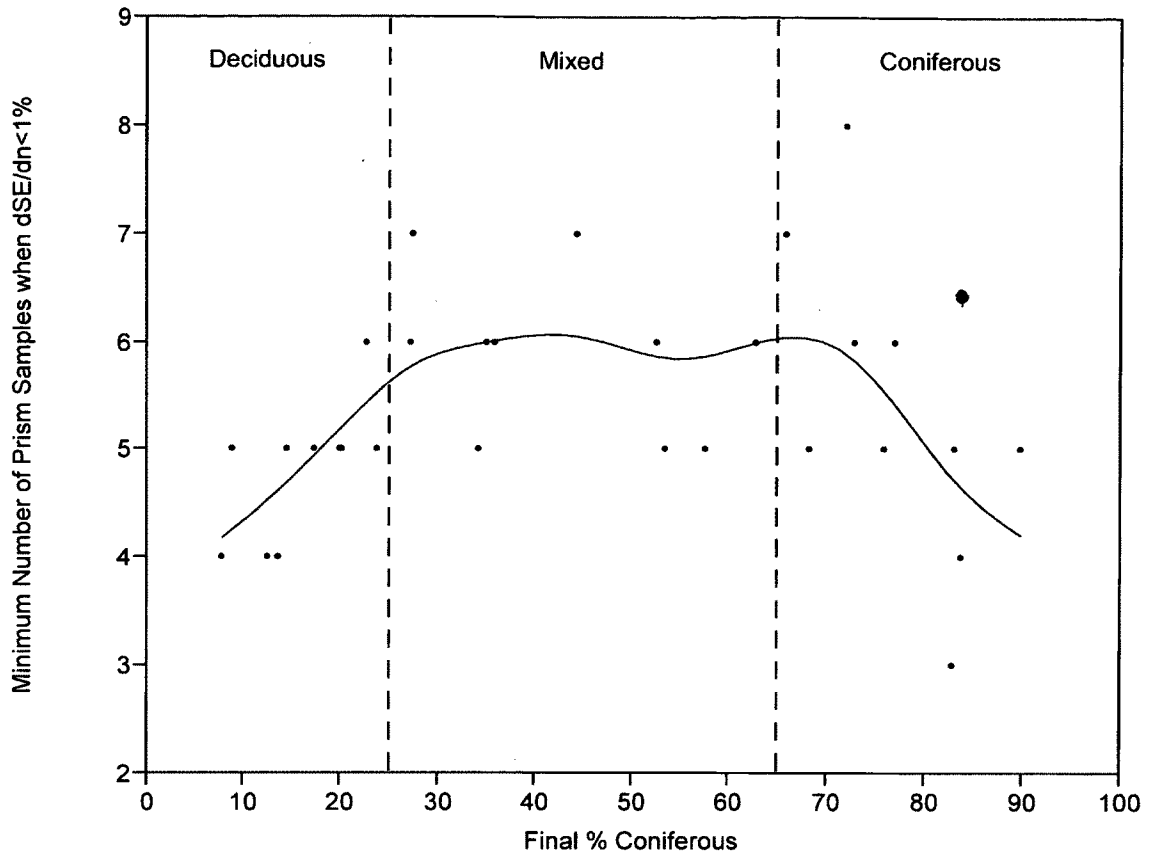


Figure 6. The minimum number of prism samples when the change in standard error (SE) for one additional prism sample is $\leq 1\%$ (the most conservative case) for each polygon plotted against the final percent coniferous (when $n=N$). The trend is shown by the grey line, and the cutoffs for deciduous, mixed, and coniferous classification are delineated by the vertical dashed lines.

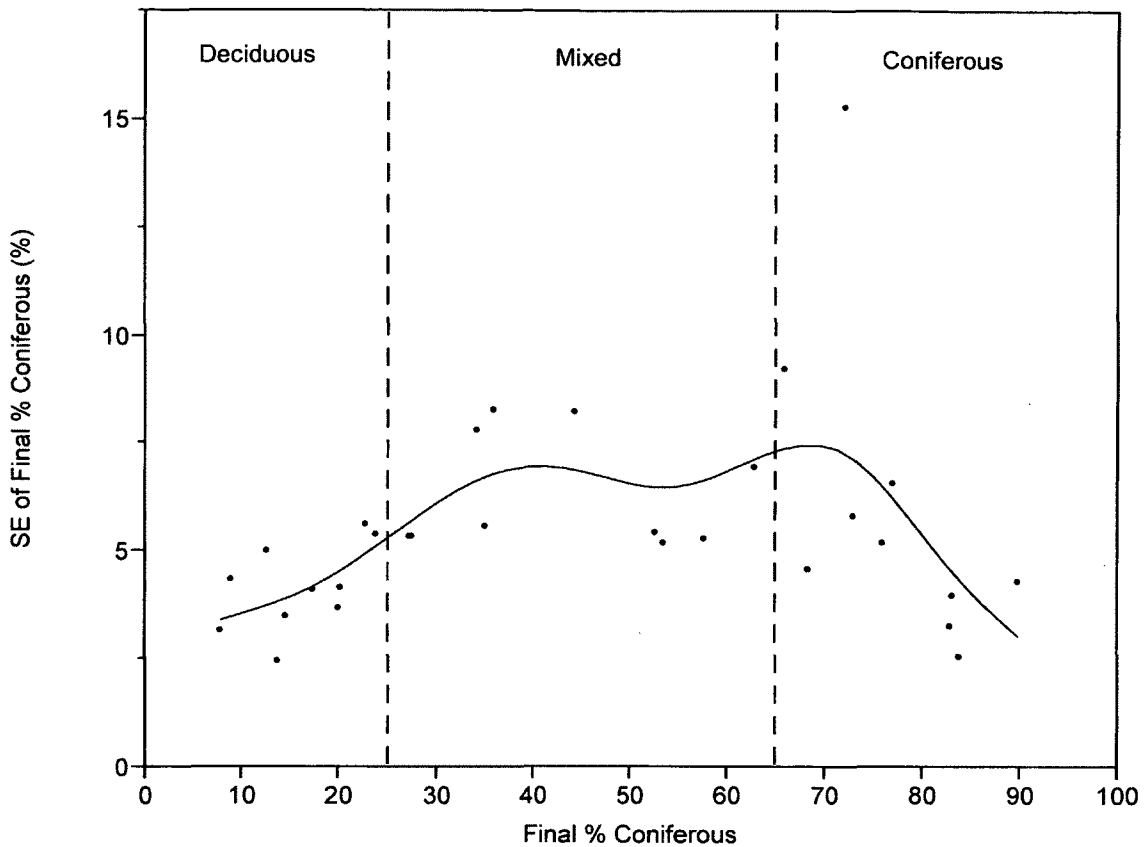


Figure 7. The standard error (SE) of the final percent coniferous (when $n=N$) for each polygon plotted against the final percent coniferous. The trend is shown by the grey line, and the cutoffs for deciduous, mixed, and coniferous classification are delineated by the vertical dashed lines.

A backward stepwise least squares multiple linear regression analysis was performed in order to determine whether any of the readily available zonal statistics from the original Landsat 5 TM image could be used as predictors of SE of percent coniferous. The hope was to identify characteristics of the polygons on the imagery that could help predict whether a polygon would be more or less difficult to classify on the ground using prism samples, since SE of percent coniferous is a positive predictor of number of prism

samples necessary. To keep the analysis simple and repeatable using most image software programs, the statistics used as predictor variables were: area of the polygon; perimeter of the polygon; perimeter/area; the mean pixel value for each band; and the standard deviation of the pixel values for each band (a measure of texture). In the stepwise regression, the model with the lowest corrected Akaike Information Criterion (AICc; Akaike, 1974; Burnham and Anderson, 2004) was chosen as the best predictive model. Since a stepwise regression was used, *p*-values are generally not interpretable for the variables in the chosen model because the best model is chosen relative to all other possible models and no significance test is completed (Burnham and Anderson, 2002). Therefore, the SE of the estimates is instead reported as well as the difference in AICc values to the next best model. Three variables provided the best model for predicting SE of percent coniferous: the mean of the first middle infrared (MIR) band (band 5); the SD of the blue band; and the SD of the NDVI derivative band (Table 4).

Table 4. Variables used in the best predictive model of SE of percent coniferous (minimum Δ AICc = 2.269).

Predictor Variable	Coefficient	SE
Intercept	9.016	2.720
Mean of Band 5 (MIR)	-0.114	0.047
SD of Band 1 (Blue)	3.244	0.746
SD of NDVI	-0.926	0.314

Conclusions

When classifying an image using an OBIA approach, it is imperative that segments be used as reference units. An advantage in using an OBIA approach in classification is that fewer overall reference units are needed to complete an accuracy assessment of a land cover map created using OBIA, as compared to when pixel reference units are used to assess a pixel-based map (Radoux *et al.*, 2011). However, labeling forested reference units by composition can be difficult without sampling, and sampling is usually costly. In this study, we found that for reference units created through the segmentation of a Landsat 5 TM image, a medium resolution image with 30 m pixels, only six prism samples were needed to label reference units as coniferous, deciduous, or mixed forest and achieve relatively high labeling accuracy ($E_{\max}=4\%$). Therefore, the relatively small number of necessary prism samples needed for proper labeling of reference units may make OBIA a potentially cost-effective tool for classification, since it may reduce the sampling effort needed to create reference units.

A stepwise linear regression was performed to create the best model for predicting the SE of the total percent coniferous. Three variables provided an improved prediction of SE of the total percent coniferous: mean of the MIR band; SD of the blue band; and SD of the NDVI derivative band; indicating that these bands may help to predict how difficult it will be to estimate the percent coniferous of a stand. In this case, the mean of the MIR band and the SD of NDVI had negative coefficients for predicting the SE of total percent coniferous, and the SD of the blue band had a positive coefficient. The most useful variable of the three may be the SD of the blue band, since positive coefficient indicates that as the SD of the blue band increases, so does the SE of the total percent

coniferous, implying that more samples may be necessary in polygons with large SD of pixel values the blue band. Therefore, the blue band could be used in the future to predict the variability of the percent coniferous within a segmented image. However, all three predictor variables can be used to determine whether more or less sampling should take place in certain polygons. Unfortunately, this relationship is likely region and imagery specific, since elevation and other factors like atmosphere can also influence image characteristics, especially the blue band.

The addition of area of the polygons in the model to predict SE of the total percent coniferous did not result in improved model fit ($\Delta AICc = 2.269$), and in a further analysis, it was also found that there was no significant correlation between area and the number of necessary prism samples at each of the three thresholds ($p < 0.05$). Therefore, area likely did not influence the number of necessary prism samples in a polygon. The number of necessary prism samples per polygon is also much lower than the current guidelines for prism sampling for conventional forest inventory purposes. The previous guidelines suggested that the number of prism samples necessary for accurate sampling within a forest stand is completely dependent on stand size; however, this study did not find this dependence, indicating that segmentation may have reduced the dependence of number of samples on stand size. The lack of size dependence is likely a result of the parameters used to define polygons during segmentation. The relatively small variance factor limited the amount of variability contained within a segment. Therefore, larger polygons are created when there is little variability in the pixels, while smaller polygons are created when contiguous pixels are more variable. The variability in pixels often relates to observable variation in species on the ground, meaning the low variance factor

limited the species heterogeneity within each polygon. Therefore, it is likely that the number of prism samples required in each polygon is more a function of the variance factor defined during segmentation rather than the size of the polygon. If the variance factor is raised, the amount of allowable variability in the pixels would also increase. Accordingly, the number of prism samples needed to label those segments should also increase. However, the average size of the polygons should also increase, decreasing the number of polygons that have to be sampled for accuracy assessment (Radoux *et al.*, 2011), creating a tradeoff between number of prism samples needed inside a reference unit, and number of reference units that must be visited.

Given the natural variability of forests in the Northeast, the minimum of six prism samples within a segment may provide a useful guideline for many forest sampling protocols using similar classification techniques. However, in situations where each additional prism sample would be very expensive to acquire (e.g. very large segments in rough terrain), as few as three prism samples may be used to attain relatively accurate reference unit labels. In no case were any less than two prism samples acceptable for labeling. Since the prism samples should be sampling across the variability of the segments in an unbiased fashion, it is important to limit the influence of subjective factors or other sources of potential bias in the distribution of samples. Appropriate sampling techniques, such as simple random sampling or stratified random sampling, should be employed to ensure proper labeling of the reference units (Congalton, 1988; Stehman and Czaplewski, 1998; Thompson, 2002).

These findings are specific to our classification scheme and segmentation parameters, but the nature of the segmentation process should allow these methods to be applied in

many different scenarios. The procedures completed in this study should be tested in other forest biomes and with different segmentation parameters to determine if greater or fewer prism samples are required for accurate reference unit labeling and if the number of prism samples required can be correlated with known landscape or imagery characteristics. As long as a single group can be used to differentiate between forest types (e.g. coniferous trees for this study), and the target variable is a percent of total (not a total area, for example), these methods should be applicable. Since all calculations were done using the total percent coniferous as estimated by using more than one prism samples, the prism BAF choice and empty samples should not impact the calculation, unless there are many empty samples. The only issue would be when all empty samples were chosen in the bootstrap estimate and the estimated percent of total was undefined, since the total tree count was zero. Hopefully, these empty samples are rare enough that past three or more samples in the bootstrap estimate, the unique situation of having all empty samples should no longer be an issue. However, when making a recommendation for sampling, six prism samples randomly located throughout a polygon should capture the majority of what is present and still be appropriate for sampling for forest composition, even if the polygon includes empty samples. Also, the choice of BAF for the prism used in sampling should be influenced by the stand structure, so the prism should compensate somewhat for very sparse or dense forest structures.

The methods presented here provide a guideline for the minimum number of prism samples needed in a mixed hardwood in the northeastern United States with a classification scheme dependent on percent coniferous to distinguish between deciduous, coniferous, and mixed forest types. However, future explorations into how labeling

strategies may change the minimum number of prism samples would be quite interesting. For instance, a labeling strategy that only has two forest types (i.e. one boundary), the E_{\max} for the standard error thresholds would be lower, therefore possibly allowing a higher threshold to be used (e.g. 2% instead of 1%), leading to a lower minimum number of prism samples. Also, a hierarchical classification system may also have different sampling needs. In a hierarchical classification, an initial classification may be based on overall percent coniferous, but a more specific label may be dependent on the percent of a specific species. In these instances, another threshold value may be used to determine the appropriate number of samples needed for the more specific label. Finally, if techniques such as fuzzy sets (Gopal and Woodcock, 1994) are used in classification, the techniques explored in this study are exceptionally useful, since the SE of the estimated total percent coniferous (as found using all of the collected samples) can be used to assign a confidence value to the classification of a particular polygon. Since the application of fuzzy classification and accuracy assessment may increase potential overall accuracy (by allowing some polygons that would otherwise be considered 'wrongly classified' to be 'partially correct'), these methods may also allow for a higher threshold value to be used when determining minimum number of necessary prism samples.

CHAPTER IV

APPLICABILITY OF MULTI-DATE LAND COVER MAPPING USING LANDSAT 5TM IMAGERY IN THE NORTHEASTERN US

Abstract

In many situations, multi-date image classification improves classification accuracies. However, with improved accuracies comes increased image processing time and effort. This work investigates the circumstances under which multi-date image classification is significantly better than single-date classification using Landsat 5TM imagery for southeastern New Hampshire. Multiple Landsat images were processed for every three years from 1986 to 2010 and classified using an object-based image analysis approach (OBIA) and a classification and regression tree (CART) technique. Two maps were created for each of the mapping years, one using a single image, and another using multiple images from that year. The multi-date classification process generally performed better than the single-date process. However, the significance of the improvement was primarily dependent on the accuracy of the single-date map. Therefore, if the accuracy of the single-date classification is acceptable, it may not be necessary to perform the multi-date classification.

Introduction

Land cover mapping is essential for effective resource management, and the use of satellite remote sensing has become a very important part of the land cover mapping process since it is a relatively inexpensive and efficient way to map land cover types. Many studies have looked into improving the accuracy of these maps through the exploration of different techniques for classifying satellite images (e.g. Conese and Maselli, 1991; Congalton *et al.*, 1993; Lunetta *et al.*, 1993; Gopal and Woodcock, 1994; Schriever and Congalton, 1995; Wolter *et al.*, 1995; Foody, 1996; Foody, 2002; Xiuwan, 2002; Drăguț and Blaschke, 2006; Lu and Weng, 2007; Duveiller *et al.*, 2008; Radoux *et al.*, 2011). One such strategy involves the use of multiple images from the same year in an attempt to capture phenological changes in vegetation, allowing the mapper to better separate vegetation classes (Liu *et al.*, 2002). Many studies have found that this multi-date classification process resulted in higher accuracies than a single-date classification when trying to separate forest types (e.g. Conese and Maselli, 1991; Schriever and Congalton, 1995; Wolter *et al.*, 1995; Liu *et al.*, 2002; Tottrup, 2004), wetlands (e.g. Lunetta and Balogh, 1999), and agricultural land cover types (e.g. Oetter *et al.*, 2000; Guerschman *et al.*, 2003). However, other studies have found the multi-date process less successful (e.g. Henry, 2008).

In multi-date classification, several images of a specific location of interest from the same year are used in the creation of a single land cover map. The potential benefit is that the added spectral information from the additional dates will result in better classification of land cover types. Multi-date classification is also used to mitigate some of the atmospheric issues, such as clouds, encountered when using satellite images. In a

multi-date classification process, individual images can be used independently to separate classes one at a time or classify areas otherwise obscured by transitional objects like clouds (e.g. Justice *et al.*, 2002), or all images can be used simultaneously in an attempt to separate all classes using a single classification algorithm (e.g. Guershman *et al.*, 2003). Typically, in order to reduce processing time and potential complications regarding sensor differences, images from the same source are used in multi-date image processing (Pohl and Van Genderen, 1998). In addition, with the now free availability of Landsat 5TM imagery, it is less expensive and more straightforward to use the multi-date classification process than it was in the past. However, it remains unclear under what conditions the multi-date process might be the most useful, and whether the potential benefit of this approach is worth the additional image processing time and effort that is required.

In this study, the multi-date classification process was tested against a single-date classification process using an object-based image analysis (OBIA) approach and a classification and regression tree (CART) technique to label each of the land cover types. The maps were created for nine mapping years, each with a different set of available images. An OBIA approach was used for this particular study to maximize the potential parameters used in classification for a particular group of pixels (Drăguț and Blaschke, 2006; Congalton and Green, 2009). The CART technique was chosen because it is a non-parametric classification algorithm that has the ability to deal with a large number of correlated variables (Breiman, 1984). All bands of multiple images of the same year were considered variables in the multi-date classification process, and these bands were generally all correlated. The CART technique was able to select from the available

bands, or variables, those that were most important for separating land cover types. All maps, created using either the multi-date process or the single-date process, used the combined OBIA and CART classification approach so that the overall accuracies of the two processes were directly comparable. This study investigated the circumstances under which multi-date classification was most appropriate, and whether the added image processing time and effort for the multi-date classification scheme significantly improved map accuracy ($p < 0.05$).

Methods

Study Area

The Coastal Watershed of New Hampshire (NH), a Hydrologic Unit Code 8-digit level (i.e. HUC-8) watershed, is located in the southeastern portion of the state, bordering Maine to the northeast (Figure 8). The watershed encompasses the only coastline of NH, as well as the Great Bay Estuary, and contains a diverse set of land cover types. For this study, eight general land cover classes were used when mapping the study area (Table 5). Development occurs in both high density city areas and very low density residential communities. There is active agriculture in the form of small family farms and a variety of natural forest community types. In general, the growing season of this region begins in April and continues through September, with a peak near infra-red (NIR) reflectance occurring toward the end of August and the beginning of September, and senescence occurring in October (Figure 9). However, there will be some variation year to year due to climatic differences (Chen and Pan, 2002).

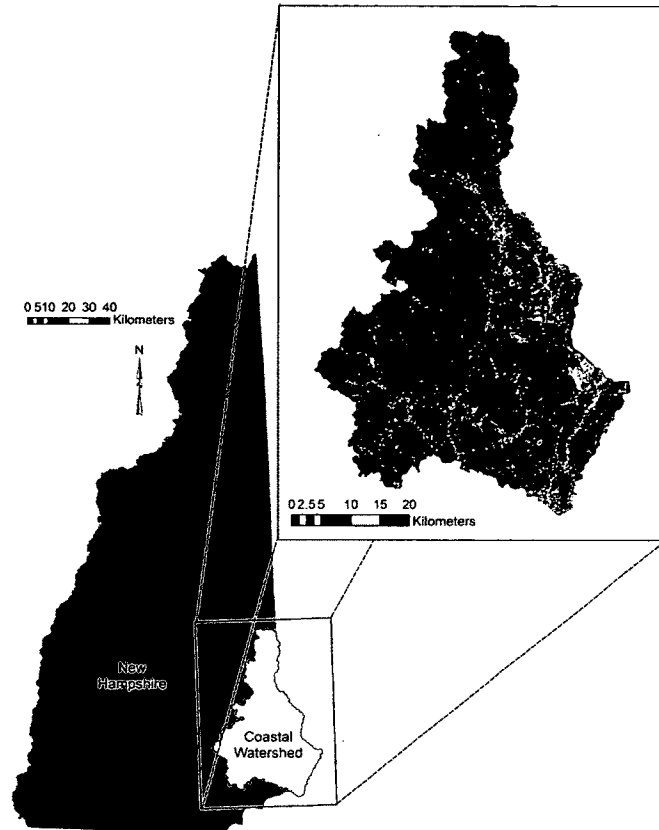


Figure 8. The Coastal Watershed study area in New Hampshire. Image is a Landsat 5TM image from 14 August 2010.

Table 5. Classification system used to map the Coastal Watershed, NH. Modified from Justice *et al.* (2002).

Class	Description
Active Agriculture	Areas dominated of row crops, hay/pasture, or orchards
Cleared/Other Open	Areas dominated by disturbed land, sand dunes, or other cleared
Developed	Areas dominated by residential/commercial/industrial development or transportation
Coniferous Forest	Forest stands comprising greater than 65% coniferous basal area per acre
Deciduous Forest	Forest stands comprising less than 25% coniferous basal area per acre
Mixed Forest	Forest stands comprising more than 25% and less than 65% coniferous basal area per acre
Open Water	Lakes, ponds, some rivers, or any other open water as defined by the U.S. Fish and Wildlife Service National Wetlands Inventory
Wetlands	Areas dominated by wetlands characteristics as defined by the U.S. Fish and Wildlife Service National Wetlands Inventory

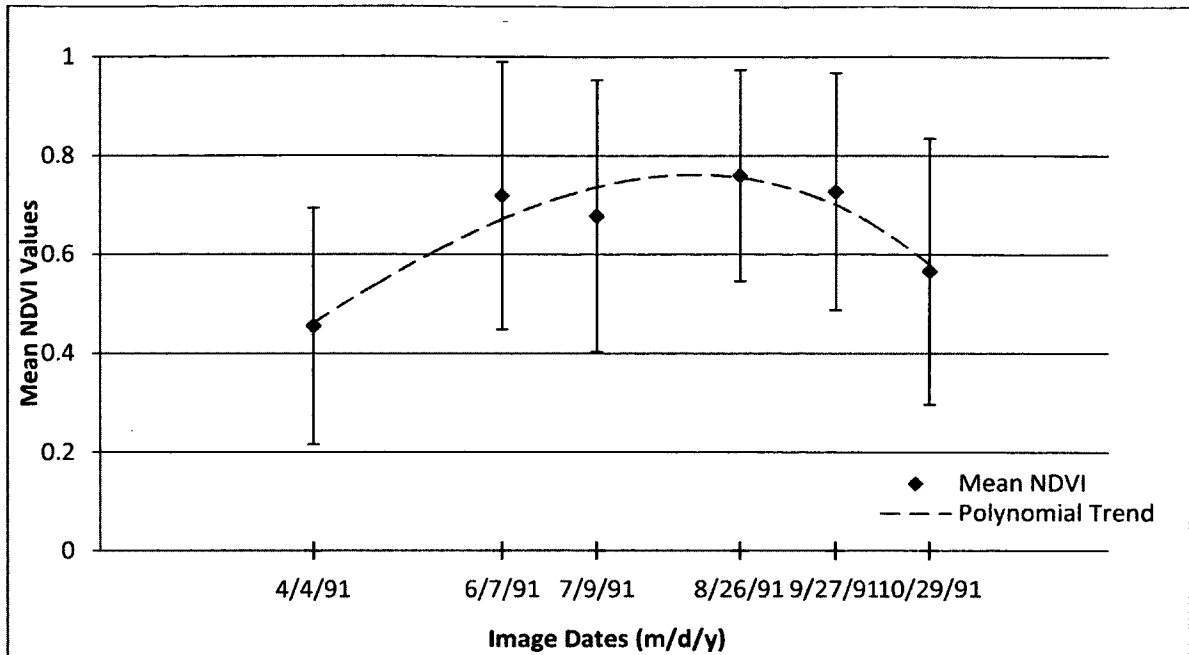


Figure 9. The trend of mean NDVI values computed using Landsat 5TM data in the Coastal Watershed for the 1991 growing season. All images have been relatively atmospherically corrected to the July date, so all NDVI values are relative and comparable. Whiskers denote standard deviation of mean NDVI values for the watershed and the dashed line is a fitted polynomial trend line.

Elevation within the watershed is relatively flat; however, the proximity to the coast can cause cloud cover issues for satellite imagery. The National Oceanic and Atmospheric Administration's (NOAA) National Climatic Data Center (NCDC) estimates that southern NH on average has 90 clear days per year (NCDC, 2008), indicating that around 75% of the time satellite imagery will contain some cloud cover. Given this, a single-date classification may not be possible for each of the nine years in this study depending on the specific atmospheric issues of that year.

Image Selection and Processing

Single date and multi-date maps were created for every third year from 1986 to 2010 using Landsat 5TM data from path 12 row 30. All images with 10% or less cloud cover for the Coastal Watershed (regardless of the cloud cover for the rest of the scene) were downloaded from USGS (all processed at Level 1T). Therefore, each mapping year has a different number of Landsat images from that year, each with a different distribution of images (Figure 10). Previous work by Guerschman *et al.* (2003) has suggested a minimum of two images from the same growing season are necessary to properly identify land cover types. All of the mapping years in this study had at least four images, but no more than seven were found for any particular year. From the available images for each year, the image with the lowest cloud cover acquired during the growing season was chosen for the single-date mapping approach. Three years had less than ideal dates for single-date imagery: 1992, 1998, and 2004. Both 1998 and 2004 had only reasonably cloud-free images from early April, near the start of the growing season. The 1992 year had only one acceptable image, and it was from the end of September, which is closer to senescence in this region.

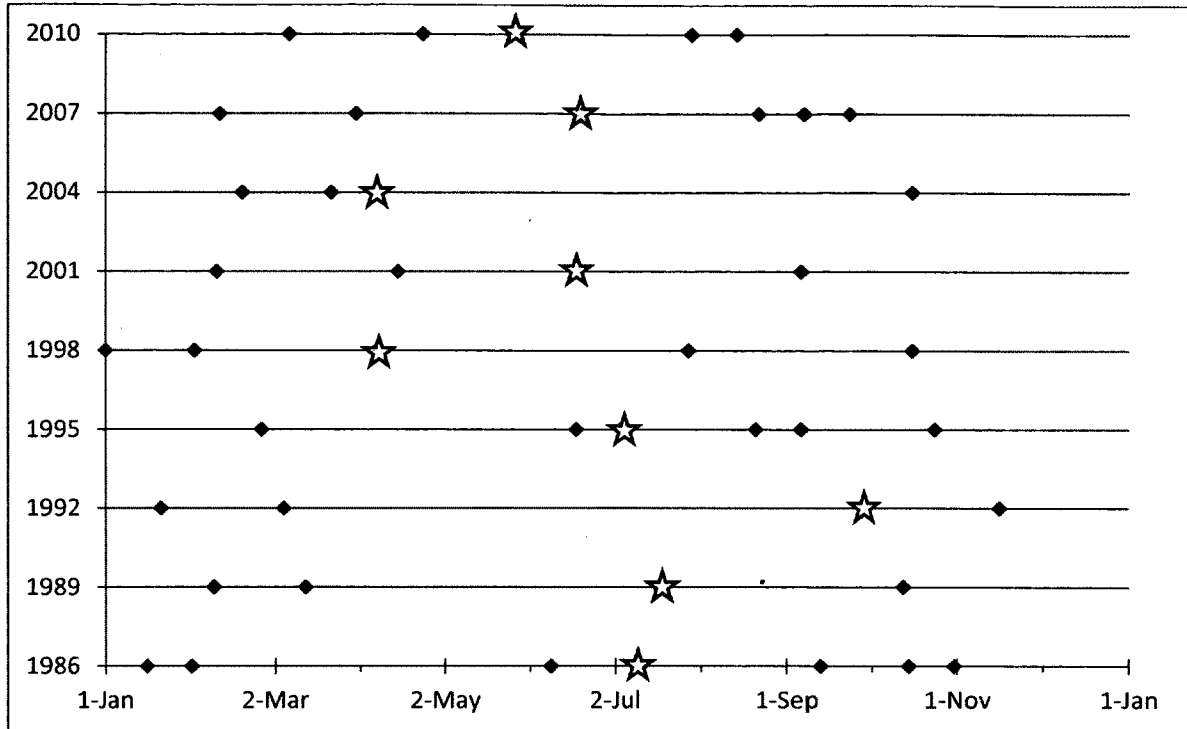


Figure 10. Distribution of Landsat 5TM images for each map year with map year along the y-axis and date in that year along the x-axis. Single-date images are represented by the stars.

Once all of the images were selected for each mapping year, the images were checked for any registration errors and clipped to the extent of the watershed. Clouds and their shadows were masked out of each image using on-screen digitizing. Each of the single-date images were then absolutely atmospherically corrected using the COST method (Chavez, 1996). The single-date images were then stretched to unsigned 8-bit and the remaining images for each mapping year were histogram matched to the corrected single-date image using Erdas Imagine software (Intergraph®). These methods ensured that the images in each mapping year were directly comparable within each year, and that differences due to haze or other atmospheric factors were minimized during processing.

Four different derivative bands were computed for each of the images: the Normalized Difference Vegetation Index (NDVI); and the first three tasseled cap bands (brightness, greenness, and wetness). These bands were then rescaled and layer stacked with the original imagery. For the multi-date process, all images from the same mapping year were also stacked together and treated like a single image for the remainder of the classification process. These stacked images are referred to here as a multi-image stack.

Image Segmentation

eCognition software (Trimble[®]) was used to segment the images prior to classification. In the single-date approach, only the single image for each mapping year and its derivative bands were segmented, while during the multi-date process the multi-image stack was segmented as a whole, treating the multi-image stack (all images available for that year and each image's derivative bands) as a single image. Each image or image stack was segmented using the same parameters within eCognition (Table 6). These parameters were determined through a series of trial and error attempts in conjunction with photo-interpretation to determine if different land cover types were sufficiently delineated, erring on the side of slightly smaller segments. The National Wetlands Inventory (NWI) was also used as an informative thematic layer to help delineate wetlands and open water, since the extent of many of these features are dependent on time of year and tidal phase (Cowardin *et al.*, 1979; Diaz *et al.*, 2004), which change throughout the imagery.

Table 6. Parameters used during segmentation.

Parameter	Value
Layer Weights	All 1
Scale	7
Shape	0.2
Compactness	0.5
Thematic Layer	National Wetlands Inventory (NWI)

Reference Data Collection

The segmented 2010 single-date image was used as the source image for collecting reference data since reference data collection began in 2010. Segments were chosen from the 2010 image to be used as reference data samples and each of the labels was determined for those segments through a combination of fieldwork and photo-interpretation. Fieldwork was performed starting in the fall of 2010 and continued through the fall of 2011. An image differencing technique was used to determine where areas of major change occurred from 1986 to 2010, and those areas were taken out of consideration as reference data locations. Therefore, it was assumed that the majority of the reference data collected should be applicable for all years from 1986 to 2010. In addition, all reference data were also visually checked after collection to ensure that they were accurate representations of the land cover for each year.

Since the forest categories were generally the most difficult to differentiate on imagery, a minimum of 30 segments (sample units) per forest class were visited on the ground. These sample units were chosen using stratified random sampling and were limited to public access properties. The segments were then labeled using six randomly located prism or Bitterlich samples within that segment (as recommended in MacLean *et*

al., 2012). Prism sampling is a quick and efficient method for estimating forest composition and is a well-tested strategy in forestry (Husch *et al.*, 2003). This method samples trees proportional to their size and is used to assess the composition of a forest stand. For this particular sampling strategy, a prism with a Basal Area Factor (BAF) of 20 ft²/acre was used during sampling, which is appropriate for forests in the Coastal Watershed (Wiant *et al.*, 1984; Ducey, 2001). Forest segments were then labeled based on their composition and the classification scheme outlined in Table 5.

The remainder of the reference data samples were collected through photo-interpretation so that each class, including those not sampled through fieldwork, had a minimum of 100 reference data samples. Reference segments were selected using stratified random sampling from throughout the study area, and NH Department of Transportation digital aerial imagery with 0.30 meter resolution was used in the photo-interpretation process. The imagery was acquired in April of 2010 with four spectral bands, three natural color bands (blue, green, and red) and one near-infrared. The labeled reference data samples were then randomly put into two groups: half were placed in the group used as training data; and the second half were placed in another group used later as accuracy assessment data.

Classification

A classification and regression tree (CART) technique was used to classify all of the images used in this analysis. The properties of the training data samples, including the traditional average Digital Number (DN) values from each of the image layers, were used to create the decision tree. Since an OBIA approach was used, each of the segments also

had a size, shape, variation in DN values, etc., that do not exist in the pixel-based approach. These segment specific properties were also used in the creation of the decision tree. Three different, unique decision trees were created for each image, or image stack, for each year, keeping the reference data consistent within the same year (Figure 11). The entire classification process was performed within eCognition (Trimble®). First, each image (either the single-date or the multi-date) was classified into two broad classes: forest and non-forest, using a single decision tree. Then, the forest segments were reclassified into more specific forest classes using one decision tree, and the non-forest segments were classified into more specific other categories using another decision tree (Figure 11). The only exceptions were the open water and wetlands categories, which were classified based upon the NWI. This hierarchical classification system resulted in much better differentiation between forest and non-forest categories, as well as less confusion between classes overall, since the decision trees were created to separate fewer categories. However, using three decision trees per classification resulted in six decision trees per year, totaling 54 different decision trees to create the nine single-date maps and nine multi-date maps.

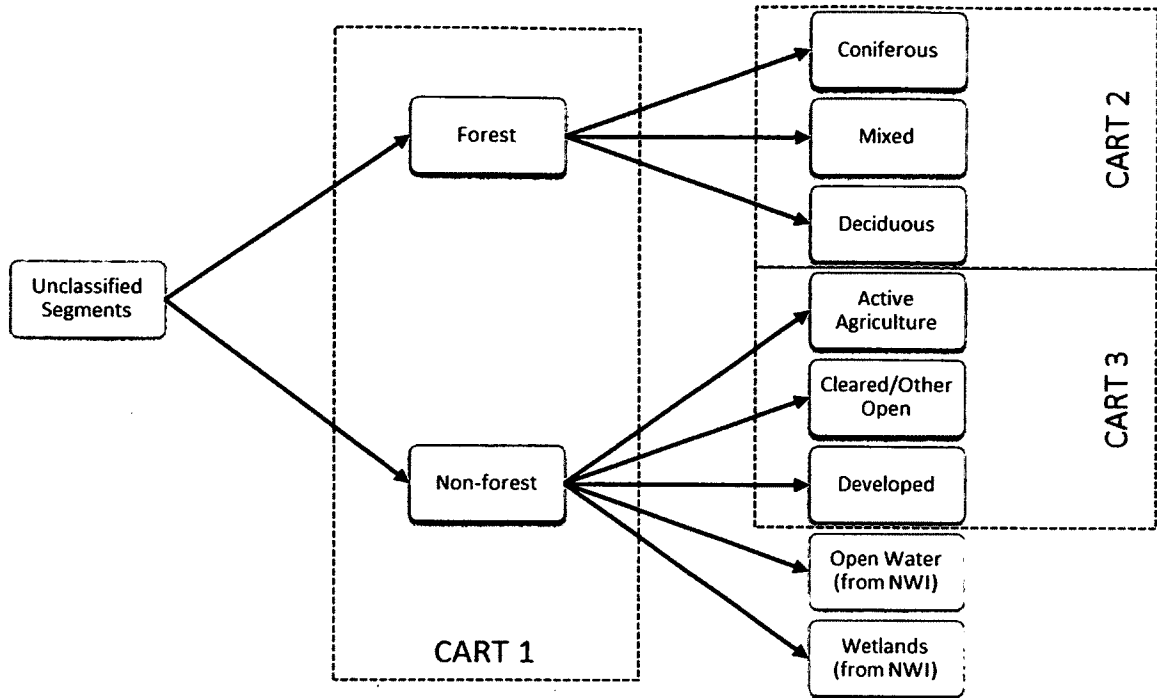


Figure 11. The hierarchical classification system used to classify each of the images. Three different decision trees were used to create each of the maps, labeled here as: CART 1, CART 2, and CART 3.

Accuracy Assessment

Each of the resulting single-date and multi-date maps were then assessed for their accuracy using both a traditional error matrix (Congalton *et al.*, 1983) as well as an area-based error matrix approach (MacLean and Congalton, 2012a). The area-based error matrix uses the same principles as the traditional error matrix, except instead of tallying each reference data sample in the correct box, the area of the segment used as a reference data sample is entered into the correct cell in the error matrix (Table 7). The accuracies of the single-date map were compared to the multi-date map for the same year, the difference in the accuracies, and a Kappa analysis was performed using the traditional

error matrices to determine if the two maps from the same year were significantly different at the 95% confidence limit (Congalton *et al.*, 1983).

Table 7. A comparison of the traditional error matrix (**a**) and the area-based error matrix (**b**), where n_{kk} is the number of reference data samples that fall in that particular cell, N is the total number of samples, and S_{kk} is the total area of all of the reference data samples that fall in that particular cell and S is the total area sampled (modified from MacLean and Congalton, 2012a).

(a) Traditional Error Matrix

		Reference Data				Row Total	User's Accuracy
		1	2	...	k		
Map Data	1	n_{11}	n_{12}	...	n_{1k}	n_{1+}	n_{11}/n_{1+}
	2	n_{21}	n_{22}	...	n_{2k}	n_{2+}	n_{22}/n_{2+}
	⋮	⋮	⋮	...	⋮	⋮	⋮
	k	n_{k1}	n_{k2}	...	n_{kk}	n_{k+}	n_{kk}/n_{k+}
Column Total		n_{+1}	n_{+2}	...	n_{+k}	N	Overall Accuracy:
Producer's Accuracy		n_{11}/n_{+1}	n_{22}/n_{+2}	...	n_{kk}/n_{+k}	$\frac{\sum_{t=1}^k n_{tt}}{N}$	

(b) Area-Based Error Matrix

		Reference Data				Row Total	User's Accuracy
		1	2	...	k		
Map Data	1	S_{11}	S_{12}	...	S_{1k}	S_{1+}	S_{11}/S_{1+}
	2	S_{21}	S_{22}	...	S_{2k}	S_{2+}	S_{22}/S_{2+}
	⋮	⋮	⋮	...	⋮	⋮	⋮
	k	S_{k1}	S_{k2}	...	S_{kk}	S_{k+}	S_{kk}/S_{k+}
Column Total		S_{+1}	S_{+2}	...	S_{+k}	S	Overall Accuracy:
Producer's Accuracy		S_{11}/S_{+1}	S_{22}/S_{+2}	...	S_{kk}/S_{+k}	$\frac{\sum_{t=1}^k S_{tt}}{S}$	

Finally, a stepwise regression was performed to determine if factors could be used to predict when the multi-date process performs better than the single-date process. The dependent variable was the difference in the traditional overall accuracies for the two maps created in a single map year. The explanatory variables included: the single-date accuracy; the total number of images; the percent of the total images used in the multi-date map taken in fall; the date of the image used for the single-date map; the percent of total images with some cloud cover; the percent of total images that were taken in the growing season; the number of images capturing senescence; and the range of dates, average date, and standard deviation of the dates for all of the images in the multi-date image stack. Since there are a low number of samples and therefore a low number of degrees of freedom, not all explanatory variables could be tested in the same model, so a forward elimination stepwise regression was performed. The model with the lowest corrected Akaike Information Criterion (AICc) was chosen as the best predictive model (Akaike, 1974; Burnham and Anderson, 2004).

Results and Discussion

Single-Date and Multi-Date Maps

Two maps were created for each map year, one using the single-date process and one using the multi-date process (Figure 12). Without any post-processing, the accuracies of the maps using the traditional error approach achieved overall accuracies in the 70 percent range (Table 8), while the accuracies computed using the area-based approach consistently achieved higher accuracies (Table 9). In all maps, the most confused classes were the cleared/other open and the mixed forest classes. Cleared/other open was

primarily confused with the active agriculture class, which for this area is quite understandable. Most agriculture in this area is hay/pasture, and the cleared/other open category encompassed areas such as golf courses and other grassy areas that are spectrally quite similar to pasture lands. The mixed forest class was confused with both the deciduous and coniferous forest categories. Given the variability of the forests in southern NH and the 30 meter pixels of the Landsat 5TM images, it is also no surprise that mixed forest was commonly confused for other types of forest.

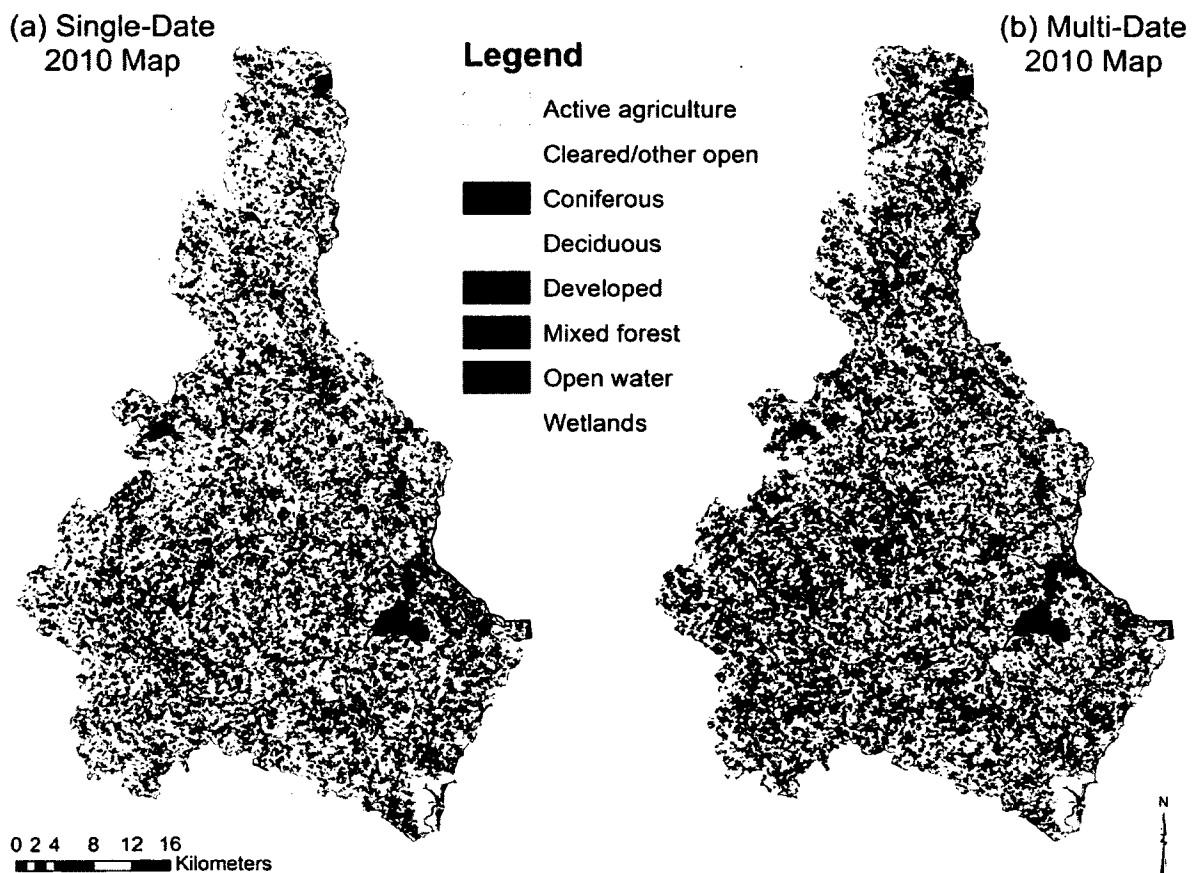


Figure 12(a). The single-date map created for 2010. **12(b).** The multi-date map created for 2010.

Table 8. The 2010 multi-date traditional error matrix. All map classes were assessed using the collected accuracy assessment data, even those map classes that were labeled using the NWI. Therefore the reported user's and producer's accuracies for Open water and Wetlands are more a reflection of the accuracy of the NWI. However, these classes are still important for assessing the overall accuracies of the land cover maps.

		Reference Data							Row Total	User's Accuracy	
		Active agriculture	Cleared/ other open	Coniferous	Deciduous	Developed	Mixed forest	Open water			Wetlands
Map Data	Active agriculture	51	20	0	0	14	0	0	0	85	60%
	Cleared/ other open	19	19	0	0	13	0	0	0	51	37%
	Coniferous	0	0	49	4	0	16	0	0	69	71%
	Deciduous	0	0	2	63	0	12	0	0	77	82%
	Developed	10	10	0	0	70	0	0	0	90	78%
	Mixed forest	0	1	14	28	0	28	0	0	71	39%
	Open water	0	0	0	0	0	0	50	0	50	100%
	Wetlands	0	0	0	0	0	0	0	50	50	100%
Column Total	80	50	65	95	97	56	50	50	543	Overall Accuracy:	
Producer's Accuracy	64%	38%	75%	66%	72%	50%	100%	100%		69.98%	

Table 9. The 2010 multi-date area-based error matrix with cell values in hectares.

		Reference Data (ha)							Row Total	User's Accuracy	
		Active agriculture	Cleared/ other open	Coniferous	Deciduous	Developed	Mixed forest	Open water			Wetlands
Map Data (ha)	Active agriculture	471.33	189.74	0.00	0.00	113.66	0.00	0.00	0.00	774.73	61%
	Cleared/ other open	161.04	165.05	0.00	0.00	114.77	0.00	0.00	0.00	440.86	37%
	Coniferous	0.00	0.00	737.08	36.70	0.00	211.97	0.00	0.00	985.75	75%
	Deciduous	0.00	0.00	14.46	1080.93	0.00	141.01	0.00	0.00	1236.40	87%
	Developed	82.08	68.95	0.00	0.00	582.35	0.00	0.00	0.00	733.38	79%
	Mixed forest	0.00	0.00	148.35	244.67	0.00	272.69	0.00	0.00	665.71	41%
	Open water	0.00	0.00	0.00	0.00	0.00	0.00	696.01	0.00	696.01	100%
	Wetlands	0.00	0.00	0.00	0.00	0.00	0.00	0.00	525.19	525.19	100%
Column Total	714.45	423.74	899.89	1362.29	810.79	625.67	696.01	525.19	6058.3	Overall Accuracy:	
Producer's Accuracy	66%	39%	82%	79%	72%	44%	100%	100%		74.79%	

Comparison of the Multi-Date and Single-Date Accuracies

In general, the multi-date mapping process performed better than the single-date process (Figure 13). However, when comparing the traditional error matrices using a Kappa analysis, the two processes were only significantly different in four of the nine years (Table 10). Only in 1992, 1995, 2001, and 2007 did the multi-date process prove to be significantly better than the single-date process. In all cases, the area-based error matrix approach did result in higher overall accuracies than the traditional error matrix approach (Figure 14).

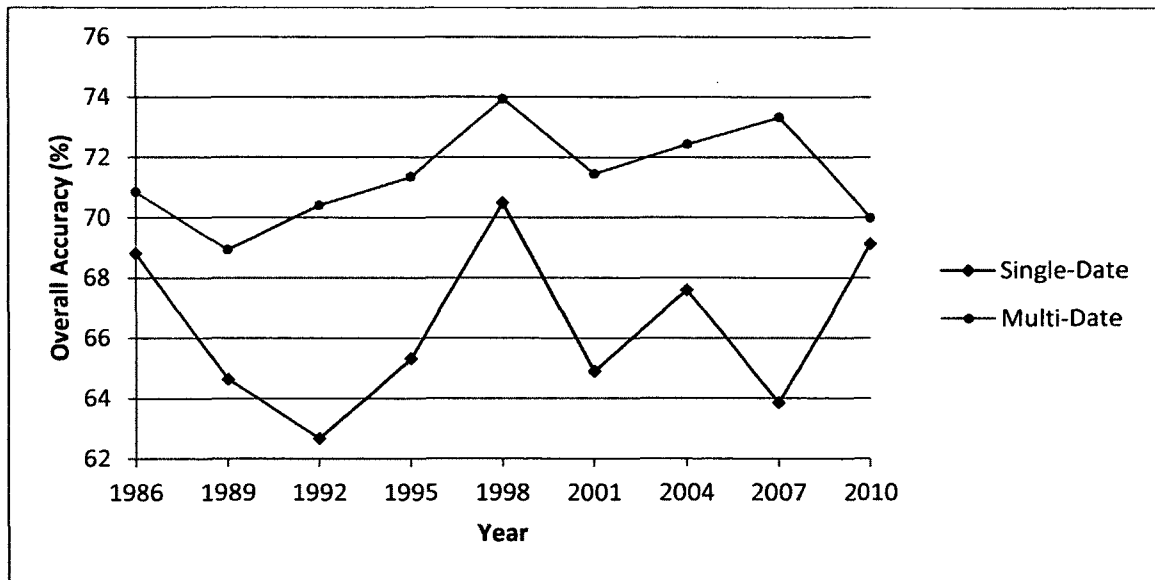


Figure 13. The single-date and multi-date overall accuracies computed using the traditional error matrix for each year.

Table 10. The traditional and area-based overall accuracies of the two maps created for each mapping year and the differences in the accuracies. The Z-statistic was computed using the traditional error matrices, where $Z_c = 1.96$ at the 95% confidence interval and the single-date and multi-date error matrices are significantly different when $Z > Z_c$. Any Z-statistics with an asterisk indicates a significant difference between single-date and multi-date classifications.

Year	Traditional Error Matrix			Area-Based Error Matrix			Z-statistic
	Single-Date Accuracy	Multi-Date Accuracy	Difference	Single-Date Accuracy	Multi-Date Accuracy	Difference	
1986	68.81 %	70.85 %	2.04 %	70.68 %	73.07 %	2.39 %	0.75
1989	64.63 %	68.93 %	4.30 %	67.41 %	72.79 %	5.38 %	1.52
1992	62.66 %	70.40 %	7.74 %	63.98 %	74.27 %	10.29 %	2.68*
1995	65.31 %	71.35 %	6.04 %	65.31 %	74.22 %	8.91 %	2.11*
1998	70.48 %	73.94 %	3.46 %	73.71 %	75.85 %	2.14 %	1.29
2001	64.89 %	71.45 %	6.56 %	68.78 %	76.12 %	7.34 %	2.27*
2004	67.59 %	72.43 %	4.84 %	72.41 %	75.02 %	2.61 %	1.78
2007	63.84 %	73.33 %	9.49 %	66.05 %	76.04 %	9.99 %	3.36*
2010	69.13 %	69.98 %	0.85 %	70.64 %	74.79 %	4.15 %	0.37

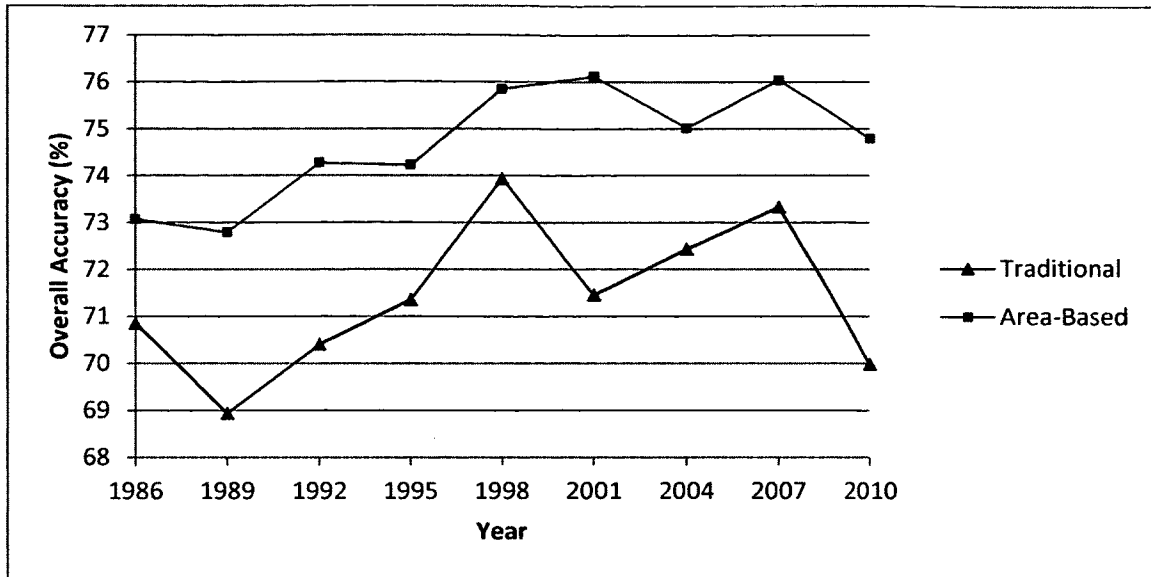


Figure 14. The multi-date overall accuracies computed using the traditional error matrix approach and the area-based error matrix approach.

In a qualitative assessment of the difference between the single-date process and the multi-date process, it was found that the lower the accuracy of the single-date map, the more likely that the difference between the accuracies of the two maps of a single year was measurable. The accuracies of the single-date maps fluctuated more than the accuracies of the multi-date maps, which were fairly consistent across years. Upon inspection of the three years with less than optimal single-date imagery (1992, 1998, and 2004), the addition of other images significantly increased the classification accuracy only in 1992. The 1992 single-date image was taken at the end of September, whereas the 1998 and 2004 images were from the beginning of April, and the 1992 single-date map contained a lot of confusion in the forest classes which were remedied in the multi-date map. Other studies have shown a similar result. For example, in a study by Schriever and Congalton (1995), September images had the lowest accuracy of three

classified dates of imagery for single-date classification of forest types in this study area, but the addition of images from other dates significantly improved the classification. However, in the current study, the single-date April images seemed to easily distinguish between forest types, but had a harder time distinguishing between active agriculture and cleared/other open. Since most crops have not started by early April, agriculture fields and any other open areas may be very similar spectrally. Therefore, accuracies were improved when including imagery from later in the growing season.

In a more quantitative assessment of the differences in accuracies, an exploration of the data found that 1998 was a statistical outlier when trying to predict the difference between the single-date and multi-date accuracies, and so was excluded from the regression analysis. The forward elimination stepwise regression analysis was performed using a standard least squares estimator, and four explanatory variables provided the best model for predicting the difference between single-date and multi-date accuracies (minimum $\Delta AICc = 169.6714$) (Table 11). The two most precise explanatory variables in the model were the accuracy of the single-date map ($SE=0.193$) and the average date of all imagery used in the multi-date maps ($SE=0.021$). There was a negative relationship between single-date accuracy and difference in multi-date and single-date accuracies, as presumed in the qualitative assessment. The average date of all of the images had a positive relationship with the difference in accuracies, indicating that the further into the year the average date was for all of the images in the multi-date map, the more likely the map was to be better than the single date map. However, the coefficient for this variable is quite small and there is likely a limit to how far into the year this relationship holds (a

limit that this study did not reach). Therefore, this variable is likely to be less useful when predicting the difference in single-date and multi-date accuracies.

Table 11. Explanatory variables used in estimating the difference between the single-date and multi-date accuracies. The standard error (SE) of the estimate is reported as a measure of the precision of the estimator. Since tests of significance are not completed when choosing a model based on AICc values, *p*-values are not reported (Burnham and Anderson, 2002).

Explanatory Variable	Coefficient	SE
Intercept	65.829	14.426
Single-date accuracy	-0.992	0.193
% of images in the fall	4.501	2.970
Senescence captured	1.384	0.466
Average date	0.037	0.021

The percent of images in the fall and the existence of a senescent image both had a positive correlation with the difference in accuracies, indicating that fall images are quite important for differentiating land cover classes and contribute to the success of the multi-date process. This result mimics the findings of previous work that have also cited the importance of fall imagery to land cover classification (Schriever and Congalton, 1995; Wolter *et al.*, 1995), and may explain why the lowest difference in accuracy was found in 2010 (the only year without a fall image). Again, since the degrees of freedom are limited in this study, the explanatory power of this model is fairly low. However, the regression analysis was primarily used as a data exploration technique and the results mimic the predictions made in a qualitative assessment of the data. This finding gives slightly more weight to the supposition that the difference in overall accuracies between

the single-date and multi-date classifications is primarily a reflection of how well the single-date process performed.

Conclusions

Overall, the multi-date classification process did perform better than the single-date process. However, only in some years did the additional image processing time and effort result in significantly better classification accuracies. The most helpful factor in determining the value of the extra processing time required for the multi-date classification process is the accuracy of the single-date classification. If the single-date classification was relatively good, it was unlikely that the additional images improved the accuracy significantly. However, if an optimal image for classification is not available, either due to cloud cover or temporal issues, the multi-date process does have the potential to produce a superior map. The potential for improved accuracy increases if fall images are used in the multi-date classification. However, as observed in the 1992 imagery, images from other times of year may help distinguish between other land cover types, particularly forest types.

While the overall accuracies for the multi-date maps are acceptable, we believe that the accuracies could be improved with the addition of ancillary data, as well as some post-processing. For example, distinguishing between forest types can be aided through the use of elevation data. Additionally, active agriculture may be differentiated from other land cover types through the use of NDVI time series analysis, from the beginning of the growing season to the end of the growing season (Moody and Johnson, 2001).

However, in this study, our aim was to directly compare the single-date and multi-date processes, so additional confounding data sources were avoided.

This study confirmed that the multi-date image classification process is a useful endeavor when a single image does not exist that meets the needs of the classification. Future work should determine whether similar results are found when classes are more specific than these used here. The classes used for this study were fairly broad, but even with these broad classes the multi-date classification outperformed single-date classification. The use of more specific classes may make it more likely that spectral information from different times of year increases classification accuracies. In these cases, the increase in accuracy from using the multi-date process may be more pronounced than in this study.

CHAPTER V

POLYFRAG: A VECTOR-BASED PROGRAM FOR COMPUTING LANDSCAPE METRICS

Abstract

The study of landscape fragmentation is important in investigating how biodiversity is changing. Several current software programs calculate metrics associated with landscape fragmentation. The most prevalent of these programs are compatible only with raster-format land cover maps. However, as classification techniques evolve, vector-format land cover maps are becoming more popular and valuable. PolyFrag is designed to compute landscape fragmentation metrics for vector-based land cover maps, is both flexible and comprehensive, and outputs metrics that are similar to those of the most widely used raster-based fragmentation programs, like FRAGSTATS. The program allows for several fragmented and fragmenter land cover classes, as well as different edge widths between interacting classes. In addition, the program is written in Python and is implemented as a tool in esri[®]'s ArcGIS 10.

Introduction

Habitat loss and fragmentation due to increasing populations and the development pressures that come with growth in population, is currently a major concern of landscape ecologists all over the world (Andrén, 1994; MacLean *et al.*, 2010). Many publications have looked at the effects of habitat loss, fragmentation, and change from anthropogenic forces on the landscape (e.g. Haila, 2002; With, 2002; Fahrig, 2003; Turner, 2005; Fischer and Lindenmayer, 2007; Wiens, 2008). In these studies, landscape modification, or the combined effects of loss and fragmentation, has been tied to losses in biodiversity, changes in carbon storage, reduction in water quality, and many other environmental issues (Andrén, 1994; Riitters *et al.*, 2002; Fischer and Lindenmayer, 2007; Vogt *et al.*, 2007). Therefore, identifying and quantifying landscape modification has become a priority for predicting how the landscape will change in the future and what species might be at risk due to these changes.

Both habitat loss and fragmentation are important factors within the study of habitat modification (With, 2002). Several studies have investigated the correlations between habitat loss and fragmentation with species richness or measures of biodiversity (e.g. Blake and Karr, 1987; Flather and Sauer, 1996; Gibbs, 1998; Rosenblatt *et al.*, 1999; Boulinier *et al.*, 2001; Damshen *et al.*, 2008; Brown and Boutin, 2009). Andrén (1994) conducted a meta-analysis of species richness in vegetation communities and concluded that above a certain threshold of habitat loss, the configuration (or fragmentation) of the landscape played an insignificant role in predicting species richness values and species richness was only correlated with habitat loss. However, below that habitat loss threshold, species richness values declined more rapidly than could be explained by

habitat loss alone. Andrén (1994) attributed the more rapid loss of species to effects from habitat fragmentation, or rather, the layout of the remaining habitat fragments. Another study, by Prugh *et al.* (2008), found that the response of species richness in forests to either habitat loss or fragmentation actually depended on the types of land cover surrounding the forest patches. The study found that the effect of the surrounding landscape on species richness within the forest patch was greatest for human modified areas. Forest patches that were created through natural processes showed very little change in species richness due to either area or isolation effects. The authors do note that patch size and isolation may be two ways of demonstrating total habitat availability for species, since isolation usually increases with habitat loss, so the authors conclude that habitat modification influences species richness values within forest patches, and the surrounding landscape can also have a profound effect (Prugh *et al.*, 2008).

Landscape processes, such as habitat modification, are primarily evaluated using land cover maps (e.g. Gustafson, 1998; McGarigal and Cushman, 2002; Fahrig, 2003; Turner, 2005; McGarigal *et al.*, 2012). Historically, when landscape change has been evaluated, studies have investigated only the amount of habitat loss and have not addressed how the amounts of habitat are spatially distributed, while others have studied only the spatial distribution of habitats (Wiens, 1989). However, both the spatial distribution and amount of a particular habitat type, as well as the interaction of these two factors, can have a pronounced impact on biodiversity (Fahrig, 2003). Habitat modification can influence population dynamics, species movement, and overall health of an ecosystem (Moran, 1984; With 2002; Fahrig, 2003; Turner, 2005; Johnson *et al.*, 2006; Fischer and Lindenmayer, 2007; Brown and Boutin, 2009). Therefore, it is important to quantify not

only the amount of a certain habitat type that is available, but also the spatial relationship between pieces of that specific habitat (Riitters *et al.*, 2002). Generally, the breaking apart of habitats and their spatial relationships are quantified using fragmentation metrics. These measures of fragmentation of a landscape can provide important information about the suitability of a landscape for a particular species or ecological community. For example, size, isolation, edge effects from surrounding land cover types, and total core area are all landscape metrics that can be used to describe a particular habitat.

Several programs have been written to compute landscape fragmentation metrics using land cover maps (e.g. Riitters *et al.*, 2002; Parent *et al.*, 2007; Vogt *et al.*, 2007; McGarigal *et al.*, 2012). Some of these programs include: FRAGSTATS (McGarigal *et al.*, 2012); Landscape Fragmentation Tool (LFT) from the Center for Land Use Education and Research (CLEAR) (CLEAR, 2009); Patch Analyst (Rempel *et al.*, 2012); the PATCH Model (Schumaker, 1998); IAN (DeZonia and Mladenoff, 2004); and Conefor (Saura and Torné, 2009). These currently or previously available programs have a wide range of capabilities. Some programs, such as FRAGSTATS, is a standalone product that primarily focuses on statistically representing the landscape using metrics like the effective mesh size of the landscape. Others, like LFT, are run within esri[®]'s ArcGIS, the most prevalent GIS software, but LFT can only be used to create a visual output of fragmentation of a landscape without computing any additional fragmentation metrics. The majority of these programs require that the input land cover maps be in raster format before an analysis can be completed. Patch Analyst is the only program of the list that will accept vector-based land cover maps, but it does not have the flexibility of many of

the other programs regarding input land cover types or edge widths, so it has limited applicability.

Traditionally, land cover maps created from digital imagery use a pixel-based method of classification. In pixel-based classification, each individual pixel is given a land cover label and the resulting land cover map remains in a raster format (or made up of equal area grid cells). However, in newer object-based image analysis (OBIA) as well as older photo-interpretation techniques, pixels are first grouped into objects and then the objects are classified as a single unit (i.e. polygon). The land cover maps created using an OBIA approach or through photo-interpretation are in a vector format (made up of points, lines, and polygons). These vector-based maps must be converted into a raster format prior to being analyzed by the current fragmentation programs. However, the conversion from vector to raster may alter some of the characteristics of the map (Congalton, 1997) making this conversion imprudent for data where the shape of the land cover units is significant or some units are significantly smaller than the average unit area. Therefore, a new fragmentation program is necessary to deal with these vector-based land cover maps, that also provides the flexibility and effectiveness of the more widely used raster-based fragmentation programs.

In order to meet the needs of a growing community of vector format land cover map users, I have created PolyFrag, a fragmentation program that is designed to use vector-based land cover maps. PolyFrag computes landscape fragmentation metrics for vector-based land cover maps. The program's script is written in Python and is compatible with esri®'s ArcGIS 10. PolyFrag outputs a fragmentation shapefile showing areas of edge, patch, and core habitat, a statistics file that contains the landscape metrics of each of the

input polygons, as well as a text file displaying all of the class and landscape metrics. The program is run as a tool in ArcMAP 10 (esri®) and can easily be implemented by users with no Python experience.

Software Specifics

PolyFrag is a unique program that computes fragmentation metrics using land cover maps in a vector format. In order to remain as user-friendly as possible, the program is packaged so that it can be added as a toolbox to ArcMAP 10 (esri®) with a well-documented input window (Figure 15). The program has the ability to compute class metrics for any number of different land cover classes, as well as accept different edge widths (here referred to as buffer widths) for different interacting classes. The land cover classes are placed in one of three categories: fragmented; fragmenter; and matrix. The fragmented classes are the classes being fragmented (e.g. forest), fragmenter classes are the classes affecting the fragmented classes (e.g. developed), and the matrix classes are the land cover types that are background or neither fragmented or fragmenter classes (e.g. water). Each interaction between a polygon of a fragmented class and a polygon of a fragmenter class can have a unique buffer width that represents the distance into the fragmented polygon the fragmenter polygon has an effect. Buffer widths are specific to each study, so the user of PolyFrag has complete control over these values.

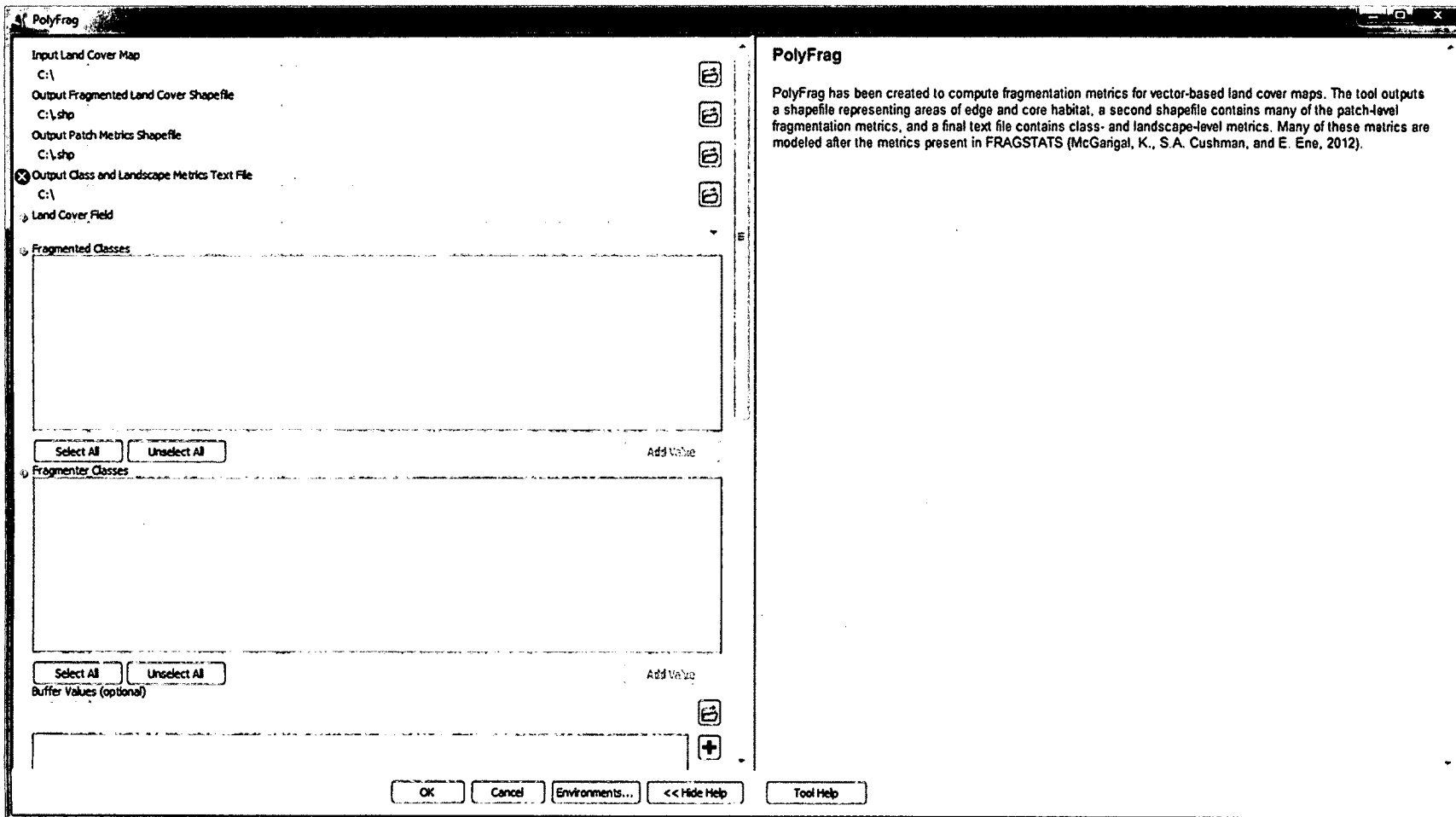


Figure 15. ArcMAP 10 tool window for PolyFrag. Each input or output has its own help screen.

The program requires that the input land cover map be a .shp file and that the two output land cover maps also be in a .shp format. The two output maps are the Output Fragmented Land Cover shapefile and the Output Patch Metrics shapefile. The program also outputs a Class and Landscape Metrics text file. The Output Patch Metrics shapefile is the output map that contains patch-level metrics for each of the input patches. Each of these metrics can be seen in Table 12 under ‘Patch Metrics’. The remainder of the metrics, both class and landscape, are output into the Class and Landscape Metrics text file. Many of these metrics are fashioned after those presented in FRAGSTATS (McGarigal *et al.*, 2012). PolyFrag computes many of the same metrics as FRAGSTATS because even though FRAGSTATS is restricted to raster datasets, it is still the foremost fragmentation program currently available. However, since PolyFrag uses vector datasets instead of raster, some of the metrics have changed somewhat from the original FRAGSTATS metrics to accommodate the change in data format.

Table 12. List of landscape metrics available in PolyFrag. The asterisks denote any metric that is optional. All metrics are modeled after those present in FRAGSTATS (McGarigal *et al.*, 2012)

Fragmentation Metric	Name	Equation	Description
Patch Metrics			
AREA	Patch Area		The area of each polygon
PERIM	Patch Perimeter		The perimeter of each polygon
PARA	Perimeter to Area Ratio	$\frac{\text{PERIM}}{\text{AREA}}$	The perimeter to area ratio for each polygon
SHAPE*	Shape Index	$\frac{\text{PERIM}}{2\pi\sqrt{\text{AREA}/\pi}}$	A measure of shape complexity of a polygon
FRAC*	Fractal Dimension Index	$\frac{2 \ln(\text{PERIM}/4)}{\ln \text{AREA}}$	Another measure of shape complexity

CIRCLE*	Related Circumscribing Circle	$1 - \frac{\text{AREA}}{\pi \left(\frac{\text{P_LENGTH}}{2}\right)^2}$ P_LENGTH = diameter of the smallest circumscribing circle	The area of a polygon divided by the smallest circumscribing circle around that polygon
CAI*	Core Area Index	$\frac{\text{CORE AREA}}{\text{AREA}} * 100$ CORE AREA = area labeled core within that polygon	The percent of the total area of a particular polygon that is actually considered core habitat
NEAR*	Euclidean nearest neighbor distance	x = centroid to centroid distance to the nearest similar polygon	The distance to the nearest neighboring patch with the same label
PROX*	Proximity Index	$\sum_{i=1}^n \frac{\text{AREA}_i}{x_i^2}$	The sum of all of the areas of polygon with the same label divided by the distance to each polygon, limited to only those polygons that fall within a maximum search distance
Class and Landscape Metrics			
CA	Class Area		Total area of each class
TA	Total Area		Total area of the landscape (all polygons)
PLAND	Percentage of Landscape	$\frac{\text{CA}}{\text{TA}} * 100$	The percentage of the total landscape each class represents
NP	Number of patches		The number of patches in each class
PD	Patch Density	$\frac{\text{NP}}{\text{TA}} * 100$	The number of patches in each class per 100 area units (hectares or acres)
LSI	Landscape Shape Index	$\frac{P}{2\pi\sqrt{A/\pi}}$ P = sum of all of the perimeters for the polygons A = sum of all of the areas for the polygons	The total perimeter of all polygons in a single class (in length units) divided by the minimum perimeter possible for the area covered by that class, as computed by the perimeter of a circle with the same area as the total area of the class
LPI	Largest Patch Index	$\frac{\text{LPA}}{\text{TA}} * 100$ LPA = the area of the largest patch for the class(es) in question	The percentage of the total landscape area occupied by the largest patch
TE	Total Edge		The sum of the areas of all of the polygons classified as 'edge'
ED	Edge Density	$\frac{\text{TE}}{\text{TA}} * 100$	The percentage of the total landscape area occupied by the edge polygons
TCA	Total Core Area		The sum of the areas of all of the polygons classified as 'core'

CPLAND	Core Percentage of Landscape	$\frac{TCA}{TA} * 100$	The percentage of the total landscape area occupied by the core polygons
PR	Patch Richness		The total number of classes in the landscape including any background classes
PRD	Patch Richness Density	$\frac{PR}{TA} * 100$	The number of classes per 100 area units
MESH*	Effective Mesh Size	$\frac{\sum_{i=1}^n AREA_i^2}{TA}$	A measure of the size of the patches (in area units) if all patches were evenly distributed throughout the landscape (including any background)
COHESION*	Patch Cohesion Index	$\frac{1 - \frac{\sum P}{\sum(P\sqrt{A})}}{1 - 1/\sqrt{TA}} * 100$	A measure of the connectedness of the class(es) in question
CONNECT*	Connectance Index	$\frac{\sum PROX_NUM}{n(n-1)/2} * 100$ PROX_NUM = the number of polygons with the same label that fall within the max search distance	The percent of the total number of patches that are patches of the same class(es) in question within a maximum search distance
SHDI*	Shannon's Diversity Index	$-\sum_{i=1}^n Pr_i * \ln Pr_i$ $Pr_i = \frac{CA_i}{TA}$	One measure of diversity used in Landscape Ecology
SIDI*	Simpson's Diversity Index	$1 - \sum_{i=1}^n Pr_i^2$	Another measure of diversity used in Community Ecology that is less sensitive to rare patches than SHDI
MSIDI*	Modified Simpson's Index	$-\ln \sum_{i=1}^n Pr_i^2$	MSIDI transforms the SIDI value into a value comparable to SHDI
SHEI*	Shannon's Evenness Index	$\frac{-\sum_{i=1}^n Pr_i * \ln Pr_i}{\ln PR}$	One measure of evenness used in Landscape Ecology
SIEI*	Simpson's Evenness Index	$\frac{1 - \sum_{i=1}^n Pr_i^2}{1 - 1/PR}$	Another measure of evenness used in Community Ecology
MSIEI*	Modified Simpson's Evenness Index	$\frac{-\ln \sum_{i=1}^n Pr_i^2}{\ln PR}$	MSIEI transforms the SIEI value into a value comparable to SHEI

The Output Fragmented Land Cover shapefile is the map that contains polygons classified as core, edge, etc., as well as some patch-level metrics for the core habitats. The input polygons in Fragmented Classes are classified as either 'core', 'edge', or

'patch' polygons. Edge polygons are classified based on the defined buffer widths, and any fragmented polygon area that falls within the designated buffer width is given an 'edge' classification. The remaining fragmented polygon area is classified as either 'core' or 'patch' based on its size and a user defined minimum core area (anything smaller than the given minimum core area size is a 'patch'). Polygons in the Fragmenter Classes are classified as 'fragmenting' and all other polygons, or background polygons, are classified as 'matrix' (Figure 16).

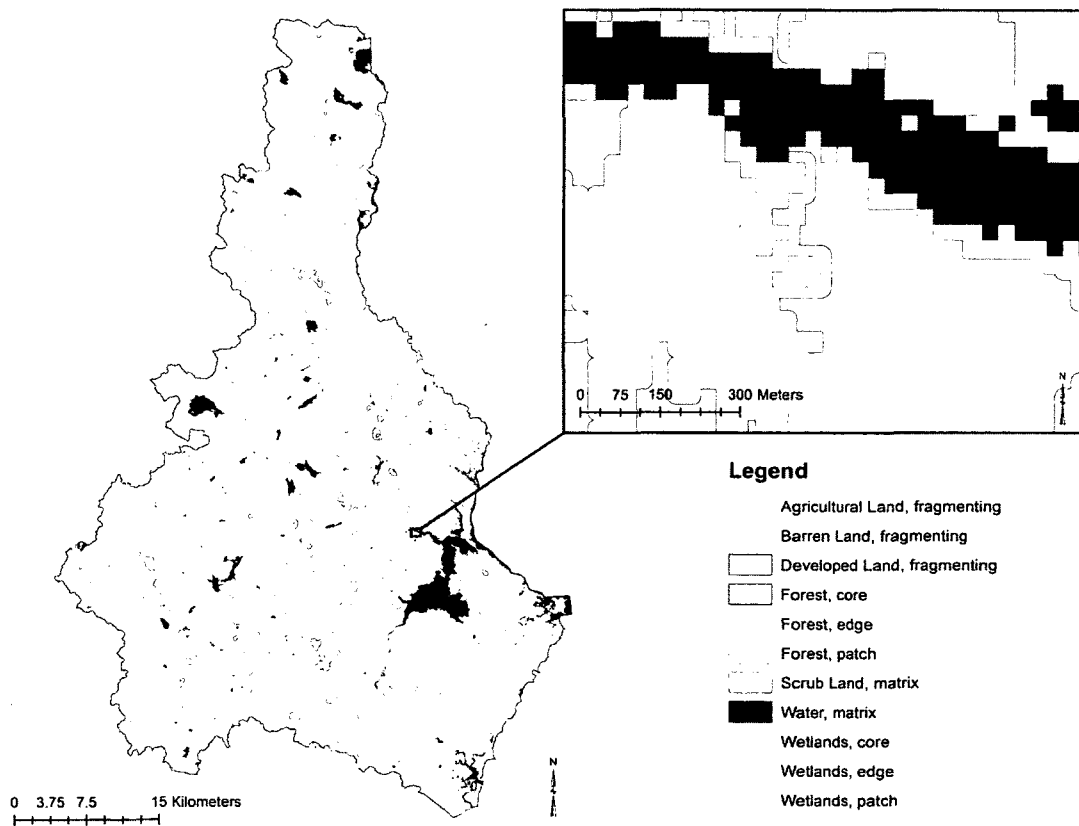


Figure 16. Example of the Output Fragmented Land Cover shapefile from PolyFrag. The map was created using a vector version of NOAA's Coastal Change Analysis

Program (C-CAP) land cover map for the Coastal Watershed of New Hampshire. The forest and wetlands classes are being fragmented by different anthropogenic classes.

Each of the output shapefiles have attribute tables describing each of the polygons (Figure 17). The attribute tables of each of the shapefiles contain the original classification of each of the polygons as well as more specific patch metrics about each of the polygons. The Output Fragmented Land Cover shapefile differs from the Output Patch Metrics shapefile in that it also has a fragmentation class for each of the polygons and many of the metrics are only computed for the core polygons (Figure 17a). The Core Area Index (CAI) metric is also only computed in the Output Patch Metrics shapefile (Figure 17b). For a manual check of the computations made by PolyFrag, please see Appendix C.

(a)

FID	Shape *	Broad_Clas	FRAG	AREA	PERIMETER	PARA	SHAPES	FRAC	P_LENGTH	CIRCLE	NEAR_DIST	PROX	PROX_NUM
5134	Polygon	Forest	patch	100	40	0.4	0	0	0	0	0	0	0
5134	Polygon	Forest	patch	3150	270	0.0857	0	0	0	0	0	0	0
5134	Polygon	Forest	patch	225	60	0.2667	0	0	0	0	0	0	0
5134	Polygon	Forest	core	14900	1816000	0.0122	41.968	1.5316	32234.751	0.8174	1591.7514	0	0
5134	Polygon	Forest	core	54330	678465.128	0.0125	25.9659	1.5078	21606.215	0.8518	1562.8874	0	0
1211	Polygon	Forest	edge	2957.1	357.5982	0.1209	0	0	0	0	0	0	0
1211	Polygon	Forest	edge	6650.2	705.5043	0.1061	0	0	0	0	0	0	0
1211	Polygon	Forest	edge	2183.4	271.1435	0.1242	0	0	0	0	0	0	0

(b)

FID	Shape *	Broad_Clas	AREA	PERIMETER	PARA	SHAPES	FRAC	P_LENGTH	CIRCLE	NEAR_DIST	PROX	PROX_NUM	CAI
3889	Polygon	Forest	2700	240	0.0889	1.3029	1.0364	84.853	0.5225	147.6482	3.1946	52	0
3889	Polygon	Forest	20700	720	0.0348	1.4117	1.0451	228.473	0.4951	124.9181	2.4965	42	94.7
3889	Polygon	Forest	7200	540	0.075	1.7952	1.1046	161.555	0.6488	122.1487	2.3674	45	0
3889	Polygon	Forest	7200	420	0.0583	1.3963	1.048	134.164	0.4907	105.2675	0.7188	62	0
3889	Polygon	Forest	2700	240	0.0889	1.3029	1.0364	84.853	0.5225	72.111	2.1337	52	0
3889	Polygon	Forest	1800	180	0.1	1.1968	1.0157	67.082	0.4907	124.673	0.8471	36	0
3889	Polygon	Forest	1800	180	0.1	1.1968	1.0157	67.082	0.4907	550.3181	0.0417	9	0
3889	Polygon	Forest	2700	240	0.0889	1.3029	1.0364	94.868	0.618	106.1323	1.8122	37	0

Figure 17. Example attribute table outputs for (a) the Output Fragmented Land Cover shapefile and (b) the Output Patch Metrics shapefile.

Software Uses

PolyFrag is useful for a variety of studies. The flexibility in defining fragmenter and fragmented classes, as well as the ability to define the edge width caused by the interactions of these classes, means that this program can be used in a plethora of environments and at many different scales. The program can be used at the landscape level, as demonstrated in Figure 16 using the Coastal Watershed of New Hampshire (NH), which is a Hydrologic Unit Code 8-digit level (HUC-8) watershed. In this example, buffer widths for the agricultural land into the forest land were smaller than

those for the developed land and the forest land. This example may be used to predict potential habitat for Autumn-olive (*Elaeagnus umbellata*), an invasive shrub in NH that prefers habitat with high sunlight and low disturbance, such as along forest edges (Johnson *et al.*, 2006). However, PolyFrag is just as applicable when looking for suitable nest sites for bumble bees (*Hymenoptera: Apidae*) on a farm in the southwestern United Kingdom (Figure 18). The program could be used by landowners concerned with managing for bumble bee habitat, to predict the most valuable habitat to keep. Mapping the interaction between land cover types is important for bumble bee management, since these bees tend to prefer sites along edges between forests and uncultivated fields, but still favor any forest edge over open fields or forests (Svensson *et al.*, 2000; Kells and Goulson, 2003).

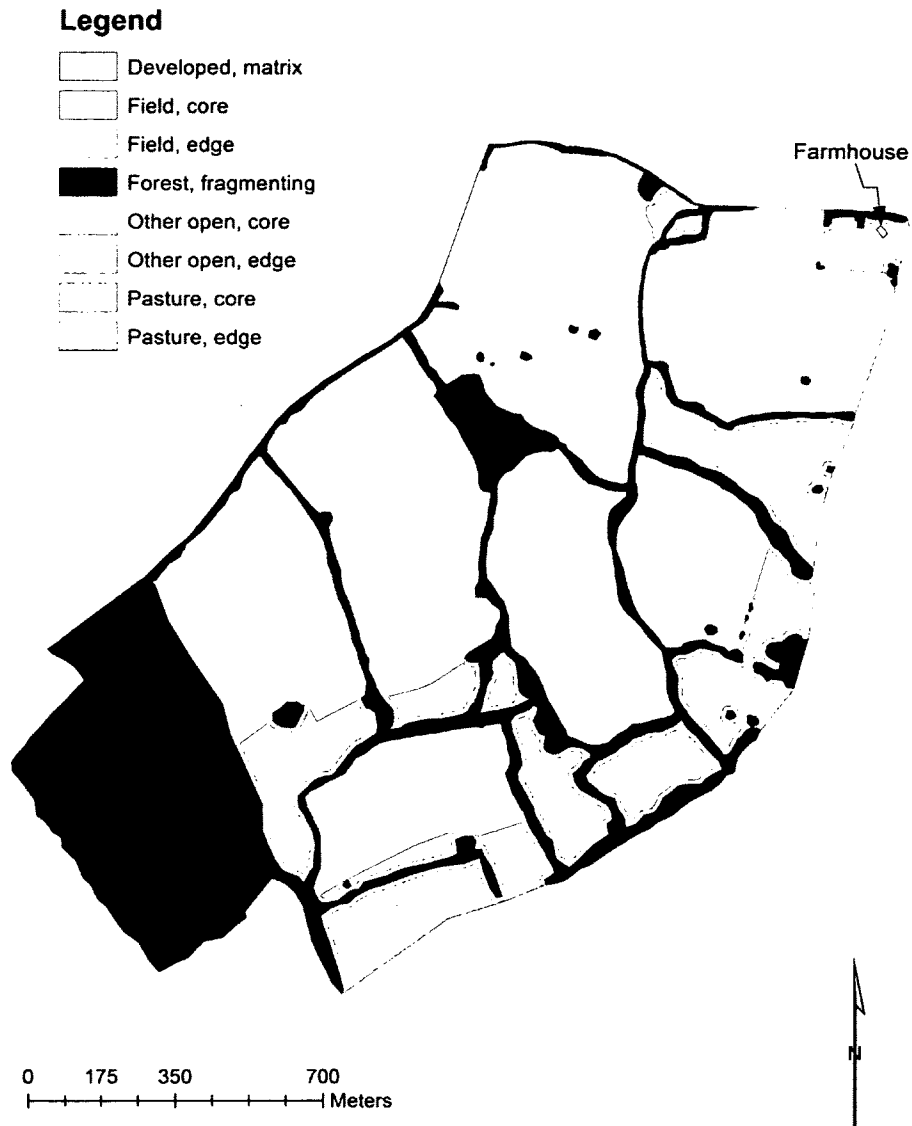


Figure 18. Example of the Output Fragmented Land Cover shapefile from PolyFrag. The map was created using a vector land cover map created using photointerpretation of digital aerial imagery from a farm in southwestern England, with the farmhouse in the northeastern corner of the property. Yellow, or edge, represents the preferred bumble bee nesting habitat on the property, with the largest edge width into the ‘Other open’ land cover class.

One of the strengths of the PolyFrag program is that inputs are relatively simple to modify to fit the particular needs of a study. Another strength of PolyFrag is that it is implemented as a tool in ArcGIS (esri®). As such, PolyFrag is meant to be a very user friendly tool for researchers looking to compute basic landscape metrics for a particular land cover map. However, it is also just a starting point. Future research in landscape fragmentation and fragmentation metrics will hopefully lead to powerful additions to PolyFrag.

Conclusions and Future Directions

With the introduction of PolyFrag, fragmentation metrics can now be computed using land cover maps in vector format. The comprehensiveness and ease of use of the program will ensure that users will not have to convert land cover maps to raster datasets, thereby avoiding the possibility of losing some precision in their data. PolyFrag is also quite flexible, and may be easily modified to meet future needs. For instance, recent studies show that some of the more traditional isolation metrics, such as NEAR or PROX may not be the most appropriate measures of isolation for a landscape (Kupfer, 2012). Therefore, more research must be done to verify newer metrics of isolation, and then incorporated into PolyFrag. The goal is that users of the program will write additional metrics that can be incorporated into the PolyFrag code. Collaboration within the Landscape Ecology community will ensure that the new and most useful metrics are integrated into PolyFrag in a timely manner.

CHAPTER VI

A REVIEW OF USING FRAGMENTATION PROGRAMS TO IDENTIFY POSSIBLE INVASIVE SPECIES LOCATIONS

Abstract

When predicting locations of woody invasive species, mapping habitat fragmentation can be an important part of the prediction process. There are many different fragmentation mapping programs, each computing a unique set of fragmentation metrics to be used in creating a model for attaining probabilities of invasive species presence. In this study, we compare the results from four prevalent, freely available, fragmentation programs: FRAGSTATS; the Landscape Fragmentation Tool; Shape Metrics; and Patch Analyst, and one new program: PolyFrag. FRAGSTATS and PolyFrag created prediction maps with the highest accuracies and were relatively easy to use. FRAGSTATS is recommended for use with raster datasets, while PolyFrag is recommended for vector datasets. Both of the programs compute similar fragmentation metrics and each model found similar metrics were significant in predicting invasive species presence. Both programs predicted that woody invasive species were less likely to be found in deciduous forests than in either mixed or coniferous forests.

Introduction

Woody invasive species have become an important concern in a number of scientific fields due to the impact of these invasive species on natural communities (Henderson *et al.*, 2006). One of the factors enhancing the spread of invasive species is the growth and mobility of the human population. Population growth and urbanization has impacted our natural systems in a number of ways, including land use change and increased habitat fragmentation. In turn, habitat fragmentation and disturbance has been linked to increased vulnerabilities of habitats to invasion by exotics (Moran, 1984; With 2002; Fahrig, 2003; Turner, 2005; Johnson *et al.*, 2006; Fischer and Lindenmayer, 2007; Brown and Boutin, 2009).

When studying landscape fragmentation, land cover maps and fragmentation programs are essential. Land cover maps are necessary to show the current and changing state of the landscape, while fragmentation programs compute fragmentation metrics to describe the state of a landscape based upon those land cover maps (Gustafson, 1998; McGarigal and Cushman, 2002; Riitters *et al.*, 2002; Turner, 2005; Parent *et al.*, 2007; Vogt *et al.*, 2007; CLEAR, 2009; MacLean and Congalton, 2012c; McGarigal *et al.*, 2012; Rempel *et al.*, 2012). There are numerous fragmentation programs that require different types of land cover maps and produce many different fragmentation metrics (e.g. Riitters *et al.*, 2002; Parent *et al.*, 2007; Vogt *et al.*, 2007; MacLean and Congalton, 2012c; McGarigal *et al.*, 2012). Each of these fragmentation programs may be useful in different settings. In this study we sought to determine which fragmentation program, of five tested, performed best when trying to predict the presence of woody invasive species.

The study area for this analysis was the Coastal Watershed in Southeastern New Hampshire. The watershed's natural landscape contains forested areas as well as many small wetlands. However, the population of this area increased 52% in the 30 year time period between 1980 and 2010, and has subsequently seen an expansion of urban areas within that time. The area has also seen a substantial increase in the spread of woody invasive species, due in part to the changing land use types (Johnson *et al.*, 2006). Many studies have shown a positive correlation between woody invasive species and disturbed landscapes, such as old agricultural fields (e.g. With, 2002; Johnson *et al.*, 2006; Brown and Boutin, 2009), or forest edges (e.g. Moran, 1994; Brothers and Spingarn, 1992), both of which are commonly a consequence of urbanization and land use change. However, little has been done in this study area to identify which particular landscape characteristics, or fragmentation types, may increase the likelihood of invasion by these exotics, as found using land cover and fragmentation mapping. Knowing the fragmentation types that increase invasion potential would be extremely helpful for conservation agencies or landowners attempting to protect their natural landscapes from invasive species. Therefore, mapping landscape fragmentation has strong potential in this area. However, determining which fragmentation types are useful in predicting invasion potential can be quite difficult for a number of reasons.

First, while land cover maps are a necessary part of fragmentation mapping, they are a source of error, since no land cover map is ever 100% accurate (Foody, 2002; Congalton and Green, 2009). Also, most land cover maps are in one of two formats: vector or raster. Some fragmentation programs will only accept raster datasets, while others will only work with vector datasets. While it is possible to convert between the

two formats, it is generally not recommended (Congalton, 1997). Therefore, the land cover map chosen for a study can restrict the fragmentation programs available for use. Second, the collection of invasive species data can be very difficult to work with, especially since sampling is often done on a presence-only basis without a statistically sound sampling protocol (Peterson, 2003). Unless a massive and costly sampling effort is undertaken, less than ideal data are often the only data available for predicting invasion presence.

Finally, the fragmentation programs themselves can be quite influential in determining the success of mapping invasion potential. Each program computes a unique set of metrics that can help to predict invasion potential. If a program is chosen that does not compute the metric that best predicts potential, some power is lost when modeling potential presence. Since each fragmentation program has unique advantages and disadvantages, and can radically influence the accuracy of a map of potential invasion, it is important that the best program be chosen. As part of the current work, we compared the outputs of five fragmentation programs to determine which of these programs performed best when identifying potential areas of woody invasive species presence within the Coastal Watershed.

Some fragmentation programs are available for purchase, while others are free to use, but may require other purchased software such as ArcGIS (esri®). The current study is limited to the programs that are either free to use, or only require ArcGIS and are otherwise free, since these programs are the most widely used and easily accessed programs for computing fragmentation metrics (e.g. Riitters *et al.*, 2002; Parent *et al.*, 2007; Vogt *et al.*, 2007; McGarigal *et al.*, 2012). The specific programs addressed here

are: FRAGSTATS; the CLEAR Landscape Fragmentation Tool (LFT); the CLEAR Shape Metrics tool; Patch Analyst; and PolyFrag. As discussed below, each of these programs has different requirements (such as data format type accepted), flexibilities (such as edge width properties), and outputs.

FRAGSTATS

FRAGSTATS is one of the more widely known and used fragmentation programs (MacLean and Congalton, 2012c). The program was first introduced in 1995 as version 2 by McGarigal and Marks (1995). Due to the wide variety of fragmentation metrics that can be computed using FRAGSTATS, and because it is free to use and independent of other programs, versions 2 and 3 have been widely used over the past decade and a half by landscape ecologists. These FRAGSTATS metrics include estimates of core area within habitat patches, proximity or isolation of patches, and many others. The program relies on the equal area grid cells of raster datasets to compute these estimates, so only land cover maps in a raster format are compatible with FRAGSTATS. Users are able to decide which metrics to run on their landscape, as well as define an edge width that is appropriate for their study. Edge widths are generally defined as how far into a given habitat effects of other habitats may be detected, and these widths can change depending on what is being studied.

In March of 2012 FRAGSTATS 4.0 was introduced with many new capabilities (McGarigal *et al.*, 2012). The new program computes essentially the same metrics as the earlier version, but with a more user-friendly graphical user interface (GUI) and with some added flexibility. The new program allows users to define unique edge widths for

different interacting land cover types. This added functionality is incredibly important for landscape ecologists, since it is unlikely that all landscapes have the same effects on the habitat of interest. However, version 4.0 still relies on raster datasets to compute fragmentation metrics, so vector datasets continue to be incompatible with FRAGSTATS. Another limitation of FRAGSTATS involves the fact that the output of the fragmentation metrics is strictly in a tabular format. Therefore, additional data manipulation is necessary to associate patch metrics with a visuospatial representation of the patches.

CLEAR Landscape Fragmentation Tool

The University of Connecticut's Center for Land Use Education and Research (CLEAR) has created a few tools for visualizing and creating fragmentation metrics. The two programs studied here are the Landscape Fragmentation Tool (LFT) and the Shape Metrics tool. Both programs are written in Python and are used as tools within esri®'s ArcToolbox in ArcGIS 9.2 or higher (CLEAR, 2009). Since the programs are used as tools in ArcGIS, ArcGIS is necessary for these programs, but an advantage is that they are fairly straightforward and easy to use for anyone familiar with ArcGIS tools. LFT has two versions, v1.0 and v2.0, and each version maps landscape fragmentation. LFT v2.0 is more widely used than v1.0 and is the version that was chosen for use in this analysis (CLEAR, 2009). LFT v2.0 is a program used to reclassify complex raster datasets into fragmentation maps using four different categories: patch; edge; perforated; and core. While the output is a raster fragmentation map with simple categories, the map does not retain the initial categories or compute any landscape metrics. LFT also lacks the ability to deal with differently sized edge widths. Unlike FRAGSTATS that can deal

with many different landscape interaction types, LFT assumes all edges between land cover types are identical.

CLEAR Shape Metrics Tool

Shape Metrics computes many landscape metrics for polygons, such as proximity index, spin index, dispersion, cohesion, etc., that have historically been difficult to compute for polygons (CLEAR, 2009). This tool computes these metrics by creating many evenly distributed sample points within each polygon and along the perimeter of the polygon, and then uses the distribution of these points to compute the metrics. However, this tool does not compute any landscape metrics and only computes patch metrics for individual polygons of interest (not all polygons). Since each polygon in the analysis must be turned into a series of points, Shape Metrics can take a great deal of processing time if many polygons are chosen for analysis. Fortunately, this tool only computes metrics that are useful in specific instances and therefore can be limited to the polygons where these shape metrics are necessary.

Patch Analyst

Like LFT and Shape Metrics tools, Patch Analyst (PA) is a program that is run as an extension to the ArcGIS (esri®) platform. Therefore, PA is user friendly and easy to employ for those familiar with ArcGIS, but ArcGIS is necessary to run PA. PA is modeled after the original FRAGSTATS program, but unlike FRAGSTATS has the ability to compute fragmentation metrics on vector shapefiles (Rempel *et al.*, 2012). Many of the fragmentation metrics generated in PA are the same as those generated in

FRAGSTATS, but in addition to these metrics, a map of the patches is output into a vector layer and attribute table. The layer can then be viewed in ArcGIS. The creation of an output spatial map of polygons with fragmentation metrics attributes makes for much easier spatial analysis of the data. In addition, PA has additional capabilities that FRAGSTATS does not, including creating hexagon regions and attribute modeling, which are useful for species specific investigations such as range and habitat mapping.

However, PA has very little flexibility when defining habitats of interest or edge widths. Even in PA 5.1 (the newest version of PA for use in ArcGIS 10, updated in April of 2012), all patches are analyzed in the same way, and only a single edge width can be defined. Also, core area and patch metrics must be computed separately. The limited flexibility of the program makes it less than ideal for complex landscapes or more elaborate studies. For example, if there are many different land cover types with different interactions, PA is limited in its ability to model these intricacies.

PolyFrag

PolyFrag was introduced in 2012 by MacLean and Congalton (2012c). The program is written in Python and is used as a new tool in esri®'s ArcToolbox for ArcGIS 10 or higher. Like both PA and the CLEAR tools, PolyFrag is very user friendly for those familiar with ArcGIS, but the ArcGIS software is necessary to use PolyFrag. Similar to PA, PolyFrag is designed to compute common fragmentation metrics on vector shapefiles. Also like PA, PolyFrag outputs shapefiles with attribute tables addressing the patch metrics, as well as a text file containing class and landscape metrics. However, the design of the PolyFrag tool is more cohesive than that used in PA, with both patch

metrics and core areas created in a single process. PolyFrag is also more comparable to FRAGSTATS than PA, computing most of the metrics available to compute for raster data in FRAGSTATS (McGarigal *et al.*, 2012). Like FRAGSTATS, PolyFrag also has the ability to define different edge widths, so different landscape interactions can be addressed.

Since PolyFrag attempts to compute many of the same metrics that are computed in FRAGSTATS, only with vector data, some modifications were made to the metrics so that polygons rather than rasters could be used. Therefore, some of the metrics are not directly comparable, although they are quite similar. Other metrics from FRAGSTATS are highly dependent on rasters and so are not computed in PolyFrag. Despite these modifications and omissions, PolyFrag is the most similar fragmentation program to the widely recognized FRAGSTATS, and the most comprehensive program for computing fragmentation metrics using vector data. The added flexibility of defining different edge widths and different fragmenting and fragmented classes makes PolyFrag much more user friendly than many of the other fragmentation programs.

Each of the five programs described above was used to compute fragmentation metrics for the Coastal Watershed. These metrics, along with woody invasive species locations, were analyzed to determine which metrics were most useful in predicting woody invasive species presence. Our aim was to quantitatively determine which of these fragmentation programs produced the best results for predicting invasive species location, as well as qualitatively assess which of these programs had the highest ease of use, especially for those researchers least familiar with creating fragmentation maps.

Methods

Land Cover Map and Invasive Species Data

A 2010 land cover map of the Coastal Watershed was used as the base land cover map for each of the five processes used to assess the current state of fragmentation of the watershed. The Coastal Watershed is a Hydrologic Unit Code 8-digit level (HUC-8) watershed and is just over 200,000 hectares in size. The map was chosen since it is the most up to date map of the Coastal Watershed and had the necessary land cover types for assessment. The land cover map was created using the same protocol as the multi-date maps created in MacLean and Congalton (2012b), and was created using five Landsat 5TM images from throughout the year 2010. These Landsat images were stacked (without the thermal band), along with NDVI and three Tasseled Cap derivative bands per image, and treated as a single multi-banded image throughout the classification process. An Object-Based Image Analysis (OBIA) approach was used to group pixels into polygons, and those polygons were then classified using a Classification and Regression Tree (CART) approach, all within eCognition (Trimble®). The classification process resulted in a vector-based land cover map with eight different land cover classes ready for analysis (Figure 19).

2010 Land Cover Map

- Active agriculture
- Cleared/other open
- Coniferous
- Deciduous
- Developed
- Mixed forest
- Open water
- Wetlands



Figure 19. The 2010 vector land cover map used to study the Coastal Watershed of New Hampshire.

The classes from the 2010 land cover map were treated as fragmented, fragmenting, or background land cover types in the fragmentation mapping process (Table 13). The fragmented, fragmenting, and background land cover types remained the same throughout each of the analyses. The fragmented land cover types are the land cover types that are being affected by the fragmenting land cover types. Edge widths are defined as the area being impacted by the fragmenting land cover types, and these areas of edge are found solely within the fragmented land cover types along the boundary between the fragmented patch and the fragmenting patch.

Few studies have conclusively determined a maximum edge width for invasive species (Moran, 1984; Brothers and Spingarn, 1992). For instance, in their study, Brothers and Spingarn (1992) found that most plant invasive species were not found at any substantial population size more than eight meters within established forest plots, and Moran (1984) found that 30 meters into an established forest, effects of anthropogenic forces were far less prevalent, with boundaries with residential areas having the highest association with invasive species. For this study, the natural landscapes of concern for invasive species are the forests and wetlands of the Coastal Watershed. Therefore, after reviewing studies attempting to determine how far within forested landscapes woody invasive species are usually found, the edge widths were defined very conservatively.

Table 13. Each land cover class was placed in one of three categories: Fragmented classes; Fragmenting classes; and Background classes; depending on how each class was interacting with the landscape.

Fragmented	Fragmenting	Background
Deciduous Forest	Active Agriculture	Open Water
Coniferous Forest	Developed	
Mixed Forest	Cleared/Other Open	
Wetlands		

The woody invasive species analyzed in this study were identified and located by The Nature Conservancy (TNC) in an effort to inventory their lands surrounding Great Bay in the Coastal Watershed (Glode, 2012). Each of the TNC properties on the northeastern side of Great Bay was surveyed and all locations of woody invasive species were recorded in detail, resulting in nearly 1000 data points representing invasive species presence (Figure 20). Twelve different woody invasive species were identified within the TNC properties (Table 14). However, since the data were recorded as presence-only, in order to create predictive models, pseudo-absence points were also created (Zaniewski *et al.*, 2002; Anderson *et al.*, 2003; Brotons *et al.*, 2004; Elith *et al.*, 2006; VanDerWal *et al.*, 2009; Barbet-Massin *et al.*, 2012). Since logistic regression was used to create the predictive models, the pseudo-absence points could be randomly located throughout TNC lands without weighting, given the assumption that TNC recorded every location of invasive species presence (VanDerWal *et al.*, 2009; Barbet-Massin *et al.*, 2012). One-thousand pseudo-absence data points were created with at least 15 meters between all of the pseudo-absence points, as well as between the pseudo-absence points and the presence points. Fifteen meters was chosen as the largest reasonable distance between points given the limited area of the TNC properties. Half of the invasive species data

points (both presence and pseudo-absence) were used as training data points for the creation of the predictive models, while the other half were set aside to be used as validation data for the accuracy assessment of the predicted fragmentation maps.

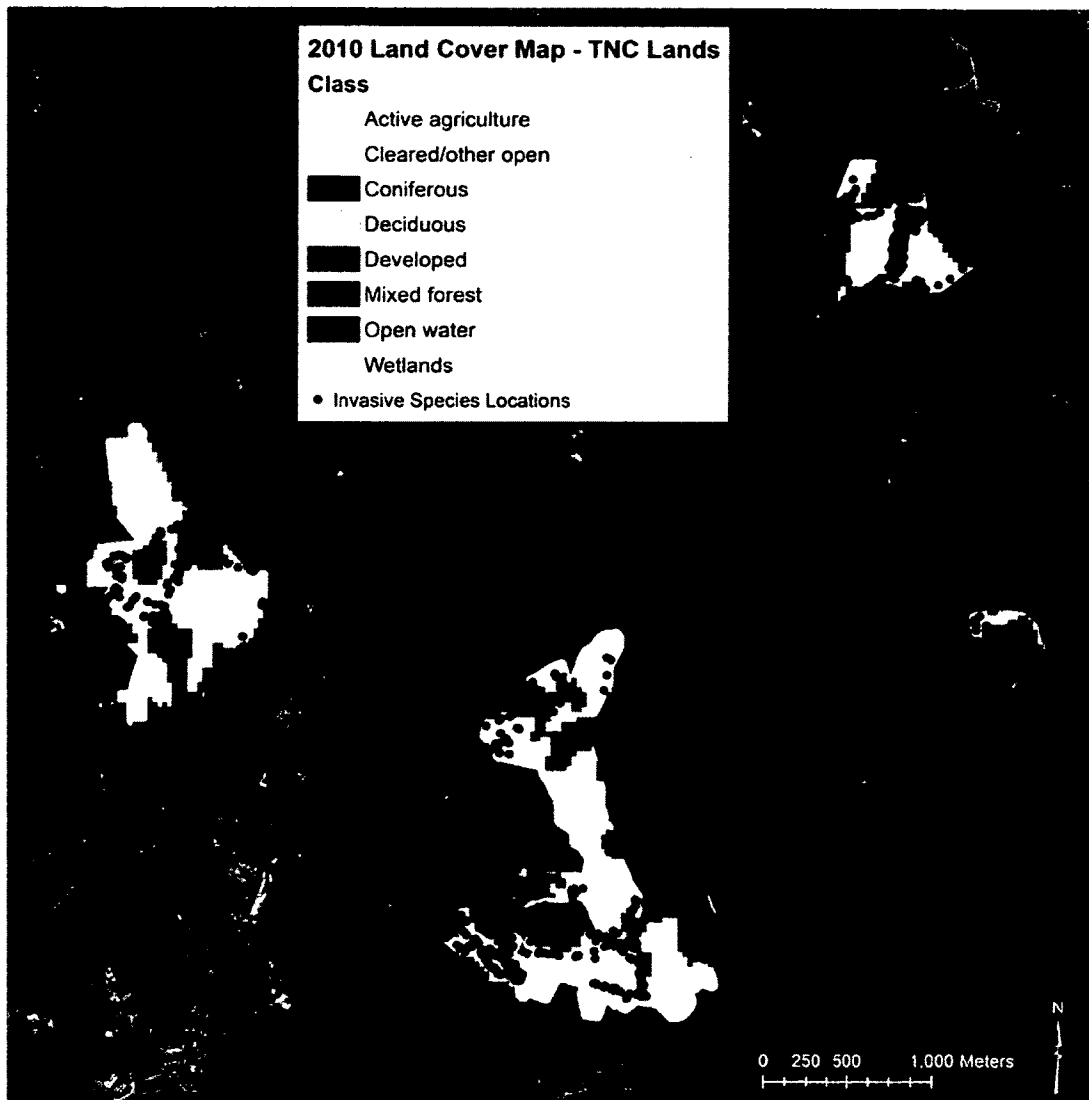


Figure 20. The 2010 land cover map clipped to the extent of the sampled TNC properties with known invasive species locations shown in red. The underlying imagery was acquired by the National Agriculture Imagery Program (NAIP) in 2006.

Table 14. The list of invasive species found within the TNC properties.

Scientific Name	Common Name (type)
<i>Acer platanoides</i>	Norway Maple (tree)
<i>Berberis thunbergii</i>	Japanese Barberry (shrub)
<i>Berberis vulgaris</i>	European Barberry (shrub)
<i>Celastrus orbiculatus</i>	Oriental Bittersweet (woody vine)
<i>Elaeagnus umbellata</i>	Autumn-olive (shrub)
<i>Euonymus alatus</i>	Winged Euonymus (shrub)
<i>Lonicera spp</i>	Honeysuckle (shrub/vine)
<i>Rhamnus cathartica</i>	Common Buckthorn (shrub)
<i>Frangula alnus</i>	Glossy Buckthorn (shrub)
<i>Robinia pseudoacacia</i>	Black Locust (tree)
<i>Rosa multiflora</i>	Multiflora Rose (shrub)
<i>Rosa rugosa</i>	Japanese Rose (shrub)

Fragmentation Map and Predictive Map Creation

The 2010 land cover map was used in each of the five tested fragmentation programs. In order to facilitate the direct comparison of results, each of the five programs was run with as similar parameters as possible, given the unique constraints of each program. All five programs were used to create a fragmentation map that in turn was used to determine which fragmentation metrics were best for predicting the presence of woody invasive species. The significant metrics from each program were determined using logistic regression, and the models were used to create maps of predicted probability of woody invasive species presence. The results of each of the fragmentation programs were judged both qualitatively and quantitatively. The usability, flexibility, and output generation were compared qualitatively, while the accuracies of the prediction maps generated by each of the models were compared quantitatively.

FRAGSTATS

Since the base land cover map is a vector dataset and FRAGSTATS will accept only raster datasets for analysis, the base map had to be converted to raster before it could be

used in FRAGSTATS. To do this conversion in ArcGIS, the polygons of the original land cover map were divided into 30 m x 30 m raster cells (the same size as the original Landsat 5TM pixels) using the majority rule to classify the resulting cells. This raster map was then used in FRAGSTATS to compute all patch, class, and landscape metrics available within the program. The analysis also took advantage of the ability to define different edge widths between land cover types (Table 15). The resulting FRAGSTATS output included three tables, one for each level of metrics: patch; class; and landscape; and a raster file numbering the groups of pixels from the original land cover map into the patches used when computing the metrics. However, a map of the areas considered 'edge' or 'core' is not output in FRAGSTATS, so only the patch metrics of the entire patch could be used in this analysis. Several steps, including processing the output tables so that they could be joined with the output raster, were necessary so that the patch metrics from FRAGSTATS could be given a spatial location.

Table 15. Edge width used between fragmented and fragmenting classes. All edges are into the fragmented class patch and do not affect the fragmenting class patch.

Fragmented Class	Fragmenting Class	Edge Width (meters)
Deciduous Forest	Active Agriculture	15
Deciduous Forest	Developed	20
Deciduous Forest	Cleared/Other Open	5
Mixed Forest	Active Agriculture	15
Mixed Forest	Developed	20
Mixed Forest	Cleared/Other Open	5
Coniferous Forest	Active Agriculture	15
Coniferous Forest	Developed	20
Coniferous Forest	Cleared/Other Open	5
Wetlands	Active Agriculture	20
Wetlands	Developed	35
Wetlands	Cleared/Other Open	5

Once the patch metrics table was joined with the output patch map from FRAGSTATS, the locations of either presence or pseudo-absence of woody invasive species were overlain onto the FRAGSTATS map. The patch metrics raster was then intersected with the known locations of woody invasive species, resulting in a vector file containing the points of presence or absence of woody invasive species as well as the values of the fragmentation metrics of those locations as computed in FRAGSTATS. A model was created in JMP (Version 7, SAS Institute Inc.) using logistic regression to determine the significant metrics for predicting invasive species presence. For FRAGSTATS, only a few of the many metrics were found to be significant in predicting invasive species presence (Table 16). A model was created using these significant metrics (Equation 1) which was then used to compute a predicted probability, using Equation 2, for each location within TNC study area, resulting in a predictive map for invasive species presence. The accuracy of the predicted fragmentation map was then assessed using the presence and pseudo-absence validation data.

Table 16. The significant predictors of the presence of woody invasive species as determined by FRAGSTATS. Where: CLASS is the land cover type, which is a categorical metric; CIRCLE is the related circumscribing circle; CONTIG is contiguity index; and ENN is functional nearest neighbor distance. For further discussion of these metrics, please see McGarigal and Cushman (2012).

Metric	Estimate	p-value
Intercept	4.334	<0.0001
CLASS[Active Agriculture]	0.623	0.0003
CLASS[Cleared/Other Open]	-0.758	<0.0001
CLASS[Deciduous Forest]	-0.460	0.0103
CIRCLE	1.907	0.0165
CONTIG	-6.579	<0.0001
ENN	-0.002	0.0016

$$a = 4.334 + 0.623 * CLASS[Active Agriculture] - 0.758 * CLASS[Cleared/Other Open] - 0.460 * CLASS[Deciduous Forest] * 1.907 * CIRCLE - 6.579 * CONTIG - 0.002 * ENN \tag{1}$$

$$P = 1 / (1 + e^{-a}) \tag{2}$$

CLEAR Landscape Fragmentation Tool

The process for creating the predictive map using LFT is similar to that used for FRAGSTATS, and the same raster land cover map was used in both analyses. LFT, however, does not have the ability to define different edge widths, and the one edge width that is chosen must be larger than a single pixel. To meet these criteria, and to equal the largest edge width defined for FRAGSTATS, an edge width of 35 meters was chosen for this application, though pixels are only defined as “edge” if their centroids fall within the edge width distance. Since LFT only outputs four different fragmentation

cover types and does not retain any of the original land cover types or compute any further statistics on its own, only the fragmentation cover types could be used to create a predictive model. In order to keep LFT analysis similar to those used for the other fragmentation maps, a logistic regression was again performed, assigning predictive values to each of the fragmentation types (Table 17), and a predictive map was created from that model.

Table 17. The significant predictors of the presence of woody invasive species as determined by LFT, where TYPE is the fragmentation cover type.

Metric	Estimate	<i>p</i> -value
Intercept	-0.138	0.3049
TYPE[fragmenting]	0.449	0.0025
TYPE[edge]	0.138	0.4916
TYPE[core]	-0.085	0.8142

CLEAR Shape Metrics Tool

A similar analysis was completed using Shape Metrics, but since Shape Metrics can accept vector-format datasets, the original vector land cover file was used. The specific metrics computed by Shape Metrics were computed for all of the fragmented land cover types, and the resulting Shape Metrics were joined with the original polygons of the TNC properties, resulting in a set of polygons representing fragmented land cover types and non-fragmented land cover types. A set of new metrics were found to be significant in predicting invasive species presence (Table 18). A separate predictive map was also created using this model.

Table 18. The significant predictors of the presence of woody invasive species as determined by Shape Metrics.

Metric	Estimate	<i>p</i> -value
Intercept	-0.138	0.0425
Perimeter	-0.0006	0.0020
Proximity	0.014	0.0009

Patch Analyst

The analysis using PA was quite similar to those using FRAGSTATS and LFT. However, because, like Shape Metrics, PA works with vector datasets, the conversion to raster was not necessary prior to creating the fragmentation map. Also, because PA runs landscape statistics separately from creating core areas, one map of patch metrics was created, as well as another map for core areas of fragmented land cover types (listed in Table 13). Like LFT, PA will only allow for a single edge width to be defined, so an edge width of 35 meters was defined to match that of LFT fragmentation map. The core map was then intersected with the fragmentation metrics map in order to create a complete map of fragmentation as well as an attribute file with all of the fragmentation metrics for the landscape. This combined map was intersected with the presence/pseudo-absence data.

Also as part of the PA analysis, another model was created using logistic regression, and a different set of metrics were found to be significant in predicting invasive presence (Table 19). A new predictive model was created, as well as a new predictive map. This new predictive map was created by giving each polygon in the fragmentation map a predicted probability of the presence of woody invasive species as computed by the model.

Table 19. The significant predictors of the presence of woody invasive species as determined by PA. Where: SI is the shape index; PAR is the perimeter-area ratio, and FD is fractal dimension. For further discussion of these metrics, please see Rempel *et al.* (2012).

Metric	Estimate	<i>p</i> -value
Intercept	55.577	0.0099
CLASS[Active Agriculture]	1.036	0.0019
CLASS[Coniferous Forest]	1.104	0.0006
CLASS[Mixed Forest]	0.790	0.0017
CLASS[Open Water]	-4.662	<0.0001
FRAG[core]	-0.803	0.0013
Area	9.37e-6	0.0016
Perimeter	-0.001	0.0019
SI	5.005	0.0025
PAR	115.232	0.0004
FD	-50.411	0.0071

PolyFrag

The steps for creating the predictive map using PolyFrag were most similar to those of FRAGSTATS. However, PolyFrag computes fragmentation metrics using vector datasets, so no conversion to raster was necessary. Similar to FRAGSTATS, PolyFrag can accept many different edge widths defined between different land cover types. Therefore, the same edge widths were used as in FRAGSTATS (Table 15). PolyFrag computes patch metrics on patches as a whole in one vector shapefile, as well as patch metrics on core areas in another shapefile. For the purposes of creating a model to predict invasive species location, both of these shapefiles were combined prior to intersecting the map with invasive species data. Once the invasive species data were intersected with the combined fragmentation map, another model was created using logistic regression and another set of metrics unique to PolyFrag were found to be

significant in predicting invasive presence (Table 20). A new predictive model was created, as well as a new predictive map for PolyFrag. The predictive map was created in the same way as with PA, but using the PolyFrag predictive model.

Table 20. The significant predictors of the presence of woody invasive species as determined by PolyFrag. Where: PARA is the perimeter-area ratio; PROX is the proximity index; and PROX_NUM is the number of nearest neighbors within a search radius. For further discussion of these metrics, please see MacLean and Congalton (2012c).

Metric	Estimate	<i>p</i> -value
Intercept	-1.142	0.0002
CLASS[Active Agriculture]	0.884	0.0006
CLASS[Coniferous Forest]	1.018	0.0009
CLASS[Mixed Forest]	1.009	0.0006
CLASS[Open Water]	-3.786	<0.0001
FRAG[core]	-1.214	0.0004
PARA	41.532	<0.0001
PROX	-1.561	<0.0001
PROX_NUM	0.323	0.0002

Results and Discussion

All five of the probability maps that were created in this analysis show slightly different results (Figure 21 through Figure 25). In a qualitative assessment of the maps, it is quickly apparent that LFT and Shape Metrics tool produced unique and startling results. Looking at the probability models, it is not surprising that these results did not match those from the other programs. LFT model could only predict invasive presence probability using fragmentation land cover types (i.e. core, edge, or fragmenting land cover types). Therefore, only three probability levels are observed, with the highest

potential for invasion in the fragmenting land cover types (Figure 22). Shape Metrics only produced metrics for the fragmented land cover types, so no information from the fragmenting land cover types was used for probability modeling. Therefore, the probability of invasive presence is purely based on the perimeter of the polygon and the proximity of all interior points of a polygon to the centroid of that polygon (a measure of compactness). Since proximity was not computed for non-fragmented polygons, their probability was purely based on the perimeter of the polygon, which is why most of the polygons fall in the same 0.25-0.5 probability range (Figure 23).

While the predictive models for both PA and PolyFrag were the most similar, the predictive maps produced are visually quite different (Figure 24 and Figure 25). Both models predicted a positive relationship between presence and Active Agriculture, Coniferous Forest, and Mixed Forest, as well as a negative relationship between presence and Open Water and Core Areas. However, the magnitude of these relationships differed among the two models. Another difference involved how edge widths were defined. It is clearly seen in the PA fragmentation map that the larger edge width, as well as area and perimeter, were quite important in determining predicted probabilities (Figure 24). The PolyFrag map (Figure 25) is more similar to the FRAGSTATS fragmentation map visually, but at first glance it appears that the FRAGSTATS model predicted a few more areas with high probability (0.75-1) of invasive presence (Figure 21). However, upon further inspection, these locations are actually areas of Open Water, which is a background fragmentation land cover type. In FRAGSTATS, no metrics are computed for background classes, so these high probability areas are actually false positives due to the absence of fragmentation metrics computed for these areas.

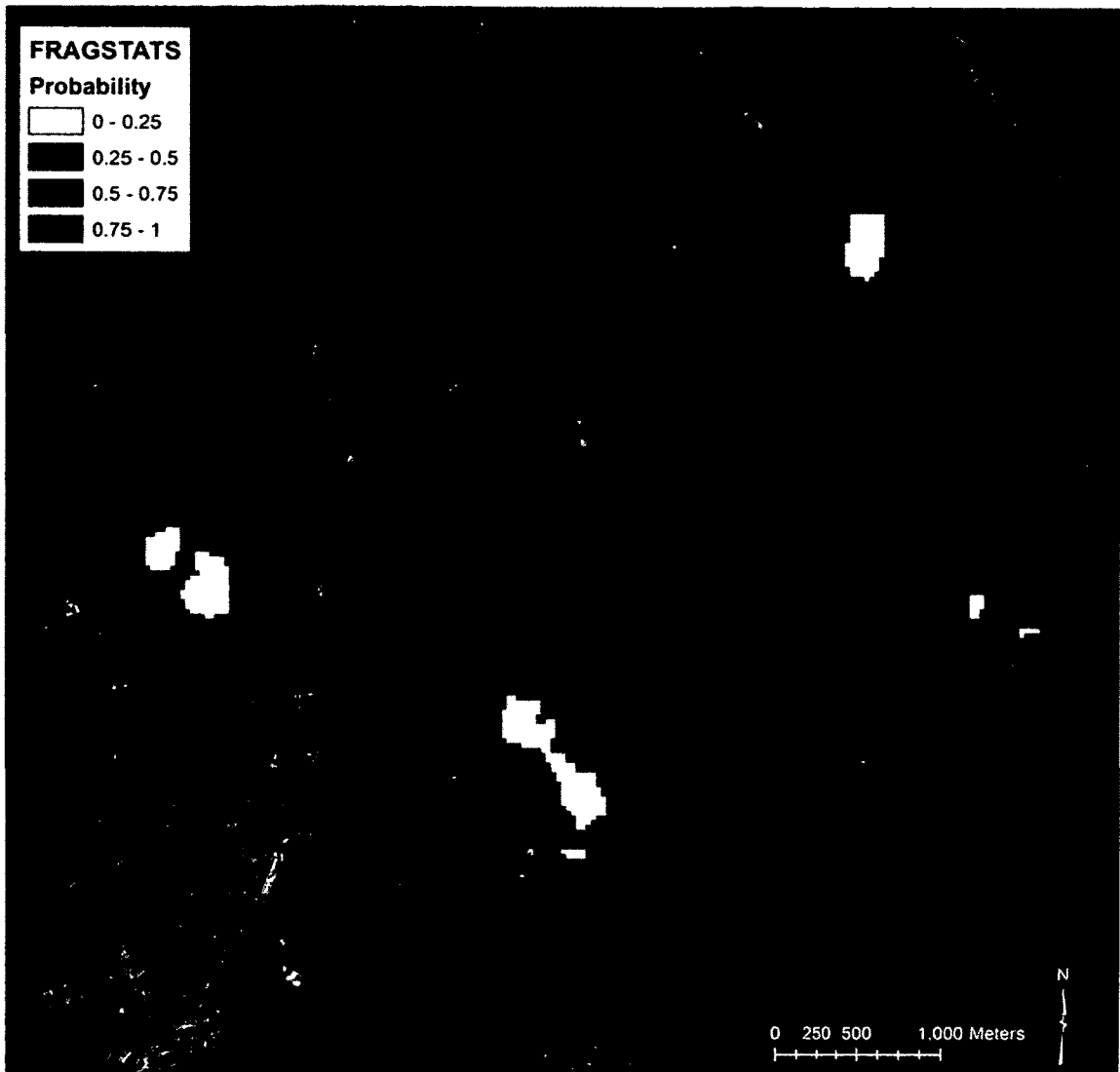


Figure 21. The predicted probability map created by FRAGSTATS for the TNC properties.

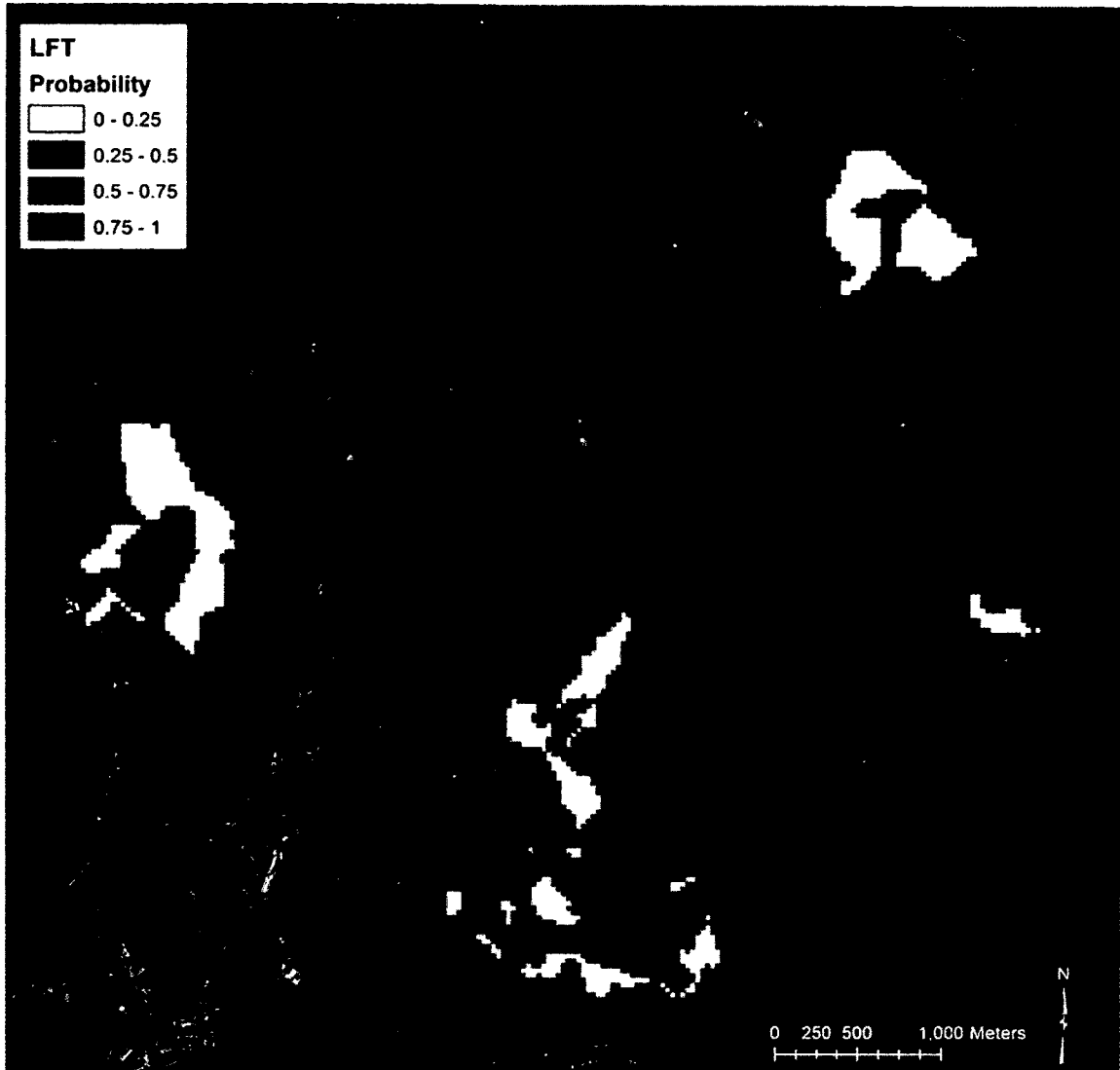


Figure 22. The predicted probability map created by LFT for the TNC properties.

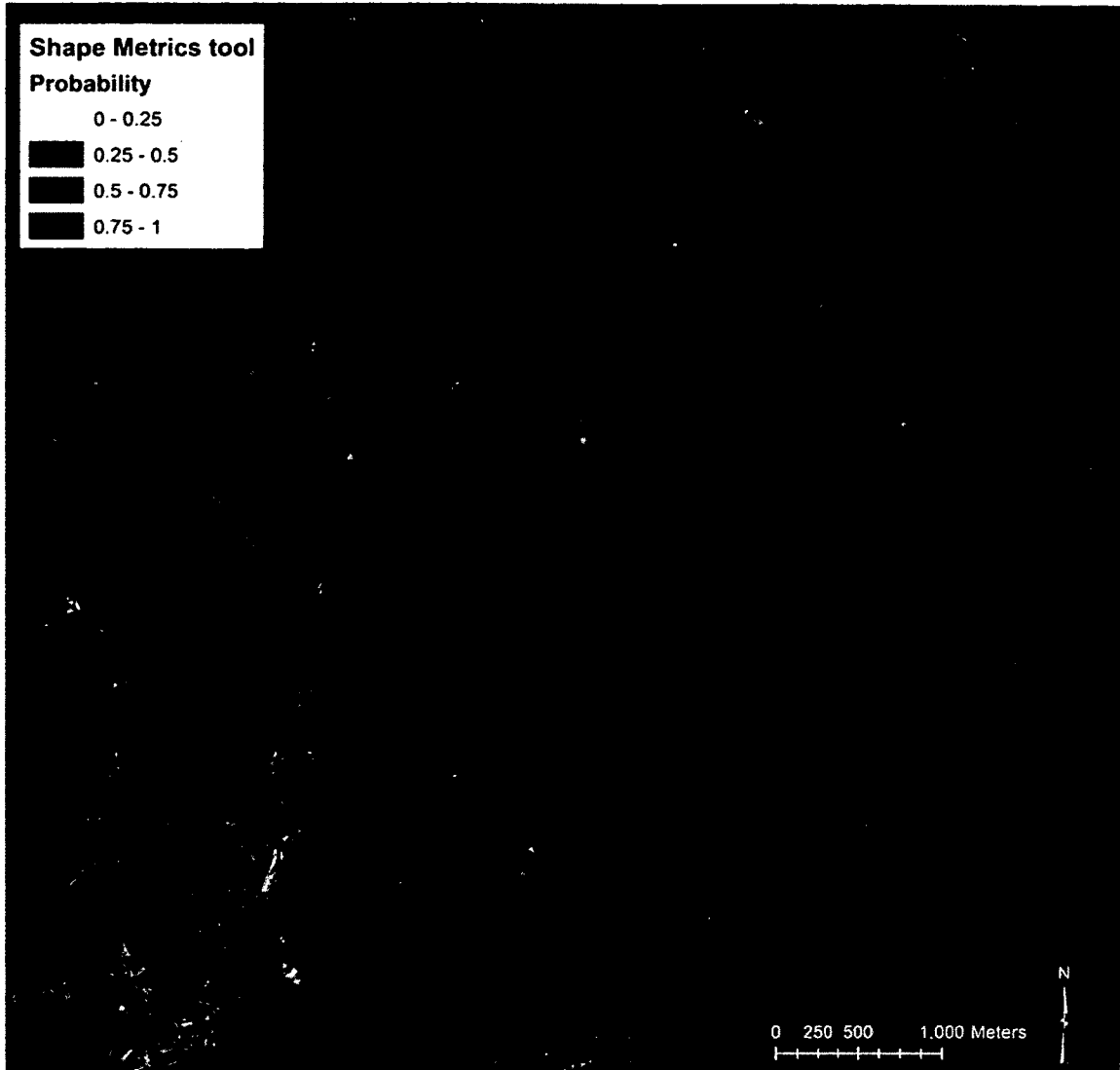


Figure 23. The predicted probability map created by Shape Metrics for the TNC properties.

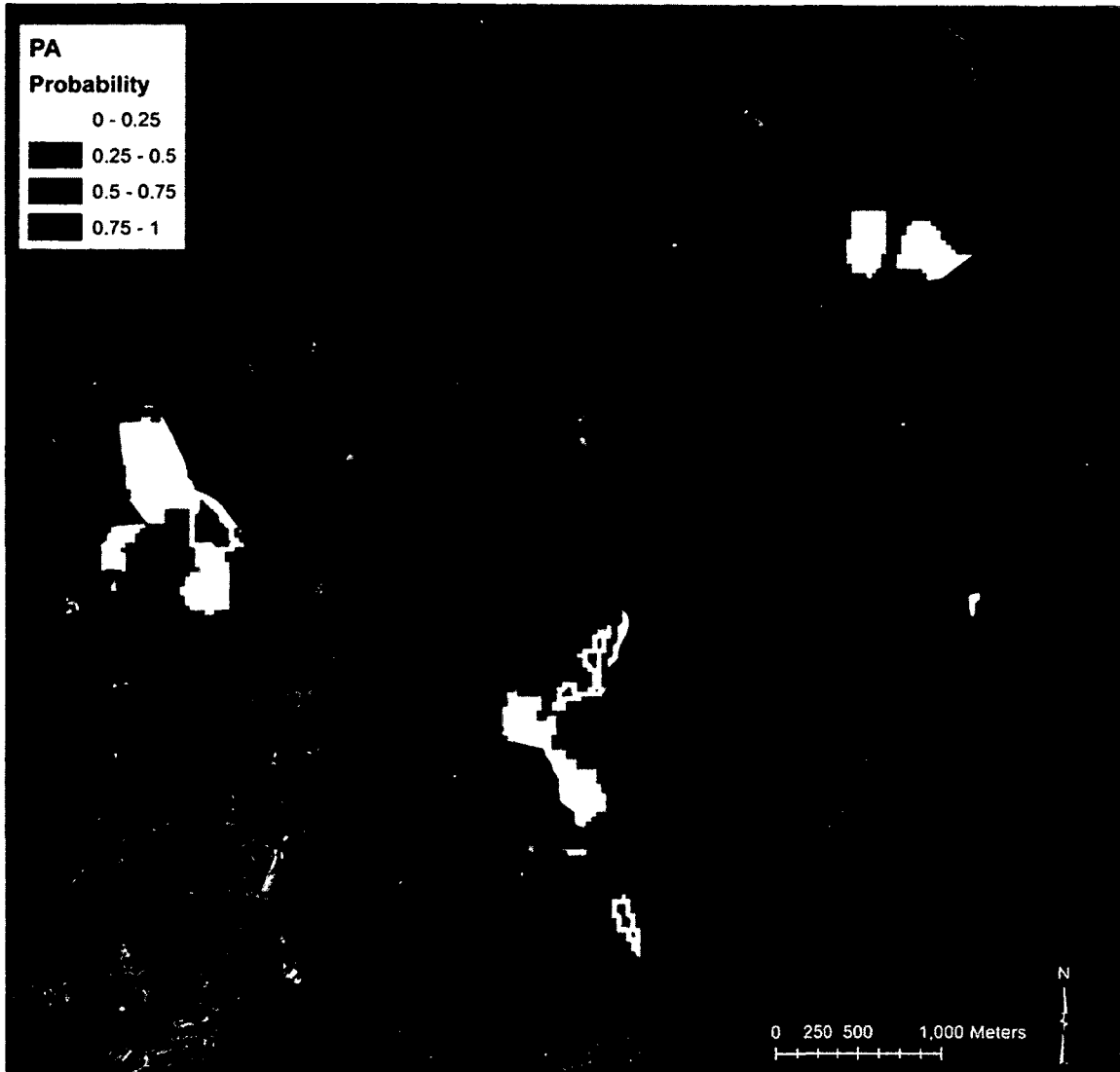


Figure 24. The predicted probability map created by PA for the TNC properties.

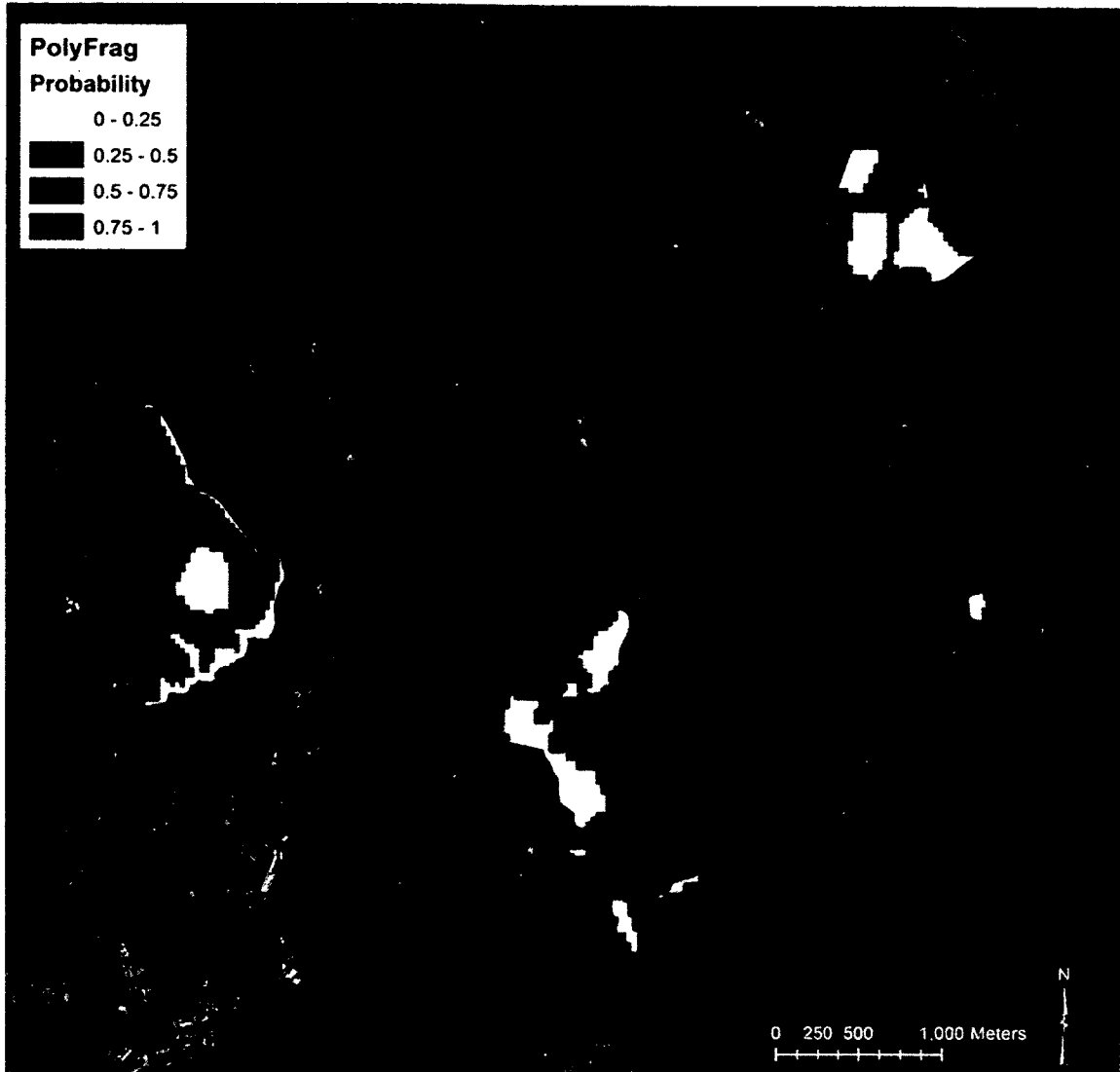


Figure 25. The predicted probability map created by PolyFrag for the TNC properties.

Quantitative Assessment of the Fragmentation Programs

An accuracy assessment was completed for each of the five fragmentation maps using traditional accuracy assessment techniques (Congalton *et al.*, 1983). With these predicted probability fragmentation maps, a probability of 0.5 was used as the threshold over which presence was assumed. Therefore, any location on the map with a predicted probability of 0.5 or higher was assumed to have a prediction of presence, and any location with a

predicted probability lower than 0.5 was assumed to be predicting absence. The known locations of presence and pseudo-absence of woody invasive species, which were previously set aside as validation data were then compared to the presence/absence maps and an error matrix was completed for each of the fragmentation maps. A Kappa analysis was also performed to determine if the maps were significantly different from one another (see Congalton *et al.*, 1983). Not surprisingly, FRAGSTATS and PolyFrag attained the highest accuracies, with FRAGSTATS significantly better than all of the other programs ($p < 0.05$; Table 21). More remarkably, PA had the lowest accuracy, though not significantly.

Table 21. The accuracies of the predictive maps in descending order. The superscript numbers represent whether the predicted maps are significantly different using a Kappa analysis. If the fragmentation programs share a number, they are not significantly different ($p < 0.05$).

Fragmentation Program	Accuracy
FRAGSTATS ¹	66.16 %
PolyFrag ²	61.47 %
LFT ^{2,3}	56.37 %
Shape Metrics tool ³	54.84 %
PA ³	52.60 %

Of the five programs, LFT actually had the fewest errors of omission for invasive presence (36%), FRAGSTATS had the second lowest (38%), and PolyFrag had the third lowest (60%). Though by tweaking the threshold of presence to a more conservative 0.45, errors of omission for PolyFrag, LFT, and FRAGSTATS improved (to 25%, 27%, and 32%, respectively). Since errors of omission are far graver than errors of

commission for predicting invasive species presence (Peterson, 2003), using a threshold of 0.45 is a useful tactic for identifying any and all areas susceptible to current or future invasion. When using a threshold of 0.45 the predictive maps created using FRAGSTATS and PolyFrag were not significantly different, but both models were significantly better than LFT ($p < 0.05$).

Qualitative Assessment of the Fragmentation Programs

The fragmentation programs used in this analysis generally fell into two categories: those that use raster datasets, and those that use vector datasets. Of the programs that compute metrics on raster datasets, FRAGSTATS performed significantly better than LFT (the only other raster format program). FRAGSTATS also had the ability to define many different edge widths, and produced many more fragmentation metrics than LFT. However, the new FRAGSTATS GUI did require more time to explore and learn than did the ArcGIS toolbox created for LFT. FRAGSTATS also required more processing time than did LFT, but neither program took more than an hour to compute metrics for the entire Coastal Watershed.

Of the fragmentation programs that compute metrics on vector datasets, PolyFrag performed significantly better than the two other fragmentation programs (Shape Metrics and PA). PolyFrag also computed metrics that were more similar to those produced by FRAGSTATS, so, given FRAGSTATS' performance, it is not surprising that PolyFrag created a more accurate model than those produced by the other two vector programs. All three of these vector programs ran within ArcGIS (esri®), but Shape Metrics and PolyFrag each ran as toolboxes, whereas PA was added as a dropdown menu in the main

window. Shape Metrics and PolyFrag both appeared more intuitive to use within ArcGIS, and both also computed all metrics with one run, whereas PA needed two runs to compute core areas and shape metrics. As far as processing time, PolyFrag took slightly more time than PA, taking a total of a few hours for the entire Coastal Watershed. However, Shape Metrics took over 10 days to complete on the same workstation and produced significantly worse resultant mapping accuracies.

Overall, FRAGSTATS and PolyFrag were the most accurate and most user friendly of the five tested fragmentation programs. To compare these two programs further, the two predictive models created using FRAGSTATS and PolyFrag were extrapolated for the entire Coastal Watershed (Figure 26). While the accuracies of these extrapolated prediction maps are likely low, the maps provide a visual for the comparison of the two models. Both models found that Active Agriculture and forest types were significant in predicting woody invasive species presence. Both models also found that some form of shape metric (CIRCLE, CONTIG, or PARA) and proximity to nearest neighbor measure (ENN, PROX, or PROX_NUM) were also significant in predicting presence. However, the extrapolated prediction maps are not visually similar. Nearly 54% of the area on the FRAGSTATS map falls into the 0.25-0.5 probability range, with only 4% of the area in the very high probability range (0.75-1.0). Conversely, the predicted probabilities on the PolyFrag map are much more variable, with more area in the very high and very low probability categories. Some of these differences may be accounted for by the addition of the core metric in the PolyFrag model, as well as the different weights of shape metrics and proximity measures.

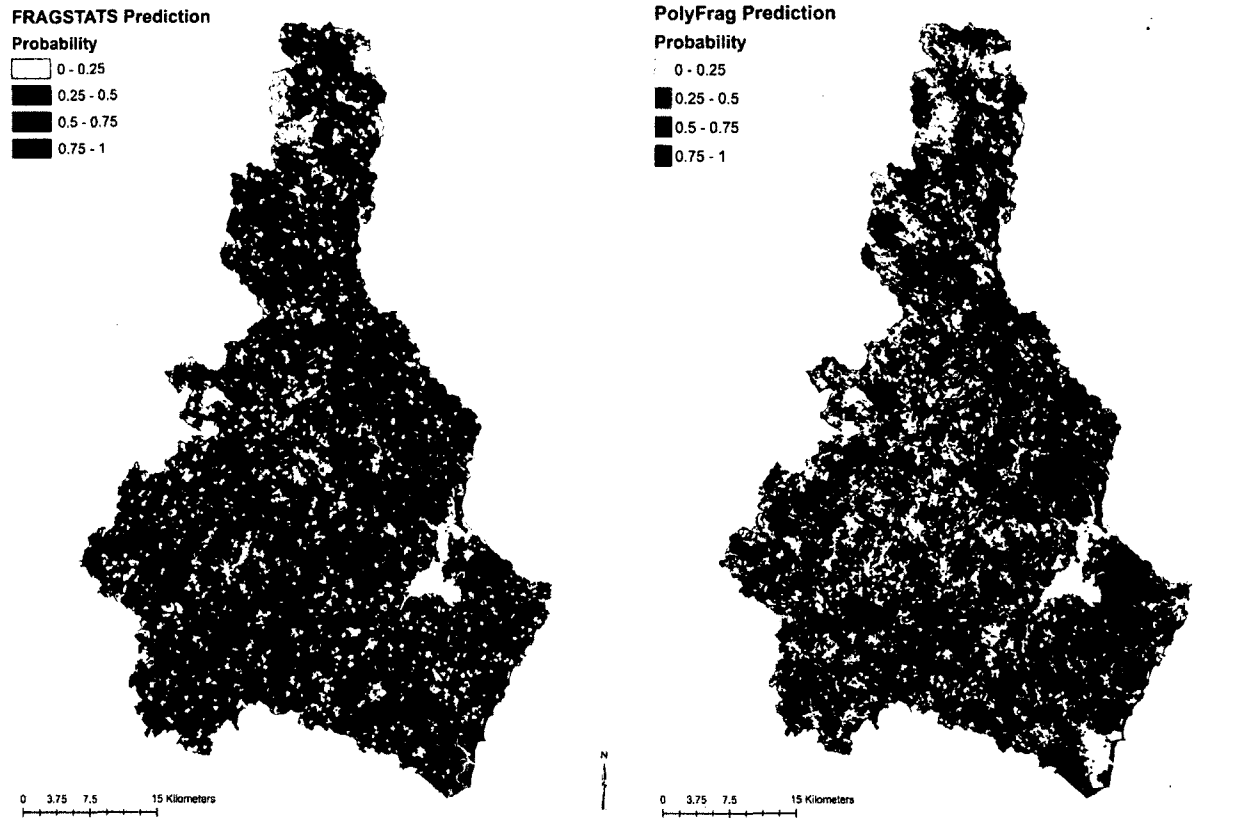


Figure 26(a). The predicted probability of the presence of woody invasive species throughout the Coastal Watershed as predicted using FRAGSTATS. The predicted probabilities of the background classes (Open Water in this case) were all changed to zero post-classification. **26(b).** The predicted probability of the presence of woody invasive species throughout the Coastal Watershed as predicted using PolyFrag.

Conclusions and Future Work

FRAGSTATS and PolyFrag produced the two most useable and accurate models for predicting woody invasive species presence. The advantages of using PolyFrag over FRAGSTATS include: increased ease of use; inclusion of fragmentation metrics, such as core or edge, in assessment; spatial map of fragmentation; and the ability to process vector datasets. The advantages of using FRAGSTATS over PolyFrag include: decreased processing time; a few additional fragmentation metrics associated with raster datasets; and the ability to process raster datasets. Given that converting from raster to vector and especially vector to raster can introduce mapping error, avoiding conversion is optimal. Therefore, it is recommended that either FRAGSTATS or PolyFrag be used with the appropriate data type. In the case of a choice of raster or vector format land cover map, and the land cover maps are equally accurate, FRAGSTATS and raster format data are recommended, since the FRAGSTATS fragmentation map was slightly more accurate than the PolyFrag fragmentation map (when using a 0.5 probability threshold).

Both the PolyFrag and FRAGSTATS probability models found that Active Agriculture was a significant predictor of woody invasive species presence and the PolyFrag model also found a negative relationship between the core areas of all forest types with presence of woody invasive species, which concurs with many previous studies' findings of woody invasive species in this area (e.g. Johnson *et al.*, 2006). In addition, both the FRAGSTATS and PolyFrag models indicated that elongated polygons, or polygons with large perimeter to area ratios, as well as relatively isolated polygons, were more likely to have woody invasive species present. Interestingly, both models also

found that Deciduous Forest was less likely to have woody invasive species than either Coniferous Forest or Mixed Forest land cover types.

The prediction that woody invasive species may be found less often in deciduous forests and more often in either mixed or coniferous forests is in alignment with some anecdotal evidence of what is found on the ground within these TNC properties as well as some theories on how past land use can influence woody invasive species presence. In this area, the positive relationship with the Coniferous Forest land cover type and woody invasive species presence and the negative relationship of presence and the Deciduous Forest land cover type may actually be a surrogate for the relationship between invasive species presence and past land use. In the Coastal Watershed, pure deciduous forests are usually the oldest continually forested areas remaining in the watershed (Foster, 1992), possibly surviving the intense deforestation of the early 1800s as a result of acting as the woodlots for the adjacent farmland.

Coniferous forests in this watershed are generally comprised of either Eastern hemlock (*Tsuga canadensis*) or white pine (*Pinus strobus*), which have very different relationships with invasive species and past land use. While finding woody invasive species under hemlock is rare, especially because it can also be associated with older forests. However, woody invasive species are often found in white pine stands. White pine is also generally considered an indicator of historically cleared sites, since often the natural reforestation of abandoned pastures and fields in the late 1800s included the establishment of white pine in these open areas (Foster, 1992). Not surprisingly, white pine is the dominant coniferous species on TNC properties used to create the predictive models, in which there was a positive correlation between woody invasive species

presence and the Coniferous Forest type. The relationship between woody invasive species presence and white pine may have been even more pronounced if the individual coniferous species had been differentiated prior to creating the predictive model, but due to the lack of an appropriate land cover map, determining the strength of the association between white pine and woody invasive species was not possible.

However, these results should be tested on a larger study area with more conclusive invasive species data. The results of this study are limited by the available invasive species data as well as the availability of land cover maps. While this study is a realistic representation of what is generally available for invasive species studies, a more comprehensive dataset may allow for greater exploration of the nuances of using vector versus raster format data for the purposes of fragmentation mapping.

CHAPTER VII

OVERALL CONCLUSIONS

The goal of this research was to create land cover maps of the Coastal Watershed, assess forest fragmentation, and estimate the probability of invasion by exotic species at different locations throughout the watershed. Additionally, many of the related issues involved in the mapping of fragmentation and invasion were examined. Initial land cover mapping of the watershed included using an object-based image analysis (OBIA) approach to classification that first groups pixels into segments with additional qualities that can aid in classification (e.g. size, shape, texture, etc.). However, when using an OBIA approach, reference data sample units must be chosen from the created segments, rather than using a three-by-three (or larger) cluster of pixels. When using segments (i.e. polygons) as reference data sample units, labeling the reference data can be more difficult, especially in forested land cover types where composition can be continually changing. Prior to this research, no literature existed recommending a sampling scheme for labeling polygon reference units within forested land cover types. I found that for the forests in the Coastal Watershed, six prism samples randomly located throughout a polygon should allow for the labeling of a reference data unit with minimal error. This recommendation, as well as the methods to determine the minimum number of necessary

prism samples, is novel and is a marked contribution to the combined fields of forestry and remote sensing.

As difficult as it is to classify forested land cover on the ground, it is just as, if not more, difficult to correctly classify forested land cover types on imagery. This is especially true when using medium spatial resolution imagery, such as Landsat 5TM imagery with 30 meter pixels, to create the land cover map. However, since Landsat is one of the most readily available (and free) sources of imagery, with quite valuable temporal resolution, it is important to maximize the usability of this data source. This research tested whether the temporal resolution of Landsat could be exploited to improve the accuracies of land cover maps created for a single year at a time, specifically targeted at differentiating between deciduous, coniferous, and mixed forest types. The multi-temporal image classification used available images from throughout the year of study to create a single land cover map for that year. An OBIA approach along with a classification and regression tree (CART) technique was used to create the land cover maps. The multiple images from throughout the year were employed to utilize the phenological changes in vegetation species to more accurately separate forest cover types. In general, the multi-date image analysis approach did perform better than the more traditional single-date approach. However, this difference was only significant ($p < 0.05$) for the years where a highly accurate single-date map could not be created due to image availability or cloud cover issues.

One of the objectives of this research was to map fragmentation within the Coastal Watershed of New Hampshire. However, a quick review of existing programs that compute fragmentation metrics revealed that a suitable program was not available for use

with vector data. Therefore, a new program, PolyFrag, was written to fill this gap. Currently, the most prevalent fragmentation program used in the literature is FRAGSTATS (McGarigal *et al.*, 2012). FRAGSTATS was updated in March of 2012 to include new features such as variable edge widths between different land cover types. An important advantage of FRAGSTATS is that it is a free program that is independent of any other programs (e.g., ArcGIS), so that the cost of use is extremely low. However, a disadvantage of the program is that it is only compatible with raster datasets. PolyFrag computes many of the same metrics as FRAGSTATS using vector data and runs within the ArcGIS (esri[®]) framework as an additional toolbox. While running within ArcGIS necessitates access to ArcGIS, it also makes the tool extremely user friendly to those who are familiar with ArcGIS tools. PolyFrag also includes the capability to define different edge widths, as in the latest version of FRAGSTATS. However, the biggest advantage of using PolyFrag over FRAGSTATS is that it outputs a spatial representation of fragmentation, rather than just a tabular representation of fragmentation metrics.

The applicability of PolyFrag for creating metrics useful in predicting woody invasive species presence was tested against four other freely available landscape fragmentation programs. PolyFrag's performance was equal to FRAGSTATS' and significantly better than the remaining programs ($p < 0.05$, when presence is defined as having a probability of 0.45 or higher) in creating a predictive map of possible invasive species presence when tested within the limited study area. Predictive maps were created using both FRAGSTATS and PolyFrag for the entire Coastal Watershed, but because the invasive species data were limited to a much smaller range, these predictive maps are unlikely to

have high accuracies. However, these maps can be used to help inform future sampling efforts that will result in better predictive models for the entire Coastal Watershed.

Overall, these maps, and especially the new program, PolyFrag, will be useful to researchers and land managers alike and will be made freely available to the NH Chapter of The Nature Conservancy (TNC), which provided the woody invasive species data, as well as other interested parties. PolyFrag will be made available to the appropriate communities and the software will be periodically updated, and useful additions requested and/or designed by users of the software will be incorporated.

The research conducted for this dissertation contributes to both the landscape ecology and remote sensing communities in a four distinct ways. First, a new sampling protocol is suggested for sampling reference units when using an OBIA approach to classification. Second, the usefulness of using a multi-date image classification approach is assessed. Third, a new fragmentation program has been created that easily allows for the analysis of vector data. Finally, this new program was used to create a probability of invasive species presence map for the Coastal Watershed that can serve as a starting point for future invasive species sampling.

LITERATURE CITED

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716-722.
- Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976. A land use and land cover classification system for use with remote sensor data. Geological Survey Professional Paper 964, 41 pp. Available online at: <http://landcover.usgs.gov/pdf/anderson.pdf> (accessed 26 June 2012).
- Andr n, H., 1994. Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review. *OIKOS*, 71:355-366.
- Baatz, M., U. Benz, S. Dehghani, M. Heymen, A. Holtje, P. Hofmann, I. Ligenfelder, M. Mimler, M. Sohlbach, M. Weber, and G. Willhauck, 2001. *eCognition User Guide*. Munich: Definiens Imaging GmbH. 310 pp.
- Barbet-Massin, M., F. Jiguet, C.H. Albert, and W. Thuiller, 2012. Selecting pseudo-absence for species distribution models: how, where, and how many? *Methods in Ecology and Evolution*, 3(2):327-338.
- Binkley, D., and T.C. Brown, 1993. Forest practices as nonpoint sources of pollution in North America. *Water Resources Bulletin*, 29(5):729-740.
- Bitterlich, W., 1947, Die Winkelzahlmessung (Measurement of basal area per hectare by means of angle measurement). *Allgemeine Forst- und Holzwirtschaftliche Zeitung*, 58: 94-96.
- Blake, J.G., and J.R. Karr, 1987. Breeding birds of isolated woodlots: area and habitat relationships. *Ecology*, 68(6):1724-1734.
- Blaschke, T. and J. Strobl, 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS*. Heidelberg: Huthig GmbH & Co. 6:12-17.
- Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65:2-16.
- Boulinier, T., J.D. Nichols, J.E. Hines, J.R. Sauer, C.H. Flather, and K.H. Pollock, 2001. Forest fragmentation and bird community dynamics: inference at regional scales. *Ecology*, 82(4):1159-1169.

- Breiman, L., J.H. Friedman, R.A. Olshen, and C. J. Stone (Eds.), 1984. *Classification and Regression Trees*. Wadsworth International, Belmont, CA, 368 pp.
- Brothers, T.S., and A. Spingarn, 1992. Forest fragmentation and alien plant invasion of central Indiana old-growth forests. *Conservation Biology*, 6(1):91-100.
- Brotans, L, W. Thuiller, M.B. Araújo, and A.H. Hirzel, 2004. Presence-absence verses presence-only methods for predicting bird habitat suitability. *Ecography*, 27:437-448.
- Brown, C.D., and C. Boutin, 2009. Linking past land use, recent disturbance, and dispersal mechanism to forest composition. *Biological Conservation*, 142:1647-1656.
- Burnham, K.P., and D.R. Anderson, 2004. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Second Edition*, Springer-Verlag New York, Inc., New York, NY, 488 pp.
- Busby, J.R., 1991. BIOCLIM – a bioclimate analysis and prediction system. In: Margiules, C.R., and M.P. Austin (eds.), *Nature conservation: cost effective biological surveys and data analysis*. CSIRO, pp. 64-68.
- Carpenter, G., A.N. Gillison, and J. Winter, 1993. DOMAIN: a flexible modeling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, 2:667-680.
- C-CAP (NOAA Coastal Change Analysis Program Regional Land Cover Database). Data collected 1995-present. Charleston, SC: National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center. Data accessed at www.csc.noaa.gov/landcover.
- Chavez, P.S., 1996. Image-based atmospheric corrections – revisited and improved. *Photogrammetric Engineering & Remote Sensing*, 62(9):1025-1036.
- Chen, X., and W. Pan, 2002. Relationships among phenological growing season, time-integrated normalized difference vegetation index and climate forcing in the temperate region of eastern China. *International Journal of Climatology*, 22:1781-1792.
- CLEAR (Center for Land Use Education and Research), 2009. Connecticut's Changing Landscape. University of Connecticut. College of Agriculture and Natural Resources. Available at the following web site: <http://clear.uconn.edu/>.
- Conese, C. and F. Maselli, 1991. Use of multitemporal information to improve classification performance of TM scenes in complex terrain. *ISPRS Journal of Photogrammetry and Remote Sensing*, 46:187-197.

- Congalton, R.G., R.G. Oderwald, and R.A. Mead, 1983. Assessing Landsat classification accuracy using discrete multivariate statistical techniques. *Photogrammetric Engineering & Remote Sensing*, 49(12):1671-1678.
- Congalton, R.G., 1988. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering & Remote Sensing*, 54(5):593-600.
- Congalton R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37:35-46.
- Congalton, R.G. and G.S. Biging, 1992. A pilot study evaluating ground reference data collection efforts for use in forest inventory. *Photogrammetric Engineering & Remote Sensing*, 58(12):1669-1671.
- Congalton, R.G., K. Green, and J. Tepy, 1993. Mapping old growth forests on National Forest and Park lands in the Pacific Northwest from remotely sensed data. *Photogrammetric Engineering & Remote Sensing*, 59(4):529-535.
- Congalton, R.G., 1997. Exploring and evaluating the consequences of vector-to-raster and raster-to-vector conversion. *Photogrammetric Engineering & Remote Sensing*, 60(4):435-434.
- Congalton, R.G., and K. Green, 2009. *Assessing the accuracy of remotely sensed data: principles and practices, Second Edition*. CRC Press, Boca Raton, FL, 208 pp.
- Cowardin, L.M., V. Carter, F.C. Golet, and E.T. LaRoe, 1979. Classification of wetlands and deepwater habitats of the United States, URL: <http://www.npwrc.usgs.gov/resource/wetlands/classwet/index.htm>, U.S. Department of the Interior, Fish and Wildlife Service, Washington, D.C. Jamestown, ND: Northern Prairie Wildlife Research Center Online (last date accessed 7 May 2012).
- Damschen, E.I., L.A. Brudvig, N.M. Haddad, D.J. Levey, J.L. Orrock, and J.J. Tewksbury, 2008. The movement ecology and dynamics of plant communities in fragmented landscapes. *PNAS*, 105(49):19078-19083.
- Desclée, B., P. Bogaert, and P. Defourny, 2006. Forest change detection by statistical object-based method. *Remote Sensing of Environment*, 102:1-11.
- DeZonia, B., and Mladenoff, D.J., 2004. IAN 1.0.23, Department of Forest Ecology & Management, University of Wisconsin, Madison, WI. Available at the following web site: <http://landscape.forest.wisc.edu/projects/ian/>.
- Diaz, R.J., Solan, M. and R.M. Valentine, 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. *Journal of Environmental Management*, 73:165-181.

- Drăguț, L., and T. Blaschke, 2006. Automated classification of landform elements using object-based image analysis. *Geomorphology*, 81:330-344.
- Du, Y., P.M. Teillet, and J. Cihlar, 2002. Radiometric normalization of multitemporal high-resolution satellite images with quality control for land cover change detection. *Remote Sensing of Environment*, 82:123-134.
- Ducey, M.J., 2001. Pre-cruise planning. pp. 26-35 in K. Bennett, technical coordinator. 2001. Forest Measurement for Natural Resource Professionals, Workshop Proceedings. University of New Hampshire Cooperative Extension Natural Resource Network Report, 71 pp. Available online at: http://extension.unh.edu/resources/files/Resource000398_Rep420.pdf (accessed 21 February 2012).
- Duveiller, G., P. Defourny, B. Desclée, and P. Mayaux, 2008. Deforestation in Central Africa: Estimates at regional, national and landscape levels by advanced processing of systematically-distributed Landsat extracts. *Remote Sensing of Environment*, 112:1969-1981.
- Dwyer, J.L., K.L. Saylor, and G.J. Zylstra, 1996. Landsat Pathfinder data sets for landscape change analysis. In: *Proceedings of the International Geoscience and Remote Sensing Symposium*, Lincoln, NE, May 27-31, vol. 1, pp. 457-550.
- Efron, B. and R. Tibshirani, 1993. *An introduction to the bootstrap*, CRC Press, Boca Raton, FL, 456 pp.
- Elith, J., C.H. Graham, R.P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J.R. Leathwick, A. Lehmann, J. Li, L.G. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J.McC. Overton, A.T. Peterson, S.J. Phillips, K.S. Richardson, R. Scachetti-Pereira, R.E. Schapire, J. Soberón, S. Williams, M.S. Wisz, and N.E. Zimmermann, 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29:129-151.
- Fahrig, L., 2003. Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 34:487-515.
- Fischer, J. and D.B. Lindenmayer, 2007. Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography*, 16:265-280.
- Fitzpatrick-Linz, K., 1981. Comparison on sampling procedures and data analysis for a land-use and land-cover map. *Photogrammetric Engineering & Remote Sensing*, 47(3):343-351.
- Flather, C.H., and J.R. Sauer, 1996. Using landscape ecology to test hypotheses about large-scale abundance patterns in migratory birds. *Ecology*, 77(1):28-35.

- Foody, G.M., 1996. Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, 17(7):1317-1340.
- Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80:185-201.
- Foster, D.R., 1992, Land-use history (1720-1990) and vegetation dynamics in central New England, USA. *Journal of Ecology*, 80(4):753-771.
- Fry, J.A., M.J. Coan, C.G. Homer, D.K. Meyer, and J.D. Wickham, 2009. Completion of the National Land Cover Database (NLCD) 1992–2001 Land Cover Change Retrofit product: U.S. Geological Survey *Open-File Report 2008–1379*, 18 p.
- Gibbs, J.P., 1998. Distribution of woodland amphibians along a forest fragmentation gradient. *Landscape Ecology*, 13:263-268.
- Glode, J.S., 2012. The Nature Conservancy, New Hampshire Chapter, Great Bay invasive species data. Unpublished raw data.
- Gopal, S., and C. Woodcock, 1994. Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering & Remote Sensing*, 60(2):181-188.
- Gustafson, E.J., 1998. Quantifying landscape spatial pattern: What is the state of the art? *Ecosystems*, 1:143-156.
- Haila, Y., 2002. A conceptual genealogy of fragmentation research: from island biogeography to landscape ecology. *Ecological Applications*, 12(2):321-334.
- Hall, F.G., D.E. Strebel, J.E. Nickeson, and S.J. Goetz, 1991. Radiometric rectification: toward a common radiometric response among multirate, multisensory images. *Remote Sensing of Environment*, 35:11-27.
- Held, M.E., and W.A. Wistendahl, 1978. The use of the Bitterlich sampling technique in an Athens County, Ohio forest. *Ohio Journal of Science*, 78(1):26-28.
- Henderson, S. T.P. Dawson, and R.J. Whittaker, 2006. Progress in invasive plants research. *Progress in Physical Geography*, 30(1):25-46.
- Henry, M.C., 2008. Comparison of single- and multi-date Landsat data for mapping wildfire scars in Ocala National Forest, Florida. *Photogrammetric Engineering & Remote Sensing*, 74(7):881-891.

- Hijmans, R.J., L. Guarino, M. Cruz, and E. Rojas, 2001. Computer tools for spatial analysis of plant genetic resources data: DIVA-GIS. *Plant Genetic Resources Newsletter*, 127:15-19.
- Hirzel, A.H., J. Hausser, D. Chessel, and N. Perrin, 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, 83(7):2027-2036.
- Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J.N., and Wickham, J. 2007. Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering & Remote Sensing*, 73(4):337-341.
- Husch, B., T.W. Beers, and J.A. Kershaw, Jr., 2003. *Forest Mensuration, Fourth Edition*. John Wiley & Sons, Inc., Hoboken, NJ, 443 pp.
- Jensen, J.R., 2005. *Introductory digital image processing: a remote sensing perspective, Third Edition*. Pearson Prentice Hall, Upper Saddle River, NJ. 526 pg.
- Johnson, V.S., J.A. Litvaitis, T.D. Lee, and S.D. Frey, 2006. The role of spatial and temporal scale in colonization and spread of invasive shrubs in early successional habitats. *Forest Ecology and Management*, 228(1-3):124-134.
- Justice, D., A. Deely, & F. Rubin, 2002. Final report: New Hampshire land cover assessment. Unpublished, 15pp.
- Kauth, R.J., and G.S. Thomas, 1976. The tasselled cap – A graphic description of the spectral-temporal development of agricultural crop as seen by LANDSAT. In *Proceedings of LARS Symposium on Machine Processing of Remotely Sensed Data*. 29 June – 1 July 1976, West Lafayette, IN.
- Kells, A.R. and D. Goulson, 2003. Preferred nesting sites of bumblebee queens (Hymenoptera: Apidae) in agroecosystems in the UK. *Biological Conservation*, 109(2):165-174.
- Kotchenova, S.Y., E.F. Vermote, R. Matarrese, and F.J. Klemm, Jr., 2006. Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: path radiance. *Applied Optics*, 45(26):6762-6774.
- Kupfer, J.A., 2012. Landscape ecology and biogeography: Rethinking landscape metrics in a post-FRAGSTATS landscape. *Progress in Physical Geography*, 36(3):400-420.
- Levey, D.J., B.M. Bolker, J.J. Tewksbury, S. Sargent, and N.M. Haddad, 2005. Response to Proches et al., 2005. *Science*, 310:782-783.

- Levins, R., 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, 15:237-240.
- Levins, R., 1970. Extinction. In: Some mathematical problems in biology, Gesternhaber, M. (ed.), *American Mathematical Society*, Providence, RI, pp. 77-107.
- Liu, Q.J., T. Takamura, N. Takeushi, and G. Shao, 2002. Mapping of boreal vegetation of a temperate mountain in China by multitemporal Landsat TM imagery. *International Journal of Remote Sensing*, 23(17):3385-3405.
- Lu, D., R. Mausel, E. Brondizio, and E. Moran, 2002. Assessment of atmospheric correction methods for Landsat TM data applicable to Amazon basin LBA research. *International Journal of Remote Sensing*, 23(13):2651-2671.
- Lu, D. and Q. Weng, 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823-870.
- Lunetta, R.S., R.G. Congalton, L.K. Fenstermaker, J.R. Jensen, K.C. McGwire, and L.R. Tinney, 1991. Remote sensing and geographic information system data integration: error sources and research ideas. *Photogrammetric Engineering & Remote Sensing*, 57(6):677-687.
- Lunetta, R.S., J.G. Lyon, J.A. Sturdevant, J.L. Dwyer, C.D. Elvidge, L.K. Fenstermaker, D. Yuan, S.R. Hoffer, and R. Werrackoon, 1993. *North American Landscape Characterization: research plan*. EPA/600/R-93/135, July, 419 pp.
- Lunetta, R.S., J.G. Lyon, B. Guidon, and C.D. Elvidge, 1998. North American Landscape Characterization dataset development and data fusion issues. *Photogrammetric Engineering & Remote Sensing*, 64(8):821-829.
- Lunetta, R.S., and M.E. Balogh, 1999. Application of multi-temporal Landsat 5 TM imagery for wetland identification. *Photogrammetric Engineering & Remote Sensing*, 65(11):1303-1310.
- MacLean, M.G. and R.G. Congalton, 2010. Mapping and analysis of fragmentation in Southeastern New Hampshire. Proceedings of the *ASPRS 2010 Fall Conference*, Orlando, FL (American Society for Photogrammetry and Remote Sensing, Bethesda, MD), unpaginated CD-ROM, 5 pp.
- MacLean, M.G., and R.G. Congalton, 2012a. Map accuracy issues when using an object-oriented approach. Proceedings of the *ASPRS 2012 Annual Conference*, 19-23 March, Sacramento, CA (American Society for Photogrammetry and Remote Sensing, Bethesda, MD), unpaginated CD-ROM.

- MacLean, M.G., and R.G. Congalton, 2012b. Applicability of multi-date land cover mapping using Landsat 5TM imagery in the Northeastern US. *Photogrammetric Engineering & Remote Sensing*, *In review*.
- MacLean, M.G., and R.G. Congalton, 2012c. PolyFrag: a vector-based program for computing landscape metrics. *In prep*.
- MacLean, M.G., M.J. Campbell, D.S. Maynard, M.J. Ducey, and R.G. Congalton, 2012. Requirements for labeling forest polygons in an object-based image analysis classification, *International Journal of Remote Sensing*. *In review*.
- Markham, B.L., and J.L. Barker, 1986. Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures. *EOSAT Landsat Technical Notes*, 1:3-8.
- McGarigal, K., and B.J. Marks. 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. USDA For. Serv. Gen. Tech. Rep. PNW-351. Available at the following web site: <http://www.umass.edu/landeco/research/fragstats/fragstats.html>.
- McGarigal, K., and S.A. Cushman, 2002. Comparative evaluation of experimental approaches to the study of habitat fragmentation effects. *Ecological Applications*, 12(2):335-345.
- McGarigal, K., S.A. Cushman, and E. Ene. 2012. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: <http://www.umass.edu/landeco/research/fragstats/fragstats.html>.
- Mitchell, W.A., H.G. Hughes, and L.E. Marcy, 1995. Prism sampling: Section 6.2.3, U.S. Army Corps of Engineers Wildlife Resources Management Manual. *Technical Report EL-95-24*, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.
- Moody, A., and D.M. Johnson, 2001. Land-surface phenologies from AVHRR using the discrete Fourier transform. *Remote Sensing of Environment*, 75:305-323.
- Moran, M.A., 1984. Influence of adjacent land use on understory vegetation of New York forests. *Urban Ecology*, 8:329-340.
- Moran, M.S., R.D. Jackson, P.N. Slater, and P.M. Teillet, 1992. Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. *Remote Sensing of Environment*, 41:169-184.
- NCDC (National Climatic Data Center), 2008. Cloudiness – mean number of days, URL: <http://lwf.ncdc.noaa.gov/oa/climate/online/ccd/cldy.html> (last date accessed 7 May 2012).

- Oetter, D.R., W.B. Cohen, M. Berterretche, T.K. Maersperger, and R.E. Kennedy, 2000. Land cover mapping in an agricultural setting using multiseasonal Thematic Mapper data. *Remote Sensing of Environment*, 76:139-155.
- Parent, J., D. Civco, and J. Hurd, 2007. Simulating future forest fragmentation in a Connecticut region undergoing suburbanization. Presented at: *ASPRS 2007 Annual Conference*, Tampa, FL. 11 pp.
- Paolini, L., F. Grings, J.A. Sobrino, J.C. Jimenez Munoz, and H. Karszenbaum, 2006. Radiometric correction effects in Landsat multi-date/multi-sensor change detection studies. *International Journal of Remote Sensing*, 27(4):685-704.
- Peterson, A.T., 2003. Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology*, 78(4):419-433.
- Pohl, C. and J.L. Van Genderen, 1998. Review article: Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing*, 19(5):823-854.
- PREP (Piscataqua Region Estuaries Partnership), 2010. *2010 Piscataqua Region Comprehensive Conservation and Management Plan*, 136 pp. Available online at: http://www.prep.unh.edu/resources/pdf/piscataqua_region_2010-prep-10.pdf (last accessed on 2 September 2012).
- Proches, S., J.R.U. Wilson, R. Veldtman, J.M. Kalwij, D.M. Richardson, and S.L. Chown, 2005. Landscape corridors: possible dangers? *Science*, 310:781-782.
- Prugh, L.R., K.E. Hodges, A.R.E. Sinclair, and J.S. Brashares, 2008. Effect of habitat area and isolation on fragmented animal populations. *PNAS*, 105(52):20770-20775.
- Pulliam, H.R., 1988. Sources, sinks, and population regulation. *American Naturalist*, 132:652-661.
- Radoux, J., R. Bogaert, D. Fasbender, and P. Defourny, 2011. Thematic accuracy assessment of geographic object-based image classification. *International Journal of Geographical Information Science*, 25(6):895-911.
- Rempel, R.S., D. Kaukinen., and A.P. Carr, 2012. Patch Analyst and Patch Grid. Ontario Ministry of Natural Resources. Centre for Northern Forest Ecosystem Research, Thunder Bay, Ontario.
- Riitters, K.H., J.D. Wickham, R.V. O'Neill, K.B. Jones, E.R. Smith, J.W. Coulston, T.G. Wade, and J.H. Smith, 2002. Fragmentation of Continental United States Forests. *Ecosystems*, 5:815-822.

- Robertson, L.D., and D.J. King, 2011. Comparison of pixel- and object-based classification in land cover change mapping. *International Journal of Remote Sensing*, 32(6):1505-1529.
- Rosenblatt, D.L., E.J. Heske, S.L. Nelson, D.M. Barber, M.A. Miller, and B. MacAllister, 1999. Forest fragments in East-central Illinois: islands or habitat patches for mammals? *American Midland Naturalist*, 141:115-123.
- Rouse, J.W., R.H. Haas, J.A. Schell, and D.W. Deering, 1974. Monitoring vegetation systems in the Great Plains with ERTS. In *Proceedings of the Third ERTS-1 Symposium*. 10-14 December 1973, NASA SP-351, Washington D.C. NASA, pp. 309-317.
- Saura, S. & J. Torné, 2009. Conefor Sensinode 2.2: a software package for quantifying the importance of habitat patches for landscape connectivity. *Environmental Modelling & Software*, 24:135-139.
- Schroeder, T.A., W.B. Cohen, C. Song, M.J. Canty, and Z. Yang, 2006. Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon. *Remote Sensing of Environment*, 103:16-26.
- Schriever, J.R., and R.G. Congalton, 1995. Evaluating seasonal variability as an aid to cover-type mapping from Landsat Thematic Mapper Data in the Northeast. *Photogrammetric Engineering & Remote Sensing*, 61(3):321-327.
- Schumaker, N.H., 1998. The PATCH Model. U.S. Environmental Protection Agency. Western Ecology Division, Corvallis, OR.
- Song, C., C.E. Woodcock, K.C. Seto, M.P. Lenney, and S.A. Macomber, 2001. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? *Remote Sensing of Environment*, 75:230-244.
- Squires, E.R., and W.A. Wistendahl, 1976. Sample unit selection for studies of herbaceous oldfield vegetation. *Ohio Journal of Science*, 76(4):185-188.
- Stehman, S.V., 1995. Thematic map accuracy assessment from the perspective of finite population sampling. *International Journal of Remote Sensing*, 16(3):589-593.
- Stehman, S.V., and R.L. Czaplewski, 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment*, 64:331-344.
- Svensson, B., J. Lagerlöf, and B.G. Svensson, 2000. Habitat preferences of nest-seeking bumble bees (*Hymenoptera: Apidae*) in an agricultural landscape. *Agriculture, Ecosystems & Environment*, 77(3):247-255.

- TNC (The Nature Conservancy), 2010. *New Hampshire's Coastal Watershed*, URL: <http://www.nature.org/ourinitiatives/regions/northamerica/unitedstates/newhampshire/howwework/land-conservation-plan-for-new-hampshires-coastal-watershed.xml> (last date accessed 2 September 2012).
- Tottrup, C., 2004. Improving tropical forest mapping using multi-date Landsat TM data and pre-classification image smoothing. *International Journal of Remote Sensing*, 25(4):717-730.
- Thompson, S.K., 2002. *Sampling, Second Edition*, John Wiley & Sons, Inc., Hoboken, NJ, 400 pp.
- Turner, M.G., 2005. Landscape ecology: what is the state of the science? *Annual Review of Ecology, Evolution, and Systematics*, 36:319-344.
- USCB (United States Census Bureau), 2011. *County Population Estimates*, URL: <http://www.census.gov/popest/counties/counties.html> (last date accessed 7 May 2012).
- USGS (United States Geological Survey), n.d. *Landsat Missions*, URL: <http://landsat.usgs.gov/> (last date accessed 7 May 2012).
- VanDerWal, J., L.P. Shoo, C. Graham, and S.E. Williams, 2009. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modeling*, 220(4):589-594.
- Vitousek, P.M., 1994. Beyond global warming: Ecology and global change. *Ecology*, 75, 1861-1876.
- Vogt, P., K.H. Riitters, C. Estreguil, J. Kozak, T.G. Wade, and J.D. Wickham, 2007. Mapping spatial patterns with morphological image processing. *Landscape Ecology*, 22:171-177.
- Warner, T.A., J.Y. Lee, and J.B. McGraw, 1998. Delineation and identification of individual trees in the Eastern Deciduous Forest. Presented in: *the International Forum on Automated Interpretation of High Spatial Resolution Imagery for Forestry*. February 10-12, 1998, Victoria, BC.
- Wiant, H.V. Jr., D.O. Yandle, and R. Andreas, 1984. Is BAF 10 a good choice for point sampling? *Northern Journal of Applied Forestry*, 2(1):23-25.
- Wiens, J.A., 1989. Spatial scaling in ecology. *Functional Ecology*, 3(4):385-397.
- Wiens, J.A., 2008. Allerton Park 1983: the beginnings of a paradigm for landscape ecology? *Landscape Ecology*, 23:125-128.

With, K.A., 2002. The landscape ecology of invasive spread. *Conservation Biology*, 16(5):1192-1203.

Wolter, P.T., D.J. Mladenoff, G.E. Host, and T.R. Crow, 1995. Improved forest classification in the northern lake states using multi-temporal Landsat imagery. *Photogrammetric Engineering & Remote Sensing*, 61(9), 1129-1143.

Xiuwan, C., 2002. Using remote sensing and GIS to analyse land cover change and its impacts on regional sustainable development. *International Journal of Remote Sensing*, 23(1):107-124.

Zaniewski, A.E., A. Lehmann, and J.McC. Overton, 2002. Predicting spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, 157:261-280.

APPENDICES

APPENDIX A

NEW HAMPSHIRE LAND COVER CLASSIFICATION SCHEME (Justice *et al.*, 2002)

Level 1	Level 2	Level 3
1	Developed land	
	10 Residential/commercial /industrial development	100 Residential/commercial /industrial development
	14 Transportation	140 Transportation
2	Active agricultural land	
	21 Cropland and pasture	
		211 Row crops
		212 Hay/pasture
	22 Orchards, fruit, and ornamental horticulture	221 Orchards
4	Forest	
	Areas dominated by trees, the majority of which are greater than 10' tall	
	41 Deciduous forest	Forest stands comprising less than 25% coniferous basal area per acre
		412 Beech/oak Deciduous stands comprising at least 30% beech and oak basal area per acre
		419 Other hardwoods All deciduous stands not meeting the beech/oak definition
	42 Coniferous forest	Forest stands comprising greater than 65% coniferous basal area per acre

			421	White/red pine	Conifer stands in which white/red pine constitutes a plurality of the coniferous basal area	
			422	Spruce/fir	Conifer stands in which spruce/fir constitutes a plurality of the coniferous basal area	
			423	Hemlock	Conifer stands in which hemlock constitutes a plurality of the coniferous basal area	
			424	Pitch pine	Coniferous stands in which pitch pine constitutes a plurality of the coniferous basal area	
	43	Mixed forest	430	Mixed forest	Forest stands comprising more than 25% and less than 65% coniferous basal area per acre	
5	Water	50	Open water	500	Open water	Lakes, ponds, some rivers, or any other open water
6	Wetlands	Areas dominated by wetland characteristics defined by the U.S. Fish and Wildlife Service National Wetlands Inventory. Basically hydric soils, hydrophytic vegetation and the hydrologic conditions that result in water at or near the surface for extended periods of the growing season.				
		61	Forested wetlands	610	Forested wetlands	Non-tidal wetlands characterized by woody vegetation 6m tall or higher
		62	Non-forested wetlands	620	Non-forested wetlands	All other non-tidal wetlands, including those dominated by shrubs, emergent, mosses, or lichens
		63	Tidal wetlands	630	Tidal wetlands	
7	Cleared/other open					
		71	Disturbed	710	Disturbed	Gravel pits, quarries, or other areas where the earth and vegetation have been altered or exposed
		73	Sand dunes	730	Sand dunes	Areas along the seacoast that are dominated by sand
		79	Other cleared	790	Other cleared	Clear cut forest, old agriculture fields that are reverting to forest, etc.

APPENDIX B

BOOTSTRAP CODE USED IN CHAPTER III – WRITTEN IN R

```
library(boot)

# Import Data
points <- read.delim(file.choose(),header=T)
points$Plot <- NULL
points

# Bootstrap

boot.se <- function(x,estimator,num.rep=400)
{
  x <- as.matrix(x)
  n <- nrow(x)
  y <- ncol(x)
  for(c in 2:n)
  {
    estimator.boot <- matrix(nrow=num.rep,ncol=y)
    average.boot <- matrix(nrow=num.rep,ncol=y)
    for(b in 1:num.rep)
    {
      inds.boot <- sample(1:n,c,replace=T)
      x.boot <- x[inds.boot,]
      #print(x.boot)
      for(a in 1:y)
      {
        estimator.boot[b,a] <- estimator(x.boot[,a])
      }
      #print (estimator.boot)
      sum.a <- rowSums(estimator.boot, na.rm = T, dims = 1)
      average.boot <- estimator.boot/sum.a*100
    }
  }
}
```



```
    }  
    #print(average.boot)  
    final.boot <- matrix(nrow=3,ncol=y)  
    for(a in 1:y)  
    {  
      final.boot[1,a] <- mean(average.boot[,a])  
      final.boot[2,a] <- sqrt(var(average.boot[,a]))  
      final.boot[3,a] <- final.boot[2,a]/final.boot[1,a]*100  
    }  
    write.table(final.boot, "C:/...txt",append=T)  
  }  
  return(final.boot)  
}  
  
boot.se(points,sum)
```

APPENDIX C

EXAMPLE MANUAL CHECK OF THE COMPUTATIONS COMPLETED WITHIN POLYFRAG

Table 22. All computations were done by hand using the C-CAP land cover map. These example patch metrics were computed either on the first patch in the attribute table, or on the first patch with less than 100% CAI, but were also checked on several other patches distributed throughout the dataset. Class metrics were computed using the Beech/Oak category. All results matched those completed by PolyFrag and all units are metric.

Fragmentation Metric	Equation	Computations	Result
PARA	$\frac{\text{PERIM}}{\text{AREA}}$	$\frac{132.857}{537.4926}$	0.2472
SHAPE	$\frac{\text{PERIM}}{2\pi\sqrt{\text{AREA}/\pi}}$	$\frac{132.857}{2\pi\sqrt{537.4926/\pi}}$	1.16166
FRAC	$\frac{2 \ln(\text{PERIM}/4)}{\ln \text{AREA}}$	$\frac{2 \ln(132.857/4)}{\ln 537.4926}$	1.11437
CIRCLE	$1 - \frac{\text{AREA}}{\pi \left(\frac{\text{P_LENGTH}}{2}\right)^2}$ P_LENGTH = diameter of the smallest circumscribing circle	$1 - \frac{537.4926}{\pi \left(\frac{57.806}{2}\right)^2}$	0.7952
CAI	$\frac{\text{CORE AREA}}{\text{AREA}} * 100$ CORE AREA = area labeled core within that polygon	$\frac{10036.3287}{14619.3273} * 100$ *computed for the first polygon with less than 100% CAI	68.6

PROX	$\sum_{i=1}^n \frac{\text{AREA}_i}{x_i^2}$	$\frac{8867.0959}{104.4474^2} + \frac{4540.3788}{370.6104^2}$ $+ \frac{5685.294}{915.1876^2}$ $+ \frac{8121.8485}{8121.8485^2}$ $+ \frac{973.6198^2}{973.6198^2}$	0.861216
PLAND	$\frac{CA}{TA} * 100$	$\frac{591.829}{15299.2} * 100$	3.87
PD	$\frac{NP}{TA} * 100$	$\frac{678}{15299.2} * 100$	4.43
LSI	$\frac{P}{2\pi\sqrt{A/\pi}}$ P = sum of all of the perimeters for the polygons A = sum of all of the areas for the polygons	$\frac{295105.9488}{2\pi\sqrt{15299.2/\pi}}$	34.22
LPI	$\frac{LPA}{TA} * 100$ LPA = the area of the largest patch for the class(es) in question	$\frac{38.0103}{15299.2} * 100$	0.25
ED	$\frac{TE}{TA} * 100$	$\frac{3.742}{15299.2} * 100$	0.02
CPLAND	$\frac{TCA}{TA} * 100$	$\frac{431.1}{15299.2} * 100$	2.818
PRD	$\frac{PR}{TA} * 100$	$\frac{19}{15299.2} * 100$	0.12
MESH	$\frac{\sum_{i=1}^n \text{AREA}_i^2}{TA}$	$\frac{29338200}{15299.2}$	1917.63
COHESION	$\frac{1 - \frac{\sum P}{\sum(P\sqrt{A})}}{1 - 1/\sqrt{TA}} * 100$	$\frac{1 - 295105.949/38837936}{1 - 1/152992000} * 100$	99.25
CONNECT	$\frac{\sum \text{PROX_NUM}}{n(n-1)/2} * 100$ PROX_NUM = the number of polygons with the same label that fall within the max search distance	$\frac{10846}{678(678-1)/2} * 100$	4.73
SHDI	$-\sum_{i=1}^n Pr_i * \ln Pr_i$ $Pr_i = \frac{CA_i}{TA}$	See Table 23	2.2827
SIDI	$1 - \sum_{i=1}^n Pr_i^2$	See Table 23	0.8531

MSIDI	$-\ln \sum_{i=1}^n Pr_i^2$	$-\ln 0.1469$	1.9180
SHEI	$\frac{-\sum_{i=1}^n Pr_i * \ln Pr_i}{\ln PR}$	$\frac{2.2827}{\ln 19}$	0.7753
SIEI	$\frac{1 - \sum_{i=1}^n Pr_i^2}{1 - 1/PR}$	$\frac{0.8531}{1 - 1/19}$	0.9005
MSIEI	$\frac{-\ln \sum_{i=1}^n Pr_i^2}{\ln PR}$	$\frac{1.9180}{\ln 19}$	0.6514

Table 23. Extra computations needed to produce diversity indices.

Polygon	Pr - Calculation	Pr - Result	Pr*ln(Pr)	Pr ²
1	591.829/15299.2	0.0387	-0.1258	0.0015
2	188.1085/15299.2	0.0123	-0.0541	0.0002
3	748.7930/15299.2	0.0489	-0.1477	0.0024
4	1170.4299/15299.2	0.0765	-0.1966	0.0059
5	87.0275/15299.2	0.0057	-0.0294	0.0000
6	4712.0230/15299.2	0.3080	-0.3627	0.0949
7	518.7635/15299.2	0.0339	-0.1147	0.0011
8	704.1808/15299.2	0.0460	-0.1417	0.0021
9	94.8578/15299.2	0.0062	-0.0315	0.0000
10	2323.8501/15299.2	0.1519	-0.2863	0.0231
11	884.8214/15299.2	0.0578	-0.1648	0.0033
12	0.6497/15299.2	0.0000	-0.0004	0.0000
13	732.1624/15299.2	0.0479	-0.1455	0.0023
14	41.0602/15299.2	0.0027	-0.0159	0.0000
15	4.2233/15299.2	0.0003	-0.0023	0.0000
16	0.6497/15299.2	0.0000	-0.0004	0.0000
17	435.9089/15299.2	0.0285	-0.1014	0.0008
18	1201.8425/15299.2	0.0786	-0.1998	0.0062
19	858.1174/15299.2	0.0561	-0.1616	0.0031
		Sum	-2.2827	0.1469