DISSERTATION

MULTI-SCALE SAMPLING OF NATIVE AND NON-NATIVE PLANT DIVERSITY: EXAMPLES OF DATA ANALYSES AND APPLICATIONS

Submitted by

Geneva W. Chong

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

for the degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2002

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

July 10, 2002

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY GENEVA W. CHONG ENTITLED "MULTI-SCALE SAMPLING OF NATIVE AND NON-NATVE PLANT DIVERSITY: EXAMPLES OF DATA ANALYSES AND APPLICATIONS" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Thomas J Stohlgren

Michael Coughenour

Dan Binkley

Dan Binkley

Department Head/Director

ABSTRACT OF DISSERTATION

MULTI-SCALE SAMPLING OF NATIVE AND NON-NATIVE PLANT DIVERSITY: EXAMPLES OF DATA ANALYSES AND APPLICATIONS

Multi-scale vegetation data were collected in Rocky Mountain National Park, Colorado, USA. Analyses were conducted on subsets of data depending on specific study objectives, the area of the Park in question, and the appropriate statistical approach.

Species-area curves and measures of heterogeneity combined with information on the area covered by each vegetation type showed that the types making the greatest contributions to biodiversity at the landscape-scale covered the smallest areas. This approach may provide an accurate and relatively rapid way to rank landscape-scale hotspots of plant diversity within regions of interest.

Butterfly diversity was sampled in conjunction with vegetation in the Beaver Meadows area of the Park. Over 180 vascular plant species were unique to the aspen vegetation type (N = 32 plots). The slope of the mean species-area curve for the aspen vegetation type was the steepest of the 10 types sampled, and aspen plots contained the greatest number of native butterfly species. However, aspen plots were the most heavily invaded by non-native plant species, which could have negative effects on native plant and butterfly species diversity.

I used stepwise multiple regression and modified residual kriging to estimate the numbers of native and exotic species, and the probability of observing a non-native species in 30 m x 30 m cells over a 54,000 ha area. These models can efficiently direct resource managers to areas in need of further inventory, monitoring, and non-native species control efforts.

I compared the abilities of trend surface analysis alone and in combination with regression tree analysis to develop full-coverage surfaces of the cover and richness of total, native, and non-native plant species in an 80,000 ha portion of the Park. The combination of approaches consistently outperformed the trend surface analysis alone with an increase in average R^2 from 0.17 to 0.54.

Future applications of this research approach could be useful for developing the concept of "scalable models" where the area modeled is adjusted to produce the most accurate models given sparse data in complex terrain. These models could then be used to estimate variables of interest in the unsampled area to direct future work.

Geneva W. Chong
Graduate Degree Program in Ecology
Colorado State University
Fort Collins, CO 80523
Summer 2002

ACKNOWLEDGEMENTS

The U.S. Geological Survey supported this research. Colorado State University provided me with two years of academic support through graduate fellowships. This work could not have been completed without excellent field crews and the logistical support of staff at Rocky Mountain National Park, the Natural Resource Ecology Laboratory, and the USGS Midcontinent Ecological Science Center.

Vegetation data were collected by (in alphabetical order): Gina Adams, Marcell Astle, Rich Bachand, Helen Fields, Jeanette Haddock, Michelle Hart, Cindy Hindes, Mohammed Kalkhan, Ingolf Kuhn, Michelle Lee, Paige Lynch, Roger Miller, Greg Newman, Annie Overlin, April Owen, Elizabeth Revenaugh (formerly Smith), Lisa Schell, Sara Simonson, Cindy Villa, and other occasional visitors in the field. Cindy Hindes and Jeanette Haddock provided invaluable plant identification skills. Sara Simonson collected and analyzed butterfly data and provided intellectual and moral support.

This type of work would be impossible without people who love to work hard all day in the field then look at plants through a microscope with dichotomous key in hand at night (if they are not making macaroni and cheese in the rain as they are being attacked by mosquitoes outside their tents).

David Barnett, Suzanne Joy, Mohammed Kalkhan, Michelle Lee, Robin Reich, Kelly Rimar, Sara Simonson, and anonymous reviewers provided helpful comments and guidance for earlier drafts of these chapters. Suzanne, Mohammed, and Robin were invaluable for the spatial analyses.

I owe many thanks to my committee, Thomas J. Stohlgren, advisor, Dan Binkley, Michael Coughenour, John Wiens, and James Detling for agreeing to work with me over these six years. I am not sure if they expected me to take so long, so I thank each one for not forgetting about me during my long spells of seeming to forget that I was a student.

There are many others who helped me attain my academic goals. Sally Dunphy kept the Graduate Degree Program in Ecology running smoothly and always had good answers to my questions. Alan Covich and Keith Paustian provided their support to a few semesters of a tropical ecology seminar, and Bob Aukerman with the American Universities International Program provided me with the opportunity to teach some tropical forest ecology in Belize. Many individuals within the USGS deserve acknowledgment including Rey Stendell, former director of the U.S. Geological Survey Midcontinent Ecological Science Center, and Doug Posson, current director.

Finally, I must thank my family and friends for always being here for me. Included in this grouping are my coworkers in the Stohlgren group- they are friends and family, my fellow NREL-ians (past, present, grad students, everyone) and all the students I have had the pleasure and privilege of interacting with.

I extend my heartfelt appreciation and thanks to all.

TABLE OF CONTENTS

ABSTRACT OF DISSERTATION iii

ACKNOWLEDGMENTS v

TABLE OF CONTENTS 1

INTRODUCTION 4

LITERATURE CITED 8

CHAPTER I: SPECIES-AREA CURVES INDICATE THE IMPORTANCE OF HABITATS' CONTRIBUTIONS TO LANDSCAPE-SCALE BIODIVERSITY 12

1.0 Abstract 12

1.1 Introduction 13

1.2 Methods 16

1.2.1 Study Sites 16

1.2.2 Plot Designs 16

1.2.3 Analyses 21

1.3 Results 23

1.4 Discussion 31

1.5 Literature Cited 35

CHAPTER II: AN EXAMPLE APPLICATION OF SPECIES-AREA CURVES FOR MANAGERS: BIODIVERSITY- ASPEN HAS THE LEAD, BUT WILL NON-NATIVE SPECIES TAKE OVER? 38

2.0 Abstract 38

2.1 Introduction 39

2.2 Methods 41

2.2.1 Study Site 41

2.2.2 Vegetation and Butterfly Sampling 42

2.2.3 Analyses 42

2.3 Results 43

2.4 Discussion 49

2.5 Literature Cited 51

CHAPTER III: NEW APPROACHES FOR SAMPLING AND MODELING NATIVE AND NON-NATIVE PLANT SPECIES RICHNESS 58

3.0 Abstract 58

3.1 Introduction 59

3.1.1 A Multi-Species, Multi-Scale, Multi-Phase Approach 60

3.1.2 Data Comparability, Analysis, and Synthesis 62

3.2 Methods 64

3.2.1 Field Data 64

3.2.2 GIS Data 64

3.2.3 Geostatistical Analysis 65

3.2.4 Spatial Integration 68

3.4 Results 68

3.4.1 Regression Models 68

3.4.2 Kriging 70

3.4 Discussion 74

3.5 Literature Cited 76

CHAPTER IV: SPATIAL MODELS COMBINING TREND SURFACE AND REGRESSION TREE ANALYSES TO PREDICT NUMBERS AND COVER OF NATIVE AND NON-NATIVE PLANT SPECIES 84

4.0 Abstract 84

4.1 Introduction 85

4.2 Methods 89

4.2.1 Study Area 89

4.2.2 Vegetation Sampling 89

4.2.3 Statistical Analysis 94

4.3 Results 96

4.4 Discussion 100

4.5 Literature Cited 103

CONCLUSION 108

EPILOGUE 111

LITERATURE CITED 116

INTRODUCTION

Although we have entered a new millennium, we have not left behind old issues and questions. E.O. Wilson's *The Future of Life* (2002) outlines his recommendations for saving the biodiversity of the Earth. One of these recommendations is to complete the mapping of the world's biodiversity. How do we do this when we do not even know what plant species occur on some of the most protected lands in the world- the U.S. National Parks (Stohlgren et al. 1995a)? Even with unlimited funding it is highly unlikely that we could complete such mapping before most of what we are interested in is severely compromised or made extinct, at least locally. Proposals such as Myers et al. (2000) that suggest areas for conservation based on their contributions to global biodiversity provide useful first-cuts for management at national scales. However, people already live within those areas, and it is not likely the global society will pay to conserve remaining unsettled areas. Thus, remaining subsets of unsettled and lightly used areas must be rapidly ranked by their suitability for biodiversity conservation with the hope that this will facilitate their protection at local, national, and global scales. Whatever we do, it must be done quickly for the greatest positive effect.

The invasion of native species assemblages by non-native species is an additional threat to biodiversity (Mack et al. 2000). Field research (Stohlgren et al.

1999a, 2002) and literature review (Lonsdale 1999) have indicated that areas rich in native plant species are more heavily invaded than areas with fewer native plant species. Determining areas at risk for non-native species invasion is becoming a National priority as one estimate of current economic losses is $33 billion/yr caused by non-native plants and $137 billion/yr caused by all non-native species (Pimentel et al. 2000). Predicting areas that are susceptible to invasion may allow resource managers to find non-native species before they successfully invade an area, so that the control of the non-native species is more tractable. Development of predictive spatial models should be a central component to this strategy.

In this dissertation I introduce several ways to analyze species richness data that were collected with the knowledge that only a small portion of any given area is likely to be sampled, yet we want to make conclusions over the larger, unsampled area.

The work presented here is focused on native and non-native plant species richness in Rocky Mountain National Park, Colorado, USA, but these approaches are finding ever-increasing use both over larger areas and across different taxonomic groups (e.g., Ferreira and Stohlgren 1999, Stohlgren et al. 1999, Simonson et al. 2001). The underlying concepts are straightforward, and their application can be designed to fit various needs- valuable aspects given the multitude of hypotheses to explain species richness (Palmer 1994).

The general research approach applied to vegetation can be described by: (1) stratify the area in question paying particular attention to the less common strata (e.g., aspen stands and riparian zones in Rocky Mountain National Park)- this requires the recognition of the effects of scale on the ability to identify strata (Stohlgren et al. 1997a); (2) sample within each stratum using a spatially explicit, multi-scale approach (e.g., the Modified-Whittaker plot, Stohlgren et al. 1995, 1998); (3) analyze the multi-scale data to develop an index of species richness in larger areas than those sampled to identify areas of interest (within smaller areas it may be logical to develop actual estimates of species richness, Stohlgren et al. 1997); and (4) develop spatial models of variables of interest (e.g., plant species richness and cover, and invasive species locations) to identify priority areas for future sampling, invasive species control, or some other management objective.

Chapter I provides a general overview of the use of multi-scale data to develop species-area curves. I then extend the use of species- and spatially-explicit, multi-scale data to develop a composite index for identifying hotspots of native plant diversity based on the slopes of species-area curves, heterogeneity of species distributions, and extent of suitable habitat. Chapter II provides a more focused example of the use of multi-scale data in relation to identifying management concerns for aspen (*Populus tremuloides* Michx.). In this example, I show that aspen stands in Rocky Mountain National Park host the largest variety of native plant and butterfly species (Simonson et al. 2001), but they also contain the greatest number of non-native plant species. This combination of information identifies several potential areas of concern for aspen management.

In Chapter III, I present a spatial modeling approach for estimating variables of interest over an entire surface (e.g., a 54,000-ha portion of Rocky Mountain National

Park). This is possible because each sample point is geo-referenced with the Global Positioning System (GPS) and thus can be correlated with data such as Landsat TM satellite imagery. In this case, the vegetation data are spatially autocorrelated, which allows the use of kriging (spatial interpolation) to interpolate values to unsampled areas. Much more variance is explained by kriging than by ordinary regression.

The final chapter presents an alternative approach to spatial modeling that is not dependent on spatial autocorrelation in the data - regression tree analysis combined with ordinary least squares regression. This is a particularly useful approach because it allows the use of sparse data from spatially independent plots (e.g., 175 plots in an 80,000-ha portion of the Park; Reich et al. 2002), and it is not as computationally demanding as kriging. As in Chapter III, I examine the ability to estimate native and non-native species richness and cover over an entire surface using simple independent variables from satellite imagery and topography (digital elevation models). The inclusion of other independent variables relies on having complete coverages of those variables in a geographic information system (GIS), and, as those coverages become available (e.g., for soil characteristics), I would expect the models to greatly improve.

The next step in this research program is the development of nested, "scalable" models. In other words, as one expands the area of interest, one would expect the types of data available (or needed) to change, and thus the appropriate modeling tools should change as well. As my research site in Rocky Mountain National Park expanded from the 750-ha Beaver Meadows area to an 80,000-ha area, the concentration of sample points thinned considerably and the heterogeneity between points within some vegetation types increased (e.g., in aspen). As the data shifted from being spatially autocorrelated to spatially independent, different interpolation analyses became appropriate. To advance the use of the sampling and analysis approaches presented here, ecologists must make better use of existing and new data, so that their full value can be used to benefit natural resources management. We must also increase our scientific understanding of underlying processes such as those influencing the persistence of biodiversity hotspots and the continued invasion of non-native species.

LITERATURE CITED

Ferreira, L.V., and T.J. Stohlgren. 1999. Effects of river level fluctuation in species richness, diversity, and plant distribution in a floodplain in Central Amazonia. *Oecologia* 120: 582 - 587.

Lonsdale, W.M. 1999. Global patterns of plant invasions and the concept of invisibility. *Ecology* 80:1522-1536.

Mack, R.N., D. Simberloff, W.M. Lonsdale, H. Evans, M. Clout, and F. Bazzaz. 2000. Biotic invasions: Causes, epidemiology, global consequences, and control. *Ecological Society of America, Issues in Ecology* 5, 20 pp.

Myers N., R.A. Mittermeier, C.G. Mittermeier, G.A.B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853-858

Palmer, M.W. 1994. Variation in species richness: towards a unification of hyptheses. *Folia Geobot. Phytotax., Praha* 29:511-530.

Pimentel, D., L. Lach, and R. Zuniga. 2000. Environmental and economic costs of nonindigenous species in the United States. *BioScience* 50:53-65.

Reich, R.M., J.E. Lundquist, and V.A. Bravo. 2002. Modeling forest fuel loadings on the Black Hills National Forest Using Landsat TM data. Proceedings...

Simonson, S.E., Opler, P.A., Stohlgren, T.J. and G.W. Chong. 2001. Rapid assessment of butterfly diversity in a montane landscape. *Biodiversity and Conservation* 10 (8): 1369-1386.

Stohlgren, T.J., J. Belnap, G.W. Chong, and R. Reich. 1999. A plan to assess native and exotic plant diversity and cryptobiotic crusts in the Grand Staircase-Escalante National Monument. Pages 269-276. In: Learning From The Land Conference Proceedings, Salt Lake City, Utah.

Stohlgren, T.J., D. Binkley, G.W. Chong, M.A. Kalkhan, L.D. Schell, K.A. Bull, Y. Otsuki, G. Newman, M. Bashkin, and Y. Son. 1999. Exotic plant species invade hot spots of native plant diversity. *Ecological Monographs* 69(1): 25-46.

Stohlgren, T. J., K.A. Bull, and Y. Otsuki. 1998. Comparison of rangeland vegetation sampling techniques in the Central Grasslands. *Journal of Range Management* 51: 164-172.

Stohlgren, T. J., G. W. Chong, M. A. Kalkhan, and L. D. Schell. 1997. Rapid assessment of plant diversity patterns: a methodology for landscapes. *Ecological Monitoring and Assessment* 48: 25-43.

Stohlgren, T. J., G. W. Chong, M. A. Kalkhan, and L. D. Schell. 1997a. Multiscale sampling of Plant Diversity: Effects of the minimum mapping unit. *Ecological Applications* 7(3): 1064-1074.

Stohlgren, T.J., G.W. Chong, L.D. Schell, K.A. Rimar, Y. Otsuki, M. Lee, M.A. Kalkhan, and C.A. Villa. 2002. Assessing vulnerability to invasion by non-native plant species at multiple spatial scales. *Environmental Management* 29:566-577.

Stohlgren, T. J., M. B. Falkner, and L. D. Schell. 1995. A Modified-Whittaker nested vegetation sampling method. *Vegetatio* 117: 113-121.

Stohlgren, T. J., J. F. Quinn, M. Ruggiero, and G. Waggoner. 1995a. Status of biotic inventories in U.S. National Parks. *Biological Conservation* 71: 97-106.

Wilson, E.O. 2002. *The Future of Life*. Alfred A. Knopf. New York. 229 pp.

CHAPTER I: SPECIES-AREA CURVES INDICATE THE IMPORTANCE OF HABITATS' CONTRIBUTIONS TO LANDSCAPE-SCALE BIODIVERSITY

1.0 Abstract

I examined species-area curves, species composition and similarity (Jaccard's Coefficients), and species richness in 17 vegetation types to develop a composite index of a vegetation type's contribution to landscape-scale species richness. The data are from 1-m^2, 10-m^2, and 100-m^2 subplots nested in 147 1000-m^2 plots sampled in Rocky Mountain National Park, Colorado, USA. I used these data to compare the ability of three species-area curve models to estimate the number of species observed in each vegetation type. The log(species)-log(area) curve had the largest adjusted coefficients of determination (r^2 values) in 12 of the 17 types, followed by the species-log(area) curve with five of the highest values. When the slopes of the curves were corrected for species overlap among plots with Jaccard's Coefficients, the species-log(area) curves estimated values closest to those observed. Combined information from species-area curves and measures of heterogeneity with information on the area covered by each vegetation type shows that the types making the greatest contributions to biodiversity at the landscape-scale covered the smallest areas. This approach may provide an accurate

and relatively rapid way to rank landscape-scale hotspots of plant diversity within regions of interest.

1.1 Introduction

The use of species-area curves to examine differences in species number has a long and contentious history (Connor and McCoy 1979, Rosenzweig 1995). Early work discussed the development of the best-fit species-area model (see Gleason 1925) with the objective of estimating the number of species found in a larger area. Post-1960 work has attempted biological interpretation of the slope and intercept of species-area curves, predominately of the power function, as reviewed by Connor and McCoy (1979). MacArthur (1965), for example, suggested the intercept measured alpha diversity and the slope beta diversity (after Whittaker 1960). Shmida (1984) described Whittaker's multi-scale vegetation sampling method and his proposal to compare species diversity and richness between different vegetation types using the slope of the semi-log (exponential) model in conjunction with more traditional diversity indices.

Current researchers and resource managers concerned with biodiversity conservation may be well served by information from species-area curves. Generally, research programs must sample from some population, censuses are rare, and statistics are then used to examine the uncertainty of inferences made about the rest of the population (Steel and Torrie 1980). Researchers in conservation biology are continually faced with the dilemma of needing to demonstrate which areas should receive conservation priority based on the diversity of species contained (see Myers et al. 2000). Various approaches for rapid assessment of species presence have been developed for application at different scales ranging from field-based, landscape-scale assessments (e.g., Conservation International's Rapid Assessment Program, Stohlgren et al. 1997) to global-scale summaries of existing information (Myers et al. 2000). Many other statistical approaches have been developed to determine how much sampling is required to adequately describe an area (e.g., Keating et al. 1998). Unfortunately, most standard methods for collecting plant species data miss many species, especially the less common or patchily distributed species (Stohlgren et al. 1998). The use of inadequate data to compare landscape-scale diversity, regardless of which indices or methods are used, may be counterproductive given limited conservation funds.

Stohlgren et al. (1997) proposed an approach for rapid assessment of landscape-scale plant species richness using semi-log species-area curves. For the relatively small study area (750 ha) described in that paper, the most accurate estimates of observed species richness were achieved by dividing the slopes of the species-log(area) regressions by Jaccard's Coefficients (an index of overlap between two species lists, Krebs 1989) calculated between plots in the same vegetation type. This approach seemed to account for the opposite of species overlap – heterogeneity. When more than one plot is sampled in a vegetation type, two plots, for example, could have 10 species each. The mean number of species found in those two plots would be 10, but if only five species are found in common to both plots, then the total number of species found on those two plots is 15, which is significantly greater than the mean of 10. Standard species-log(area) curves underestimated the total number of species observed because they are based on the mean number of species observed at each sample scale without taking into account different species identities. Jaccard's Coefficient summarizes heterogeneity in species composition between plots in a given vegetation type, so I hypothesized that dividing the slope of a vegetation type's species-area curve with its Jaccard's Coefficient could result in a more accurate estimate of the total number of species observed in a vegetation type.

The species-accumulation curve is an alternative to the species-area curve approach I present, but it is less useful for the relative comparisons that I propose here because it is subject to non-linear changes in behavior (slope) as sample scales increase (Plotkin and Levin 2001). In addition, if species richness data were collected in different ways from different places differences in slopes could arise from the effects of sampling methods (Stohlgren et al. 1995) and would not provide good comparisons across studies. Species-accumulation curves should be useful for determining if sampling is adequate to capture most species in a larger area (i.e., if the curve levels off) and for comparing the contributions of vegetation types that are equally sampled. When only small sample sizes are available in various vegetation types, as in this work, species-accumulation curves may have large errors associated with extrapolated values. I discuss both of these applications of species-accumulation curves in the Epilogue of this dissertation.

The objectives of the work presented in this chapter were to: (1) determine the best-fit species-area models for vegetation data from a 54,000-ha area; and (2) use the species-area, Jaccard's Coefficient, and observed species richness information as indices of a vegetation type's contribution to landscape-scale plant diversity (species richness).

1.2 Methods

1.2.1 Study Sites

From 1995 through 1998, 147 Modified-Whittaker plots were established using stratified, random sampling in 17 vegetation types in a 54,000-ha portion of Rocky Mountain National Park, Colorado, USA (Figure 1-1). Vegetation cover types were identified on aerial photos (1987, color; 1:15840 scale; after Stohlgren et al. 1997; Table 1-1). Twenty of the aspen plots were located based on a previous bird study (stands were similarly randomly located; Natasha Kotliar, personal communication). Forty-two forest plots were established in conjunction with a forest ecotone study (Stohlgren et al. 2000).

1.2.2 Plot Design

The standard Modified-Whittaker plot for vegetation sampling consists of a 20 m x 50 m plot (1000 m^2) that contains 10 1-m^2 subplots, two 10-m^2 subplots (in opposite corners), and one 100-m^2 subplot (in plot center). The original plot design placed the 10 1-m^2 subplots around the inside of the plot perimeter (Stohlgren et al. 1995). However, after 1996 four of the 1-m^2 subplots were moved to the outside perimeter of the 100-m^2 subplot to decrease the linearity of the 1-m^2 subplots (Stohlgren et al. 1998; Figure 1-2).

The 1-m^2 subplots were placed to maximize the distance between them and to facilitate relocation for future monitoring. Within each 1-m^2, 10-m^2, and 100-m^2 subplot, all vascular plant species were identified. Finally, the entire 1000-m^2 plot was surveyed and any previously unrecorded (in the subplots) species were recorded.

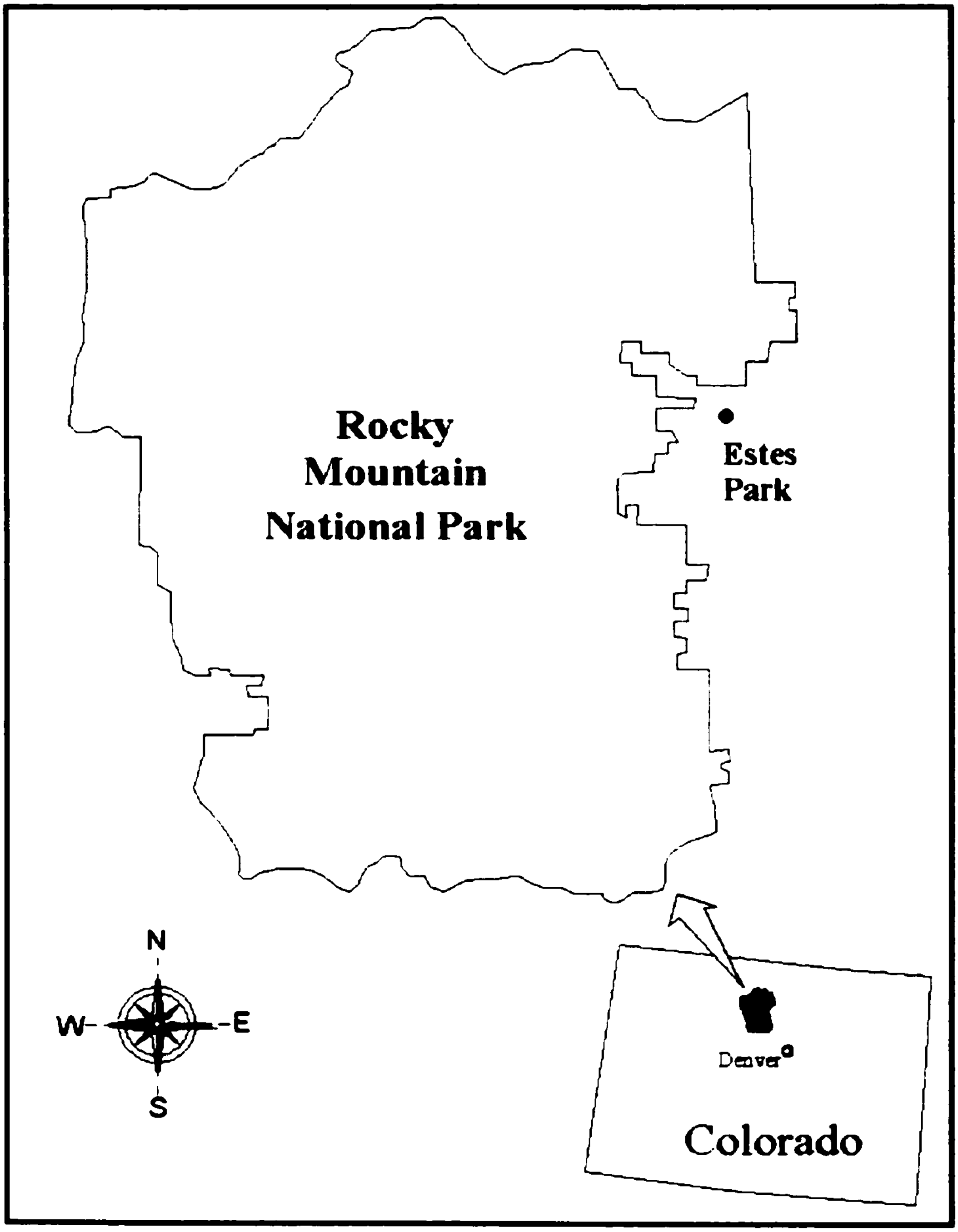

Figure 1-1. Rocky Mountain National Park, Colorado, USA. The study was conducted in the central third of the Park from the east to the west boundaries.

Table 1-1. Vegetation types and their area sampled (m^2) in Rocky Mountain National Park. Types were identified based on dominant species' cover.

Vegetation Type	**Area Sampled**	**Dominant Species**
Alpine Tundra	4000	mixed forbs and grasses
Aspen	32000	*Populus tremuloides* Michx.
Douglas Fir (DF)	6000	*Pseudotsuga menziesii* (Mirb.) Franco
DF/Lodgepole*	3000	*Pseudotsuga menziesii*, *Pinus contorta* Dougl.
Dry Meadow	9000	mixed forbs, shrubs, and grasses
Limber Pine (LIM)	3000	*Pinus flexilis* James
Lodgepole (LP)	23000	*Pinus contorta* Dougl.
LP/LIM*	3000	*Pinus contorta*, *Pinus flexilis*
LP/Ponderosa*	3000	*Pinus contorta*, *Pinus ponderosa* Laws.
LP/Spruce/Fir*	2000	*Pinus contorta*, *Picea engelmannii* (Parry) Engelm., *Abies lasiocarpa* (Hook.) Nutt.
Mixed Conifer	5000	mixed conifer species below 3000 m
Ponderosa Pine (PP)	12500	*Pinus ponderosa* Laws.
PP/DF*	3000	*Pinus ponderosa*, *Pseudotsuga menziesii*
Spruce/Fir	6000	*Picea engelmannii*, *Abies lasiocarpa*
Subalpine	9000	mixed conifer species from 3000 m to treeline
Wet Meadow	13000	mixed forbs and grasses
Willow	9000	*Salix* species

*Sites are ecotones between the two vegetation types.

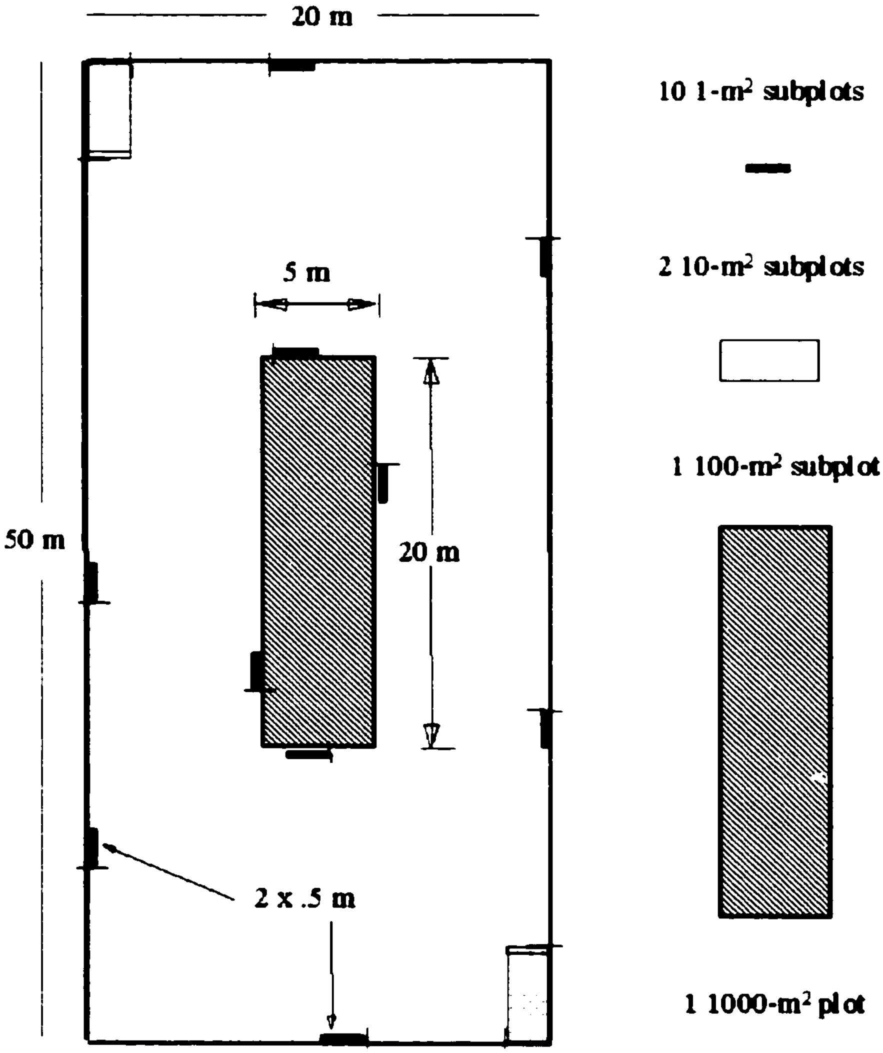

Figure 1-2. The Modified-Whittaker multi-scale vegetation sampling plot. See the website for more information: www.nrel.colostate.edu/projects/stohlgren/howtomodwhitt.htm

[illegible] Analysis

I calculated the mean number of species in the 1- and 10-m^2 subplots of each plot, the total number of species found in the 100-m^2 subplot of each plot, and the total number of species found in each plot for construction of species-area curves. Use of plot size instead of area sampled (as for species-accumulation curves) follows Shmida (1984) and dictates that the mean number of species are used for the plot sizes where more than one plot of that size is sampled.

Species lists were compiled for each vegetation type from the plot species lists. All plant specimens identified to species follow the National Plants Database nomenclature (USDA NRCS 1999). Some plants (< 5% each year) could not be identified to species because of phenology or condition, so they were labeled with a unique identifier. If these "unknowns" could be identified to genus, then the genus was included in the descriptive name. For analyses I erred on the side of low-diversity by combining difficult unknowns together (e.g., many grasses and small, basal rosettes of composites with no or inadequate voucher specimens).

I compared untransformed species-area, species-log(area) (i.e., exponential), and log(species)-log(area) (i.e., power law) curves for each vegetation type. The general form of the curves is the standard equation of a line: the number of species (y) is equal to the slope of the line (m) multiplied by the area (A) plus the constant (b) ($y = m(x) + b$). I used quadrat area to construct the curves (Shmida 1984). All analyses were conducted using SYSTAT v. 8.0 (SPSS Inc. 1998). Curves were evaluated based on their adjusted squared multiple R (r^2, a more conservative value that accounts for the comparison of multiple models or the addition of more predictive variables, which always increases R^2) and their ability to predict the number of species found in the total area of each vegetation type sampled.

Jaccard's Coefficient (Krebs 1989), an index of overlap between two lists, was used to correct the slope of the species-log(area) and log(species)-log(area) curves (Stohlgren et al. 1997). Jaccard's coefficient is defined as:

$$J=A/(A+B+C)$$

Where A = the number of species shared by both plots, B = the species on plot one but not on plot two, and C = the species on plot two but not on plot one. Jaccard's Coefficient ranges from zero (no overlap; no zeros in this data set) to one (complete overlap). Jaccard's Coefficients were used to quantify how different plots within a vegetation type are from each other, on average, based on all possible pair-wise comparisons of plot species lists.

Estimates of species number were compared between non-transformed and transformed slope (slopes divided by Jaccard's Coefficients) values for the species-log(area) and log(species)-log(area) curves. Dividing the slope by Jaccard's Coefficient will increase the slope in proportion to the overlap between species lists. Low overlap (values closer to zero) indicates heterogeneity between plots in the same vegetation type and will increase the slope of the species-area curve to take the different identities into account – identities that are lost when average numbers of species are used to construct the curves. The numbers of unique species (species that were only recorded from one vegetation type) and the mean number of species per plot were calculated for each vegetation type.

To develop an index of each of the 17 vegetation types' contribution to landscape-scale diversity, I ranked the vegetation types according to each of four criteria: their mean numbers of species per plot (high to low), slopes of the species-log(area) curves (high to low), Jaccard's Coefficients (low to high), and mean numbers of unique species per plot (high to low). Although these values are correlated in the sense that they all result from the species lists for each vegetation type, each value indicates a unique property for a given vegetation type that is not quantified by a species list alone. The index consists of the sum of the rankings (i.e., in this case the maximum index value would be 4 x 17 = 68). Vegetation types that share the same value within a category are given an equal ranking. I plotted the relationship between the index and the estimated area of coverage for the 9 vegetation types that had area data available (Alpine, Aspen, Dry Meadow, Douglas Fir, Lodgepole, Ponderosa, Spruce/Fir/Subalpine, Wet Meadow, and Willow).

1.3 Results

The slopes of the species-area curves only ranged from 0.02 to 0.05 because the range of the non-transformed plot size (area) was so large relative to the range of species numbers (Table 1-2). The slopes of the species-log(area) and log(species)-log(area) curves are more interesting as indices because they have a larger spread (Table

1-2). The log(species)-log(area) model had the greatest explanatory power (i.e., explained the greatest variance) with the highest adjusted r^2 in 12 of the 17 vegetation types while the species-log(area) model had the highest r^2 in 5 cases (Table 1-2). Based on Jaccard's Coefficient, the wet meadow vegetation type is the most heterogeneous because, on average, plot species lists only overlapped by 12% while the Lodgepole/Spruce/Fir ecotone was the most homogeneous with 40% overlap (Table 1-2).

In all cases, the untransformed model (species-area) overestimated and the species-log(area) model underestimated the number of observed species. The log(species)-log(area) model underestimated the number of observed species in 15 cases (one less than the observed in 3 cases), but it overestimated the number of species in alpine tundra (Table 1-3). As dividing the slope by Jaccard's coefficient results in an increase in the number of species estimated, it did not make sense to try correcting the slope of the species-area curve. For the species-log(area) models, dividing the slope by the Jaccard's coefficient resulted in an overestimation in 15 vegetation types (Table 1-3). For the log(species)-log(area) models, dividing the slope by Jaccard's coefficient resulted in an extreme overestimation in all types (not reported). Although the log(species)-log(area) and the slope-corrected (i.e., the slope, m, divided by the Jaccard's Coefficient, J, or m/J) species-log(area) models had the closest estimates (in 8 and 7 cases, respectively), the slope-corrected model was only off by a total of 972 while the log(species)-log(area) model was off by 15,658 (mostly from the alpine tundra estimate, Table 1-3).

Plots of the estimated number of species against the observed number of species show that the slope-corrected models provide the best estimates overall with a slope of 0.79, which was the closest to one (Figure 1-3). Although the species-area curve (S-A) explains the most variance ($R^2 = 0.91$), the steepness of the slope (slope = 2.9) results in large over-estimations of species richness as more species are observed (Figure 1-3).

The index of a vegetation type's contribution to landscape-scale diversity, based on the ranking of the vegetation types' mean numbers of species per plot, slopes of the species-log(area) curves, Jaccard's Coefficients (low to high values), and mean numbers of unique species per plot, resulted in the following overall ranking (index value is in parentheses): willow (63), aspen (62), wet meadow (51), lodgepole/ponderosa pine (48), ponderosa pine (47), dry meadow and alpine tundra (both 45), ponderosa pine/Douglas fir (43), spruce/fir (40), mixed conifer (38), Douglas fir (35), lodgepole (34), subalpine (33), Douglas fir/lodgepole (27), limber pine (24), lodgepole/limber pine (21) and lodgepole/spruce/fir (13) (Table 1-4).

The plot of the relationship between the indices and the areas of 9 of the vegetation types suggests that the types that contribute the most to landscape-scale diversity cover the least amount of area ($y = 101.5(x)^{-0.098}$, $R^2 = 0.41$, $P<0.001$; Fig. 1-4). The spruce/fir and subalpine types were combined by averaging the indices of the two types so that I could use the area described on the Rocky Mountain National Park vegetation map as spruce/fir/subalpine.

Table 1-2. Species-area curves and mean Jaccard's Coefficients (J) with standard errors (s.e.) for 17 vegetation types in Rocky Mountain National Park, Colorado. The number of plots (N) x 1000 results in the number of square meters sampled. The constant is labeled "C", and "r^2" is the adjusted squared multiple R. P values are ≤ 0.02. The curves are: species-area (S-A), species-log(area) (S-ln(A)), and log(species)-log(area) (ln(S)-ln(A)).

Vegetation Type (N)	Curve	Slope	C	r^2	J (s.e.)
Alpine Tundra (4)	S-A	0.03	22.4	0.80	0.35 (0.03)
	S-ln(A)	12.0	13.7	0.87	
	ln(S)-ln(A)	0.39	2.76	0.89	
Aspen (32)	**S-A**	**0.05**	**22.8**	**0.67**	**0.20 (0.00)**
	S-ln(A)	**18.6**	**7.6**	**0.79**	
	ln(S)-ln(A)	**0.30**	**1.02**	**0.78**	
Douglas Fir (DF) (6)	S-A	0.03	14.9	0.45	0.30 (0.02)
	S-ln(A)	12.7	4.0	0.60	
	ln(S)-ln(A)	0.36	0.63	0.69	
DF/Lodgepole (3)	**S-A**	**0.03**	**7.8**	**0.69**	**0.38 (0.04)**
	S-ln(A)	**10.6**	**-0.08**	**0.83**	
	ln(S)-ln(A)	**0.46**	**0.25**	**0.88**	
Dry Meadow (9)	S-A	0.03	18.7	0.63	0.29 (0.02)
	S-ln(A)	13.3	7.7	0.76	
	ln(S)-ln(A)	0.23	1.0	0.82	
Limber Pine (LIM) (3)	**S-A**	**0.02**	**11.2**	**0.47**	**0.28 (0.03)**
	S-ln(A)	**8.9**	**3.5**	**0.66**	
	ln(S)-ln(A)	**0.31**	**0.61**	**0.70**	
Lodgepole (LP) (23)	S-A	0.02	8.4	0.60	0.17 (0.01)
	S-ln(A)	9.2	1.0	0.68	
	ln(S)-ln(A)	0.39	0.36	0.70	
LP/LIM (3)	**S-A**	**0.02**	**9.3**	**0.54**	**0.29 (0.01)**
	S-ln(A)	**9.1**	**1.5**	**0.73**	
	ln(S)-ln(A)	**0.40**	**0.38**	**0.75**	
LP/Ponderosa (3)	S-A	0.03	18.2	0.43	0.27 (0.03)
	S-ln(A)	14.1	6.5	0.54	
	ln(S)-ln(A)	0.31	0.79	0.45	
LP/Spruce/Fir (2)	**S-A**	**0.02**	**10.1**	**0.56**	**0.40 (n.a.)**
	S-ln(A)	**6.9**	**3.9**	**0.91**	
	ln(S)-ln(A)	**0.26**	**0.67**	**0.86**	
Mixed Conifer (5)	S-A	0.03	8.2	0.45	0.18 (0.02)
	S-ln(A)	9.7	0.5	0.50	
	ln(S)-ln(A)	0.41	0.32	0.67	
Ponderosa Pine (PP) (14)	**S-A**	**0.03**	**15.4**	**0.57**	**0.24 (0.01)**
	S-ln(A)	**13.4**	**4.2**	**0.71**	
	ln(S)-ln(A)	**0.33**	**0.73**	**0.67**	
PP/DF (3)	S-A	0.03	23.5	0.36	0.31 (0.05)
	S-ln(A)	14.9	9.4	0.73	
	ln(S)-ln(A)	0.32	0.88	0.65	
Spruce/Fir (6)	**S-A**	**0.03**	**15.0**	**0.50**	**0.21 (0.03)**
	S-ln(A)	**11.7**	**4.7**	**0.72**	
	ln(S)-ln(A)	**0.29**	**0.79**	**0.77**	
Subalpine (9)	S-A	0.02	10.4	0.57	0.27 (0.02)
	S-ln(A)	9.3	3.1	0.68	
	ln(S)-ln(A)	0.32	0.59	0.73	
Wet Meadow (13)	**S-A**	**0.03**	**15.6**	**0.68**	**0.12 (0.01)**
	S-ln(A)	**12.0**	**6.5**	**0.71**	
	ln(S)-ln(A)	**0.23**	**0.96**	**0.82**	
Willow (9)	S-A	0.04	18.2	0.71	0.18 (0.01)
	S-ln(A)	15.3	5.9	0.81	
	ln(S)-ln(A)	0.25	0.98	0.88	

Table 1-3. Observed and estimated (based on the total area sampled in a vegetation type) species counts based on three species-area regressions and the Jaccard's Coefficient corrected species-log(area) curve (slope/J). Absolute values of the observed – estimated values are in parentheses and are summed below.

Vegetation Type	Obs.	S-A	S-ln(A)	ln(S)-ln(A)	slope/J
Alpine Tundra	104	142 (38)	57 (47)	14615 (14511)	115 (11)
Aspen	**527**	**1623 (1096)**	**91 (436)**	**235 (292)**	**427 (100)**
Douglas Fir (DF)	116	195 (79)	52 (64)	98 (18)	164 (48)
DF/Lodgepole	**59**	**98 (39)**	**36 (23)**	**71 (12)**	**96 (37)**
Dry Meadow	169	289 (120)	60 (109)	81 (88)	189 (20)
Limber Pine (LIM)	**60**	**71 (11)**	**35 (25)**	**49 (11)**	**114 (54)**
Lodgepole (LP)	223	468 (245)	41 (182)	115 (108)	237 (14)
LP/LIM	**60**	**69 (9)**	**33 (27)**	**59 (1)**	**111 (51)**
LP/Ponderosa	100	108 (8)	56 (44)	74 (26)	188 (88)
LP/Spruce/Fir	**35**	**50 (15)**	**27 (8)**	**34 (1)**	**61 (26)**
Mixed Conifer	108	158 (50)	36 (72)	69 (39)	200 (92)
Ponderosa Pine (PP)	**204**	**390 (186)**	**59 (145)**	**121 (83)**	**233 (29)**
PP/DF	99	114 (15)	61 (38)	98 (1)	177 (78)
Spruce/Fir	**134**	**195 (61)**	**49 (85)**	**77 (57)**	**215 (81)**
Subalpine	128	190 (62)	40 (88)	72 (56)	139 (11)
Wet Meadow	**294**	**406(112)**	**56(238)**	**81(213)**	**418(124)**
Willow	234	378(144)	66(168)	93(141)	342(108)
Sum of the Obs.-Exp.		**2290**	**1799**	**15658**	**972**

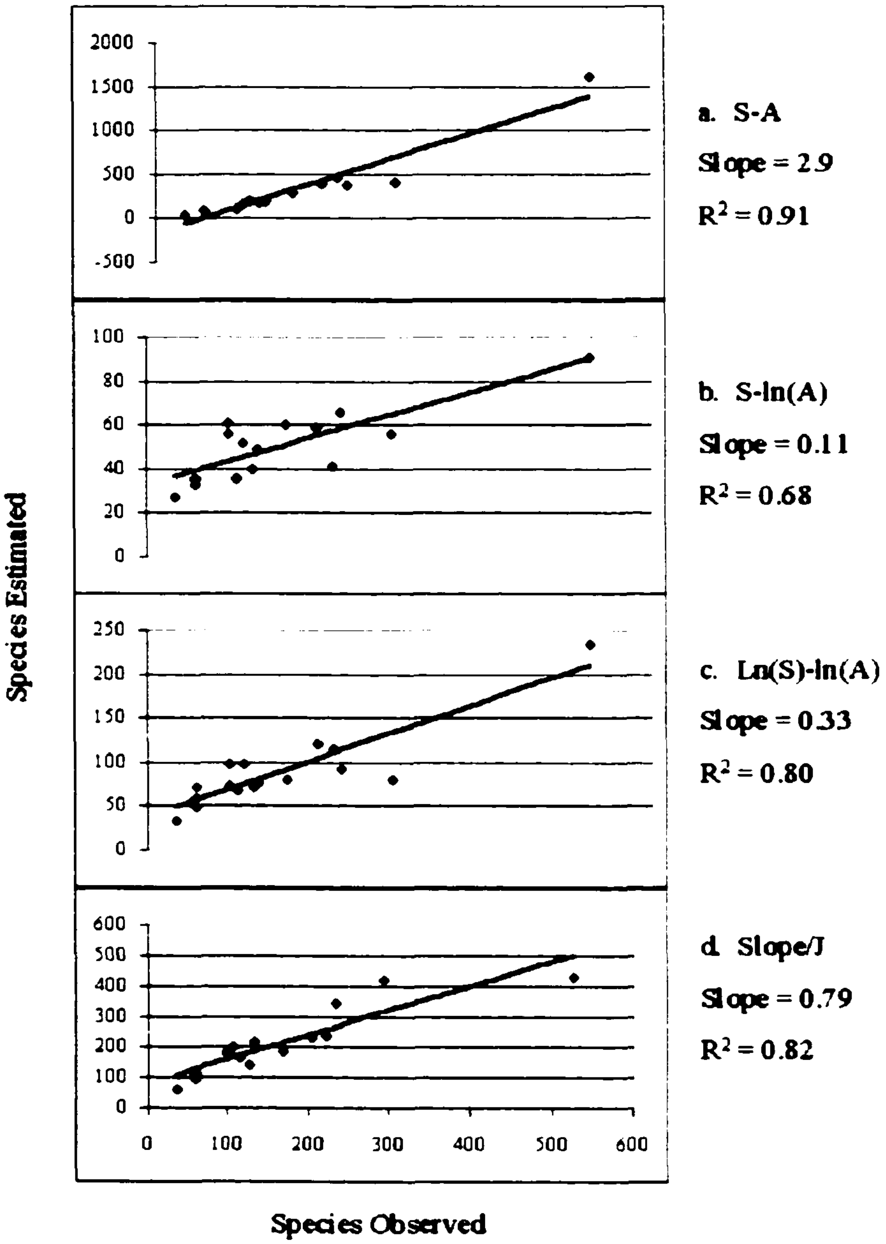

Figure 1-3. The relationships between the estimated versus the observed numbers of species from the four species-area models.

Table 1-4. Vegetation types ranked by the mean number of species observed in a plot (value in parentheses), slope (high to low, Table 1-2), Jaccard's Coefficient (low to high, Table 1-2), and the mean number of unique species (species only recorded in one vegetation type, high to low, value in parentheses) found in each plot in a vegetation type. Entries with equal superscripts have equal values.

Mean number Species/Plot	Slope S-ln(A)	Jaccard's Coefficient	Mean Unique Species/Plot
Aspen (67)	Aspen	Wet Meadow	Alpine Tundra (7)
Willow (55)	**Willow**	**Lodgepole**	**Willow (6)**
Alpine Tundra (53)	PP/DF	Mixed Conifer^2	Aspen (5)
PP/DF (52)	**LP/PP**	**Willow^2**	**Dry Meadow (4)**
LP/PP (51)	Ponderosa (PP)	Aspen	Wet Meadow (4)
Dry Meadow (50)	**Dry Meadow**	**Spruce/Fir**	**Mixed Conifer (2)**
Ponderosa (47)	Douglas Fir (DF)	Ponderosa Pine	Ponderosa Pine (2)
Wet Meadow (46)	**Alpine Tundra^1**	**LP/PP^3**	**Spruce/Fir (2)**
Douglas Fir (44)	Wet Meadow^1	Subalpine^3	DF/LP (1)
Spruce/Fir (40)	**Spruce/Fir**	**Limber Pine**	**LP/PP (1)**
DF/LP (33)	DF/LP	Dry Meadow^4	Subalpine (1)
Subalpine (33)	**Mixed Conifer**	**LP/LIM^4**	**Douglas Fir (<1)**
Mixed Conifer (32)	Subalpine	Douglas Fir	Limber Pine (<1)
Limber Pine (31)	**Lodgepole (LP)**	**PP/DF**	**Lodgepole (<1)**
Lodgepole (31)	LP/LIM	Alpine Tundra	LP/Spruce/Fir (<1)
LP/LIM (30)	**Limber Pine (LIM)**	**DF/LP**	**PP/DF (<1)**
LP/Spruce/Fir (25)	LP/Spruce/Fir	LP/Spruce/Fir	LP/LIM (0)

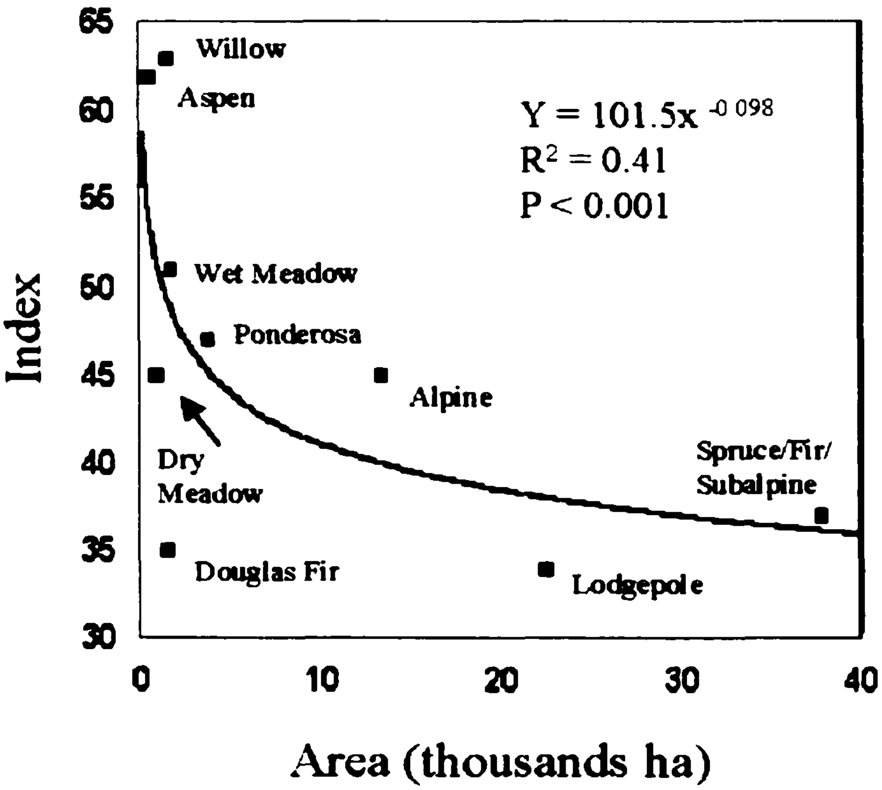

Figure 1-4. The relationship between the area a vegetation type covers in Rocky Mountain National Park, Colorado, USA, is negatively related to the index of a vegetation type's contribution to landscape-scale biodiversity (species richness). The index is composed of the sum of the rankings of the vegetation types (mean number of species/plot, slope of the species-log(area) curve, Jaccard's Coefficient, and mean unique species/plot).

1.4 Discussion

Species-area models (Plotkin and Levin 2001) and species lists (Palmer et al. 2002) are critical tools for biodiversity conservation because they provide the foundation for understanding what is to be conserved. Although both of these tools may seem straightforward initially, underlying assumptions (e.g., random distribution of individuals) and lack of field data (e.g., few, subjectively placed plots) may make their application difficult (Plotkin and Levin 2001, Palmer et al. 2002). The use of species-area models presented here to compare the relative contributions of vegetation types to diversity at landscape-scales may be a useful application for conservation biology because information from the curves allows vegetation types to be compared without the use of abundance and distribution data, which are difficult to obtain, and without absolute knowledge of all species present in the larger, unsampled areas, which may be logistically impossible to know.

The slopes of the curves allow vegetation types to be compared with respect to their contributions to species richness of an area based on their relative richness contributions. The accuracy of the relative contributions (ranking) may be increased using Jaccard's Coefficient to correct the slope of the species-log(area) curve because the coefficient accounts for heterogeneity (species identity differences) between plots within a given vegetation type. Those species identity differences are lost when species-area curves are constructed based on the average number of species alone (Palmer 1990). Regardless of the methods used to compare habitats' contributions to diversity, adequate sampling is essential. The use of 0.1 ha, multi-scale plots located using a stratified random design ensures that data will not be skewed to only the most common species (Stohlgren et al. 1997, 1998) as is the case when smaller plots are used or when visually homogenous areas are sought for sampling.

The first richness index produced by a species-area model is the slope. Steeper slopes indicate that increases in area will result in greater additions of species. The use of species-area models to estimate the real number of species that would be expected to occur in a larger area than the area sampled seems less likely to be useful for conservation decision-making because these estimates may not be predictably accurate. However, estimations may be useful as indices of relative contributions to species richness (Palmer 1990). Although the slope adjustments of the species-log(area) curve with Jaccard's Coefficients resulted in overestimation in 15 cases (Table 1-3), Jaccard's Coefficient may provide a way to improve the accuracy of such indices because the models using the slope divided by Jaccard's Coefficient resulted in the best overall predictions based on the slopes of the plots of observed versus estimated numbers of species (slope = 0.79; Figure 1-3).

Examination of vegetation types' contributions of unique species is interesting, but they are influenced by sample size. For example, a disproportionately large number of plots were sampled in aspen stands, which resulted in the highest total number of unique species but only the third highest number of unique species per plot (Table 1-4). The use of the mean number of species per plot may be more useful with unequal sample sizes.

Generally, rankings of vegetation types by the individual indices (Jaccard's Coefficients, mean unique species/plot, the observed mean species/plot, or the slope of the species-log(area) curves) did not mirror each other (Table 1-4). However, they offer additional insight into the patterns of species richness when they are examined individually and then combined into an overall index. For example, alpine tundra plots averaged a high number of total and unique species per plot, but the alpine tundra slope and mean Jaccard's Coefficient were intermediate and high, respectively, indicating that one would not expect to encounter many more total or unique species if more plots were sampled. Alpine tundra was ranked intermediately for its contribution to landscape-scale diversity in the Park when the individual ranks were combined into the single index. However, because it occupies a restricted range, alpine tundra would probably warrant greater protection than that based on its biodiversity contributions alone.

Thus, I propose a comprehensive approach to evaluating vegetation types' contributions to landscape-scale species richness. The approach starts with stratified, random sampling using relatively large (0.1 ha), multi-scale plots (non-contiguous), then uses species-area curves, and measures of heterogeneity (Jaccard's Coefficient, mean number of species per plot, and unique species) to rank the vegetation types based on their measured and estimated contributions to species richness. Based on this, the vegetation types in this study are ranked by the mean number of species observed by plot, the slopes of the species-log(area) curves, the Jaccard's Coefficients, and the mean number of unique species by plot. This ranking is of interest to resource managers of this study area because the vegetation types that contribute the most to landscape-scale plant species diversity (alpine tundra, aspen, wet meadow, and willow) are also the vegetation types that cover the least amount of area (Fig. 1-4) and are most likely to be impacted by human and large ungulate use.

As the conservation community continues to embrace the use of rapid assessments, remote sensing, and geographic information systems (GIS) to make long-term decisions, the time is right to re-examine the use of one of the oldest "rules" in ecology, which says that more species are found as the area examined is increased (Rosenzweig 1995). Although it is intuitive that larger areas will contain more species than smaller areas, small areas of some vegetation types contain more species than much larger areas of other types. This quality is generally reflected in the slope of the species-area curve being steeper in species-rich types. However, an important goal for conservation is protecting rare and unique species as well as species-rich areas. Jaccard's Coefficient may provide a simple means to incorporate landscape-scale uniqueness (i.e., heterogeneity) into an index of richness (the species-area curve).

Using these indices in relation to known areas, such as with a GIS, provides much more potential information about landscape-scale diversity than species lists or maps of individuals' locations alone. For example, Myers et al. (2000) present an approach to maximizing species conservation through protecting regional hotspots of diversity and endemism. However, each region is relatively large and diversity and availability of land for protection within each region varies. The identification of hotspots within hotspots could be based on the methods and indices proposed here that give value to species richness as well as uniqueness (heterogeneity) within and between local vegetation types.

1.5 Literature Cited

Connor, E.F., and E.D. McCoy. 1979. The statistics and biology of the species-area relationship. *The American Naturalist* 113:791-833.

Gleason, H.A. 1925. Species and area. *Ecology* 6:66-74

Keating, K.A., J.F. Quinn, M.A. Ivie, and L.L. Ivie. 1998. Estimating the effectiveness of further sampling in species inventories. *Ecological Applications* 8:1239-1249.

Krebs, C.J. 1989. *Ecological Methodology*. Harper & Row, New York, NY.

MacArthur, R. 1965. Patterns of species diversity. *Biological Reviews* 40:510-533.

Myers N., R.A. Mittermeier, C.G. Mittermeier, G.A.B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403:853-858.

Natural Resources Conservation Service, USDA. 1999. The PLANTS database [WWW document]. URL http://plants.usda.gov/plants

Palmer, M.W. 1990. The estimation of species richness by extrapolation. *Ecology* 71:1195-1198.

Palmer, M.W., P.G. Earls, B.W. Hoagland, P.S. White, and T. Wohlgemuth. 2002. Quantitative tools for perfecting species lists. *Environmetrics* 13:121-137.

Plotkin, J.B. and S.A. Levin. 2001. The spatial distribution and abundances of species: lessons from tropical forests. *Theoretical Biology* 6:251-278.

Rosenzweig, M.L. 1995. *Species Diversity in Space and Time*. Cambridge University Press, Cambridge, UK.

SPSS, Inc. 1998. SYSTAT 8.0. SPSS Inc., 233 S. Whacker Drive, 11th Floor, Chicago, IL 60606-6307.

Steel, R. and J. Torrie. 1980. *Principles and Procedures of Statistics*. 2nd edn. McGraw Hill, Inc., New York, NY.

Shmida, A. 1984. Whittaker's plant diversity sampling method. *Israel Journal of Botany* 33:41-46.

Stohlgren, T.J., K.A. Bull, and Y. Otsuki. 1998. Comparison of rangeland vegetation sampling techniques in the Central Grasslands. *Journal of Range Management* 51:164-172.

Stohlgren, T.J., G.W. Chong, M.A. Kalkhan, and L.D. Schell. 1997. Rapid assessment of plant diversity patterns: a methodology for landscapes. *Environmental Monitoring and Assessment* 48:25-43.

Stohlgren, T.J., M.B. Falkner, and L.D. Schell. 1995. A Modified-Whittaker nested vegetation sampling method. *Vegetatio* 4:1-8.

Stohlgren, T.J., A. Owen, and M. Lee. 2000. Monitoring shifts in plant diversity in response to climate change: a method for landscapes. *Biodiversity and Conservation* 9:67-86.

Whittaker, R.H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* 30:279-338.

CHAPTER II: AN EXAMPLE APPLICATION OF SPECIES-AREA CURVES FOR MANAGERS: BIODIVERSITY- ASPEN HAS THE LEAD, BUT WILL NON-NATIVE SPECIES TAKE OVER?

2.0 Abstract

Vascular plant and butterfly diversity were investigated in Rocky Mountain National Park. I identified a total of 188 native and non-native vascular plant species unique to the aspen vegetation type in 32 0.1-ha plots. The slope of the mean species-area curve for the aspen vegetation type was the steepest of the 10 types sampled, which indicates that an increase in aspen area could have much greater positive impacts on total plant species richness than an increase in area of other vegetation types. Aspen plots contained the greatest number of butterfly species and ranked highest with all diversity indices tested. However, aspen plots were the most heavily invaded by non-native plant species. Aspen plots contained 38 of the 42 non-native species identified in all plots, and 15 of the 38 (36% of all non-natives found) were not found in any other vegetation type. Non-native species could have negative effects on native plant and butterfly species diversity.

2.1 Introduction

The importance of aspen (*Populus tremuloides* Michaux) for maintaining biodiversity in western landscapes is well introduced by DeByle et al. (1985), where aspen stands are noted for their own genetic diversity, as well as providing habitat for insects, birds, and mammals. Stohlgren et al. (1997a, 1997b) found a disproportionately high number of vascular plant species in aspen stands in relation to their coverage in the Beaver Meadows area (750 ha) of Rocky Mountain National Park, Colorado. There, aspen covered only 1.2% of the landscape, but it contained 45% of the plant species sampled.

Resource managers in Rocky Mountain National Park (the Park) are concerned that elk (*Cervus elaphus* Nelsoni) may be harming vegetation in portions of the Park (Berry et al. 1997). Localized studies have reported little or no aspen regeneration in elk winter range (Baker et al. 1997, Olmstead 1997), while a more extensive study did find successful regeneration at landscape scales in areas of low elk use (Suzuki et al. 1999). Similar concern and controversy over regeneration exists in the Greater Yellowstone Ecosystem (Krebill 1972, Gruell and Loope 1974, Weinstein 1979, Bartos et al. 1991, Romme et al. 1995, Barnett and Stohlgren 2001). Another potential harmful effect on aspen stands and the diversity that they support is invasion by non-native plant species. Work in other species-rich habitat types has found that hotspots of native plant diversity are being invaded by non-native plant species (Stohlgren et al. 1998b, 1999a, 1999b, 2001). Non-native plant species invasions are listed as the second greatest threat to native biodiversity, behind habitat loss (Wilcove et al. 1998). The presence of non-native species may have long-term, negative consequences for native diversity especially in vegetation types such as aspen that are small, scattered, and rare on the landscape in parts of their range.

Species-area curves allow comparisons across vegetation types, and even other studies, because the slopes of the curves can be calculated and compared without the difficulties posed by other diversity indices that often require abundance data (for evenness) and vary greatly depending on study design (Ludwig and Reynolds 1988). In addition, species-area models allow one to estimate the number of species expected in an area larger than the area sampled. Although the accuracy of these estimates may not be sufficient to be taken as a true value, they may be used as an index of diversity (richness) to rank the vegetation types sampled in terms of their relative contributions to diversity (Palmer 1990).

My objective was to assess a variety of vegetation types' contributions to plant and butterfly species richness in Rocky Mountain National Park. I used species-log(area) curves (Gleason 1925, Shmida 1984, Rejmanek and Ejvind 1992) to compare the relative contributions to vascular plant species richness made by 10 different vegetation types in the Park. To test the use of plant richness to predict the diversity of other taxonomic groups, the significant contributions of aspen to butterfly species diversity and the relations between butterfly species richness and plants in the Beaver

Meadows area were investigated (Simonson et al. 2001). In addition. I examined the number of species that only occur in one vegetation type (unique to a vegetation type) and the number of non-native plant species found in a set of vegetation plots.

2.2 Methods

2.2.1 Study Site

From 1995 through 1998. 104 Modified-Whittaker plots (Figure 1-2) were established (using stratified. random sampling) in 10 vegetation types in a 54,000-ha portion of Rocky Mountain National Park. Colorado (after Stohlgren et al. 1997b). Vegetation cover types were identified on aerial photos (1987. color: 1:15840 scale) and included aspen (*Populus tremuloides* Michaux: 32 plots). willow (*Salix* spp.: 8 plots). dry meadow (various species: 12 plots). wet meadow (various species: 12 plots). ponderosa pine (*Pinus ponderosa* Dougl. ex Laws.:12 plots). lodgepole pine (*Pinus contorta* Dougl.: 12 plots). mixed conifer (various species: 4 plots). subalpine (various species: 4 plots). alpine tundra (various species: 4 plots). and spruce/fir (*Picea engelmannii* Parry ex Engelm. and *Abies lasiocarpa* (Hook.) Nutt: 4 plots). Twenty of the aspen plots were located based on a previous bird study (stands were similarly randomly located: Natasha Kotliar. U.S. Geological Survey. personal communication). Butterfly data were collected in 1996 on the 24 Modified-Whittaker plots (4 plots in each of 6 vegetation types: aspen. burned ponderosa. dry meadow. lodgepole pine.

ponderosa pine, and wet meadow) in the 750-ha Beaver Meadows Study area (Simonson et al. 2001).

2.2.2 Vegetation and Butterfly Sampling

Vegetation sampling is described in section 1.2.2 of this document.

For butterfly sampling, the 20 m x 50 m plot contained six 10-m^2 subplots around the inside of the plot perimeter and one 100-m^2 subplot in the plot center (Simonson 1998). The butterfly plot was overlain directly on the vegetation plot. Butterfly diversity was measured based on systematic surveys of the subplots and plot (Simonson 1998). Butterflies were identified to species, and abundances were also recorded.

2.2.3 Analyses

Species lists were compiled for plants and butterflies to determine the total numbers of species found in each vegetation type as well as the species unique to a vegetation type. Non-native plant species were noted. All plant specimens identified to species follow the National Plants Database nomenclature (USDA NRCS 1999). Some plants could not be identified to species because of phenology or condition, so they were labeled with a unique identifier. If these "unknowns" could be identified to genus, then the genus was included in the descriptive name. For analyses, I erred on the side of caution by combining unidentifiable (no or inadequate voucher specimen) unknown species together (e.g., many grasses and small, basal rosettes of composites).

Species-area curves, were constructed (SYSTAT version 8.0; SPSS, Inc. 1998) using the mean number of plant species found in each subplot for the 1-m^2 and 10-m^2 subplots and the total number of plant species found in the 100-m^2 subplot and the full 1000-m^2 plot. These curves are the result of linear regression where the number of species is a function of the area sampled (log(plot size); Shmida 1984). Thus, the curve rises more rapidly in species rich habitat types because more species are encountered as area is increased (the slope of the line is steeper). Species-area curves were developed for each vegetation type based on all the plots sampled in that type. Butterfly data were analyzed using rarefaction curves, analysis of variance, diversity and evenness indices, and regression (Simonson et al. 2001).

2.3 Results

One hundred and eighty-eight vascular plant species occurred only in aspen plots (N=32 plots). The remaining vegetation types (N=72 plots) combined contained 264 species that only occurred in one vegetation type. Thus, the aspen plots, on average, contain 6 unique species (species that were only found in aspen) while any other plot only contains 4 unique species, on average. The slope of the mean species-area curve for the aspen vegetation type was the steepest of the nine vegetation types sampled (Table 2.1, Figure 2-1). The slope indicates the rate of accumulation of new species as the area sampled is increased. Aspen plots also contained 38 of the 42 non-native species identified in all plots, and 15 of the 38 were not found in any other vegetation type (Table 2-2).

In the Beaver Meadows portion of the Park, four plots were sampled for plant and butterfly diversity in each of six vegetation types. Aspen covered the smallest area with a total of 8.8 ha scattered throughout the 750-ha area. Thirty-three butterfly species, a total of 252 individuals, were recorded in aspen stands, and seven of those species were found only in aspen (Simonson 1998). For both richness and diversity indices, aspen consistently ranked the highest of the vegetation types sampled for butterfly diversity (Simonson et al. 2001, Table 2-3). Butterfly species richness was strongly positively correlated with native plant species richness ($r=0.64$; $p<0.01$), but the best predictors of butterfly species richness were exotic plant species richness ($r=0.70$; $p<0.01$) and exotic plant species cover ($r=0.70$; $p<0.01$; Simonson et al. 2001).

Table 2.1 Species-area curves for vascular plant species in 10 vegetation types from a 54,000-ha portion of Rocky Mountain National Park, Colorado. The equations follow the general equation of a line: $N=m(x) + b$. Where N = the number of species; m = the slope of the line; x = log(plot size); and b is a constant (the intercept). Larger values of the slope (m) indicate a greater accumulation of species as area is increased. The adjusted squared multiple R (r^2) is more conservative than R^2 because it accounts for comparison of multiple models or the addition of more predictive variables, which always increases R^2. In all cases, $p < 0.01$. The number of plots (#) is shown.

Vegetation Plots	Equation	Adjusted R^2	#
Aspen	N = 17.18(log(area)) + 7.85	0.76	32
Willow	N = 15.11(log(area)) + 6.31	0.80	9
Dry Meadow	N = 13.14(log(area)) + 8.20	0.75	9
Spruce/Fir	N = 12.81(log(area)) + 5.78	0.74	4
Ponderosa	N = 12.47(log(area)) + 3.70	0.85	8
Tundra	N = 11.69(log(area)) + 14.40	0.86	4
Wet Meadow	N = 11.60(log(area)) + 6.55	0.71	8
Mixed Conifer	N = 9.36(log(area)) + 1.33	0.48	5
Lodgepole	N = 9.23(log(area)) + 2.82	0.74	8
Subalpine	N = 9.02(log(area)) + 3.80	0.66	8

Table 2-2. Non-native plant species found in 32 aspen plots. Species identified as "N" were found only in aspen while species identified as "N*" were found in other vegetation types as well.

Status	Family	Scientific Name	Common Name
N*	Poaceae	*Agrostis gigantea*	Redtop
N	Poaceae	*Alopecurus pratensis*	Meadow Foxtail
N*	Brassicaceae	*Alyssum alyssoides*	Pale Madwort
N*	Brassicaceae	*Arabis glabra*	Tower Mustard
N	Brassicaceae	*Brassica juncea*	India mustard
N*	Poaceae	*Bromus tectorum*	Cheatgrass
N*	Brassicaceae	*Camelina microcarpa*	False Flax
N	Apiaceae	*Carum carvi*	Caraway
N	Asteraceae	*Carduus nutans*	Musk Thistle
N*	Chenopodiaceae	*Chenopodium album*	Lambsquarters
N*	Asteraceae	*Cirsium arvense*	Canadian thistle
N	Convolvulaceae	*Convolvulus arvensis*	Field Bindweed
N*	Apiaceae	*Conium maculatum*	Poison Hemlock
N	Poaceae	*Cynodon dactylon*	Bermudagrass
N	Boraginaceae	*Cynoglossum officinale*	Gypsy Flower
N	Poaceae	*Dactylis glomerata*	Orchardgrass
N*	Brassicaceae	*Descurainia Sophia*	Flixweed
N	Poaceae	*Elytrigia repens* var. *repens*	Quackgrass
N*	Poaceae	*Festuca ovina*	Sheep Fescue
N*	Apiaceae	*Heracleum sphondylium*	Eltrot
N	Brassicaceae	*Lepidium campestre*	Field Pepperweed
N	Brassicaceae	*Lepidium perfoliatum*	Clasp. Pepperweed
N*	Poaceae	*Lolium pratense*	Ryegrass
N	Fabaceae	*Melilotus officinalis*	Yellow Sweetclover
N*	Poaceae	*Phleum pratense*	Timothy
N	Poaceae	*Poa compressa*	Canada Bluegrass
N*	Poaceae	*Poa pratensis*	Kentucky Bluegrass
N*	Polygonaceae	*Rumex acetosella*	Sheep Sorel
N*	Polygonaceae	*Rumex crispus*	Curly Dock
N	Caryophyllaceae	*Silene vulgaris*	Maidenstears
N*	Brassicaceae	*Sisymbrium altissimum*	Tall Tumblemustard
N*	Asteraceae	*Taraxacum officinale*	Common Dandelion
N*	Brassicaceae	*Thlaspi arvense*	Pennycress
N*	Asteraceae	*Tragopogon dubius*	Salsify
N*	Fabaceae	*Trifolium hybridum*	Alsike Clover
N*	Fabaceae	*Trifolium repens*	White Dutch Clover
N	Scrophulariaceae	*Verbascum thapsus*	Common Mullein

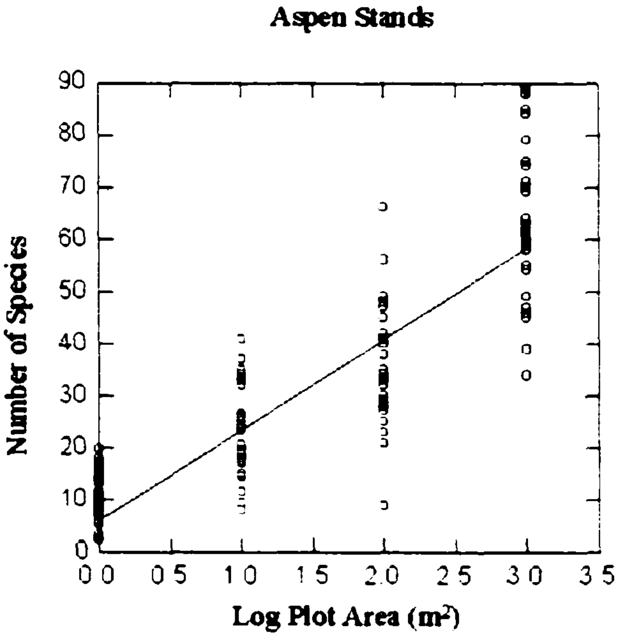

Figure 2-1 An example species-log(area) curve. Data are from 32 Modified-Whittaker plots placed in 32 aspen stands. The slope of the regression line is an index of species richness: the steeper the slope, the more new species are expected to be added as the area sampled is increased. I used plot size (m^2) for area. The equation of the aspen regression is:

$$N = 17.18*\ln(\text{area}) + 7.85$$

(adjusted $r^2 = 0.76$, $p < 0.01$), where N is the predicted number of species for a given area, the slope is 17.18, and the constant (i.e., intercept) is 7.85.

Table 2-3. Richness (cumulative species and estimated species) and diversity indices for the number of butterfly species/plot in aspen (A), wet meadow (WM), Ponderosa pine (PP), dry meadow (DM), burned conifer (BC), and lodgepole pine (LP) vegetation types in the Beaver Meadows study area, Rocky Mountain National Park, Colorado (Simonson et al. 2001).

	Vegetation Types					
Butterflies	**A**	**WM**	**PP**	**DM**	**BC**	**LP**
Richness						
Observed Species	33	27	21	19	15	13
Estimated Species[a]	19	15	14	11	14	13
Diversity						
Simpson's A	0.08	0.11	0.12	0.17	0.12	0.13
Shannon's H'	2.92	2.56	2.29	2.10	2.32	2.16
Hill's N1	18.60	12.89	9.83	8.13	10.21	8.71
Hill's N2	12.27	9.54	6.45	5.96	8.55	7.80

[a] Based on rarefaction, which allows the comparison of species numbers between vegetation types where sample sizes (number of individuals observed) were unequal. This estimate provides an index of richness.

2.4 Discussion

Although aspen stands cover a small proportion of Rocky Mountain National Park (2% based on one Park map or 5% based on recent work by Kaye et al., 2001), they contribute a disproportionate amount to plant and butterfly species richness. For example, aspen comprised only 1.2% of the vegetation cover in the Beaver Meadows study area, yet the four plots sampled in aspen contained 150 plant species (45% of the plants observed on all 24 plots). Of the plant species, 50 were unique to the aspen type (25% of the unique species observed in the six vegetation types in that study, Stohlgren et al. 1997b). Beaver Meadows aspen contained more unique butterfly species than any other vegetation type. Thirty-three of the 49 butterfly species observed were seen in aspen, and seven of those were recorded only in the aspen type (Simonson et al. 2001). In the Beaver Meadows study area (750 ha, Stohlgren et al. 1997b) and the larger study area (54,000 ha, present study) the slopes of species-area curves for plants in aspen were steeper than those for any other vegetation type sampled. Using slope steepness as an index of a vegetation type's contribution to species richness, aspen stands are clearly important for maintaining landscape biodiversity. Resource managers are justified in their concern about aspen's persistence on the landscape.

Managers must add invasive, non-native species to their list of potential threats to the integrity of aspen ecosystems. I observed 42 non-native plant species in the Park, 38 of those occurred in aspen stands, and 15 of those were not found in the plots in any other vegetation type. This is partially explained by the large number of plots in aspen

(N = 32 plots), but it is still extremely high compared to the 72 non-aspen plots. Especially alarming was the presence of noxious, agricultural and urban weeds (e.g., field bindweed, *Convolvulus arvensis* L.) in seemingly remote, undisturbed aspen stands. None of our sites that appeared relatively undisturbed had high cover of non-native species, but their presence indicates that seed sources are available. With seed available, any disturbance is likely to result in increased cover of non-native plant species. The ability of the non-native species to form dense stands can prevent native plants from persisting or establishing (Whitson et al. 1996).

Other potential negative effects involve pollinator interactions between native and non-native plants. For example, many butterflies were observed (S. Simonson, personal communication) on the flowers of musk thistle (*Carduus nutans* L.) and Canada thistle (*Cirsium arvense* L.). Even though these non-native species provide nectar, they are not suitable host plants for most butterfly larva, which often require specific native hosts. If non-natives are being pollinated this may increase their invasion success. In addition, many native plants are believed to be pollinator-limited (Burd 1994) even without competition for pollinators from non-natives. If pollination of natives is reduced and results in decreased reproduction, this will exacerbate their displacement and could impact butterfly diversity by reducing the populations of required host plants.

The results of this work indicate that emphasis on maintaining aspen on the landscape may be justified to maintain landscape-scale biodiversity. In addition to managing natural processes such as elk herbivory and fire, resource managers may need to detect invasive species early, monitor their effects, and control non-native plant invasions to maintain the native diversity supported by aspen.

Attempts to manage for intact aspen stands as a component of forest ecosystems must consider the potential negative impacts of non-native plant species on native species richness across taxonomic groups. Controlling invasive non-native species in aspen stands must be carefully done because these areas contain more unique assemblages of native species. Researchers and managers may need to be especially attentive to processes that encourage aspen regeneration or establishment (e.g., fire, disturbance) because these processes also facilitate non-native species establishment. Understanding the connectivity of aspen and other vegetation cover types in relation to non-native plant species movement and establishment will be an essential component to pro-active management of native species and aspen stands.

2.5 Literature Cited

Baker, W.L., J.A. Munroe, A.E. Hessl. 1997. The effects of elk on aspen in the winter range of Rocky Mountain National Park. *Ecography* 20:155-165.

Barnett, D.T., and T.J. Stohlgren. 2001. Aspen persistence in the Southern Greater Yellowstone Ecosystem. *Landscape Ecology* 16:569-580.

Bartos, D.L., W.F. Mueggler, and R.J. Campbell. 1991. *Regeneration by suckering on burned sites in western Wyoming*. Research Paper INT-448. Odgen, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 10p.

Berry, J., D. Decker, J. Gordon, R. Heitschmidt, D. Huff, D. Knight, W. Romme, and D. Swift. 1997. *Rocky Mountain National Park: Science-Based Assessment of Vegetation Management Goals for Elk Winter Range*. Fort Collins, CO: Environment and Natural Resources Policy Institute, Colorado State University. 16p.

Burd, M. 1994. Bateman's principle and plant reproduction: the role of pollinator limitation in fruit and seed set. *The Botanical Review* 60:83-111.

DeByle, N.V. and R.P. Winokur (eds.) 1985. *Aspen: Ecology and Management in the Western United States*. General Technical Report RM-119. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 283p.

Gleason, H.A. 1925. Species and area. *Ecology* 6:66-74.

Gruell, G.E. and L.L. Loope. 1974. Relationships among aspen, fire, and ungulate browsing in Jackson Hole, Wyoming. Moose, USDI National Park Service: 33p.

Kaye, M.W., K. Suzuki, D. Binkley, and T.J. Stohlgren. 2001. Landscape-scale dynamics of aspen in Rocky Mountain National Park, Colorado. Pp. 39-46 in Sheppard, W. D. et al. (compilers) Sustaining Aspen in Western Landscapes: Symposium Proceedings; 13-15 June 2000; Grand Junction, CO. Proceedings RMRS-P-18. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 460 p.

Krebill, R.G. 1972. *Mortality of aspen on the Gros Ventre winter range*. Research Paper INT-129. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 16p.

Ludwig, J.A. and J.F. Reynolds. 1988. *Statistical Ecology*. New York: John Wiley & Sons. 337p.

Olmstead, C.E. 1997. Twenty years of change in Rocky Mountain National Park elk winter range aspen. Greeley, CO: University of Northern Colorado. 45p.

Palmer, M.W. 1990. The estimation of species richness by extrapolation. *Ecology* 71:1195-1198.

Rejmanek, M. and R. Ejvind. 1992. Influence of colonizing shrubs on species-area relationships in alvar plant communities. *Journal of Vegetation Science* 3:625-630.

Romme, W.H., M.G. Turner, L.L. Wallace, and J.S. Walker. 1995. Aspen, elk, and fire in northern Yellowstone National Park. *Ecology* 76:2097-2106.

Shmida, A. 1984. Whittaker's plant diversity sampling method. *Israel Journal of Botany* 33:41-46.

Simonson, S. 1998. Rapid Assessment of butterfly diversity: a method for landscape evaluation. M.S. Thesis. Fort Collins, CO: Dept. of Bioagricultural Sciences and Pest Management, Colorado State University. 50p.

Simonson, S.E., P.A. Opler, T.J. Stohlgren, and G.W. Chong. 2001. Rapid assessment of butterfly diversity in a montane landscape. *Biodiversity and Conservation* 10:1369-1386.

Spackman, S., B. Jennings, J. Coles, C. Dawson, M. Minton, A. Kratz, and C. Spurrier. 1997. *Colorado Rare Plant Field Guide*. Fort Collins, CO: Colorado Natural Heritage Program.

SPSS, Inc. 1998. SYSTAT version 8. Chicago, IL: SPSS Inc.

Stohlgren, T.J., D. Binkley, G.W. Chong, M.A. Kalkhan, L.D. Schell, K.A. Bull, Y. Otsuki, G. Newman, M. Bashkin, and Y. Son. 1999a. Exotic plant species invade hot spots of native plant diversity. *Ecological Monographs* 69:25-46.

Stohlgren, T.J., K.A. Bull, Y. Otsuki. 1998a. Comparison of rangeland vegetation sampling techniques in the Central Grasslands. *Journal of Range Management* 51:164-172.

Stohlgren, T.J., K.A. Bull, Y. Otsuki, C.A. Villa, and M. Lee. 1998b. Riparian zones as havens for exotic plant species in the central grasslands. *Plant Ecology* 138:113-125.

Stohlgren, T.J., G.W. Chong, M.A. Kalkhan, and L.D. Schell. 1997a. Multi-scale sampling of plant diversity: effects of minimum mapping unit size. *Ecological Applications* 7:1064-1074.

Stohlgren, T.J., G.W. Chong, M.A. Kalkhan, and L.D. Schell. 1997b. Rapid assessment of plant diversity patterns: a methodology for landscapes. *Environmental Monitoring and Assessment* 48:25-43.

Stohlgren, T.J., G.W. Chong, L.D. Schell, K.A. Rimar, Y. Otsuki, and C. Villa. 1999b. Assessing vulnerability to invasion by exotic plant species at multiple spatial scales. *Biological Invasions* 29:566-577.

Stohlgren, T.J., M.B. Falkner, and L.D. Schell. 1995. A Modified-Whittaker nested vegetation sampling method. *Vegetatio* 4:1-8.

Stohlgren, T.J., Y. Otsuki, C.A. Villa, M. Lee, and J. Belnap. 2001. Exotic Plant Invasions in Southeast Grand Staircase-Escalante National Monument, Utah. *Biological Invasions* 3:37-50.

Suzuki, K., H. Suzuki, D. Binkley, and T.J. Stohlgren. 1999. Aspen regeneration in the Colorado Front Range: differences at local and landscape scales. *Landscape Ecology* 14:231-237.

USDA NRCS. 1999. The PLANTS database (http://plants.usda.gov/plants). Baton Rouge, LA: National Plant Data Center.

Weinstein, J. 1979. The condition and trend of aspen along the Pacific Creek trail in Grand Teton National Park. Pages 79-82 in M.S. Boyce and L.D. Hayden-Wing (eds): *North American Elk: Ecology, Behavior, and Management*. Laramie, WY: University of Wyoming.

Whitson, T.D., L.C. Burrill, S.A. Dewey, D.W. Cudney, B.E. Nelson, R.D. Lee, and R. Parker. 1996. *Weeds of the West*. 5th ed. Laramie, WY: University of Wyoming. 630p.

Wilcove, D.S., D. Rothstein, J. Dubow, A. Phillips, and E. Losos. 1998. Quantifying threats to imperiled species in the United States. *BioScience* 48:607-615.

CHAPTER III: NEW APPROACHES FOR SAMPLING AND MODELING NATIVE AND NON-NATIVE PLANT SPECIES RICHNESS

3.0 Abstract

A new multi-phase, multi-scale approach is presented for sampling and modeling native and non-native plant species to predict the spread of invasive species and aid in control efforts. The test site was a 54,000-ha portion of Rocky Mountain National Park, Colorado, USA. This work is based on previous research where vegetation sampling techniques to identify hotspots of diversity, important rare habitats, and locations of invasive plant species were developed. Statistical modeling tools were used to assess current patterns of native and non-native plant species to determine which habitats are most vulnerable to invasion by non-native species. I used stepwise multiple regression and modified residual kriging to estimate the numbers of native species and non-native species and the probability of observing a non-native species in 30 m x 30 m cells. The final models accounted for 62% of the variability observed in the number of native species, 51% of the variability observed in the number of non-native species, and 47% of the variability associated with observing a non-native species. Important independent variables used in developing the models included

geographical location, elevation, slope, aspect, and Landsat TM bands 1-7. These models can help resource managers prioritize areas for further inventory, monitoring, and non-native species control efforts.

3.1 Introduction

Habitats with high native plant species diversity can be more vulnerable to non-native plant species invasions than less species-rich areas (Lonsdale 1999, Stohlgren et al. 1998b, 1999a, 2002). To efficiently address the threats posed by non-native species to native biodiversity, non-native plant species must be detected early. However, once an invasion has occurred, those populations and surrounding areas must be monitored to provide resource managers with the information needed to contain and control the non-native species.

Only a small portion of any landscape can be affordably measured (usually <1%), so predicting species occurrences or other features over the remainder of the landscape requires accurate multi-scale techniques (Stohlgren et al. 1997d). Most native and non-native plant species, rare habitats, and hot spots of diversity are patchy on most landscapes, so they are usually missed by single-phase, single-scale transects and small plots (Stohlgren et al. 1998a, Palmer et al. 2002). In addition, many inventory and monitoring attempts are hampered by unknown sources and amounts of error. For example, inventories based on resource maps must include an assessment of what information would be gained by using a map of higher resolution (smaller minimum mapping unit; Stohlgren et al. 1997b) as many coarse-scale maps fail to recognize rare but important habitats (Palmer et al. 2002).

The objective of this paper is to describe a sampling and spatial modeling approach that can provide resource managers with a clearer picture of what areas and habitats are vulnerable to invasion by non-native plant species. This information can improve resource management decisions for control of non-native species and the inventory and monitoring of native and non-native plant species.

3.1.1 A Multi-Species, Multi-Scale, Multi-Phase Approach

The importance of conserving biological diversity is recognized worldwide, and recognition of the benefits of conserving biodiversity at the ecosystem level, rather than the individual species level (Noss 1983, Agee and Johnson 1988, LaRoe 1993), has resulted in the need to identify areas for protection based on their biodiversity. Stohlgren et al. (1997d) outlined an approach to a landscape-scale assessment of plant diversity to complement the National GAP Analysis Program (Scott et al. 1993). The approach recognizes that the resolution of investigation is the first source of potential error in identifying important habitats for conservation. Common minimum mapping units (MMU's) of 100 ha, 50 ha, and 2 ha failed to identify rare but important habitats such as aspen stands in the Beaver Meadows area of Rocky Mountain National Park, Colorado (Stohlgren et al. 1997b). Aspen stands in this area have unique and rich assemblages of plants (Stohlgren et al. 1997c), birds (T. Mabee, *personal* *communication*), and butterflies (Simonson 1998). A key feature of this approach is to sample rare and common habitats with a stratified random sampling design. Unbiased vegetation sampling sites are selected in each stratum (Stohlgren et al. 1997b).

A second source of error in assessing patterns of biodiversity results from single-scale sampling techniques. A comparison of several common sampling techniques demonstrated that small, single-scale plot and linear transect techniques missed many locally rare species- both native and non-native (Stohlgren et al. 1998a). In addition, it is hard to extrapolate to larger areas from single-scale data. The Modified-Whittaker nested vegetation sampling plot (Figure 1-2) provides a framework for collection of multi-scale data, which allow one to estimate the relative contributions of vegetation types to landscape-scale biodiversity (species richness) (Palmer 1990).

Stohlgren et al. (1997c) tested a rapid biodiversity assessment using multi-phase, multi-scale sampling in the Beaver Meadows area mentioned above. Multi-phase sampling refers to using ground-truth plots (Modified-Whittaker), aerial photos, and satellite images to sample some characteristic, such as vegetation cover, at overlapping locations (Kalkhan et al. 1995). These multiple layers of data allow the assessment of the accuracy of the satellite image and air photo vegetation classifications, and classifications can be improved from the multiple layers of data (Kalkhan et al. 1998). The multi-scale, Modified-Whittaker vegetation plot sampling design allowed identification of hot spots of biodiversity and a reasonable estimate of the total number of plant species expected to be found in a relatively small (750 ha) study area.

3.1.2 Data Comparability, Analysis, and Synthesis

In addition to identifying where species of interest and hot spots of diversity occur, multiple threats to native species diversity must be recognized so that appropriate management strategies can be developed. Using comparable sampling methods allows both local and regional analyses and monitoring of species diversity, for example, across management units. Sampling designs and methods must be able to accurately assess the effects of a particular management action or potential resource threat.

For example, data collected using the Modified-Whittaker plot have proven valuable for assessing impacts and outcomes. A grazing study in Rocky Mountain grasslands demonstrated that vegetation composition differences inside and outside of grazing exclosures could not be attributed to the effects of grazing alone because of landscape heterogeneity in vegetation distributions that had not been sampled in earlier studies (Stohlgren et al. 1999b). Vegetation sampling in the U.S. Central Grasslands and Rocky Mountains showed that non-native plant species are invading areas with the highest native plant species richness and cover (Stohlgren et al. 1998b, 1999a, 2002). Modified-Whittaker plots arranged along transects that cross forest ecotones in Rocky Mountain National Park, Colorado provided information on understory species richness and species distributions (Stohlgren et al. 2000) and may provide a means to monitor changes in regional climate (Stohlgren et al. 1998c).

In recognition of the strengths of multi-scale sampling, the Forest Health Monitoring Program of the U.S. Forest Service has modified their single-scale understory vegetation sampling method so that it is comparable to the multi-scale Modified-Whittaker plot (Busing et al. 1999). Grand Staircase-Escalante National Monument, Utah is using the multi-phase, multi-scale approach to inventory its vascular plant diversity and soil crust development (Stohlgren et al. 1997a). The Smithsonian Institution's Biodiversity Program has adopted the Modified-Whittaker vegetation sampling design, and they successfully used the methods in Peru's Amazon basin (Stohlgren and Chong 1997). Many other Federal and non-Federal resource managers are adopting multi-scale approaches to inventory and monitor biodiversity.

Predictive models developed from multi-scale data are an excellent example of data synthesis for resource management (Kalkhan et al. 2000). Modeling small-scale variability in landscape characteristics requires the generation of full-coverage maps depicting characteristics measured at points in the field (Reich and Bravo 1998). While many spatial datasets describing land characteristics have proven reliable for macro-scale ecological monitoring, these relatively coarse-scale data fall short in providing the precision required by more refined ecosystem resource models (Gown et al. 1994). Spatial statistics and geostatistics provide a means to develop spatial models that can be used to correlate coarse scale geographical data with multi-scale field measurements of biotic and abiotic variables (Kalkhan and Stohlgren 2000).

In summary, an inventory and monitoring approach where the resulting data are useful for many different applications at various scales was developed. In this paper, I present results from the spatial modeling of native and non-native plant species in a 54,000 ha portion of Rocky Mountain National Park as an example of this approach.

3.2 Methods

3.2.1 Field Data

To evaluate the model procedures discussed in this paper, I used Modified-Whittaker vegetation data from a 54,000-ha portion of Rocky Mountain National Park, Colorado, USA (Figure 1-1). Ninety-four 1000-m^2 plots had all the information required for the spatial modeling. Sample points were located based on stratified random sampling in 11 vegetation cover types: alpine tundra, aspen, Douglas fir, dry meadow, lodgepole, mixed conifer, ponderosa pine, spruce/fir, subalpine forest, wet meadow, and willow (the procedure is described in Stohlgren et al. 1997c). Point locations were recorded using the global positioning system (GPS), and coordinates were taken using the universal transverse mercator (UTM) system (provides (x,y) coordinates in meters from a regional reference point). This dataset was used to develop spatial models to predict species richness (native and non-native) and the presence (or absence) of non-native species in 30 m x 30 m cells.

3.2.2 GIS Data

The GIS database used to develop the models contained several coverages of independent variables thought to influence the variability in species richness and the presence of non-native species. These included a 30-m resolution Digital Elevation Model (DEM; Department of Interior, U.S. Geological Survey), which was used to create a 30-m grid overlay of percent slope and aspect (GRID, ArcInfo, ESRI 1997).

The database also included 30-m resolution overlays of Landsat TM bands 1 through 7. For each point that was sampled in the field with a Modified-Whittaker plot, the elevation, slope, aspect, and the digital numbers associated with the seven Landsat bands were extracted. Plant species data analyzed were from the 10 1-m^2 subplots in each of the 94 Modified-Whittaker plots.

3.2.3 Geostatistical Analysis

Multiple regression analysis was first used to explore the variation in species richness and the presence of non-native species as a function of the geographical location, elevation, slope, aspect, and Landsat TM bands 1-7 (Figure 3-1). Forward stepwise regression was used to identify the best linear combination of independent variables (geographical location, elevation, slope, aspect and Landsat TM reflectance) for describing the variables of interest (numbers of native and non-native plant species and the presence of at least one non-native plant species).

Residuals of the regression models were computed and used for modeling their semivariograms to determine the effects of distance between points on variation between points (Reich 1999). Model parameters were estimated using weighted least squares (Cressie 1985). The residuals were also analyzed for spatial autocorrelation and cross-correlation (Czaplewski and Reich 1993, Reich et al. 1994, Bonham et al. 1995) with the geographical variables. Because there was no cross-correlation, ordinary kriging using the four nearest neighbors was used to model the residuals. Inverse distance sampling was used to define the spatial weights matrix, which is used to define the relationships between adjacent points (Reich 1999).

Estimates of species richness (native and non-native) and the presence of non-native species were obtained by adding the regression estimates based on elevation, slope, aspect, etc. and the estimated residuals computed using ordinary kriging. The modified residuals kriging models were cross-validated (Guisan and Zimmerman 2000) to assess the variability in the prediction errors. Cross-validation included deleting one observation from the data set and predicting the deleted observation using the remaining observations in the data set. This process was repeated for all observations in the data set. Summary statistics of the estimated values were computed. The accuracies of the kriging models were assessed using the relative mean squared error suggested by Havesi et al. (1992).

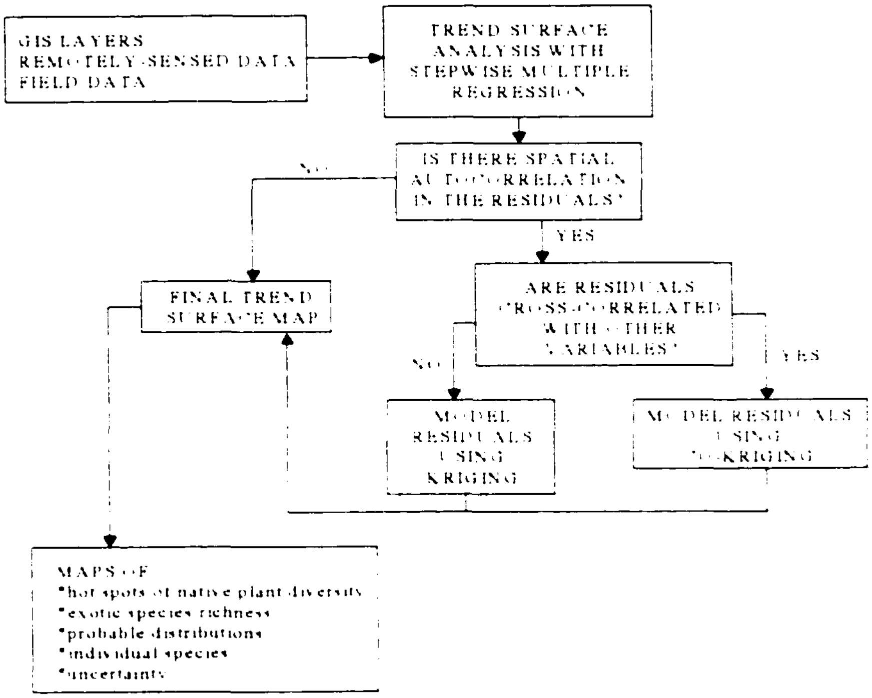

Figure 3-1. Flow diagram of statistical procedures.

3.3.4 Spatial Integration

The ability to spatially model field data allows integration over any specified geographical region (i.e., point- and plot-level field data, management unit, watershed, region) to obtain a point estimate and associated standard error of prediction. This is accomplished by integrating the three dimensional response surface representing the variable of interest over the area of interest and dividing by the area. Since the spatially modeled response surfaces can be represented as a grid in ARC/INFO (ESRI 1997), any specified region will contain a finite number (n) of grid cells of uniform size (i.e., 30 m x 30 m). The point estimate of a resource in some bounded region A, is obtained by summing the point estimates associated with each cell, Φ_i, and dividing by the number of cells in the bounded region. It is also possible to obtain estimates of the variance. Resource managers can use this information to determine what areas warrant further field data collection to increase model accuracy.

3.4 Results

3.4.1 Regression Models

The regression model developed to estimate the number of native species in a 30 m x 30 m cell (describes the large-scale spatial variability in the number of native species over the 54,000 ha area) included the geographical location (UTM coordinates), elevation and Landsat bands 1, 3, 5, 6, and 7. The model described 21% of the variance with a standard error of 4.2 (n = 940 1-m^2 subplots). The positive correlation between elevation and the number of native species (coefficient = 0.01, $p < 0.05$) suggests that species richness increases with increasing elevation. The number of native species was higher in the northern and western portions of the study area as shown by the negative correlation with the UTM easting (describes location in meters on an east-west axis; coefficient = -0.0002, $p < 0.05$), and the positive correlation with the UTM northing (describes location in meters on a north-south axis; coefficient = 0.0001, $p < 0.05$). The significant Landsat TM bands provide information about differences in the vegetation and soils throughout the study area based on reflectance values (Jensen 1996).

The regression model to estimate the number of non-native species in a 30 m x 30 m cell included the geographical location, elevation, slope, aspect, the number of native species, and Landsat bands 2, 3, 5, 6, and 7. The model explained 31% of the variance (standard error = 0.8, n = 940 1-m^2 subplots). Non-native species were more prevalent in the eastern (UTM easting coefficient = 0.00003) and southern (UTM northing coefficient = -0.00002) portions of the study area and at lower elevations (elevation coefficient = -0.0011; all $p < 0.05$). A positive correlation with slope and a negative correlation with aspect indicate that non-native species are more prevalent on steeper and more northerly exposures. A positive correlation with the number of native species indicates that the non-native species are invading areas with high native plant species richness. This result agrees with the findings of Stohlgren et al. (1998b, 1999a, 2002), Lonsdale 1999, and Kalkhan and Stohlgren (2000).

The regression model to estimate the presence of non-native species in a 30 m x 30 m cell was similar to the one developed for the number of non-native species. The factors that influenced the number of non-native species also influenced the probability of observing a non-native species: geographical location, Landsat bands 2, 3, 5, and 7, slope, aspect, elevation, and the number of native species ($R^2 = 0.38$, standard error = 0.4, n = 940 1-m^2 subplots). The directions of the relationships between the independent variables and the presence of a non-native species (signs of the coefficients) were the same as those for estimating the number of non-native species.

3.4.2 Kriging

Semivariograms are plots that describe how spatial continuity of a variable changes as a function of distance and direction (variation is plotted on the y-axis against distance on the x-axis; Reich 1999). Generally, as the distance between two points increases the variation will increase to a maximum and then reach a plateau (that distance is the range of variation, and the plateau is the sill). Usually, because of sampling error and real variability, the variation is greater than zero even at the smallest distances between points (the origin of the graph), and that is the nugget effect.

The semivariogram models used to describe the spatial correlation of the residuals were Gaussian for the number of non-native species and the presence of a non-native species and exponential for the number of native species. The large range associated with the residuals for the native species (327.0×10^4) suggests the presence of large-scale spatial continuity in the number of native species across the study area. In contrast, the small range associated with the non-native species models (number of non-native species range = 151 and presence of a non-native species range = 215) indicates that the non-native species occur in patches throughout the study area. The large nugget effect relative to the sill for the two non-native species models (number of non-native species nugget = 0.76, sill = 1.12 and presence of a non-native species nugget = 0.08, sill = 0.19) also suggests that there is a considerable amount of variation among these patches.

Ordinary kriging was used to refine the models because the residuals of the regression models were positively spatially autocorrelated ($\alpha = 0.05$), but there was no significant cross-correlation observed between the residuals and the independent variables used in developing the models. The residuals were approximately normally distributed.

The kriging approach improved the estimates of the numbers of native and non-native species and the presence of non-native species (Table 3-1). The modified residual kriging model for the number of native species had a relative mean squared error of 8.4 native species ($R^2 = 0.63$) while, in comparison, the regression model had a relative mean squared error of 17.8 native species ($R^2 = 0.21$). Kriging the residuals reduced the relative mean squared error by 53%. The modified residual kriging model for the number of non-native species had a relative mean squared error of 0.50 ($R^2 = 0.51$). This represents a reduction in relative mean squared error of 29%. Similar mean squared errors were observed for the probability model of non-native species (mean squared error = 0.12, R^2 = 0.47).

Table 3-1. Relative mean square errors and explanations of variance (R^2) associated with estimating the number of native and non-native species and the probability of observing a non-native species in a 30 m x 30 m cell.

Regression Model	Estimation Technique	Relative Mean Squared Error	R^2
# Native	Regression	17.8	0.21
Species	Kriging	8.4	0.63
# Non-native	Regression	0.70	0.31
Species	Kriging	0.50	0.51
Presence	Regression	0.14	0.38
Non-native	Kriging	0.12	0.47

3.4 Discussion

I have outlined a comprehensive approach to sampling and modeling native and non-native plant species for natural resources management. The approach provides an alternative to individual-based reaction-diffusion and spatially explicit simulation models and their assumptions and limitations (see Higgins et al. 1996). For example, the use of full-coverage, fine-scale variables (e.g., Landsat TM data with a 30 m x 30 m pixel resolution) is a valuable addition to spatial modeling and addresses problems relating to the lack of empirical data and inappropriate scales that affect the previously mentioned types of models. Also, the approach is based on current species locations, so no direct assumptions are made about dispersal or autecology (see Kot et al. 1996).

The combination of spatial statistics and stepwise multiple regressions greatly increased the predictive capabilities of the models for estimating the numbers of native and non-native species and the probability of encountering a non-native species in 30 m x 30 m cells. This improvement occurred even though the sample area was large (54,000 ha) relative to the number of sample points (94 1000-m^2 plots, each with 10 1-m^2 subplots).

The models could be further improved with more sample points in under-sampled areas and with the addition of more full-coverage data, such as soil characteristics (Kalkhan and Stohlgren 2000). The under-sampled areas could be identified with maps of "uncertainty" based on sub-sampling the data with Monte-Carlo simulations (Kalkhan et el. 2000). Uncertainty maps could also provide land managers with a spatial representation of the confidence of the model and the completeness of the plot data. The multi-phase sampling approach (i.e., data from ground-truth plots, air photos, and Landsat TM images) provides additional ways to assess vegetation classification accuracy and where more ground-truth plots are needed (Kalkhan et al. 1998).

The semivariograms related to non-native species indicate that the current models are not optimal because the proportion of the nugget (essentially error) to the sill (the greatest variation measured) is very high (68% for the number of non-native species, 43% for the presence of a non-native species, and, for reference, 1% for the number of native species). The larger errors (nugget values) associated with the non-native species models are due primarily to the small-scale spatial heterogeneity associated with the occurrence and density of non-native species. This small-scale spatial heterogeneity makes it difficult to predict the spatial variability in the presence/absence or number of non-native species at the 30 m x 30 m cell size or the 1000-m^2 plot size. To improve the models related to non-native species in particular, sampling schemes could be developed to increase the probability of encountering non-native species in sample plots. For example, sampling could be stratified based on areas known to concentrate non-native species such as roads (Greenberg et al. 1997) or riparian corridors (Stohlgren et al. 1998b).

Because this approach provides estimates of species richness (native and non-native) on a pixel by pixel scale, where each cell is associated with a specific UTM location, resource managers may be able to use this information to set priorities for searching for non-native plant species and for non-native control efforts.

Similar models could be developed for individual species in more restricted areas (with a greater density of sample points) to better understand their ecology (where they are able to occur) and patterns of spread. Spatial models for common invasive plant species could help managers in Rocky Mountain National Park, for example, to better understand the effects of grazing and browsing, natural and prescribed fire, and rapid climate change on invasive plant species.

3.5 Literature Cited

Agee, J.K. and D.R. Johnson. 1988. *Ecosystem management for parks and wilderness.* University of Washington Press, Seattle, WA. 237 pp.

Bonham, C.D., R.M. Reich, and K.K. Leader. 1995. A spatial cross-correlation of *Bouteloua gracilis* with site factors. *Grasslands Science* 41:196-201.

Busing, R., K. Rimar, K.W. Stolte, and T.J. Stohlgren. 1999. Forest Health Monitoring: Vegetation Pilot Field Methods Guide. USDA Forest Service, P.O. Box 96090, Washington, DC 20090. 55 pp.

Cressie, N. 1985. Fitting variogram models by weighted least squares. *Mathematical Geology* 17:563-586.

Czaplewski, R.L. and R.M. Reich. 1993. Expected value and variance of Moran's bivariate spatial autocorrelation statistic under permutation. Research Paper RM-309. U.S. Department of Agriculture, Rocky Mountain Experimental Range Station, Fort Collins, CO.

ESRI. 1997. Environmental Systems Research Institute, Inc. 380 New York St., Redlands, CA 97373 USA.

Gown, S.N., R.H. Waring, D.G. Dye, and J. Yang. 1994. Ecological remote sensing at OTTER: Satellite Macroscale Observation. *Ecological Applications* 4:322-343.

Greenberg, C.H., S.H. Crownover, and D.R. Gordon. 1997. Roadside soils: a corridor for invasion of xeric scrub by nonindigenous plants. *Natural Areas Journal* 17:99-109.

Guisan, A. and N.E. Zimmerman. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147-186.

Havesi, J.A., J.D. Istok, and A.L. Flint. 1992. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: structural analysis. *Journal of Applied*

Meterology 3:661-676.

Higgins, S.I., D.M. Richardson, and R.M. Cowling. 1996. Modeling invasive plant spread: the role of plant-environment interactions and model structure. *Ecology* 77:2043-2054.

Jensen, J.R. 1996. *Introductory digital image processing: a remote sensing perspective*. 2nd edition. Prentice-Hall, Inc., Upper Saddle River, NJ. 316 pp.

Kalkhan, M.A., R.M. Reich, and T.J. Stohlgren. 1998. Assessing the accuracy of Landsat Thematic Mapper classification using double sampling. *International Journal of Remote Sensing* 19:2049-2060.

Kalkhan, M.A. and T.J. Stohlgren. 2000. Using multi-scale sampling and spatial cross-correlation to investigate patterns of plant species richness. *Environmental Monitoring and Assessment* 64:591-605.

Kalkhan, M.A., T.J. Stohlgren, and M.B. Coughenour. 1995. An investigation of biodiversity and landscape-scale gap patterns using double sampling: a GIS approach. Proceedings of the Ninth Conference on Geographic Information Systems, Vancouver, BC, Canada.

Kalkhan, M.A., T.J. Stohlgren, and Y. Otsuki. 2000. Landscape-scale assessment of plant diversity: integration of spatial information and spatial statistics. Page 130 *in* the Abstracts of the Ecological Society of America 85th Annual Meeting, August 6-10, 2000, Snowbird, Utah.

Kot, M., M.A. Lewis, P. van den Driessche. 1996. Dispersal data and the spread of invading organisms. *Ecology* 77:2027-2042.

LaRoe, E.T. 1993. Implementation of an ecosystem approach to endangered species conservation. *Endangered Species Update* 10:3-6.

Lonsdale, W.M. 1999. Global patterns of plant invasions and the concept of invasibility. *Ecology* 80:1522-1536.

Noss, R. 1983. A regional landscape approach to maintain diversity. *BioScience* 33:700-706.

Palmer, M.W. 1990. The estimation of species richness by extrapolation. *Ecology* 71:1195-1198.

Palmer, M.W., P.G. Earls, B.W. Hoagland, P.S. White, and T. Wohlgemuth. 2002. Quantitative tools for perfecting species lists. *Environmetrics* 13:121-137.

Reich, R.M. 1999. Spatial statistical modeling of ecosystem resources and the environment. Department of Forest Sciences, Colorado State University, Fort Collins, CO 80523. 285pp.

Reich, R.M. and V.A. Bravo. 1998. Integrating spatial statistics with GIS and remote sensing in designing multi-resource inventories. Proceedings of the North American Symposium Toward a Unified Framework for Inventorying and Monitoring Forest Ecosystem Resources, Guadalajara, Mexico.

Reich, R.M., R.L. Czaplewski, and W.A. Bechtold. 1994. Spatial cross-correlation in growth of undisturbed natural shortleaf pine stands in northern Georgia. *Journal of Environmental and Ecological Statistics* 1:201-217.

Scott, M.J., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Erchia, T.C. Edwards, J. Ulliman, and R.G. Wright. 1993. GAP analysis: a geographic approach to protection of biological diversity. *Wildlife Monographs* 123:1-41.

Simonson, S. 1998. Rapid Assessment of Butterfly Diversity: a Method for Landscape Evaluation. Masters thesis, Dept. of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO.

Stohlgren, T.J., J. Belnap, G.W. Chong, and R. Reich. 1997a. A Plan to Assess Native and Exotic Plant Diversity and Cryptobiotic Crusts in the Grand Staircase-Escalante National Monument. Proceedings of Learning from the Land: Grand Staircase-Escalante Science Symposium, Southern Utah University, Cedar City, UT.

Stohlgren, T.J., D. Binkley, G.W. Chong, M.A. Kalkhan, L.D. Schell, K.A. Bull, Y. Otsuki, G. Newman, M. Bashkin, and Y. Son. 1999a. Exotic Plant Species Invade Hot Spots of Native Plant Diversity. *Ecological Monographs* 69:25-46.

Stohlgren, T.J., K.A. Bull, and Y. Otsuki. 1998a. Comparison of rangeland vegetation sampling techniques in the central grasslands. *Journal of Range Management* 51:164-172.

Stohlgren, T.J., K.A. Bull, Y. Otsuki, C.A. Villa, and M. Lee. 1998b. Riparian zones as havens for exotic plant species in the central grasslands. *Plant Ecology* 138:113-125.

Stohlgren, T.J., T.N. Chase, R.A. Pielke, Sr., T.G. Kittel, and J.S. Baron. 1998c. Evidence that local land use practices influence regional climate, vegetation, and stream flow patterns in adjacent natural areas. *Global Change Biology* 4:495-504.

Stohlgren, T.J. and G.W. Chong. 1997. Assessment of Biological Diversity and Long-Term Monitoring Plan for the Lower Urubamba Region. Pages 41-44 *in* F. Dallmeier and A. Alonso, editors, Biodiversity Assessment and Monitoring of the Lower Urubamba Region, Peru. SI/MAB Series 1, Smithsonian Institution, Washington, DC.

Stohlgren, T.J., G.W. Chong, M.A. Kalkhan, and L.D. Schell. 1997b. Multi-scale sampling of plant diversity: effects of minimum mapping unit size. *Ecological Applications* 7:1064-1074.

Stohlgren, T.J., G.W. Chong, M.A. Kalkhan, and L.D. Schell. 1997c. Rapid assessment of plant diversity patterns: a methodology for landscapes. *Environmental Monitoring and Assessment* 48:25-43

Stohlgren, T.J., G.W. Chong, L.D. Schell, K.A. Rimar, Y. Otsuki, M. Lee, M.A. Kalkhan, and C.A. Villa. 2002. Assessing vulnerability to invasion by nonnative plant species at multiple spatial scales. *Environmental Management* 29:566-577.

Stohlgren, T.J., M.B. Coughenour, G.W. Chong, D. Binkley, M.A. Kalkhan, L.D. Schell, D.J. Buckley, and J.K. Berry. 1997d. Landscape analysis of plant diversity. *Landscape Ecology* 12:155-170.

Stohlgren, T.J., M.B. Falkner, and L.D. Schell. 1995. A Modified-Whittaker nested vegetation sampling method. *Vegetatio* 4:1-8.

Stohlgren, T.J., A.J. Owen, and M. Lee. 2000. Monitoring Shifts in Plant Diversity in Response to Climate Change: a Method for Landscapes. *Biodiversity and Conservation* 9:65-86.

Stohlgren, T.J., L.D. Schell, and B. Vanden Heuvel. 1999b. How Grazing and Soil Quality Affect Native and Exotic Plant Diversity in Rocky Mountain Grasslands. *Ecological Applications* 9:45-64.

CHAPTER IV: SPATIAL MODELS COMBINING TREND SURFACE AND REGRESSION TREE ANALYSES TO PREDICT NUMBERS AND COVER OF NATIVE AND NON-NATIVE PLANT SPECIES

4.0 Abstract

I compared the abilities of trend surface analysis alone and in combination with regression tree analysis to develop full-coverage estimate surfaces of the cover and richness of total, native, and non-native plant species in 30 m x 30 m cells in an 80,000-ha portion of Rocky Mountain National Park, Colorado, USA. The combination of approaches consistently outperformed the trend surface analysis alone with increases in variance explained (R^2) from: 7% to 51% for the total number of plant species and 2% to 59% for the number of native plant species. Mean squared errors (MSE) decreased with the combination of approaches from: 234.5 to 124.7 for the total number of plant species and 157.8 to 65.7 for the number of native plant species. The combined models generally used a larger number of independent variables. Future applications of this approach could be useful for developing the concept of "scalable" models, where the area modeled is adjusted to produce the best model given existing data. These models could then be used to estimate variables of interest in the unsampled area to direct

future sampling efforts (e.g., sampling in areas estimated to be species-rich because these areas may be more prone to invasion by non-native plant species).

4.1 Introduction

Land managers are consistently challenged to describe the biotic and abiotic resources of their management units, but they rarely have funding for adequate inventory sampling let alone funding for follow-up monitoring sampling, data analysis and data management (Stohlgren et al. 1995b). Current large-scale, national efforts (e.g., USDA Forest Service national Forest Health Monitoring; Bull et al. 1998) concentrate on systematic sampling of a large number of plots to achieve statistical power in conventional, parametric analyses (Stohlgren et al. In Preparation). However, the results gleaned from this type of data are limited to a coarse, summary scale (U.S. Forest Service 1999).

Ideally, one would like to sample an area quickly and cheaply and then be able to develop predictive, spatial models that estimate the variables of interest in unsampled areas with known accuracy. I describe a system for field sampling and subsequent data analysis that combines trend surface and regression tree analyses to provide vegetation information for a large area with a relatively small number of sample points. The field sampling employs a stratified, random approach for plot selection (sample points are chosen based on overstory vegetation classification from air photos and satellite images; Stohlgren et al. 1997b, 1997c), and field plots are sampled at multiple scales (Stohlgren et al. 1995, 1998, Figure 1-2), which allows coarse- and fine-scale spatial modeling (i.e., regional and local scales; Agterberg 1984, Reich et al. 2002) and the development of species-area curves for non-spatial analyses (Chapters I and II).

The combination of trend surface and regression tree analyses, can, in some cases, explain significantly more variance than trend surface analysis alone (Joy and Reich 2002a, b), and, because spatial autocorrelation is not required as it is for kriging (e.g., Chong et al. 2001), many fewer sample points are required (i.e., points can be more widely dispersed on the landscape). Although larger sample sizes may be recommended for dealing with data sets that have non-linear interactions (e.g., > 200 plots, Joy and Reich 2002a), decision trees can be used to describe non-linear interactions, and the reasons for the splits can be examined for easy understanding (Figure 4-1).

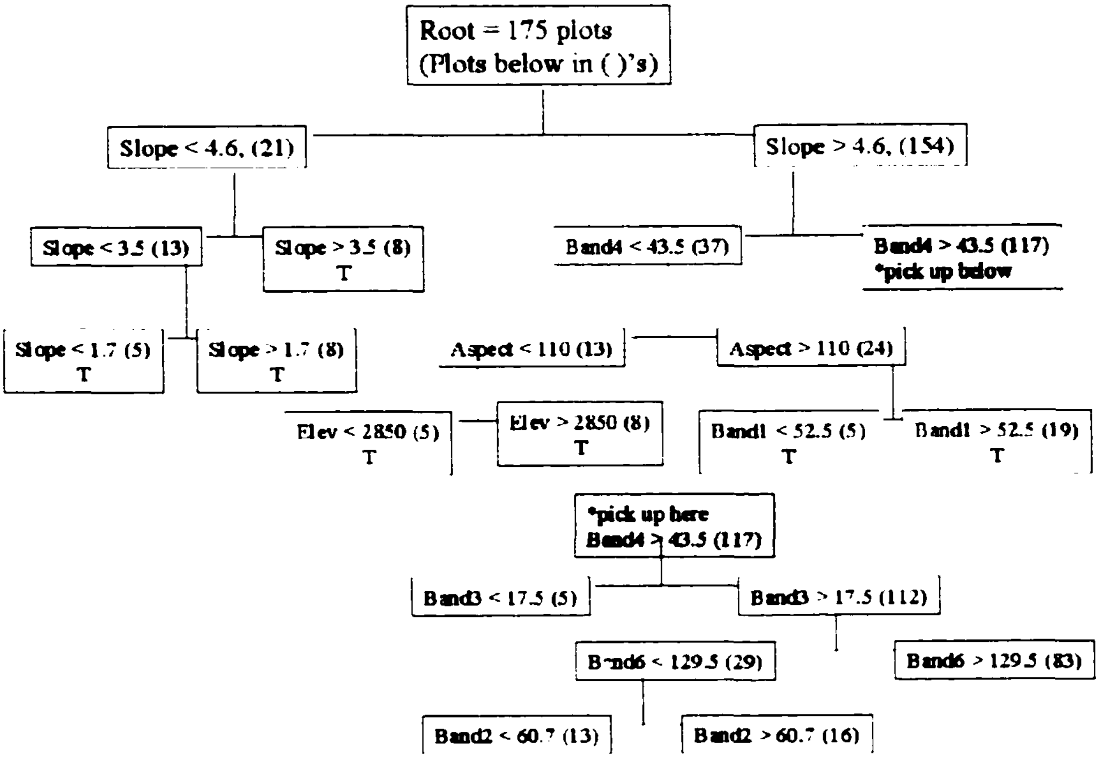

Figure 4-1. A portion of the decision tree, which shows the criteria used to separate pixels (30 m x 30 m) into different classes to estimate the total plant cover estimated to occur on that pixel based on the sample points. The full tree has 19 terminal nodes, which are noted with a "T" in this figure.

Rocky Mountain National Park, Colorado, USA, provides a challenging landscape for testing these methods. Although the vegetation overstory can be described relatively simply by six dominant tree species, the vegetation species diversity lies in the understory, and it is greatly influenced by complex topography and subsequent understory habitat variation. Small vegetation patches are difficult to detect from remotely sensed data (e.g., wet meadows, riparian zones, aspen patches; Chapter II, Stohlgren et al. 1997). In addition, non-native, invasive plant species are a concern to managers, and these species are predominantly herbaceous and found in the forest understory. I sampled 175 plots in 20 vegetation types in an 80,000-ha portion of the Park with the objective of identifying patterns in native and non-native plant species richness and cover to provide managers with information useful for native biodiversity conservation and for non-native plant species control.

The objective of this paper is to describe the use of a combination of trend surface and regression tree analyses to estimate the richness and cover of total, native, and non-native plant species in 30 m x 30 m cells in an 80,000 ha portion of Rocky Mountain National Park.

4.2 Methods

4.2.1 Study Area

Rocky Mountain National Park covers approximately 107,500 ha at elevations ranging from approximately 2,300 m to over 4,200 m above sea level and straddles the continental divide of the Rocky Mountains in north-central Colorado, USA. Dominant tree species include, generally from low to high elevation, ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), Douglas fir (*Pseudotsuga menziesii*), Engelmann spruce (*Picea engelmannii*), subalpine fir (*Abies lasiocarpa*), and limber pine (*Pinus flexilis*) (Peet 1988). Aspen (*Populus tremuloides*) stands and stems are scattered throughout open areas and forests (Kaye and Stohlgren 2002); more open vegetation types include wet and dry meadow areas and alpine tundra (Mohammed A. Kalkhan, personal communication); and riparian areas contain willows (*Salix* spp.) and alder (*Alnus incana*) (Stohlgren et al., unpublished data).

4.2.2 Vegetation Sampling

I sampled 175 Modified-Whittaker, multi-scale plots (Stohlgren et al. 1995, 1998) between 1995 and 1999 in the central 80,000-ha portion of Rocky Mountain National Park. Plot placement was random within 20 vegetation types (Table 4-1) that were identified on natural color air photos (1:15840 scale, taken September 1987;

Mohammed A. Kalkhan, personal communication). The general approach is described in Stohlgren et al. (1997b). All but four of the alpine tundra plots were one quarter the size of the standard Modified-Whittaker plot and thus covered 250 m^2 as opposed to the usual 1000 m^2. This size was adopted because significantly more species were not encountered as the plot size was increased in the alpine tundra (Ingolf Kuhn, personal communication).

Plots were located in the field using U.S. Geological Survey 7.5" quadrangle maps, aerial photos, compasses and the geographic positioning system (GPS). All plot locations were documented with a GPS reading at the lower right corner (Figure 1-2, the origin of the plot where both meter tapes that define the plot are anchored), drawings and photos, and by marking them on the topographic maps.

Table 4-1. Vegetation types included in the trend surface and regression tree analyses. The area sampled is in square meters.

Vegetation Type	Number of Plots	Area Sampled
Alpine Meadow	1	1,000
Alpine Tundra	33	11,250
Aspen	30	30,000
Douglas Fir	6	6,000
Douglas Fir/Lodgepole Pine*	2	2,000
Dry Meadow	9	9,000
Limber Pine	3	3,000
Lodgepole Pine	23	23,000
Lodgepole Pine/Limber Pine*	3	3,000
Lodgepole Pine/Ponderosa	3	3,000
Lodgepole Pine/Spruce/Fir*	2	2,000
Mixed Conifer	4	4,000
Ponderosa Pine	14	14,000
Ponderosa Pine/Douglas Fir*	3	3,000
Riparian	1	1,000
Rocks	1	1,000
Spruce/Douglas Fir*	8	8,000
Subalpine	10	10,000
Wet Meadow	12	12,000
Willow	9	9,000

*Ecotone study plots

In each plot, all vascular plants were identified to species, if possible. On average, 10% of the species list for each year consisted of unknowns - plants that could not be identified to species because of developmental stage. Scientific names are based on the U.S. Department of Agriculture, Natural Resource Conservation Service, Plants Database (NRCS 2001). Species are classified as native or non-native based on the Plants Database except when they have classified species as naturalized native, in which cases I classified them as non-native (these are typically widespread agricultural species such as Kentucky bluegrass; Table 4-2). Species codes are included in Table 4-2 to ease tracking of species name changes over time.

Plants that could not be identified to species are classified as unknown. Because of the relatively small number of non-native species and the resulting relative ease of identifying them, unknowns were included as native species for this analysis on the basis that field taxonomists would be able to recognize the non-native species in almost all cases.

In each 1- m^2 subplot (10 per plot) the aerial cover of each species was estimated to the nearest percent (a 10 cm x 10 cm area). Many species cover <1% of the area in a subplot, so they are assigned a cover value of 0.5% in those subplots. Cover of a species in a plot is based on the average cover from the 10, 1-m^2 subplots (Stohlgren et al. 1997b).

Table 4-2. Non-native plants found in Rocky Mountain National Park. All codes are from the Plants Database: http://plants.usda.gov/

NRCS Code	Scientific and Common Names
aggi2	*Agrostis gigantea*, redtop
alal3	*Alyssum alyssoides*, pale madwort
alge2	*Alopecurus geniculatus*, water foxtail
alpr3	*Alopecurus pratensis*, meadow foxtail
brin2	*Bromus inermis*, smooth brome
brju	*Brassica juncea*, india mustard
brte	*Bromus tectorum*, cheatgrass
caca19	*Carum carvi*, caraway
cami2	*Camelina microcarpa*, littlepod falseflax
canu4	*Carduus nutans*, nodding plumeless thistle
cefov2	*Cerastium fontanum* ssp. *vulgare*, big chickweed
ciar4	*Cirsium arvense*, canadian thistle
coar4	*Convolvulus arvensis*, field bindweed
coma2	*Conium maculatum*, poison hemlock
coor	*Conringia orientalis*, hare's ear mustard
cyda	*Cynodon dactylon*, bermudagrass
cyof	*Cynoglossum officinale*, gypsyflower
dagl	*Dactylis glomerata*, orchardgrass
deso2	*Descurainia sophia*, herb sophia
elre3	*Elytrigia repens*, creeping quackgrass
eues	*Euphorbia esula*, wolf's milk (leafy spurge)
feov	*Festuca ovina*, sheep fescue
leca5	*Lepidium campestre*, field pepperweed
lepe2	*Lepidium perfoliatum*, clasping pepperweed
lopr7	*Lolium pratense*, meadow ryegrass
mavu	*Marrubium vulgare*, horehound
meof	*Melilotus officinalis*, yellow sweetclover
phpr3	*Phleum pratense*, timothy
poav	*Polygonum aviculare*, prostrate knotweed
poco	*Poa compressa*, canada bluegrass
popr	*Poa pratensis*, kentucky bluegrass
potr2	*Poa trivialis*, rough bluegrass
ruac3	*Rumex acetosella*, common sheep sorrel
rucr	*Rumex crispus*, curly dock
sial2	*Sisymbrium altissimum*, tall tumblemustard
sivu	*Silene vulgaris*, maidenstears
taof	*Taraxacum officinale*, common dandelion
thar5	*Thlaspi arvense*, field pennycress
trdu	*Tragopogon dubius*, yellow salsify
trhy	*Trifolium hybridum*, alsike clover
trre3	*Trifolium repens*, white clover
veth	*Verbascum thapsus*, common mullein

4.2 Spatial Analysis

I compared the use of multiple regression (i.e., trend surface) analysis (OLS; Reich and Davis 1998; S-PLUS, Statistical Sciences 2000) alone and in combination with regression tree analysis (S-PLUS, Statistical Sciences 2000) to predict plant species richness and cover for native, non-native and total plant species over an 80,000-ha area. Trend surface analysis was used to examine coarse-scale variability in the species richness and cover as a function of Landsat TM reflectance values (bands 1-7), slope, aspect and elevation (Joy and Reich 2002b, Reich et al. 2002). Regression tree analysis was used to describe the fine-scale spatial variability of the species richness and cover by modeling the residuals from the trend surface analysis. Full-coverage surfaces (grids) for species richness and cover were generated from the trend surface and regression tree analyses. The final, predicted surfaces for total, native, and non-native species richness and cover resulted from summing the trend surface and regression tree grids for each (Reich et al. 2002). Kriging and co-kriging were not options for these data because the data were not spatially autocorrelated.

Tree-based models are exploratory and can uncover structure in data as an alternative to linear and additive regression models (Mathsoft, Inc. 1999). The regression tree is a non-parametric approach to regression that uses a binary partitioning algorithm to maximize the dissimilarities among groups by comparing all possible splits of independent, continuous variables. The algorithm recursively splits the data in each group until the subset is homogenous or there are fewer than 5 observations in the subset (Mathsoft, Inc. 1999).

Hansen et al. (1996) describe classification trees:

> Classification trees use a set of independent variables to predict class memberships. A tree is constructed by recursively partitioning a data set into purer, more homogenous subsets. The method uses a deviance measure, the likelihood ratio statistic, to compare all possible splits of the data to find the one split that maximizes the dissimilarity among the resulting subsets. Possible splits of each independent variable are examined, and the particular split within a particular variable that produces the largest deviance measure is chosen to partition the dependent data. Once the tree partitions the data into new subsets, entirely different relationships using other predictor variables can be defined to split the new subsets. In addition, predictor variables which have already been used higher up in the tree may be reexamined and possibly reintroduced into the tree structure. In this manner, hierarchical, nonlinear relationships within the data are derived.

I included the independent variables: elevation, slope, and aspect (derived from a digital elevation model), and Landsat Thematic Mapper (TM) bands 1-7. I compared means (observed and predicted), R^2, and mean square errors (MSE) between the trend surface model alone and the trend surface model combined with the regression tree. I also compared the independent variables used in each model.

To evaluate the performance of the models, I used a 10-fold cross-validation procedure (Efron and Tibshirani 1993, Guisan and Zimmerman 2000, Joy and Reich 2002b). The data were randomly split into 10 parts, and, for each of the 10 parts, the trend surface and regression tree models were fitted to the remaining 9 parts of the data. The fitted model was used to predict the one part of the data that was excluded from the fitting. This was done 10 times so that each observation was excluded from the model construction and its response predicted.

4.3 Results

The combined trend surface and regression tree models consistently outperformed the trend surface model alone based on the amount of variance explained and mean square errors (MSE) (Table 4-3). The combined models used more independent variables than OLS models alone except in the case of estimating exotic plant species richness (Table 4-4).

After pruning, there were 19 terminal nodes (from 29) on the tree for the total plant cover (Figure 4-1). Pruning successively removes the least important splits, where the importance is determined by a cost-complexity measure that takes into account the deviance of the subtree, the number of terminal nodes of the subtree, and a cost complexity parameter (Mathsoft, Inc. 1999, p 385). This is important because tree size is not limited during the growing process. The pruned decision tree is used to develop estimates of total plant cover for each Landsat TM pixel (a full coverage map results; Figure 4-2).

Table 4-3. Comparison of Ordinary Least Squares (OLS) regression analysis and the combination of OLS and Regression Tree (RT) analysis. Smaller mean square errors (MSE) indicate better model fit as do larger R^2 values.

	Mean obs.	OLS MSE	**OLS R^2**	OLS p-value	RT MSE	**RT R^2**	RT est. (mean)	RT est. std dev
Total plant species	44	234.5	**0.07**	0.01	124.7	**0.51**	44	12.6
Total plant cover (%)	60	904.3	**0.15**	0.00	433.5	**0.59**	59	22.9
Native plant species	37	157.8	**0.02**	0.51	65.7	**0.59**	58	39.2
Cover of natives	58	846.1	**0.12**	0.00	444.7	**0.54**	57	21.8
Non-native plant species	3	5.3	**0.43**	0.00	3.9	**0.59**	3	2.4
Cover of non-natives	3	27.8	**0.20**	0.00	19.9	**0.42**	3	4.4

Table 4-4. Independent variables used in the two different model approaches. B1 indicates Landsat TM band 1 digital values. Slope (Slp.), aspect (Asp.), and elevation (Elv.) were derived from digital elevation models.

	Model	B1	B2	B3	B4	B5	B6	B7	Slp.	Asp.	Elv.
Total plant species	OLS				O						O
	OLS + RT		T	T	T	T	T		T	T	T
Total plant cover (%)	OLS			O	O		O	O			
	OLS + RT	T			T	T			T	T	T
Native plant species	OLS			O	O		O				
	OLS + RT	T		T		T	T	T	T	T	T
Cover of natives	OLS			O	O		O				
	OLS + RT	T			T		T		T	T	T
Non-native plant species	OLS				O				O		O
	OLS + RT						T				
Cover of non-natives	OLS				O						O
	OLS + RT						T		T		T

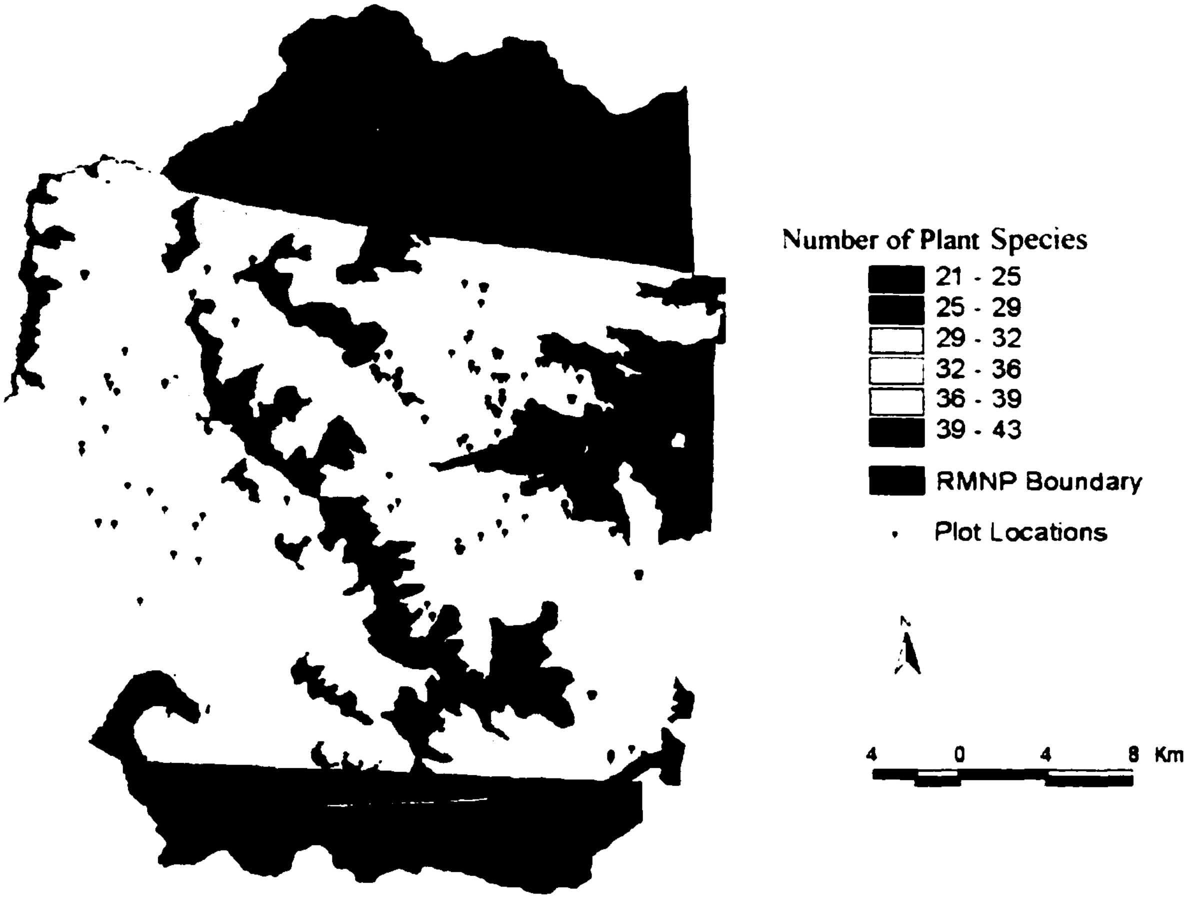

Figure 4-2. An example map of total estimated number of plant species derived from combined OLS and regression tree analyses for an 80,000 ha portion of Rocky Mountain National Park, Colorado, USA. The model used Landsat Thematic Mapper bands 2 through 6, slope, aspect, and elevation as independent variables to estimate the number of plant species in a 30 m x 30 m cell ($R^2 = 0.51$).

To test the accuracy of the map resulting from the decision tree, I used 10-fold cross validation (Efron and Tibshirani 1993, Guisan and Zimmerman 2000). For example, the mean of the observed data for total plant cover was 60% (standard deviation = 35.4), while the mean of the cross-validated predicted values was 59% cover (standard deviation = 23.6). A comparison of the mean square errors (MSE) and R^2 values of the OLS and regression tree analyses shows a general decrease in errors and an increase in R^2 values (Table 4-3).

More independent variables were used in the combined models than in the OLS models alone (Table 4-4). All but one combined model (cover of natives) used Landsat band 6, which is the thermal infrared (10.4-12.5 μm). This band is typically used for vegetation classification, vegetation stress analysis, soil moisture studies, and capturing unique information on differences in aspect in mountainous areas (Jensen 1996).

4.4 Discussion

Although the regression tree combined with ordinary regression outperformed regression alone, the models could still be improved. One way to immediately improve the model would be to reduce the area analyzed and thus decrease the number of Landsat pixels that are not correlated to any of the sampled pixels. These models would also be improved by including other independent variables that would provide more discriminatory power than simple numbers of percent cover and species richness. For example, vegetation types (categorical), a moisture index, or a soil fertility or texture index would start to address the underlying mechanisms leading to different values of native and non-native species richness and cover (and different values for some of the Landsat TM bands). However, to use such variables, one needs to have full-coverage maps for each of them (as is provided by the Landsat data). Without full-coverage maps, one is forced to model those independent variables then use those estimated values in subsequent models (Chong et al. 2001), and, each time that is done, errors are propagated within and between models.

Another way to improve these models would be to use Landsat TM data that are more synchronous with field data. Due to availability and cost a 1994 satellite image for the area was used with vegetation data predominantly collected in 1996 and 1997. Variations in moisture and season between the image and the periods when the data were collected could result in weaker correlations between field point data and reflectance values. For example, correlations could remain fairly strong for more stable reflectance signatures such as those from rock outcrops, those dominated by soil types or topography, or those dominated by perennially wet areas, but correlations would be weak for areas with vegetation that responds rapidly to available moisture. The vegetation variables of interest here are more ephemeral in many ways- especially those related to herbaceous understory where spring emergence and subsequent growth are greatly affected by seasonal moisture and temperature conditions.

A smoothing process could be used to develop a surface of average pixel values and thus reduce potential registration and vegetation type identification errors. Joy and

Reich (2002a) used this approach to resample their study area to a 10 m x 10 m resolution (the plot area they sampled on the ground). This approach could be useful with the Rocky Mountain National Park data because a significant portion of the plots (24) represented ecotones, which are areas of transition between overstory cover types. Another portion (30 plots) represented the aspen type that is often not distinguishable on satellite imagery at the 30 m x 30 m resolution. In both of these situations averaged pixel values may be more strongly correlated to the points sampled on the ground.

The application of regression trees that I have presented is an attempt at a finer scale analysis than has been attempted previously (Hansen et al. 1996; Joy and Reich 2002a, b). In previous analyses, researchers were identifying and modeling forest overstory and landcover, while in this analysis I was attempting to model numbers and cover of understory species (which are not directly measured by Landsat TM sensors). An interesting future application would be the modeling of the cover of specific species of non-native plants. In one sense, modeling only species richness and cover (native and non-native) is very difficult because a given number (or cover) of species, say 20, could occur on many plots that are in turn related to pixels that are substantially different from one another in terms of pixel reflectance values. This would result in very low correlations for the standard regressions and very low discriminatory, or splitting, power for the regression tree process.

Another possible application for the current work is identification of potential "hotspots" of species richness because these areas are more likely to be affected by non-native plant species (Lonsdale 1999; Stohlgren et al. 1999). The model could be used to direct future sampling for non-native plant species and for strengthening subsequent versions of the model (i.e., sample areas that currently have a low density of plots).

Finally, the OLS and Regression Tree modeling approach is much less computer intensive than the kriging and co-kriging approach (minutes vs. weeks for a single run). This lends the approach to scalable modeling where the researcher could conceivably fine-tune the area modeled given existing field data points to optimize the model's capabilities. The optimized model could be used to estimate the variables of interest in the area excluded from the model, and those results could be used to direct future sampling (e.g., if one modeled species richness as an indicator of likelihood of non-native species invasions, future sampling could be directed to those areas estimated to have high species richness). As increased field data become available models could be refit to larger areas and the process continued.

4.5 Literature Cited

Agterberg, F.P. 1984. Trend Surface Analysis. Pp. 147-171 in: *Spatial Statistics and Models*. Gaile, G.L. and C.J. Willmott (eds.). Reidel, Dordrecht, The Netherlands.

Bull, K.A., K.W. Stolte, et al. 1998. Forest Health Monitoring: Vegetation Pilot Field Method Guide. Washington, DC, USDA Forest Service.

Chong, G.W., R.M. Reich, M.A. Kalkhan, and T.J. Stohlgren. 2001. New approaches for sampling and modeling native and exotic plant species richness. *Western North American Naturalist* 61:328-335.

Effron, B. and R.J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York. 436 pp.

Guisan, A.G. and N.E. Zimmerman. 2000. Predictive habitat distribution models in ecology. *Ecological Modeling* 135:147-186.

Hansen, M., R. Dubayah, and R. Defries. 1996. Classification trees: an alternative to traditional land cover classifiers. *Journal of Remote Sensing* 17:1075-1081.

Jensen, J.R. 1996. *Introductory digital image processing: a remote sensing perspective*. 2nd ed. Prentice-Hall, Inc., Upper Saddle River, NJ. 316pp.

Joy, S.M. and R.M. Reich. 2002a. A non-parametric, supervised classification of vegetation types on the Kaibab National Forest using decision trees. *International Journal of Remote Sensing*. In press.

Joy, S.M. and R.M. Reich. 2002b. Modeling Forest Structure on the Kaibab National Forest in Arizona. *Forest Science*. In review.

Kaye, M.W. 2002. Forest ecology of quaking aspen in Rocky Mountain National Park: a landscape survey, a reconstruction, and a model. PhD dissertation. Colorado State University, Fort Collins, CO.

Lonsdale, W.M. 1999. Global Patterns of Plant Invasions and the Concept of Invasibility. *Ecology* 80:1522-1536.

Mathsoft, Inc. 1999. *S-PLUS 2000 Guide to Statistics, Volume 1*. Data Analysis Division, Mathsoft, Seattle, WA.

NRCS. 2001. The PLANTS Database, Version 3.1, National Plant Data Center, Baton Rouge, LA 70874-4490 USA.

Peet, R.K. 1988. Forests of the Rocky Mountains. Pp. 63-101 in: *North American Terrestrial Vegetation*. M.G. Barbour and W.D. Billings (eds.). Cambridge University Press, New York, NY

Reich, R.M. and R.A. Davis. 1998. On-line Spatial Library for the S-PLUS© Statistical Software Package. Colorado State University, Fort Collins, CO.

Reich, R.M., J.E. Lundquist, and V.A. Bravo. 2002. Spatial models for examining the effects of diseases and other disturbances on patterns of fuel loading. *International Journal of Wildland Fire*. In review.

Statistical Sciences. 2000. S-PLUS© 4.0 Statistical Software Package for Personal Computers. StatSci Division, Mathsoft, Inc., Seattle, WA.

Stohlgren, T.J., D. Binkley, G.W. Chong, M.A. Kalkhan, L.D. Schell, K.A. Bull, Y. Otsuki, G. Newman, M. Bashkin, and Y. Son. 1999. Exotic Plant Species Invade Hot Spots of Native Plant Diversity. *Ecological Monographs* 69:25-46.

Stohlgren, T.J., K.A. Bull, and Y. Otsuki. 1998. Comparison of rangeland vegetation sampling techniques in the Central Grasslands. *Journal of Range Management* 51:164-172.

Stohlgren, T.J., G.W. Chong, M.A. Kalkhan, and L.D. Schell. 1997. Multi-scale sampling of plant diversity: effects of minimum mapping unit size. *Ecological Applications* 7:1064-1074.

Stohlgren, T.J., G.W. Chong, M.A. Kalkhan, and L.D. Schell. 1997b. Rapid assessment of plant diversity patterns: a methodology for landscapes. *Environmental Monitoring and Assessment* 48:25-43.

Stohlgren, T.J., M.B. Coughenour, G.W. Chong, D. Binkley, M.A. Kalkhan, L.D. Schell, D.J. Buckley, and J.K. Berry. 1997c. Landscape analysis of plant diversity. *Landscape Ecology* 12:155-170.

Stohlgren, T. J., M.B. Falkner, and L.D. Schell. 1995. A Modified-Whittaker nested vegetation sampling method. *Vegetatio* 4:1-8.

Stohlgren, T.J., J.F. Quinn, M. Ruggiero, and G. Waggoner. 1995b. Status of Biotic Inventories in U.S. National Parks. *Biological Conservation* 71:97-106.

Stohlgren, T.J., K.A. Rimar, and K.W. Stolte. (In Preparation). Monitoring native and exotic plant diversity in the National Forests of the U.S.

U.S. Forest Service. 1999. FHM regional technical data summaries. http://www.na.fs.fed.us/spfo/fhm/fhh/fhmusamap.htm

CONCLUSION

I started working on the Landscape-scale GAP Analysis project shortly after it was funded in 1994. The National Park Service initiated the project out of concern that existing vegetation maps were not at a resolution fine enough for making resource management decisions. Over the course of the study new sampling methods (comparable multi-phase, multi-scale sampling and the Modified-Whittaker plot) were developed, tested, and improved, and they are now in use by many other researchers (Federal, state, private, student, and international). Although individual research projects are driven by a variety of questions and objectives, an overarching question will provide continuous, exciting challenges: what more can be gleaned from existing data?

With this dissertation, I presented the tip of the iceberg of the analysis of the Rocky Mountain National Park vegetation data. Although these data are place specific, I hope that the analyses provide useful examples for other place specific projects as well as ideas for larger-scale analyses that combine pre-existing data sets (such as current work by Stohlgren et al. to look at regional and statewide patterns of non-native plant species). Near future work with these data includes experimenting with "scalable" models as introduced in the discussion of Chapter IV and ongoing ground-truthing for new vegetation maps (e.g., the Colorado Natural Heritage Program in conjunction with the National Park Service and the U.S. Geological Survey).

Chapters I and II presented applications of multi-scale plot data through the use of species-area curves and an initial look at linking species richness (diversity) with area. Chapter III introduced a computationally expensive (kriging) but information-rich analysis approach for spatial extrapolation of point data. Chapter IV presented an alternative spatial analysis approach (trend surface analysis combined with regression tree analysis) to that presented in Chapter III. The analysis in Chapter IV is much less computationally expensive, and may have more potential utility because fewer data points are needed.

Much of the current work I am involved with still examines patterns of plant species richness, but there is a new emphasis on invasive plant species and their potential negative effects on native biodiversity. New applications of the research approach developed to gather the data for this dissertation include investigating the effects of wildfire severity on subsequent revegetation and non-native species invasion. Plant species composition and cover may have profound effects on post-fire soil erosion, water run-off and infiltration, and future fire severity (e.g., through the production of different types of fuels). Development of full surface maps of forest fuels, for example, may be invaluable for directing local and regional wildfire and invasive species mitigation work.

Globally, current natural resource crises are tied to our demands for resources. Scientists must remain impartial to maintain credibility, while providing quantitative, unbiased information to others who have the responsibility of wise stewardship of our natural resources. We cannot muster interest and financial resources to protect what we do not know we have. As global citizens, we must provide scientific tools and resources to protect natural resources, such as the lowland forests in Indonesia, wetlands, and old-growth forests, worldwide. I must believe that every contribution of information that can be used to aid biodiversity conservation is important. To that end, I will continue to seek ways to improve rapid inventory and assessment methods for sampling and analytical tools so that the unbiased data gathered are useful to natural resource managers, researchers, policy makers, and the global public.

EPILOGUE

Research always leads to new questions. I constructed a sample of species-accumulation curves from three groups of the plots sampled in Rocky Mountain National Park. A species-accumulation curve is a plot of the total number of species encountered (y-axis) as a new plot area is sampled (x-axis). I used software developed for use in Microsoft Access with a Modified-Whittaker database structure, both developed by Rick Shory at the Natural Resource Ecology Laboratory, Colorado State University. The objectives of this epilogue were to examine the shapes of the example species accumulation curves to determine if they began to level off with the given sampling effort and to compare the relative contributions of two vegetation types to plant species richness.

The first species-accumulation curve was developed from the 149 full-size (1000-m^2) Modified-Whittaker plots that were sampled in the Park from 1995-1999 (Figure E-1). The second plot is a comparison of the species-accumulation curves from the lodgepole pine and aspen vegetation types (Figure E-2). All curves are an average of three random draws from all possible plots in the respective groups (i.e., the first group was all 149 plots, the second was 23 lodgepole pine plots, and the third was 23 aspen plots).

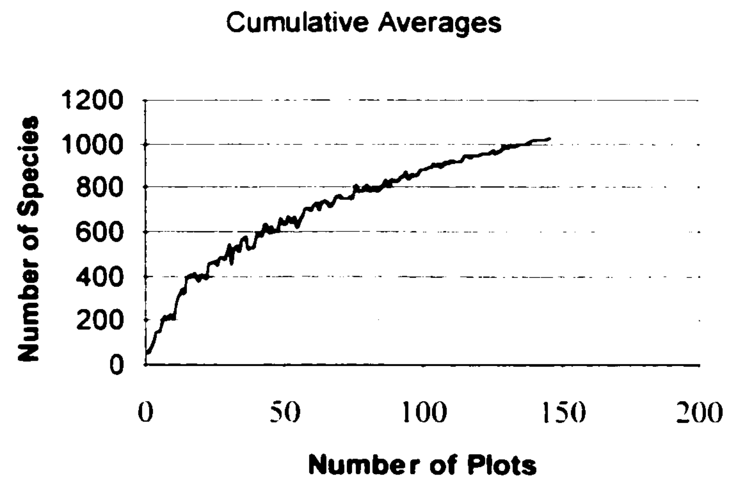

Figure E-1. The cumulative number of species encountered (based on the mean from three random draws) in 149 1000-m^2 Modified-Whittaker plots sampled in Rocky Mountain National Park.

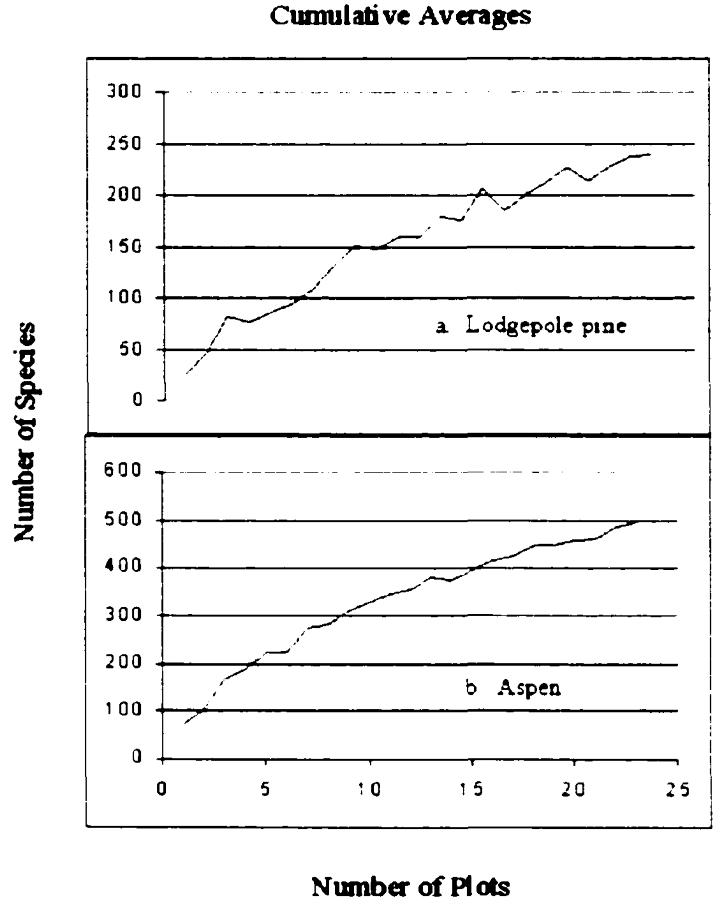

Figure E-2. The cumulative number of species encountered (based on the mean from three random draws) in (a) 23 1000-m^2 Modified-Whittaker plots sampled in lodgepole pine stands and (b) 23 1000-m^2 Modified-Whittaker plots sampled in aspen stands.

The leveling off of a species-accumulation curve should be the result of the underlying distribution of species, which was often assumed to be random (e.g., MacArthur's 1965 and MacArthur and Wilson's 1967 island biogeography work). Plotkin and Levin (2001) provided a good evaluation of how that assumption may be appropriate over very small scales (<1 ha) but not at larger scales where species clumping becomes measurable.

I predicted that the species-accumulation curves would not level off because the area sampled is still very small relative to the actual area of these vegetation types within their geographic and climatic range. As expected, the species-accumulation curve for the 149 plots does not appear to level off (Figure E-1). The logarithmic equation (number of species = 251.6*ln(area) - 296, $R^2 = 0.95$) may underestimate the number of species at the smallest and largest areas (numbers of plots) and overestimate for the intermediate areas. This behavior is also observed for the lodgepole pine (23 plots; number of species = 74.3*ln(area) -11, $R^2 = 0.94$) and aspen (23 plots; number of species = 145.9*ln(area) + 5, $R^2 = 0.96$) vegetation types (Figure E-2). As expected, the aspen coefficient is much greater than the lodgepole pine coefficient. However, to ensure a fair comparison between vegetation types, I was only able to use 23 of 32 aspen plots because there were only 23 lodgepole pine plots. That essentially resulted in a loss of existing information that could improve my ability to extrapolate from the sampled to the unsampled aspen area (32 aspen plots; number of species = 156.1*ln(area) - 19.7, $R^2 = 0.96$).

I maintain that species-area curves as presented in Chapters I and II are useful for comparing vegetation types' contributions to species richness, and they may be useful for estimating the number of species expected to be found in relatively small, well-sampled areas (Stohlgren et al. 1997). Species-accumulation curves as presented in the Epilogue may be useful for examining species' distributions (e.g., Plotkin and Levin 2001) and for assessing the completeness of species lists. Quantitative, spatially explicit plant species data are essential for optimizing future research. For example, the aspen and lodgepole species-accumulation examples (Figure E-2) indicate that far more species could be encountered with increased sampling in the aspen vegetation type, even though it is rare on the landscape, than with increased sampling in the lodgepole vegetation type.

The story of plant diversity in Rocky Mountain National Park is far from complete. There may be several hundred more plant species as yet undiscovered in the Park. The Park list likely has < 1000 species on it after duplicate species are removed. Sampling an additional 150 1000-m^2 plots in rare habitats would likely add several hundred species to the list and increase understanding of the Park's flora.

LITERATURE CITED

MacArthur, R.H. 1965. Patterns of species diversity. *Biology Review* 40:510-533.

MacArthur R.H. and E.O. Wilson. 1967. *Island Biogeography*. Princeton University Press, Princeton, New Jersey.

Plotkin, J.B. and S.A. Levin. 2001. The spatial distribution and abundances of species: lessons from tropical forests. *Comments on Theoretical Biology* 6:251-278.

Stohlgren, T.J., G.W. Chong, M.A. Kalkhan, and L.D. Schell. 1997. Rapid assessment of plant diversity patterns: a methodology for landscapes. *Ecological Monitoring and Assessment* 48:25-43.