DISSERTATION

SYNERGISTIC METHODS TO GENERATE PREDICTIVE MODELS AT LARGE SPATIAL EXTENTS AND FINE RESOLUTION

Submitted by

Catherine S. Crosier

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 2004

COLORADO STATE UNIVERSITY

June 24, 2004

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY CATHERINE S. CROSIER ENTITLED "SYNERGISTIC METHODS TO GENERATE PREDICTIVE MODELS AT LARGE SPATIAL EXTENTS AND FINE RESOLUTION" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Thomas J. Stohlgren
Advisor

Jack
Co-Advisor

Dan Binkley
Department Head/Director

ABSTRACT OF DISSERTATION

SYNERGISTIC METHODS TO GENERATE PREDICTIVE MODELS AT LARGE SPATIAL EXTENTS AND FINE RESOLUTION

Non-native plant species must be mapped and contained because they have negative economic impacts and degrade native ecosystems and wildlife habitats. By combining data from many different sources, known as data synergy, knowledge of their locations is greatly improved. Different data types (two species lists, quarter quad survey, weed mapping, and vegetation plots from 45 datasets) added different information, and the number of species recorded per county in Colorado increased on average by 30% even in the most intensively surveyed areas. These combined data were then used to create predictive models for Colorado for non-native and native species richness and a probability of occurrence model for *Euphorbia esula*, leafy spurge, a poisonous, non-native weed. The non-native species richness model had an r^2 of 0.36, suggesting that significant general patterns could be identified. The most invaded areas in the Colorado riparian areas and the least invaded areas were high elevation vegetation types. The generalized linear model (GLM) for *E. esula* explained 62% of the variance in occurrence. The final model, including spatial autocorrelation, did not discriminate well when compared to a quarter quad level statewide survey of county weed managers for estimates of *E. esula* presence, but this result occurred, at least in part, from differences in the datasets (estimated current distribution from survey versus model of potential distribution of a spreading species) rather than actual model performance.

Overall, data synergy holds promise to increase knowledge of non-native species current locations and their predicted distributions and richness over large geographic areas.

Catherine Shirin Crosier
Graduate Degree Program in Ecology
Colorado State University
Fort Collins, CO 80523
Summer 2004

ACKNOWLEDGEMENTS

The U.S. Geological Survey supported this research. Colorado State University provided me with one year of academic support through graduate fellowships. This work could not have been completed without the logistical support of the Natural Resource Ecology Laboratory and the USGS Fort Collins Science Center.

Assistance in dataset collection and data processing was provided by Jim Graham, Greg Newman, and Rick Shory. Jim and Rick especially have been invaluable in their teachings on data management. I would also like to extend my thanks to all those willing to share their datasets with me- without their hard work in originally collecting data and willingness to share the products of these efforts, this project would not have been possible. Geneva Chong, Greg Newman, Kate Searle, Alycia Waters, and Todd Wojtowicz provided helpful comments and guidance for earlier drafts of these chapters.

I owe many thanks to my committee, Thomas J. Stohlgren, advisor, James Detling, co-advisor, Mohamed Kalkhan, Barry Noon, and David Theobald for agreeing to work with me. I appreciate all the time and energy they spent helping me along the way.

Finally, I must thank my family and friends for always supporting and assisting me. This includes my coworkers in the Stohlgren group- they are good friends, my fellow NRELians (especially my fellow grad students and office mates) and all the other students I have met and worked with during my time at CSU. I would also like to thank those at USGS who have helped me along the way.

Thank you to all- I appreciate all the help and support.

TABLE OF CONTENTS

INTRODUCTION

Non-native species cause economic impacts, threaten human health, and impact biodiversity (Mack et al. 2000). Impacts of non-native species in general cost the U.S. an estimated $137 billion/yr, and non-native plant species in particular cost about $34 billion/yr (Pimentel et al. 2000). Non-native species are now the number two threat to endangered and threatened species, second only to habitat loss (Wilcove et al. 1998). Thus, determining areas vulnerable to invasion and predicting the spread of potential invaders is becoming a national priority. Many studies have focused on identifying factors related to the establishment and spread of invasive species.

Vulnerability of habitats to invasion

Various studies have focused on identifying factors that influence the vulnerability of systems to invasion. These studies have been conducted at different scales, have concentrated on different factors, and have found varying results (Appendix A). A commonly cited factor first proposed by Charles Elton (1958) is low species richness. However, some researchers have found a positive correlation between species diversity and susceptibility to invasion (Robinson et al. 1995, Planty-Tabacchi et al.

1996, Wiser et al. 1998, Smith and Knapp 1999, Stohlgren et al. 1999a), while others have found a negative correlation (Elton 1958, Fox and Fox 1986, Tilman 1997, Levine and D'Antonio 1999, Dukes 2001). The conflicting results may arise because these studies were conducted at different scales and in different communities. In a theoretical examination, Shea and Chesson (2002) suggested that looking across broad spatial scales encompasses heterogeneous extrinsic factors leading to a positive relationship, but the relationship would be negative at small scales encompassing homogeneous extrinsic factors. Brown and Peet (2003) found a graded relationship from negative to positive with increasing scale when examining patterns across five to six scales in Appalachian riparian zones.

Research has also evaluated the relationship between disturbance regimes and invasion vulnerability. Exogenous disturbances, or disturbance regimes altered from those historically found in an area, have been found to increase the vulnerability to invasion of an area (Hobbs and Huenneke 1992). Similarly, Fox and Fox (1986) determined that a departure from the endogenous disturbance regime is necessary for invasion to occur. Looking at fire in the Kanza Prairie, Smith and Knapp (1999, 2001) found that deviation from the natural fire regime increased vulnerability to invasion. They found a similar trend for grazing, but Stohlgren et al. (1999b) did not find any significant effect from grazing in the Central Grasslands at large spatial scales. In contrast, Larson et al. (2001) found that the most common invasive species in Theodore Roosevelt National Park did not follow disturbance patterns. This study only incorporated anthropogenic physical disturbance such as roadways, however, and not departures from natural disturbance regimes such as fire suppression.

In Glacier National Park, Tyser and Worley (1992) reported a correlation between non-native species and disturbance from roads and trails. Greenburg et al. (1997) found that roadways facilitated non-native species spread, especially when roadside material differed from the native soil type. In a global study that accounted for size differences (Lonsdale 1999) and a regional study (MacDonald et al. 1989), high visitation to reserves was positively correlated with high invasion rates.

Soil type has also been proposed as a factor related to vulnerability to invasion. In the serpentine grasslands of California, Harrison (1999) and Huenneke et al. (1990) determined that invasion was greater on soils with greater levels of nutrients. Stohlgren et al. (1999a, b) found a correlation between soil percent nitrogen and non-native species richness and cover at various spatial scales in plots throughout the central grasslands and Colorado Rockies.

Certain community types have also proved more invaded than others. In several studies in different areas of the world and at different spatial scales, riparian areas were more invaded than upland sites (Planty-Tabacchi et al. 1996, Kotanen 1997, Stohlgren et al. 1998b, Levine 2000, Larson et al. 2001). Similar studies have shown that rare habitats are more heavily invaded (Stohlgren et al. 1999a).

These studies indicate that factors influencing the abundance and distribution of non-native species vary at different spatial scales. For example, soil nitrogen trends are found at a much smaller scale than anthropogenic disturbance. These factors that influence vulnerability to invasion could also be used to predict what areas will be most heavily invaded.

Modeling invasions

Higgins and Richardson (1996) review traditional models predicting non-native plant spread including simple-demographic models, spatial-phenomenological models, and spatial-mechanistic models to aid in determining which model type to use given specific data at hand. Simple demographic models are suitable when population density relates to the area invaded (density dependence) or when determining if a species will successfully establish. Spatial phenomenological models formulate predictions based on the past. Mechanistic models do not require an empirical precedent. Mechanistic models discussed include reaction-diffusion, metapopulation, and individual-based models. Metapopulation models are less demanding commutatively, but individual-based models should be employed when patterns of environmental heterogeneity are not captured with a metapopulation model, fine-scale ecological heterogeneity is important in invasion success, and interactions among individuals (e.g., competition) is important. The model chosen to represent a specific invasion should be chosen based on the characteristics of that particular invasion.

Higgins et al. (1996) assert that reaction-diffusion (R-D) models are probably the most widely applied invasion models, but have generally been used to model animal invasions (Higgins and Richardson 1996). The assumptions of this model prevent use of interactions between plant attributes and the environment (e.g., higher biomass occurs with higher nutrient levels), therefore relegating environmental heterogeneity and stochasticity as unimportant. These authors propose the use of a spatially explicit individual-based simulation (SEIBS) model as an alternative that integrates space, ecological processes, and stochasticity. They compared the two models and determined that there is a qualitative and quantitative difference in the predictions.

Higgins et al. (2000) scaled up a SEIBS model to create a spatially explicit landscape extent simulation (SELES) model to predict the rate of spread of non-native plant species. This model used the predicted probability of occurrence of a specific species from previous work (Higgins et al. 1999) to predict the spread of these occurrences. The native plant data were collected at a coarser scale (1 km^2) than the grain chosen for the model (200 m), so they converted the data by assuming that if a species was present at 1 km^2, then it would be present in each of the 25 squares (200 m x 200 m). The potential non-native plant distribution was based on a logistic regression model of environmental preferences (Higgins et al. 1999). This variable was used along with plant dispersal, mortality recruitment, and control efforts to create the SELES model. A comparison of the two models, SEIBS and SELES, did not reveal scaling artifacts.

Wadsworth et al. (2000) modeled the spread of two non-native plant species in England using MIGRATE, a spatially explicit model. They used 13 different reproductive and dispersal parameters in the model. Six different management strategies were simulated to determine the effectiveness of control strategies.

Peterson and Vieglias (2001) used the Genetic Algorithm for Rule-set Prediction, or GARP, to model species distributions. This model used ecological niches to predict invasion by a specific non-native species. The program ran many iterations of multiple regressions to predict presence or absence or intersection of ranges and compared these results to sets of resampled points from known occurrences. From these analyses, a set of five to 50 rules that together defined the species' ecological niche were obtained. These rules were then used to define the species distribution. Variables included in the model were vegetation type, slope, aspect, elevation, and soil type. Peterson and Vieglias based their model on the concept that 'ecological niches are stable and determine the set of possible conditions under which a species is able to invade a particular region'. Most data used for the model were derived from data sets obtained from museum collections. The model provided information on the distribution of species; no information on abundance or cover was provided. Additionally, most models that have been generated using GARP are at a meso-scale of around 1 km^2.

The studies described above model the distribution of a single species, predicting its presence or absence. None look at the abundance or cover of a single species or attempt to model the total number of non-native species distributed across an area (non-native species richness). These models do not attempt to identify hot spots of invasion or areas vulnerable to invasion. Additionally, these models ignore spatial information that could be particularly important in predicting the spread of non-native species across a landscape.

Iverson and Prasad (1998) developed a new methodology that models native plant species richness in less sampled areas using data from better sampled areas. They took a county level dataset for Illinois and used stepwise regression to develop a model using variables from environmental and topographic variables from a Geographic Information System (GIS). They then examined the response surface for possible spatial autocorrelation using Moran's I and Geary's C. In this way, they would have included fine-scale variability from neighbor interactions in addition to the coarse scale trend, but they did not find spatial autocorrelation in the response surface.

However, a spatial modeling process described by Reich and Bravo (1998) makes use of a unique combination of spatial statistical methods to model landscape structure. The input data include field data, remotely sensed imagery, and GIS data. Data are extracted from the GIS layers and images for each field point. A regression model can then be used to describe the coarse-scale variability in the data. If the residuals contain spatial autocorrelation a model can be developed to describe the small-scale variability in the data using kriging or co-kriging. Finally, the two surfaces (the coarse scale model and the co-kriged residuals) are combined to create a final trend surface model that best describes the spatial distribution of non-native species richness or the spatial distribution of an individual non-native species.

Kalkhan et al. (2000a, b, 2004) and Chong et al (2001) used Reich and Bravo's methods to predict the distribution, presence and pattern of native and exotic plant species and soil characteristics. The spatial statistical models produced had higher R^2 values than the simple regression models. The use of full-coverage, fine-scale data (e.g., Landsat TM data) ameliorates the problems associated with the other types of models arising from lack of empirical data and scale issues. These methods also allow for the development of maps of uncertainty based on subsampling. This surface provides land managers or others using the data with spatially represented confidence for the model, and also can direct further sampling efforts to areas with low confidence.

These methods integrating field data, remotely sensed data, GIS, and spatial statistical techniques provide a method that captures the patterns of species distributions.

They also require less intensive data collection than individual based models reliant on demographic data. The predictions can cover large areas without increasing spatial grain. For determining species distribution, abundance and patterns, these models appear the most appropriate for high accuracy without extensive (and cost prohibitive) field sampling.

In these examples, ordinary least squares regression was used to capture the coarse-scale variability. General linear models (GLM), a technique similar to classical regression but that does not assume a normal distribution, have recently become more widely used in ecology (Guisan et al. 2002). A review of the literature seems to indicate that at least for count data such as invasive species distributions that approximate a Poisson distribution, log-linear models are favored. Nicholls (1989a) used GLM methods to extrapolate existing information from a specific area to a region to improve knowledge of species distributions. Margules and Stein (1989) used logistic regression, the binary distribution equivalent of log-linear regression, to predict the presence of canopy tree species using environmental variables. McIntyre and Lavorel (1994) and Luoto et al. (2002) used log-linear regression to predict the presence of rare species because the rare species distributions were very skewed toward zero. In a review of modeling techniques, Guisan and Zimmerman (2000) advocate the use of GLM techniques for species abundance data. In a table listing statistical approaches and models, they list GLM for individual species counts and species richness but do not include least squares regression as an appropriate statistical technique. A special issue of *Ecological Modelling* resulting from an international conference had several articles discussing the application of these techniques to species distribution modeling (Guisan et al. 2002). GLM techniques in combination with kriging of residuals may be a very powerful technique to create potential distribution models for invasive species and non-native species richness.

Overview

In Chapter 1, I illustrate the benefits of data synergy. There are many different studies with a wealth of data that have been conducted, and by combining these datasets knowledge of the location and distribution of non-native species is improved. Chapter 2 then describes how this composite dataset can be used to predict species richness over large geographic areas which can in turn be used to identify hotspots of invasion. Chapter 3 also makes use of the large dataset, providing an example of a probability of occurrence model for a single non-native species of interest.

Literature Cited

Brown, R. L., and R. K. Peet. 2003. Diversity and invasibility of southern Appalachian plant communities. Ecology 84:32-39.

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.

Dukes, J. S. 2001. Biodiversity and invasibility in grassland microcosms. Oecologia 126:563-568.

Elton, C. 1958. The ecology of invasions by plants and animals. Meuthuen and Company, LTD, London.

Fox, M. D., and B. J. Fox. 1986. The susceptibility of natural communities to invasion. Pages 57-66 in R. H. Groves and J. J. Burdon, editors. Ecology of biological invasions: and Australian perspective. Cambridge University Press, Cambridge.

Greenburg, 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., T. C. J. Edwards, and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157:89-100.

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

Harrison, S. 1999. Local and regional diversity in a patchy landscape: native, alien, and endemic herbs on serpentine. Ecology 80:70-80.

Higgins, S. I., and D. M. Richardson. 1996. A review of models of alien plant spread. Ecological Modelling 87:249-265.

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.

Higgins, S. I., D. M. Richardson, and R. M. Cowling. 2000. Using a dynamic landscape model for planning the management of alien plant invasions. Ecological Applications 10:1833-1848.

Higgins, S. I., D. M. Richardson, R. M. Cowling, and T. H. Trinder-Smith. 1999. Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. Conservation Biology 13:303-313.

Hobbs, R. J., and L. F. Huenneke. 1992. Disturbance, diversity, and invasion: implications for conservation. Conservation Biology 6:324-337.

Huenneke, L. F., S. P. Hamburg, R. Koide, H. A. Mooney, and P. M. Vitousek. 1990. Effects of soil resources on plant invasion and community structure in Californian serpentine grassland. Ecology 71:478-491.

Iverson, L. R., and A. Prasad. 1998. Estimating regional plant biodiversity with GIS modelling. Diversity and Distributions 4:49-61.

Kalkhan, M. A., G. W. Chong, R. M. Reich, and T. J. Stohlgren. 2000a. Landscape-scale assessment of mountain plant diversity: integration of remotely sensed data, GIS, and spatial statistics. in ASPRS 2000 Proceedings: Start the 21st century: launching the geospatial information age, Washington, D.C.

Kalkhan, M. A., T. J. Stohlgren, G. W. Chong, L. D. Schell, and R. M. Reich. 2000b. A predictive spatial model of plant diversity: integration of remotely sensed data, GIS, and spatial statistics. in J. D. Greer, editor. Eighth Biennial Remote Sensing Applications Conference: Remote Sensing and Geospatial Technologies for the New Millennium. American Society for Photogrammetry and Remote Sensing, Albuquerque, NM.

Kalkhan, M. A., E. J. Martinson, P. N. Omi, T. J. Stohlgren, G. W. Chong, and M. A. Hunter. 2004. Integration of spatial information and spatial statistics: a case study of invasive plants and wildfire on the Cerro Grande fire, Los Alamos, New Mexico. in R. T. Engstrom and W. J. de Groot, editors. Proceedings of the 22nd Tall Timbers Fire Ecology Conference: Fire in Temperate, Boreal, and Montane Ecosystems, Tallahassee, FL.

Kotanen, P. M. 1997. Effects of experimental soil disturbance on revegetation by natives and exotics in coastal Californian meadows. Journal of Applied Ecology 34:631-644.

Larson, D. L., P. J. Anderson, and W. Newton. 2001. Alien plant invasion in mixed-grass prairie: Effects of vegetation type and anthropogenic disturbance. Ecological Applications 11:128-141.

Levine, J. M. 2000. Species diversity and biological invasions: relating local process to community pattern. Science 288:852-854.

Levine, J. M., and C. D'Antonio. 1999. Elton revisited: a review of evidence linking diversity and invasibility. Oikos 87:15-26.

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

Luoto, M., T. Toivonen, and R. K. Heikkinen. 2002. Prediction of total and rare plant species richness in agricultural landscapes from satellite images and topographic data. Landscape Ecology 17:195-217.

MacDonald, I. A., L. L. Loope, M. B. Usher, and O. Hamann. 1989. Wildlife conservation and the invasion of nature reserves by introduced species: a global perspective. Pages 215-255 in J. A. Drake, H. A. Mooney, F. di Castri, R. H. Groves, F. J. Kruger, M. Rejmanek, and M. Williamson, editors. Biological invasions: a global perspective. John Wiley and Sons, New York.

Mack, R. N., D. Simberloff, W. M. Lonsdale, H. Evans, M. Clout, and F. A. Bazzaz. 2000. Biotic invasions: Causes, epidemiology, global consequences, and control. Ecological Applications 10:689-710.

Margules, C. R., and J. L. Stein. 1989. Patterns in the distributions of species and the selection of nature reserves: An example from Eucalyptus forests in southeastern New South Wales. Biological Conservation 50:219-232.

McIntyre, S., and S. Lavorel. 1994. Predicting Richness of Native, Rare, and Exotic Plants in Response to Habitat and Disturbance Variables across a Variegated Landscape. Conservation Biology 8:521-531.

Nicholls, A. O. 1989. How to make biological surveys go further with generalised linear models. Biological Conservation 50:51-75.

Peterson, A. T., and D. A. Vieglais. 2001. Predicting species invasion using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. BioScience 51:363-371.

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

Planty-Tabacchi, A.-M., E. Tabacchi, R. J. Naiman, C. DeFerrari, and H. Decamps. 1996. Invasibility of species-rich communities in riparian zones. Conservation Biology 10:598-607.

Reich, R. M., and V. A. Bravo. 1998. Integrating spatial statistics with GIS and remote sensing in designing multiresource inventories. Pages 202-207 in North America Science Symposium: Toward a unified framework for inventorying and monitoring forest ecosystem resources. USDA Rocky Mountain Research Station Proceedings, RMRS-P-12, Guadalajara, Jalisco, Mexico.

Robinson, G. R., J. F. Quinn, and M. L. Stanton. 1995. Invasibility of experimental habitat islands in a California winter annual grassland. Ecology 76:786-794.

Shea, K., and P. Chesson. 2002. Community ecology theory as a framework for biological invasions. Trends in Ecology & Evolution 17:170-176.

Smith, M. D., and A. K. Knapp. 1999. Exotic plant species in a C-4-dominated grassland: invasibility, disturbance, and community structure. Oecologia 120:605-612.

Smith, M. D., and A. K. Knapp. 2001. Size of the local species pool determines invasibility of a C-4-dominated grassland. Oikos 92:55-61.

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

Stohlgren, T. J., Y. Otsuki, C. A. Villa, M. Lee, and J. Belnap. 2001. Patterns of plant invasions: a case example in native plant species hotspots and rare habitats. Biological Invasions 3:37-50.

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

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.

Tilman, D. 1997. Community invasibility, recruitment limitation and grassland biodiversity. Ecology 78:81-92.

Tyser, R. W., and C. A. Worley. 1992. Alien flora in grasslands adjacent to road and trail corridors in Glacier National Park, Montana (U.S.A.). Conservation Biology 6:253-262.

Wadsworth, R. A., Y. C. Collingham, S. G. Willis, B. Huntley, and P. E. Hulme. 2000. Simulating the spread and management of alien riparian weeds: are they out of control? Journal of Applied Ecology 37:28-38.

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.

Wiser, S. K., R. B. Allen, P. W. Clinton, and K. H. Platt. 1998. Community structure and forest invasion by an exotic herb over 23 years. Ecology 79:2071-2081.

CHAPTER 1 : IMPROVING BIODIVERSITY KNOWLEDGE THROUGH DATASET SYNERGY: A CASE STUDY OF NON-NATIVE VASCULAR PLANTS IN COLORADO

1.0 Abstract

Species lists are important tools for managing biodiversity, including managing non-native species, but they are incomplete or lacking for many areas. Synergy of disparate datasets can increase knowledge of species distributions, while minimizing further expenditures of cost and time. I examined this increase in knowledge by comparing five different data types of non-native plant locations at the county level comprised of over 30 datasets covering Colorado to illustrate the benefits of data synergy. Species lists captured the most data, but they missed many of the noxious weeds recorded by weed mapping data. The number of species recorded per county increased on average by 30% even in the most intensively surveyed areas, and each data type seemed to follow the same pattern of survey intensity, leaving some areas in the state consistently un-surveyed or under-surveyed. With the county level data type as a baseline, there was on average a 44% increase in number of non-native plant species recorded per county when the other four data types were included. Overall, inclusion of more data types resulted in greatly increased knowledge of the non-native species in Colorado. Therefore,

harnessing the synergy of disparate datasets seems to be a cost-effective first step to increase knowledge of species richness in an area of interest.

1.1 Introduction

Biodiversity conservation is increasing in its importance as a global concern (Wilson 1988). However, knowledge of biodiversity is generally lacking, including the species richness component of biodiversity. Management to conserve biodiversity is hampered without knowledge of where species occur, but species lists do not exist for many public land units (e.g., national wildlife refuges). Besides information on native species diversity, information on non-native species is important because non-native plants can negatively impact agricultural land, rangeland, and forests; alter ecosystem function; and potentially reduce native biodiversity (Vitousek et al. 1997, Mack et al. 2000). These impacts can be important for economic, ecological, and ethical reasons.

Ecological studies provide important information on biodiversity but are often conducted over short temporal scales in small geographic areas. This spatially and temporally small scale has been criticized by many as a weakness of ecology and does not lead to comprehensive data collection or effective monitoring of biodiversity at broader scales. There has been a call for research projects that cover broader extents and longer time periods. However, both long-term and large scale, intensive surveys are very costly. Thus, various researchers have proposed quantitative methods to improve species lists for geographic areas (Nicholls 1989b, Iverson and Prasad 1998, Guisan and Zimmermann 2000, Palmer et al. 2002). These have included various statistical methods and auxiliary variables, which can be complex and time intensive although avoiding the costs and time associated with new field studies.

Biodiversity data for native and non-native species collected by different groups are patchy, diverse, and incomplete. Much information on non-native plant species exists in the form of weed mapping (a point location for a specific weedy species), while plant community information (native and non-native species biodiversity) is commonly a product of some form of small-scale, snapshot-in-time vegetation survey. These various smaller-scale data types are then often used to create species lists for larger geographic areas such as a national park. Thus, non-native plant species richness information not only exists in multiple data sets, but these data sets greatly differ in their biases as a result of their contrasting research objectives and field methods.

Approximately 35% of the land area in Colorado is publicly owned. Agencies and other organizations managing these lands often do not share information, even though it is important to know what species occur on adjacent lands because they can act as source populations for invading species (D'Antonio et al. 2001). Species found on adjacent lands would therefore be likely to occur in the management unit of interest. Increased data sharing between agencies and landowners may be an important way to aid managers in control of non-native species and monitoring of native species.

In this paper, I explore a quick, simple means of exploiting existing datasets by merging them into one dataset. These techniques can be used to create a central repository of information that can facilitate data sharing between various management groups. With synthesis of these smaller datasets, overall knowledge of species lists might improve and the number and complexity of research questions that can be addressed may increase. Using pre-existing datasets also capitalizes on resources already expended, reducing time and money constraints. My objective was to merge disparate datasets into a cohesive form to illustrate the utility of data synergy. Specifically, I wanted to determine the relative contribution of different types of datasets to overall knowledge and illustrate how this knowledge might improve with the combination of these different data types. I considered only non-native plant species data as an example as I had many different data types for this subset of species and non-native plant species locations are an important management concern.

1.2 Methods

1.2.1 Datasets

Datasets detailing information on non-native plant locations in Colorado were obtained from 54 sources including federal, state, and local agencies, university professors, graduate students, and non-profit organizations (Figure 1-1; Appendix B). These datasets were merged into a spatial database using SQL server and ArcGIS that captured the various attributes associated with each dataset, including the location of non-native plant species. Plant species names were standardized using the National Resource Conservation Service's Plants Database lists of accepted and synonymous scientific names (Natural Resources Conservation Service 2003). Data from the combined 54 datasets were then classified by methodology into five broad data collection categories including the Biota of North America Program (BONAP) county level species lists, quarter quad survey data, species lists for large geographical regions, vegetation plot data, and weed mapping Global Positioning System (GPS) data. Each classification will be referred to as a data type throughout this paper.

BONAP data consist of county level species lists for all 63 Colorado counties created from herbarium records. Quarter quad survey data were generated from a survey conducted by the state of Colorado in 2002 in which county weed managers estimated acreage per quarter quad for each of 20 non-native plant species for all quarter quads in the state. The species lists dataset was composed of species lists for large geographic units including national parks, national fish and wildlife refuges, state parks, and Long Term Ecological Research stations. Plot data included vegetation surveys that used various methods including the multi-scale Modified Whittaker plot covering 1000 m^2 to the Daubenmire plot covering only 0.1 m^2. GPS data included point, line, or polygon features representing locations for a specific non-native species. Occasionally, secondary and tertiary species present were recorded with the primary species of interest for these locations. GPS data and quarter quad data are limited to those species considered noxious weeds (a legal status for species that cause great negative economic impact) or species of concern by the state or by a county. BONAP data, species lists, and plot data include all non-native species that have been identified in the area that they cover. For this paper, the non-native designation of species is taken from the U.S. Department of Agriculture Natural Resources Conservation Service plants database classification (Natural Resources Conservation Service 2003). Data were also divided based on status as a federal noxious weed or state of Colorado noxious weed to determine whether trends for these categories differed from total non-native species.

1.2.2 Statistical analyses

Data were analyzed at the county level. BONAP data were used as a baseline data set because it was at the largest spatial scale (county level) and covered all 63 counties in Colorado. A separate species list was created for each county based on each of the other different data types, resulting in five different species lists for counties represented by all five data types. For example, if 20 vegetation plots were found in Larimer County, then a composite species list of unique species from all 20 plots would be the plot county species list for Larimer County. In the species list data type, only lists that were from geographic areas encompassed in a single county were included in the analysis. The number of species recorded per county was used as opposed to density of species per county. Species richness and species density were correlated in the state, and the same trends existed for both measures. However, species richness is more intuitive for weed managers, so this measure was used.

The first analysis compared each data type, examining the average total number of non-native species and the average number of unique non-native species recorded for each data type per county. A unique species was defined as a species found in only one data type for the county. This comparison was done for total non-native species and for only those species listed as noxious by the state of Colorado. Only counties containing the specific data type of interest were included in the average values for that data type, so sample sizes differed among types.

Using BONAP data as a baseline, the contribution of each of the other four data types and the patterns of survey intensity in the state were examined. For the purpose of these analyses, counties were divided into three groups (low, middle, and high survey intensity) by dividing counties into thirds based on the number of species recorded per county by the BONAP data type (Table 1-1). Despite the fact that counties undoubtedly do differ greatly in non-native species richness, the number of non-native species recorded per county by this data type (2 – 248) seemed to follow the expected pattern of research intensity based on the human demographics of the state. Therefore, this number is assumed to reflect the survey intensity in a county and thus a researcher's knowledge of species in that county. Other factors such as habitat or environmental heterogeneity per county may be a factor, but the range from two to 248 species per county is probably the result of survey intensity in concert with these other factors. Similar trends existed when data were divided into three categories based on species density, but those are not shown here. Species richness is more intuitive to land managers and land owners than density, and because the same patterns were found when area was accounted for, this metric was used.

Data were analyzed by data type to determine differences in the contribution of each to the baseline BONAP data. The average number of both total non-native species and noxious species added to the BONAP dataset by the other four data types were compared. The percent increase in the number of species from adding a second data type to the BONAP data type was also examined to determine the relative increase at the different survey intensity levels.

I examined the importance of combining multiple data types. I determined the relationship between number of data types present in a county and the survey intensity of the BONAP data by calculating the average number of data types found in a county by level of survey intensity. I also calculated the number of non-native species added to the BONAP data type by adding more data types to each survey intensity level. Finally, I compared species density maps of the BONAP data type and all data types together.

I calculated species accumulation curves for the data types. The curve describes the contribution of each data type to the total non-native species list generated by all five data types combined. Data types were added in the order a weed manager would be likely to obtain the data. The first data type was the GPS data, which is the weed mapping data weed managers collect themselves. I next added the quarter quad weed estimates that weed managers created for the state. Species lists for geographic areas and the county species lists of the BONAP data were then added. Plot data from researchers were the last data set to be added to the species accumulation curve.

1.3 Results

Within each data type, the range in data was great (Table 1-2). For example, the baseline BONAP data type ranged from two non-native species recorded in Hinsdale County to 248 species recorded in Boulder County. The ranges of number of non-native species recorded per county of the other data types fell within the BONAP range of two to 248. The number of units per county had a wide range, too, reflecting the clustering of certain data types within single counties (e.g., one county contained 1050 plots while others contained none). Also, there were uneven sample sizes as several counties did not contain all five data types.

I found that 35% of the counties had four data types available, including the BONAP data type, and only 29% had all five data types represented. The high survey intensity counties had the largest number of datasets available for them, while the low knowledge counties had the least number of datasets available (Figure 1-2a). Counties containing the most data types added the most species to the original BONAP list, regardless of the original sets rank of high, mid, or low survey intensity (Figure 1-2b).

However, the high intensity survey counties had more absolute number of non-native species added with one, two, or four additional data types, even though they also had a higher number of non-native species to begin with. For example, of those counties containing all five data types, high intensity surveyed counties added 48 species on average compared to 41 and 34 for mid and low survey intensity counties.

The species list data type captured the most non-native plant species and the most unique species (49 and 25 non-native species, respectively; Table 1-3). However, BONAP and quarter quad data recorded the most noxious species (10 and 11 non-native species, respectively) and recorded the same number of unique noxious species as species lists (five non-native species). The GPS data type had the lowest values for all categories except mean noxious species richness where the plot data type recorded the fewest. The values for all data types together were almost two times higher than for any individual data type.

Species lists added many more non-natives species to the baseline BONAP dataset than the other three data types (29 compared to next highest of 10 species; Table 1-4a and b). However, the species list data type missed many noxious species that the quarter quad data captured, adding two compared to eight noxious species on average. When examining by survey intensity groups, there were not large differences in the number of total non-native species added to the three intensity levels by data type (e.g., the range of species added by species list was only from 28 to 32 non-native species for different survey intensities; Table 1-4a and b). However, the overall trends between survey intensity groups differed.

The high- and mid-intensity survey counties had similar numbers of non-native species added to the BONAP list overall (33 and 31 non-native species respectively), but the low survey intensity counties had half as many total non-native species added to the BONAP list (14 non-native species). However, the percent increase in recorded species was highest in the low knowledge species lists. Even Boulder County, with a BONAP species list of 248 non-native species, had 23 species added to it (a 12% increase), including two noxious weeds. The use of additional data types increased the number of species recorded per county so much that following the original intensity division a majority of Colorado counties now fell in the high intensity range. Examining species density per county from the BONAP data type and from all data types displays the same pattern, with counties originally in the lower or middle third ranges falling in the higher range as defined by the BONAP data type's species richness (Figure 1-3). These results reiterate the improvement in the overall knowledge of species occurrence at the county level by using multiple datasets.

The species accumulation curve had the steepest slope for the BONAP data type and the species lists, indicating that they added the greatest number of non-native species (Figure 1-4). The curves had a similar shape for all intensity groups and for all species together, and the curves always had a positive slope. However, the low survey intensity curve was much more linear.

1.4 Discussion

I expected that combining data sets would improve information on non-native species richness and distributions. However, the magnitude of the increase was greater than expected and the pattern of increase was unexpected. The species list data type covers the largest extent, and, therefore, given the species-area relationship, would be expected to contribute the most new species to county lists. The GPS data type only records a few specific weeds of concern in a select area, and thus would be expected to add the least number of new species. However, data types such as GPS and quarter quad are important in capturing the noxious weeds, as these data types target those species instead of attempting to capture overall diversity, and the species lists data type missed many of the noxious weeds.

I was surprised to see that more non-native species were added to the high survey intensity counties than the mid or low. It could be that these counties are simply more heterogeneous and therefore more diverse (accounting for area by using density did not change this result). However, this result may not reflect that these counties actually contain more non-native plant species, but could result from other data types following a similar pattern of survey intensity to that assumed for the BONAP data type. Thus, low intensity counties from the BONAP data would also have low survey intensity with the other data types. Boulder County, which had the most species recorded in the BONAP data type, is the site of the state herbarium and the University of Colorado. Hinsdale County, with the least species recorded by the BONAP data type, is in southwestern Colorado isolated from universities and with a population of only 740 people. Boulder County had all five data types, and ranked among the top 10 counties for number of non-native species recorded for each data type. Hinsdale County, meanwhile, only had three data types represented and was among the counties with the fewest non-native species recorded for each data type. These data seem to indicate that areas that are poorly studied according to the BONAP data type are also poorly studied by the other data types, so survey bias seems to be consistent across data types. Research, no matter which data collection methods are used or what objectives exist, seems to focus on the same areas.

The comparison of the number of data types found in each county of different survey intensity also indicates that there are similarities in the concentration of survey efforts across the state. Areas of low intensity according to the BONAP data type had fewer data types present on average than areas of high intensity.

All data types together added many more species than only one data type with the BONAP data type, and as new data types were added the number of non-native species recorded for the county increased. Therefore, using multiple datasets appears to increase knowledge of species' distributions and richness across a landscape.

Although this analysis examined non-native vascular plant species data, similar trends may be found for data including native species and other taxa. Synergy of datasets could provide a cost-effective first step to improve existing species lists of any sort. Other quantitative methods could then be used to further improve the lists. For example, Iverson and Prasad (1998) used regression analysis to predict how many species were missing from county lists in less surveyed regions in Illinois. The results of this analysis would probably be improved if the number of species recorded in the well-surveyed areas were first supplemented by incorporating different available datasets.

These results provide recommendations for improving knowledge of species richness and distributions for data collectors and managers. 1) Data sharing greatly improves "completeness" of knowledge of species richness and distributions. 2) Different data types should be utilized for different objectives: weed mapping data and survey data such as the quarter quad data type are important for noxious weeds, while community level data like vegetation plots or species lists are most important for capturing total numbers of species and non-native species that are not currently considered noxious (but they may be in the future). 3) Survey intensity seems to be biased across all data types, so future data collection efforts should be concentrated in areas of low intensity surveying to fill in these gaps. A comprehensive, systematic data set such as BONAP could serve as a surrogate to determine these low survey intensity areas. 4) Combining data sets is a first step in selecting priority non-native species for control based on their distributions. The combined data also indicates what species most groups are concentrating efforts to map. 5) Data synergy is an important step in creating predictive models of habitats vulnerable to invasion and individual species distributions by increasing the geographic area and environmental types covered by data types. The increased scope of knowledge could change the environmental envelope defining a species distribution.

Much can be gained by sharing data. As illustrated in this paper, land managers data synergy can provide land managers with "watch lists" of species around their land management unit and identify areas lacking survey data. Thus, if our knowledge greatly increases through data synergy, why does it not occur more often? The computing capabilities currently exist to make this type of project a simple task. Collaboration between different agencies and groups could vastly increase the state of our knowledge with minimal effort by all in both time and funds. Greater collaboration and more synthesis of existing datasets are needed for a better understanding of species distributions and richness patterns.

1.5 Literature Cited

D'Antonio, C. M., J. M. Levine, and M. Thomsen. 2001. Ecosystem resistance to invasion and the role of propagule supply: A California perspective. Journal of Mediterranean Ecology 2:233-245.

ESRI. 2002. ArcGIS. Environmental Systems Research Institute, Redlands, California.

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

Iverson, L. R., and A. Prasad. 1998. Estimating regional plant biodiversity with GIS modelling. Diversity and Distributions 4:49-61.

Mack, R. N., D. Simberloff, W. M. Lonsdale, H. Evans, M. Clout, and F. A. Bazzaz. 2000. Biotic invasions: Causes, epidemiology, global consequences, and control. Ecological Applications 10:689-710.

Natural Resources Conservation Service. 2003. Plants Database. Natural Resources Conservation Service, U.S. Department of Agriculture.

Nicholls, A. O. 1989. How to Make Biological Surveys Go Further with Generalized Linear-Models. Biological Conservation 50:51-75.

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.

Vitousek, P. M., C. M. D'Antonio, L. L. Loope, M. Rejmanek, and R. Westbrooks. 1997. Introduced species: a significant component of human caused global change. New Zealand Journal of Ecology 21:1-16.

Wilson, E. O. 1988. Biodiversity. National Academy of Sciences/Smithsonian Institution, Washington, D.C.

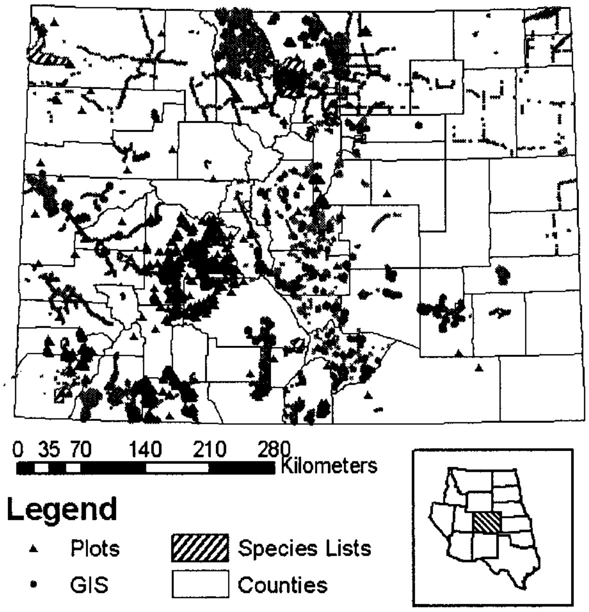


Figure 1-1. Location of data in the state of Colorado. Quarter Quads (QQs) are not pictured, but they form a uniform grid throughout the state.

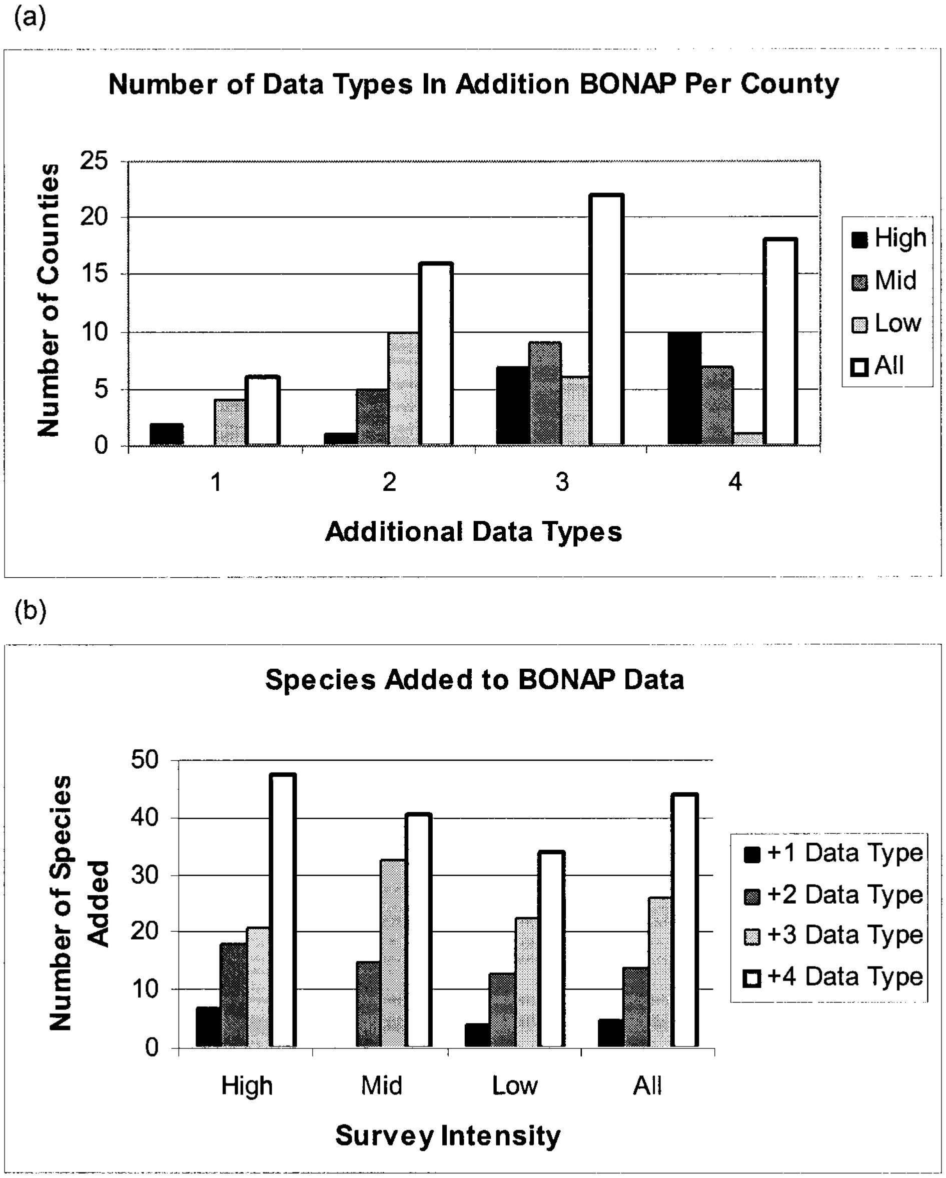
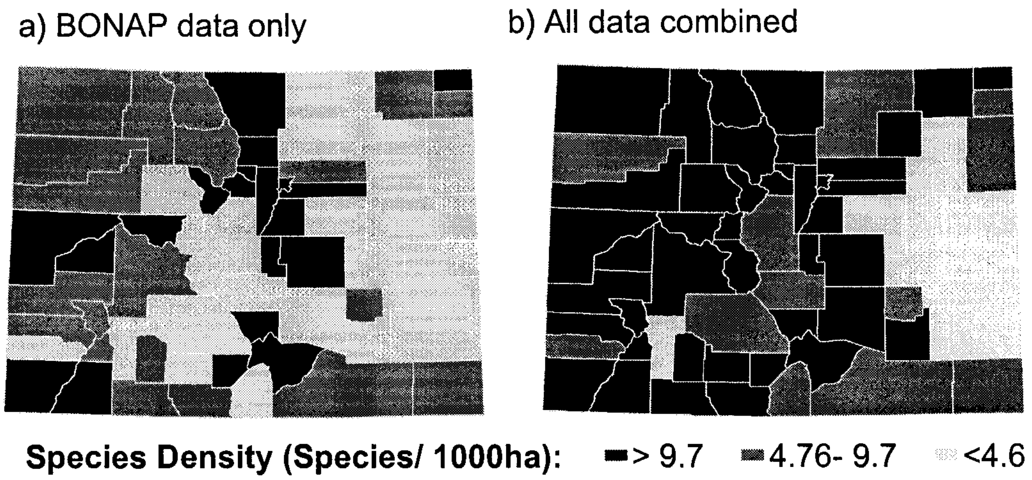


Figure 1-2. (a) Number of data types in addition to the BONAP data type available for each county with counties divided into high, mid, and low survey intensity and all counties together and (b) number of non-native species added to the BONAP list by different numbers of additional data types.


a) BONAP data only

b) All data combined

Figure 1-3. (a) Non-native species density for Colorado counties from the BONAP data type and (b) non-native species density in the counties from all data types combined with counties subdivided into thirds based on density from the BONAP data type. More counties in (b) are in the original upper third range.

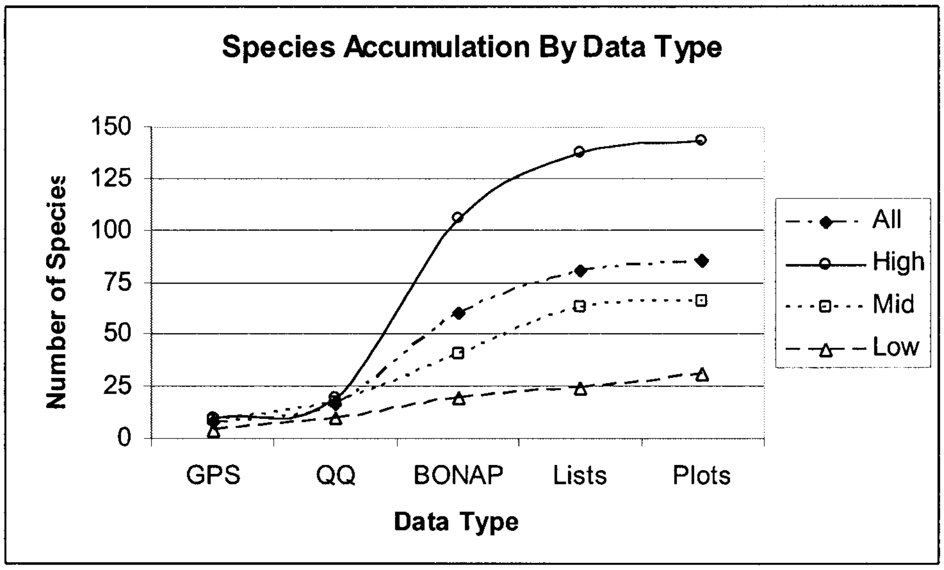


Figure 1-4. Species accumulation curves for different survey intensities of non-native species with the five data types added according to the order a weed manager would generally obtain them for the three survey intensity levels and for all counties together. (GPS = weed mapping data, QQ = state quarter quad survey, BONAP = Biota of North America Program county level species lists, Lists = species lists for large geographic areas, and Plots = vegetation survey plot)

Table 1-1. Summary data for BONAP groupings including number of counties in the group, range of number of non-native species recorded per county in the group, and the mean number (standard deviation) of non-native species recorded for counties in the group, and the standard deviation of this number for each group.

Survey Intensity	Sample Size (n)	Species Number	Mean (standard deviation)
Low	21	2-15	12 (4)
Middle	22	16-37	27 (6)
High	20	41-248	94 (48)
TOTAL	63	2-248	37 (40)

Table 1-2. Summary statistics for the five different data types including range of geographic area covered by the data type, mean number of data type units per county, mean number of units of the data type per county(range), mean number of non-native species recorded per county list (range), and the number of counties represented by the data type.

Data Type	Area range	Mean Number of Units Per County	Mean Species Recorded /County *	Number of Counties Represented
BONAP	28,223–1,235,566 ha	1	37	63
			(2 - 248)	
Quarter Quad	3642 ha	61	13	61
		(5 - 190)	(1 – 37)	
Species List	13 – 21,569 ha	1	49	31
		(0 – 4)	(2 – 102)	
Plot	0.25 -0.1 ha	46	11	34
		(0 - 1050)	(1 – 51)	
GPS	Point – 47.3 ha	3	6	50
		(0 – 17)	(1 – 18)	

* Counties not containing a specific data type were left out of the analysis

Table 1-3. Sample size (N) and mean (range) values for non-native species richness, noxious species richness, unique non-native species, and unique noxious species richness for all five data types and all data types together. Only counties containing a particular data type were included in the individual data type analysis, and all counties were used for the combined data.

Data Type	N	Mean Species Richness	Mean Noxious Richness	Mean Unique Species	Mean Noxious Unique Species
BONAP	63	37	10	24	5
		(2-248)	(0-55)	(2-168)	(0-22)
Quarter Quad	61	13	11	6	5
		(1-37)	(0-35)	(0-25)	(0-23)
Species List	31	49	6	25	5
		(2-102)	(0-21)	(1-50)	(0-17)
Plot	34	11	3	7	1
		(1-51)	(0-13)	(0-22)	(1-6)
GPS	50	6	5	1	1
		(1-18)	(0-15)	(0-7)	(1-7)
ALL	63	63	20	n/a	n/a
		(9-277)	(2-57)		

Table 1-4. A) Sample size (N), mean non-native species richness (range), and mean noxious species richness (range) for each data type added to the BONAP data type by low, mid, and high sampling intensity. B) Sample size (N), percent increase in non-native species richness, and percent increase in noxious species richness for each data type added to the BONAP data type.

A)		Low			Mid			High		All	
Data Type	N	Species	Noxious	N	Species	Noxious	N	Species	Noxious	Species	Noxious
Quarter Quad	20	7	6	21	12	10	20	9	7	10	8
		1-14	0-11		4-30	2-27		1-18	0-16	1-30	0-27
Species List	3	32	3	14	28	1	14	29	3	29	2
		25-45	0-10		1-57	0-13		5-52	0-10	1-57	0-7
Plot	9	8	1	11	5	2	14	7	1	7	1
		2-20	0-4		0-19	0-7		0-22	0-4	0-22	0-14
GPS	14	4	3	19	5	5	17	3	2	4	3
		1-7	0-7		0-17	0-14		0-9	0-8	0-17	0-13
Average Total	22	14	7	21	31	13	20	33	9	24	10
		1-50	0-16		4-71	4-27		2-66	0-23	0-71	0-27

B)		Low			Mid			High		All	
Data Type	N	Species	Noxious	N	Species	Noxious	N	Species	Noxious	Species	Noxious
Quarter Quad	20	43	61	21	31	57	20	13	27	29	49
Species List	3	74	21	14	46	6	14	29	14	41	11
Plot	9	44	25	11	16	19	14	8	4	20	14
GPS	14	25	47	19	17	38	17	6	13	16	32
Average Total	22	51	61	21	51	63	20	30	33	44	53

CHAPTER 2 : PREDICTING SPECIES RICHNESS AND IDENTIFYING HOTSPOTS OF INVASION AT FINE RESOLUTION FOR THE STATE OF COLORADO

2.0 Abstract

Knowledge of species rich areas of both native and non-native species is important for conservation purposes and prioritizing invasive species control efforts. Data were obtained from a database containing over 30 datasets for the state of Colorado with 586 locations with species richness information. These data along with environmental and topographic data at a 30-m^2 resolution were input into an ordinary least squares regression for native species and a generalized linear model for non-native species. Spatially autocorrelated residuals from both were then kriged. The final surface was a combination of the two. The native species richness model had an r^2 of 0.21 whereas the non-native species richness model had a D^2 of 0.32. Overlaying the non-native richness model with vegetation classes of the state indicated that populated areas were the most highly invaded. For natural areas, moist areas such as wetlands and riparian zones were the most highly invaded, while high elevation sites were the least invaded. These techniques, in combination with synergized datasets, can provide important information on patterns of invasion at large spatial extents, while maintaining a relatively fine spatial grain.

2.1 Introduction

Introductions of non-native species are increasing in frequency and extent as human movements become more global and international trade increases. This globalization has caused a breakdown in the regional distinctiveness of the Earth's flora and fauna. Humans have become the most important mechanism for plant dispersal and carry plants to areas where they would not arrive without anthropogenic means (Mack and Lonsdale 2001). Additionally, species known to have dispersed great distances without human aid have not become detrimental (Mack and Lonsdale 2001). Many non-native species are not harmful, but some can become the cause of serious ecological and economic degradation. Non-native species invasions decrease human economical wealth by impacting agricultural land, rangeland, and forests; altering ecosystem function; and threatening native biodiversity that is important for economic, environmental, and ethical reasons (Vitousek et al. 1997, Mack et al. 2000).

For example, Pimentel et al. (2000) estimate that the economic losses and environmental damage from invasive non-native species total over $137 billion per year in the United States. This number and the magnitude of the associated ecological effects is difficult to calculate, but even if it is off by a factor of five or ten, the amount is still very large. In Hawaii Volcanoes National Park, the non-native nitrogen fixing plant, *Myrica faya* Ait., has been shown to alter ecosystem properties by changing the nutrient dynamics of the system, perhaps allowing plants that would have been excluded due to limited nitrogen to enter the area and more successfully compete with native species that are better adapted to lower nitrogen availability (Vitousek et al. 1987, Vitousek 1990). Non-native species can also dramatically alter the fire regime of a system, as *Bromus* *tectorum* L. *(*cheatgrass) has done in the western United States (D'Antonio and Vitousek 1992). *Euphorbia esula* L. (leafy spurge), a major problem in areas such as Theodore Roosevelt National Park, North Dakota, is a non-native plant species that has become established in the Northern Great Plains. This plant displaces native species, thereby decreasing herbivore habitat, and causes economic losses to farmers and ranchers whose land it occupies. *Euphorbia esula* can out compete native species because of its extensive root system (Ringwall et al. 2000), grazing tolerance mechanisms (Olson and Wallander 1999b), grazing avoidance mechanisms (Trammell and Butler 1995, Lym 1998), and possible allelopathic effects (Steenhagen and Zimdahl 1979). Many non-native species such as the latter two form monocultures that decrease biodiversity. The impact of non-native species introductions on our ability to conserve native species is clear; almost half of the threatened and endangered species listed under the Endangered Species Act are thought to be listed due to competition with or predation by non-native species (Wilcove et al. 1998).

The change in the rate of introductions because of increased human mobility exacerbates these problems because new introductions continue to occur before a system has time to adapt to previous introductions of non-native species. If the time between disturbance events decreases to an interval shorter than the time to recovery, then major components of the entire system may change. It is this increase in the rate of disturbances and the distance of non-native species introductions that is currently alarming many scientists.

Only a small fraction of introduced species survive, and of those an even smaller percent become dominant species that can alter community structure and function (Mack et al. 2000). Numerous studies have shown that non-native species richness is highest in areas of native species richness (Robinson et al. 1995, Planty-Tabacchi et al. 1996, Stohlgren et al. 2001, 2003), indicating that non-native species prosper under the same environmental conditions as native species. However, high species richness does not necessarily equate with high impact. Greater non-native species richness is correlated with higher non-native species cover. If non-native species survive best in areas with high water and nutrients, then these areas may serve as the original source for the spread of the high impact species. Additionally, as more species are added to an area, the probability of one arriving that is a high impact species increases (Huston 1997).

Because lag times often exist before a species becomes invasive, it is also important to monitor non-native species to catch an invasion as it begins. With early detection eradication may be possible and control is less costly (Rejmanek 2000). Therefore, knowledge of where species rich areas are located is important because they may be the most likely place for a new high impact species to initially establish, specific traits of species may be selected for, and they may harbor species in their lag period.

In this chapter, I will describe predictive models for non-native species richness and native species richness for the state of Colorado. I suspect that using data from different sources will make it easier to generate a useful model of species richness for a large area. I will then use these models to examine patterns of invasion across the state, identify hot spots of invasion, and determine vegetation types vulnerable to invasion. I hypothesize that the areas most invaded will be human population centers because they can act as sources for new invasive species and that high elevation areas will be least invaded due to harsh environmental conditions and isolation from sources.

2.2 Methods

2.2.1 Floristic data

Datasets containing information on non-native plant species were collected from many research groups throughout the state of Colorado – local to national, governmental and non-governmental - and were synthesized based on common attributes into a spatial database using SQL Server and ArcSDE (See Chapter 1, Appendix B). The Natural Resource Conservation Service (NRCS) plant codes from the U.S. Department of Agriculture plants database were used to standardize species names as different datasets included synonyms for the same species (USDA NRCS 2004).

Species area relationships indicate that as the survey area increase, the number of species recorded increase. However, the rate of change is dependent on the specific area being surveyed because some areas are more species rich than others and more heterogeneous than others. Therefore, for this analysis only vegetation plot data were used that could reasonably be extrapolated to a 30-m x 30-m grid of the state. This is the largest approximate grid cell size that plant species community data was collected. For example, Modified Whittaker plots (Stohlgren et al. 1995) measure 20-m x 50-m, so their species richness could be extrapolated to a 30-m x 30-m grid. However, Daubenmire plots (Daubenmire 1959) are only 0.1-m^2 each, so extrapolating information from them to 30m x 30m cannot be assumed to be accurate for either the total number of non-native species or the absence of individual species. Therefore, the small plot size would result in overlooking some species due to patchy species distributions (Stohlgren et al. 1998a). Thus, the number of species found in this plot type would be an inaccurate representation of species richness for the grid cell.

Five hundred eighty-six vegetation plots located throughout Colorado met the above criteria (Figure 2-1). Total number of native and non-native species and geographic coordinates (with projection Universal Transverse Mercator Zone 13 North, datum NAD 1983) were extracted from the database for each plot using NRCS definitions for nativity to the United States (USDA NRCS 2004).

2.2.2 GIS data and remotely sensed imagery

Ancillary data were chosen for the models based on an extensive review of invasion literature (Appendix A). Geographic Information System (GIS) layers representing the ancillary data were collected from various sources (Appendix C). These layers were processed to 30-m rasters for the state containing the variable of interest (Table 2-1, Figure 2-2 and 2-3). Some of the ancillary data were originally at a coarser resolution than 30-m because most large extent datasets are not available at the fine resolution of the vegetation data. In all cases the best available data were used.

Landsat 7 Thematic Mapper data were also obtained for the whole state and used to create various indices related to vegetation at a 30 m resolution (Figure 3). Storing individual bands requires a large amount of storage space, and using these indices reduced the data size. These indices also are biologically meaningful representations of the satellite data captured in the different bands. A mosaic of images from 13 August 1999 to 20 September 2002 was used to generate tassel cap 1 to 4 (weighted sums of thermal bands) and Normalized Difference Vegetation Index (NDVI) by NASA Goddard Space Flight Center, Greenbelt, MD. Tassel cap band 1 is interpreted as brightness (information on bare soil) with negative values indicating no bare soil and very large values indicating lack of vegetation. Tassel cap band 2 is designated as greenness (information on abundance and vigor) with negative values indicating lack of vegetation and very high values representing lush areas. Tassel cap band 3 is yellowness or wetness (reflecting the difference between the red and green bands) with low values representing low moisture content and high values high moisture content (Campbell 2002). NDVI is calculated using the infrared and red bands that contain information on vegetation as a measure of photosynthetic activity and biomass.

For all maps and analyses, the Universal Transverse Mercator (UTM) zone 13 north projection with North American Datum 1983 was used, as a majority of the state falls withint this zone. Data for each of the 586 raster cells with plant species community data were extracted from each of the 12 ancillary data coverages.

2.2.3 Statistical analyses

The data were analyzed using S-plus statistical software (Statistical Sciences 2000) with Reich and Davis's (Reich and Davis 1998) spatial library and ArcGIS (ArcGIS 2002, Appendix F). I used trend surface analysis and a geostatistical method to capture both large and small-scale variability. By using these methods, the problems arising with classical statistics caused by spatial autocorrelation in the data were avoided and the natural spatial dependency was taken into account. The coarse resolution, large scale trend is typically captured by some form of regression. The geostatistical surface calculated using the residuals from the trend surface analysis captures the fine scale patterns. In this case, ordinary kriging or co-kriging was used as a surrogate for species' dispersal, which is particularly important for non-native species that are still spreading and have not reached their full potential distribution. The trend surface removes the large-scale trends in the data, alleviating problems with the kriging assumption of stationarity in the data. This removal is especially important for a model covering a broad spatial extent, where environmental relationships may change across the landscape. Additionally, by using these two methods processes affecting species distributions at different scales are captured – the coarse scale trend across the landscape and the fine scale variability found in the spatial relationships between data points (kriging) or in those spatial relationships between data points relationship with particular ancillary variables (co-kriging).

The trend surface analysis methods differed for native and non-native species richness because the two variables had different distributions associated with them. For native species richness inputs to the trend surface analysis included the spatial weights matrix, which weights points closer together via Euclidean distance less than points farther apart, and covariates selected from table 2-1 using combinatorial screening. The spatial distance weights matrix is a matrix of distances between all point locations that is used in calculating spatial autocorrelation in the data. It is calculated as the Euclidean distance between two points to the -1 power. Combinatorial screening compares all possible combinations of predictor variables and selects the combination with the minimum Aikaike's Information Criterion for small sample sizes (AICC, Burnham and Anderson 1998). This value weights the likelihood that one variant better explains the variation in the data than another. Examination of a histogram of the distribution of native species richness values indicated an approximately normal distribution (MathSoft 1999), so ordinary least squares (OLS) regression was used to create the trend surface for native species richness. The distribution was not a perfect bell shaped curve, but it was not noticeably skewed in either direction and ecological data generally does not follow a perfect normal distribution.

A generalized linear model (GLM) for a Poisson distribution with a logarithmic link function was appropriate for non-native species richness. This type of model is favored for count data such as non-native species richness that are highly skewed towards zero (nearly 20% of plots had a zero value and the histogram was skewed left, McIntyre and Lavorel 1994, Luoto et al. 2002). GLM assumes transformation of an expected response is a linear function, but unlike least squares methods allows for different distributions such as Poisson to be specified. The log link function is simply a logarithmic transformation of the regression equation. The step procedure for GLM in S-plus was used to select the best set of predictors to be used in the model using AICC. Estimates of regression parameters are generated using maximum likelihood produced by iteratively reweighted least squares. Examination of a histogram of the residuals from this process indicated an approximately normal distribution and therefore were treated the same as those from the ordinary least squares regression for native species richness.

If the residuals from the trend surface analysis were autocorrelated as shown by the Moran's *I* statistic, then ordinary kriging or co-kriging was performed using the ArcGIS kriging function (ArcGIS 2002). This statistic rather than Geary's *c* was used as Moran's *I* is more robust to departures from the normal distribution (Cliff and Ord 1973). The Moran's *I* statistic typically ranges in values between -1.0 and 1.0, with values of zero indicating little spatial autocorrelation and a random pattern, positive values indicate a smooth clustered pattern with similar values close together, and negative values indicating a contrasting, checkerboard pattern with similar values scattered across an area. ArcGIS was used for the purposes of this example because of current computational limitations of statistical packages like S-plus. This software was the only software package I found capable of performing the calculations at this extent and resolution. Other readily available algorithims crashed with this size dataset. Co-kriging would be used when the residuals are cross-correlated with an ancillary variable. Moran's *I* is actually a special case of the cross-correlation statistic looking at autocorrelation for a single variable, whereas the cross-correlation statistic examines the spatial relationship between two variables (Kalkhan and Stohlgren 2000). However, the non-native species residuals and native species residuals were not cross-correlated, so this will not be discussed further.

Inputs to the kriging process, including the nugget, sill, and range of the variogram and the number of nearest neighbors to use (see table 2-2) were determined using Splus. A variogram was estimated for each of the dependent variables and the type of variogram (i.e., Gaussian, spherical, or exponential) was chosen using AICC. The difference between the Gaussian and spherical models were minimal (AICC differed by less than 0.01 for non-native species residuals; table 2-3), so the Gaussian model was used. The number of nearest neighbors used in kriging was determined by minimizing the residual error and maximizing the R^2 value the using cross validation function in Reich and Davis's spatial library for S-plus, a process that predicts a random point of the data using all the other data, repeating for all points (see Table 2-4). The final surface was created by adding together the trend surface and the kriged surface rasters using the ArcGIS raster calculator. The trend surface was generated using raster calculator and the regression equation with the regression coefficients multiplied by the predictor variable surface.

A Colorado vegetation map was used to determine habitats most vulnerable to invasion. Theobald et al. (2003) generated a 90-m resolution vegetation type grid for the state from the National Land Cover Dataset and ancillary coverages related to vegetation. The state's vegetation was classified into 48 different types (Appendix D). Zonal statistics in ArcMap's spatial analyst extracted the mean, maximum, and standard deviation for non-native species richness from the final surface for each vegetation type.

2.2.4 Model Evaluation

I performed a rough model evaluation, but it is difficult to evaluate a predictive model for non-native species because many species have not reached their full distribution potential and are still spreading. Thus, datasets of current distribution (the only datasets currently available for evaluation) might differ from the future distribution being predicted. Three datasets obtained after the project was begun were used to evaluate the model output. These datasets included 20 Modified Whitaker vegetation plots from the Hayman burn area (burnt summer 2001), 37 Modified Whitaker vegetation plots from the Gunnison basin area, and 13 Modified Whitaker vegetation plots from the Front Range foothills. All plots were surveyed in during summer 2003. Predicted values were extracted from the both the GLM model and from the spatial GLM model for each observed location, and the values were compared.

2.3 Results

2.3.1. Model parameters

The summary statistics for the variables cover a wide range, with the mean value being close to the middle of the range for most variables (Table 2-1). Comparing the summary statistics for field data points (Table 2-1) and the whole state (Appendix C), the 586 data points used to generate the model appear to cover the extent of variation in the auxiliary variables in the state. Field data are biased towards areas close to roads (mean of 8.5 m compared to state mean of 1500 m). Also, areas with high soil fractions especially of clay were absent from sampling (compare maximum values for the state (54% - Appendix C) and for the sample (38% - Table 2-1). Sampling was also biased towards natural areas with low population densities.

2.3.2 Species richness models

The variables selected for regression differed for native and non-native species, including the number of variables chosen (Table 2-2). There were 11 coefficients selected for non-natives richness, with average annual precipitation (-0.16), clay soil fraction (0.023), and human population density (0.017) having the strongest effect. These three parameters in the model were at a coarser resolution than the dependent data set. The deviance explained in the non-native species richness GLM was 31%. Seven variables were selected for native species richness, and slope and sand soil fraction had the strongest effect on native species richness. The r^2 value for the native species richness OLS was 0.13. Although significant, this model does not explain much of the variation.

The residuals from the general linear model for non-native species and ordinary least squares regression for native species approximated a normal distribution and exhibited significant positive spatial autocorrelation (Moran's I = 0.2 and 0.21, respectively; p-value < 0.001; Table 2-4), so kriging of the residuals was performed. The residuals were not cross-correlated with any of the independent variables not used in the regression modeling (Appendix G).

The semi-variogram model parameter estimates reveal information about the spatial dependency of species richness across the landscape (Table 2-3). Both native and non-native species richness had relatively low ranges, while the nugget value for native species was higher than non-natives. The low range values indicate small, patchy distributions of richness across Colorado, and the relatively low nugget value indicates that there is not much unexplained noise in the data for non-native and native species richness.

The final surface for non-native species richness had values ranging from zero to more than 32 non-native species. The areas with greatest concentration of non-native species were populated areas, particularly along the Front Range, while those areas with the fewest species were concentrated in the high elevation mountain regions. The r^2 value from the regression increased with kriging, while the error decreased for native species richness (Table 2-4). The final model for native species richness still explained only 20% of the variance. Because of the poor performance of this model, no further analyses were performed on it.

2.3.3 Zonal Statistics

The invasion of vegetation types of Colorado indicates that there are differences in invasion success thus far in the state (Appendix D). The mean values reported ranged from 19 species in low level residential areas to zero species in snow/ ice areas. The most heavily invaded types are those associated with human population centers, while the least invaded areas are alpine areas followed by sub-alpine areas. The greatest invasion in natural systems occurred in wetland zones and other high moisture areas according to the predicted surface. A visual analysis of the final surface clearly indicates the greatest non-native species richness in population centers, particularly along the Front Range, with the observer being able to pick out individual large cities.

2.3.4 Model Evaluation

The predicted values from the model on average differed from the 70 observed values by 2.0 non-native species for both the GLM and the spatial GLM (Table 2-5; Appendix E). Both the spatial and non-spatial models underestimated the number of non-native species, with the observed maximum from the plots being 13 and the two GLMs predicting a maximum of six non-native species. The standard deviations for all were similar, and the averages were fairly close. However, the greatest difference in predictions by the two models was in the average, with the GLM predicting 2.6 species and the spatial GLM predicting 4.0. The observed average was 3.2.

2.4 Discussion

The parameters with the greatest influence on the non-native species richness model were interesting. Non-native species richness is correlated with moisture as evidenced by the high level of invasion in wetland and riparian areas (Planty-Tabacchi et al. 1996, Stohlgren et al. 1998b). However, this variable may have been chosen for a different reason. In Colorado, high elevation areas receive the most precipitation and some of the driest areas are those low in elevation, and non-native species invasion is inversely related to elevation. Clay soil fraction was another influential predictive factor. Clay soils retain more moisture than sandy or silty soils because of the small particle size (Brady and Weil 1999). Thus, clay soil fraction may be more important in relating soil moisture to invasion potential rather than actual precipitation's relationship to invasion potential.

The examination of model parameters indicated a bias towards some variables (more sample sites close to roads being the most biased). These biases may have some effect on the model. However, the regression coefficient for distance from roads was quite small, indicating that this factor influenced the final model very little. Additionally, random sampling is not an assumption in spatial statistical modeling as it is for classical statistical techniques. In fact, for spatial statistics it is important to capture the heterogeneity across the landscape and this is often accomplished with subjective sampling. Sampling intensity needs to be greater in heterogeneous areas than in homogeneous areas. With regards to non-native species, road corridors are probably more heterogeneous because they are corridors to invasion.

The final model of non-native species richness for the state of Colorado fits the results from different studies of invasion. Stohlgren et al. (1998b, 1999a, 2001, 2002) found that areas with the highest non-native species richness were those most heavily invaded (riparian areas) and low elevation sites were much more invaded than high elevation sites in observational studies in Colorado and elsewhere. The most heavily invaded natural areas in the modeled surface are vegetation types with the greatest species richness such as wetland areas. These areas constitute natural area hotspots of invasion.

The greatest hotspots of invasion in Colorado as predicted by the models were mainly population centers along the Front Range, where humans have planted and cultivated many non-native species. These hotspots of invasion are found in areas that are also generally highly disturbed. These areas immediately around population centers should be monitored for indications of new high impact invasive species, both ones indicating high invasive tendencies in other states and new species exhibiting invasive characteristics here. These areas have the potential to act as sources for the spread of new invasive species in the state.

The model for non-native species richness was fairly accurate when compared with the 70 test plots. Although it did not produce the exact number of species found, it was typically off by a maximum of two species. The basic statistics describing the observed dataset and the predicted dataset were all very close with the exception of the maximum number of species recorded. The models under-estimated the number of species at the most invaded points in the observed dataset. However, the evaluation dataset covered only three very small areas of the state. It is unknown how well the model performed in other areas of the state, particularly under-sampled areas such as the Eastern plains. Again, however, it must be remembered that evaluation of a predictive model for non-native species richness is difficult as many of there species are continuing to spread and there are other potential invaders that have not yet reached Colorado. Thus, the relative difference in number of species predicted rather than the actual numbers predicted may be the most important result from the model.

The model of native species richness did not explain much of the variation in the data. The poor performance of the model could result from important explanatory variables being left out. Additionally, the factors associated with variation in native species richness could be acting at a different spatial scale than the one being used here. Distributions of native species have been determined by long evolutionary periods, unlike recently introduced non-native species, and therefore their distributions across the state may be greater (Ricklefs 2004). Other attempts to model species richness across large extents at fine resolution have not been successful, although there are many examples at coarser resolutions such as $0.25km^2$ by Luoto et al. (2002) and $1km^2$ by Andreone et al. (2003). This failure may be due to processes controlling native plant distributions at these scales not being captured by the predictor variables being used or the scale of the predictor variables not being appropriate. Additionally, because of their long evolutionary histories, it may be more difficult to find variables based off relatively current environmental conditions to predict their occurrence.

The models would probably be improved with higher resolution ancillary data. For example, soils are typically a fine-scale indicator of plant species richness, but the resolution of the available data was very coarse. This problem, along with the coarse resolution of other possible predictor variables, may be improved with the use of fine-resolution remotely sensed imagery parameters such as tassel cap and NDVI and the different bands themselves. However, fine resolution data are currently costly and highly computationally intensive. The resolution of more readily available data like Landsat

Thematic Mapper is still 30-m^2, which may not be the most ecologically meaningful scale to model species richness.

Additionally, plot data more representative of the entire state might improve the model. The eastern plains of Colorado had very sparse field data (Figure 2-1), and although this area is also one of the more homogeneous areas of the state and may not require as many field sites to capture the variability present, more data points than currently available may help model performance. The model was weakened by including the Eastern plains, but this was done to examine species richness trends throughout the whole state and to help identify areas in the state for future field survey efforts. These models, especially native species richness which performed very poorly, might best be used to guide future sampling efforts in the state. Areas with the greatest model uncertainty should first be targeted for sampling and then an iterative procedure followed. These initial models were not expected to be completely accurate, but were intended help focus statewide efforts to combat the problems of non-native species. Once further sampling has been conducted in areas identified in the initial models, the models can be re-run and the same analysis process followed with each new set of models being refined based on new field data collected based on the previous model version.

The database containing datasets from numerous studies was an important first step in being able to create a model of this resolution and extent. It also assists the iterative process, enabling the model to be easily updated as more information is collected by various groups and input into the database. The updated models again indicate poorly sampled areas with high uncertainty were researchers and managers should focus further sampling efforts. At least for non-native species richness, these results indicate that existing datasets can be merged and together create a model to examine patterns of invasion across the state, help determine hotspots for future monitoring, and identify areas for future data collection efforts.

2.5 Literature Cited

Andreone, F., F. Glaw, R. A. Nussbaum, C. J. Raxworthy, M. Vences, and J. E. Randrianirina. 2003. The amphibians and reptiles of Nosy Be (NW Madagascar) and nearby islands: a case study of diversity and conservation of an insular fauna. Journal of Natural History 37:2119-2149.

ArcGIS. 2002. ArcGIS. Environmental Systems Research Institute, Redlands, California.

Brady, N. C., and R. R. Weil. 1999. The nature and properties of soils, 12th edition. Prentice Hall, Upper Saddle River, N.J.

Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer, New York.

Campbell, J. B. 2002. Introduction to remote sensing, 3rd edition. Guilford Press, New York.

Cliff, A. D., and J. K. Ord. 1973. Spatial autocorrelation. Pion, London.

D'Antonio, C. M., and P. M. Vitousek. 1992. Biological invasions by exotic grasses, the grass/fire cycle, and global change. Annual Review of Ecology and Systematics 23:63-87.

Daubenmire, R. F. 1959. Canopy coverage method of vegetation analysis. Northwest Science 33:43-64.

Huston, M. A. 1997. Hidden treatments in ecological experiments: Re-evaluating the ecosystem function of biodiversity. Oecologia 110:449-460.

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.

Luoto, M., T. Toivonen, and R. K. Heikkinen. 2002. Prediction of total and rare plant species richness in agricultural landscapes from satellite images and topographic data. Landscape Ecology 17:195-217.

Lym, R. G. 1998. The biology and integrated management of leafy spurge (Euphorbia esula) on North Dakota rangeland. Weed Technology 12:367-373.

Mack, R. N., and W. M. Lonsdale. 2001. Humans as global plant dispersers: getting more than we bargained for. BioScience 51:95-102.

Mack, R. N., D. Simberloff, W. M. Lonsdale, H. Evans, M. Clout, and F. A. Bazzaz. 2000. Biotic invasions: Causes, epidemiology, global consequences, and control. Ecological Applications 10:689-710.

MathSoft. 1999. S-PLUS 2000 Guide to Statistics, Volume 1. Data Analysis Products Division. MathSoft, Inc., Seattle, Washington.

McIntyre, S., and S. Lavorel. 1994. Predicting Richness of Native, Rare, and Exotic Plants in Response to Habitat and Disturbance Variables across a Variegated Landscape. Conservation Biology 8:521-531.

Olson, B. E., and R. T. Wallander. 1999. Carbon allocation in Euphorbia esula and neighbors after defoliation. Canadian Journal of Botany 77:1641-1647.

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

Planty-Tabacchi, A.-M., E. Tabacchi, R. J. Naiman, C. DeFerrari, and H. Decamps. 1996. Invasibility of species-rich communities in riparian zones. Conservation Biology 10:598-607.

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.

Rejmanek, M. 2000. Invasive plants: approaches and predictions. Austral Ecology 25:497-506.

Ricklefs, R. E. 2004. A comprehensive framework for global patterns in biodiversity. Ecology Letters 7:1-15.

Ringwall, K., M. Biondini, and C. Grugiel. 2000. Effects of nitrogen fertilization in leafy spurge root architecture. Journal of Range Management 53:228-232.

Robinson, G. R., J. F. Quinn, and M. L. Stanton. 1995. Invasibility of experimental habitat islands in a California winter annual grassland. Ecology 76:786-794.

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

Steenhagen, D. A., and R. L. Zimdahl. 1979. Allelopathy of leafy spurge (Euphorbia esula L.): breeding system analysis. Canadian Journal of Botany 77:1684-1688.

Stohlgren, T. J., D. T. Barnett, and J. T. Kartesz. 2003. The rich get richer: patterns of plant invasions in the United States. Frontiers in Ecology 1:11-14.

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., 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. Falkner, and L. D. Schell. 1995. A Modified-Whittaker Nested Vegetation Sampling Method. Vegetatio 117:113-121.

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

Stohlgren, T. J., Y. Otsuki, C. A. Villa, M. Lee, and J. Belnap. 2001. Patterns of plant invasions: a case example in native plant species hotspots and rare habitats. Biological Invasions 3:37-50.

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

Theobald, D. M., N. Peterson, and W. H. Romme. 2003. The Colorado Vegetation Model: Using National Land Cover Data and Ancillary Spatial Data To Produce A High Resolution, Fine-Classification Map of Colorado (v1.0). Unpublished Report, Natural Resource Ecology Laboratory, Colorado State University.

Trammell, M. A., and J. L. Butler. 1995. Effects of exotic plants on native ungulate use of habitat. Journal of Wildlife Management 59:808-816.

USDA NRCS. 2004. The PLANTS Database, Version 3.5 (http://plants.usda.gov). National Plant Data Center, Baton Rouge, LA 70874-7790 USA.

Vitousek, P. M. 1990. Biological invasions and ecosystem processes: towards an integration of population biology and ecosystem studies. Oikos 57:7-13.

Vitousek, P. M., C. M. D'Antonio, L. L. Loope, M. Rejmanek, and R. Westbrooks. 1997. Introduced species: a significant component of human caused global change. New Zealand Journal of Ecology 21:1-16.

Vitousek, P. M., L. R. Walker, L. D. Whitaker, D. Mueller-Dombois, and P. A. Matson. 1987. Biological invasion by Myrica faya alters ecosystem development in Hawaii. Science 238:802-804.

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.

Table 2-1. Summary statistics, including minimum and maximum values, mean and standard deviation for the dependent and independent variables per grid cell at the point locations (n=586).

Attribute	Minimum	Maximum	Mean	Standard Deviation
Number of native species per plot	1	79	32.1	14.9
Number of non-native species per plot	0	32	3.7	4.1
Absolute aspect (°)	0.2	180	85.8	53.5
Slope (°)	0	66	11.9	9.4
Elevation (m)	1394	3762	2502.1	470.9
Sand fraction (%)	0	75	54.7	9.9
Silt fraction (%)	0	43	31.8	5.1
Clay fraction (%)	0	38	12.9	5.6
Average annual precipitation classes	1	10	6.0	1.4
Population density per zip Code (people/km)	0	88	8.2	7.0
Distance from road (m)	< 30	7625	1045.5	1060.8
Distance from stream or river (m)	< 30	2160	482.0	389.3
Normalized Difference Vegetation Index	-109	295	134.1	51.9
Tassel Cap 1	-51	321	77.8	54.4
Tassel Cap 2	-13	253	137.5	40.7
Tassel Cap 3	-17	305	164.7	61.1

Table 2-2. Coefficients for regression for (a) non-native species richness and (b) native species richness for the whole state of Colorado.

(a)

Variable	Coefficient	t-value
Intercept	2.88	9.8
Absolute aspect	0.001	1.5
Slope	-0.0095	-1.9
Elevation	-0.00052	-3.6
Clay	0.023	4.1
Average annual precipitation	-0.16	-3.6
Population density	0.017	3.4
Distance from roads	-0.0001	-2.2
Distance from stream or river	-0.0003	-2.8
Tassel cap 2	-0.0028	-1.7
Tassel cap 3	0.0019	2.5
NDVI	0.0031	2.83

(b)

Variable	Coefficient	P-value
Intercept	29.74	0.0001
Slope	0.13	0.04
Sand	0.21	0.002
Distance from roads	0.0015	0.01
Tassel Cap 1	-0.09	0.002
Tassel Cap 2	0.076	0.005
Tassel Cap 3	-0.065	0.01
NDVI	-0.037	0.04

Table 2-3. Semi-variogram values for the regression residuals for non-native and native species richness.

Regression Model	Nugget	Sill	Range	Semivariogram Model	AICC
Non-native	0.5	2.1	377.1	Gaussian	110.2
species	0.5	2.1	450.6	Spherical	110.2
richness	0.2	2.1	360.2	Exponential	132.9
Native	4.3	120.7	122	Gaussian	286.7
species	8.7	125.7	322.7	Spherical	286.3
richness	0	133.9	198.5	Exponential	306.7

Table 2-4. Results from the regression analysis and from the final surface including regression and kriged residuals for non-native and native species richness.

Non-native Species:	
Null deviance	2230
Residual deviance	1512.3
D^2 (deviance reduction)	0.32
Moran's *I* (residuals)	0.2 (p>0.001)
Nearest Neighbors	10

Native Species:	
R^2	0.13
Residual Standard Error	13.9
Lagrange Multiplier (p-value)	186.4 (p>0.001)
Moran's *I* (residuals)	0.21
Nearest Neighbors	4
R^2 (kriged)	0.21
Standard Error (kriged)	13.2

Table 2-5. Results from comparing species richness data from 71 vegetation plots to the predicted numbers from the Spatial GLM and GLM. Mean difference is the difference between the value predicted by the model and the observed value by the 70 plots at their grid cells.

	Standard Deviation	Maximum	Minimum	Mean	Mean Difference
Observed	2.1	13.0	1.0	3.2	N/A
Spatial GLM	2.3	6.0	0	4.0	2
GLM	1.9	6.0	1.0	2.6	2

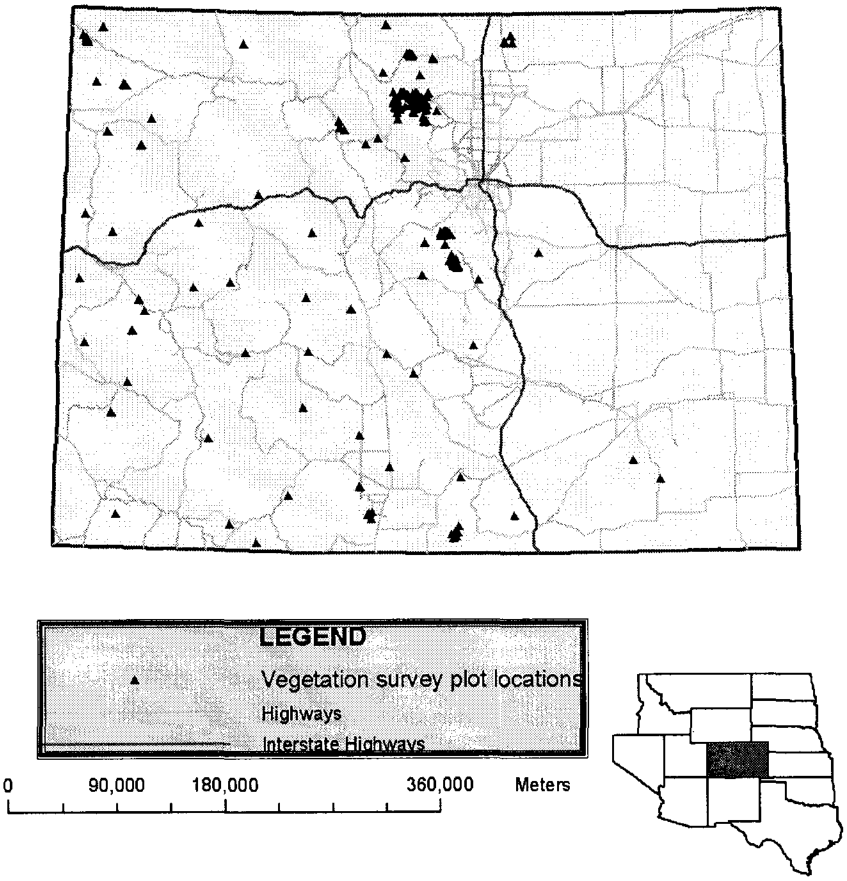
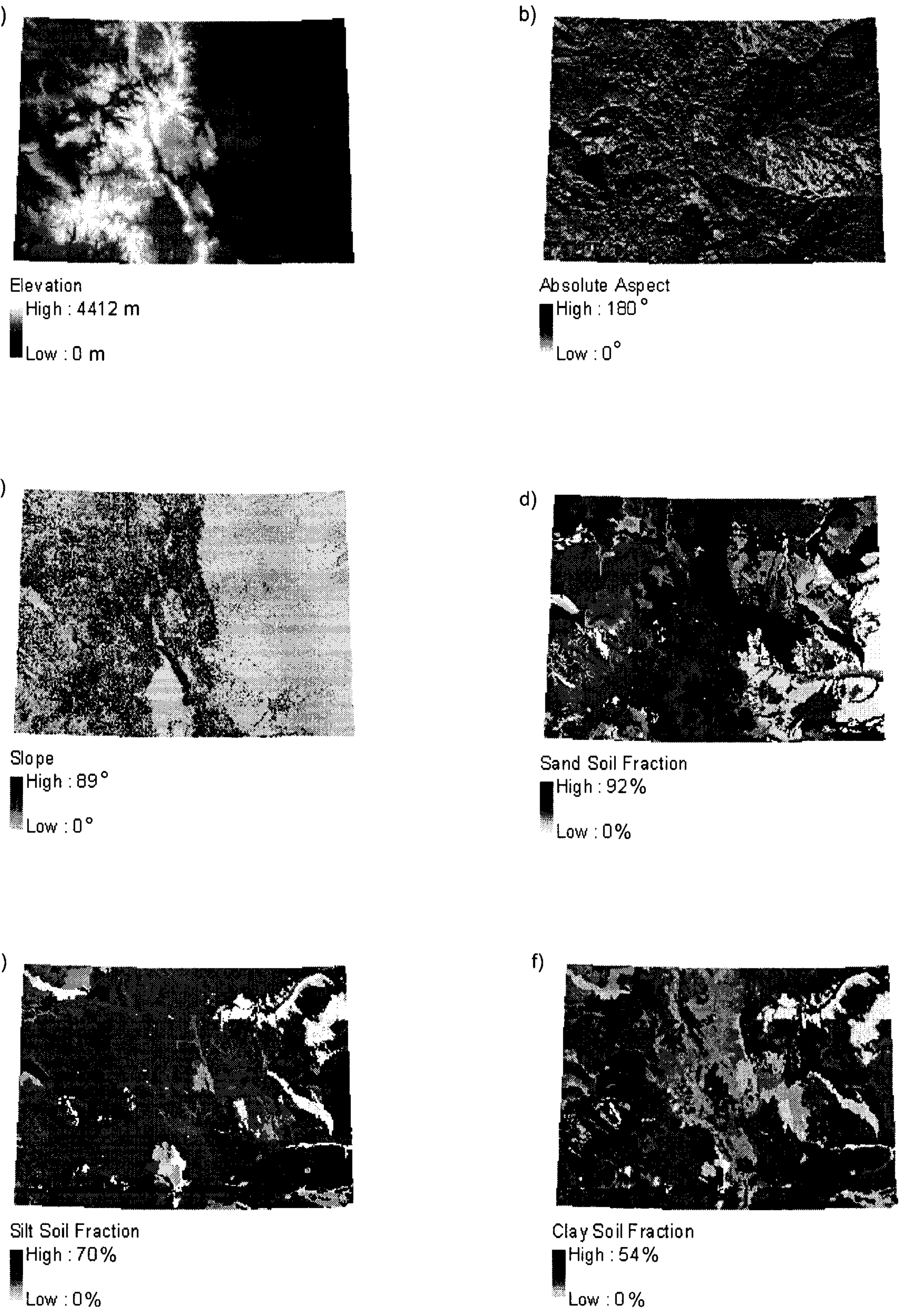
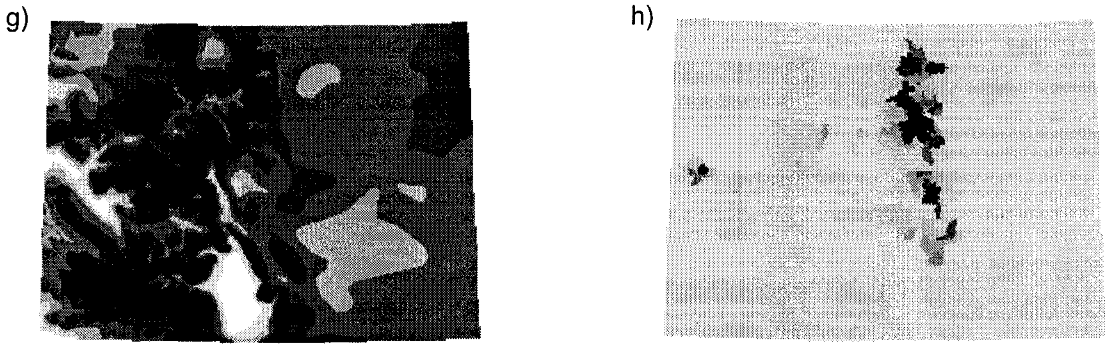


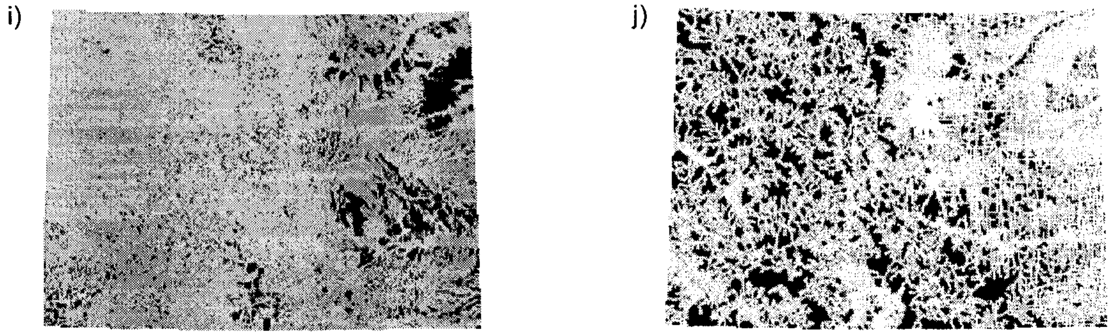
Figure 2-1. Locations of vegetation survey plots in Colorado with native and non-native plant species richness information.


a)
Elevation
High : 4412 m
Low : 0 m
b)
Absolute Aspect
High : 180°
Low : 0°
c)
Slope
High : 89°
Low : 0°
d)
Sand Soil Fraction
High : 92%
Low : 0%
e)
Silt Soil Fraction
High : 70%
Low : 0%
f)
Clay Soil Fraction
High : 54%
Low : 0%



Average Annual Precipitation
High : 60 - 70 cm
Low : 0 - 7 cm

Population Density by Zip Code
High : 6567 people/km^2
Low : 0 people/km^2



Distance from rivers and streams
High : 26859 m
Low : 0 m

Distance from roads
High : 23265 m
Low : 0 m

Figure 2-2. Topographic and environmental GIS layers for Colorado.

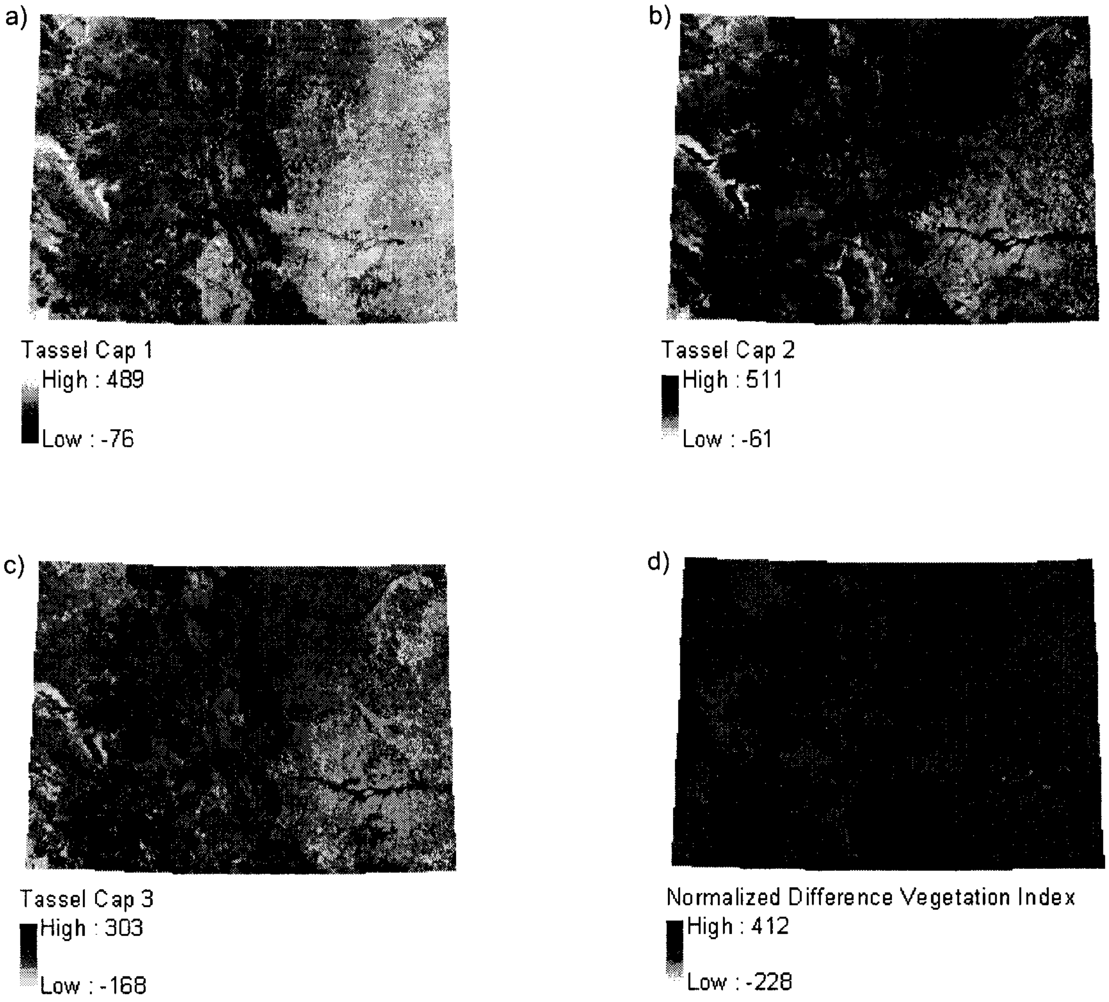


Figure 2-3. Remotely sensed imagery used as parameters in the model including tassel cap 1 to 3 (a through c) and normalized difference vegetation index (NDVI; d).

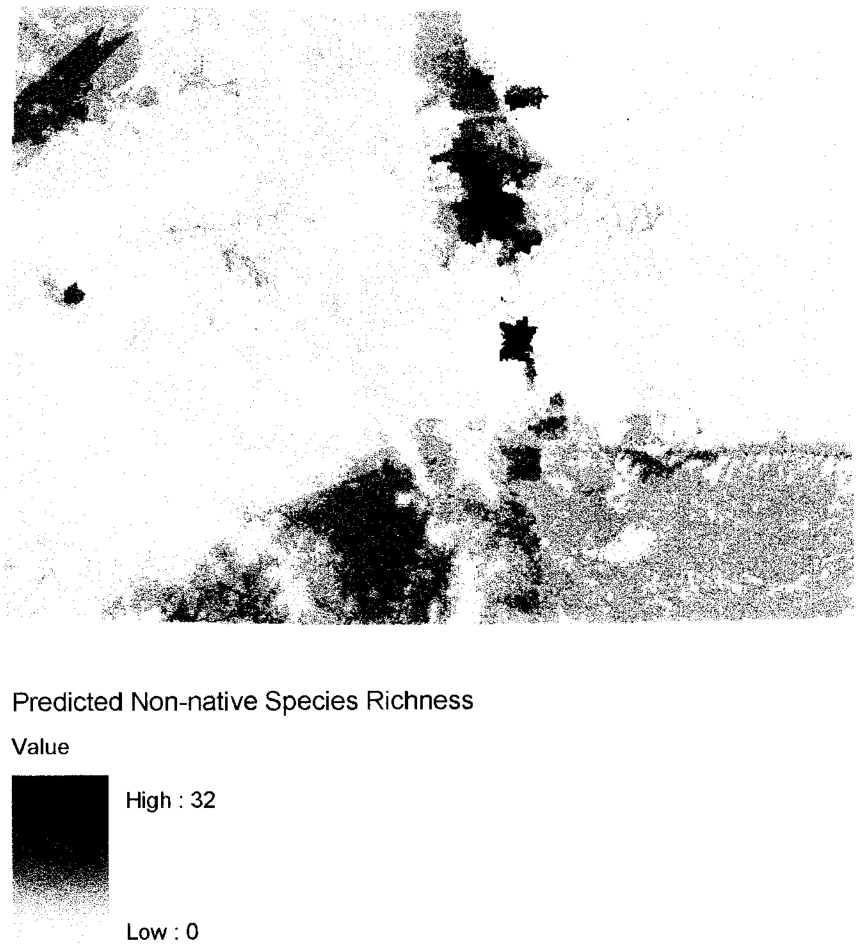


Figure 2-4. Predicted non-native species richness from the GLM and kriged residuals.

CHAPTER 3 : A SPATIAL MODEL TO PREDICT THE OCCURRENCE OF AN INVASIVE SPECIES, *EUPHORBIA ESULA*, IN COLORADO

3.0 Abstract

Potential distribution maps for many non-native species are unavailable, making early detection a difficult task for land managers. In this chapter, I used a database containing data from across the state of Colorado with presence/absence locations for leafy spurge, *Euphorbia esula,* to produce a map of probability of occurrence. Fourteen ancillary variables were tested, and ten were used in the final model with clay soil fraction having the greatest effect (coefficient of 0.055). The generalized linear model (GLM) explained 62% of the deviance. The residuals were kriged to capture the latent spatial autocorrelation in the data, and the final surface was generated by adding the GLM and kriged surfaces together. However, a Mann-Whitney test between the model and a statewide survey of county managers of *E. esula* distribution at the quarter quad level indicated poor discrimination between presence/absence sites. Before discrediting the model completely, it should be noted that the survey was based on county manager's incomplete knowledge of current distributions, not potential ones, and was conducted at a coarser resolution. This modeling effort can help guide future survey efforts for the

species and can be iteratively improved as new data are collected an incorporated into the database. Additional higher resolution ancillary data may improve model performance. The composite database allows for this type of large extent modeling effort that promises to be a valuable tool to combat the spread of invasive, non-native species.

3.1 Introduction

Invasive species already cost the United States an estimated $137 billion per year (Pimentel et al. 2000), and the number of species arriving in the United States is continuing to grow (Mack et al. 2000). There are also many different mechanisms for the dispersal of species to new areas by humans (Mack and Lonsdale 2001), and it will be impossible to prevent the movement of many species with increasing globalization. Thus, knowledge of where species might spread is important in helping combat this problem.

In general, potential distribution maps are not available for non-native species that are potential invaders. Therefore, managers do not know where to focus their limited funds and efforts to control the spread of non-native plant species, and the control methods are reactive, not proactive. The work required for eradication and the size of the initial infestation are highly correlated, indicating that early detection is very important in eradication efforts (Rejmanek 2000). The costs associated with eradication also increase with size of infestation. However, it is impossible to survey everywhere. Modeling potential distributions of non-native species is a strategy that may aid in the early detection of invasive species as they spread into new areas. The United States is a large, heterogeneous area, and preventing the expansion of established non-native species into new regions may be aided by early detection efforts.

Many different methods exist for modeling species distributions. Traditional regression techniques required normally distributed data and ignored spatial dependency in the datasets. General linear models (GLM) are a more recent branch of regression that can be used with non-normal datasets. Recently, spatial autocorrelation has been incorporated into statistical models (e.g., Lichstein et al. 2002). In fact, Austin (2002) declared that testing residuals for spatial autocorrelation should be a required step in species distribution modeling. This paper demonstrates the utility of using GLM and spatial statistical methods with readily available, general predictor variables to create potential distribution maps for non-native species. In this paper, I explore a method using readily available data to predict the occurrence of a highly invasive noxious weed in the state of Colorado.

3.1.1 The Species: Euphorbia esula

Euphorbia esula, or leafy spurge, is a non-native invasive species originally introduced to North America in 1827 as a contaminant in seed grain. Since that time it has spread throughout the United States and has become a major problem in the north central plains, invading rangeland and displacing native grassland species. In recent years it has become more widespread in Colorado. It currently occupies over 2 million ha in North America, and is distributed in 35 states and all but one Canadian province (Selbo and Carmichael 1999). *Euphorbia esula* is currently found on the noxious weed lists of 35 states (USDA NRCS 2004). In Colorado, it is considered one of the top ten invaders that are the most widespread and have the greatest economic impact in the state. It is also one of the four weeds that must be managed statewide in accordance with the Colorado Weed Management Act.

Euphorbia esula has many effects on native ecosystems. An increase in its cover is associated with a decrease in the cover of native plant species (Trammell and Butler 1995). There is also a positive correlation with its presence and that of other invasive plant species including *Poa pratensis* (Kentucky Bluegrass) and *Bromus inermis* (Smooth Brome, Belcher and Wilson 1989). Presence of *E. esula* results in the decreased biomass of native forage species of native ungulates and cattle, and bison, elk, and deer have been observed to avoid areas where it is present (Trammell and Butler 1995).

The ecology of *E. esula* affords it its competitive ability to displace native species. Its milky latex deters herbivory by native and domestic herbivores (Selleck et al. 1962, Upadhyay et al. 1978, Lym 1998), and, if herbivory does occur, *E. esula* has mechanisms to recover such as the release of adventitious roots and crown buds (Cyr and Bewley 1989). It emerges earlier in the spring than many native species (Morrow 1979) and grows rapidly, being able to reproduce vegetatively as early as ten days after emergence (Lym 1998) and shade out other slower growing and later emerging species (Belcher and Wilson 1989).

It also has a large reproductive capacity, reproducing both by vegetative buds that mature rapidly (Selleck et al. 1962, Olson and Wallander 1999a) and by seeds that allow it to disperse greater distances, typically from wind dispersal (Selleck et al. 1962, Morrow 1979, Selbo and Carmichael 1999). This ability to reproduce both sexually and asexually aids its spread by allowing a single plant to create a new infestation. It is able to self-pollinate as well, which is also advantageous in colonizing new areas (Selbo and Carmichael 1999).

Its extensive root system, however, seems to confer the most advantages. *Euphorbia esula* can compete well in both patchy nutrient distributions due to the plasticity of its root systems (Ringwall et al. 2000) and in areas with well-distributed nutrients due to the extent of its root systems. Most roots are in the top 12 cm but they can extend down to the water table (Lym 1998) with a lateral spread up to five m (Ringwall et al. 2000). The roots also have great storage reserves that help the plant in times of stress (Lym and Messersmith 1987, Schimming and Messersmith 1988, Cyr and Bewley 1989, 1990). All of these factors make *E. esula* a formidable competitor to native grassland species.

3.2 Methods

3.2.1 Floristic data

Presence/absence data for *E. esula* were extracted from a spatial database containing information collected from different organizations and individual researchers and land managers in the state of Colorado. This database included disparate data ranging from vegetation plots from ecological studies to weed managers' weed maps (Appendix B). *Euphorbia esula* was chosen because location information for this species covered a large area of the state, it is one of twenty species identified by the state as being of top concern, and there are also various methods available for the control of the species. The resolution of the field was at 30-m because this scale was closest to that available for the ancillary data. Presence data were derived from weed mapping data for the species and vegetation plots of all sizes (0.1-m^2 Daubenmire to 0.1-ha Modified Whitaker). Absence data were only derived from vegetation plots that cover approximately a 30m X

30m grid cell and would have recorded *E. esula* if present. There were 2,684 points including 1,504 absence locations and 1,180 presence locations (Figure 3-1).

3.2.2 Geographical Information System and Remote Sensing Data

Potential variables were determined from a set created from an extensive review of invasion literature (Appendix A). Variables chosen were those with readily available datasets that were somewhat limited by the scale of the model (Table 3-1, Figure 3-2 and 3). For example, State Soil Geographic (STATSGO) soil data were available statewide, but the resolution was very large. Consequently, the use of these variables was limited to three soil variables in the modeling process. Also, although distance to roads is correlated with invasion (Gelbard and Belnap 2003, Watkins et al. 2003), this variable was not used because sampling was highly biased towards areas close to roadsides. The proposed variables were used as predictor variables in logistic regression analysis.

Topographic variables included elevation, slope and absolute aspect derived from a Digital Elevation Model (DEM; Appendix C). Absolute aspect was used instead of aspect because it is a linear rather than circular variable maximizing the difference between north and south facing slopes. Other predictors included distance from rivers or streams, average annual precipitation, soil fraction of sand, silt, and clay, and human population density.

Landsat 7 Thematic Mapper data were also obtained for the whole state and used to create various indices related to vegetation at a 30-m resolution (Figure 3). Storing individual bands requires a large amount of storage space, and using these indices reduced the data size. These indices are also transformations of the bands into biologically meaningful variables. A mosaic of images from 13 August 1999 to 20

September 2002 was used to generate tassel cap 1 to 4 (weighted sums of thermal bands) and Normalized Difference Vegetation Index (NDVI) by NASA Goddard Space Flight Center, Greenbelt, MD. Tassel cap band 1 is interpreted as brightness (information on bare soil) with negative values indicating no bare soil and very large values indicating lack of vegetation (Campbell 2002). Tassel cap band 2 is designated as greenness (information on abundance and vigor) with negative values indicating lack of vegetation and very high values representing lush areas. Tassel cap band 3 is yellowness or wetness (reflecting the difference between the red and green bands) with low values representing low moisture content and high values high moisture content. Tassel cap band 4 is a summary of all the spectral data. NDVI is calculated using the infrared and red bands that contain information on vegetation as a measure of photosynthetic activity and biomass.

3.2.3 Statistical Analyses

I used trend surface analysis in conjunction with spatial statistical techniques to capture both the broad and fine scale variability. GLM was used for the trend surface analysis in this paper. Logistic regression is a type of GLM appropriate for data with a binary distribution such as species presence or absence (McCullagh and Nelder 1989). All GLM contain three components: a random component (Y), a systematic component (Xs), and a link (functional relationship between Xs and the mean of Y). The outcome of the regression analysis is a model of probability of occurrence. For *E. esula* logistic regression was performed using a logit link function assuming a binomial distribution in S-plus (Statistical Sciences 2000, Appendix F). GLM assumes transformation of an expected response is a linear function, but unlike least squares methods allows for different distributions such as Binomial to be specified. The logit link function is simply a transformation of the regression equation using

$$(1) \qquad Log\left(\frac{\mu}{1-\mu}\right)$$

where μ is the original value at the point (MathSoft 1999). Estimates of regression parameters are generated using maximum likelihood produced by iteratively reweighted least squares and selection is made by the program using AICC. Variable selection was performed using the stepwise procedure for GLM in S-plus.

The fine scale variability was captured using kriging, a geostatistical interpolation technique. Kriging estimates unknown values using known values and a semivariogram describing the relationship of values at different distances. This method was chosen as it mimics the dispersal of the species by weighting the influence of near points greater than far points. Dispersal is particularly important to capture when predicting the distribution of a species like *E. esula* that is near the initial stages of an invasion.

Examination of a histogram of the residuals from the final GLM model indicated an approximately normal distribution (MathSoft 1999) and a Moran's *I* test indicated the presence of spatial autocorrelation (Cliff and Ord 1973). Moran's *I* typically varies between negative one and one, with negative values indicating a checkerboard pattern where like values are distributed across a landscape, values around zero indicate a random pattern, and positive values indicate clumping of like values. Spatial autocorrelation is necessary for kriging to be a valid technique.

A semivariogram was fit for the data in S-plus (Statistical Sciences 2000) using Reich and Davis's spatial library (Reich and Davis 1998). The semivariogram type (Guassian, spherical, or exponential) was determined by minimizing the AICC and the range, sill, and nugget were estimated. These values and the residuals for the variogram were imported into ArcGIS to use as input to the krig function (ArcGIS 2002). This software was used for the purposes of this example because of current computational limitations of statistical packages like S-plus. ArcGIS was the only software package I had capable of performing the calculations at this extent and resolution. Other programs crashed with this size dataset before completion and results that were produced before crashing for small sections of the state were anomalous, producing values magnitudes larger than the input range.

The final surface was generated with the Raster Calculator in ArcGIS (ArcGIS 2002) using the predictor variable rasters with the statistical output from S-plus and the kriged surface. The resulting cell values were in the logit scale and were therefore back-transformed to the original scale of the probability surface using:

$$p = \frac{e^{(LP)}}{1 + e^{(LP)}} \tag{2}$$

where p is the probability and LP is the linear predictor.

3.2.4 Model evaluation

When generating predictive models, some sort of evaluation of model performance is necessary to determine if the model holds accurate predictions. However, for a model of non-native species distributions, this can be especially difficult when a species is still spreading into new areas. Therefore, I used two different methods to attempt to evaluate the model, although poor results from these tests do not necessarily mean that the model performed poorly.

To evaluate the model, both percent deviance explained (which is similar to an R^2 value) and measure of discrimination were calculated. Percent deviance explained was calculated as

(3) $$Percentdeviance = \frac{NullDev - \mathrm{Re}\,sDev}{NullDev} X100$$

where NullDev is the null deviance of the evaluation data and ResDev is the residual deviance of the evaluation data in relation to probabilities predicted by the model (Everitt 1994). This measurement is of overall goodness of fit of the model to the known observations. The measure of discrimination was calculated using the Mann-Whitney U test, which is the same as the Wilcoxon test for non-parametric data (Sokal and Rohlf 1995). The model's ability to discriminate between presence and absence is dependent on the threshold used to assign probability of occurrence to one of these categories. Receiver operating characteristics (ROC) allow discrimination for any threshold between 0 and 1 for probability of occurrence (Fielding and Bell 1997). For each threshold, there are sites predicted correctly and incorrectly, resulting in four categories: true positives, true negatives, false positives, and false negatives. From these, true positive fractions (i.e., proportion of correctly predicted presences to falsely predicted presences) and false positive fractions (i.e., proportion of correctly predicted absences to falsely predicted absences) can be calculated.

The ROC curve is defined by the smooth curve created from a graphical display of true and false positive fractions. The area under this curve corresponds to the discrimination ability of the model. With no ability to discriminate, the value would be 0.5, while a value of 1.0 would indicate perfect discrimination. Hanley and McNiel (1982) showed that the area under the ROC curve could be approximated by the Mann-

Whitney test statistic when it is standardized by dividing by the product of the two sample sizes. The result is between the values of 0.5 and 1.0, with values below 0.7 indicating poor discrimination, values between 0.7 and 0.9 indicating reasonable discrimination, and values above 0.9 indicating excellent discrimination (Swets 1988, Pearce and Ferrier 2000).

A preliminary, statewide coverage of *E. esula* was available for the state from a survey of county weed managers conducted by the state of Colorado where managers were asked to estimate acreage per quarter quad. For the discrimination analysis, each quarter quad with some acreage reported was designated as present and all others were assigned absence. The maximum, minimum, mean, and sum of probability of occurrence from the modeled 30-m grid surface of the state were extracted for each quarter quad using ArcGIS spatial analyst zonal statistics option (ESRI 2002). The Mann-Whitney test was then performed on these data for both the GLM model and the spatial GLM model.

3.3 Results

Comparing the summary statistics for the field data points (Table 3-1) and the whole state (Appendix C), the field points reflect the range of attributes in the state fairly well. There is a bias towards less populated areas (e.g. maximum population density of 945 people/km^2 versus 6567 people/km^2), and the very high values for soil fractions are absent (e.g. clay maximum of 38% compared to 54%). For most variables the range of values for predictor variables was much greater for absence locations than for presence locations.

Stepwise variable selection resulted in a model containing ten of the 14 variables tested (Table 3-2). Only one of these, percent clay of the soil fraction, was at a much coarser resolution than the dependent variable, although slope and absolute aspect were at a resolution of around 90-m. The coefficients were both positive and negative, with clay having the strongest effect (0.055). These coefficients reflect the logit function and not the original scale.

A test of the residuals from the model revealed significant spatial autocorrelation (Moran's $I = 0.53$, $p < 0.001$; Table 3-3), so a variogram was fit to the residuals and ordinary kriging was performed. The variogram curve chosen was the Gaussian model and the parameters indicate that there is not much noise left in the data (nugget 0.0946; Table 3-4).

The final surface showed high probability of invasion in the Eastern plains of the state and the far western edge (Figure 3-4). The high elevations throughout the central north-south corridor of the state are not likely to be invaded. The kriged error surface for the state indicated the greatest error in the Eastern plains.

With a null deviance of 3,682 and residual deviance of 1,216, the GLM model explained 67% of the deviance (Table 3-3). The standardized Mann-Whitney test statistic for the mean, maximum, minimum, and sum probability per quarter quad was between 0.55 and 0.59 for all four statistics for the spatial GLM and between 0.55 and 0.63 for the GLM, indicating poor discrimination (Table 3-5). The values differed for each statistical category for the two models, and the best discrimination occurred in the maximum value for the GLM model.

The distribution of presence locations used in model development was restricted to the northern third of the state. However, the quarter quad survey indicated the presence of *E. esula* throughout the state. Additionally, field data points were very sparse in the Eastern plains. Elevation had a negative effect on probability of occurrence. The field data used to generate the model did not include any presence locations above 2896 m elevation, although quarter quads with greater elevations reported the presence of *E. esula*. Other variables may have more restricted ranges than reality because of the field data locations.

3.4 Discussion

The change in deviance of the model suggests that the model is good, especially considering the relatively broad extent and fine grain of the dependent variable. This test of the model does not include improvements in model performance resulting from the addition of the kriged surface that captures the small-scale spatial variability. Thus, the variables used in the model seem to be good predictors of *E. esula* distributions, at least for the sites sampled and preliminary occurrence data available. However, the results of the ROC test indicate that the model poorly discriminates between sites where *E. esula* is present and absent. Also, the GLM model performed better than the spatial GLM model according to this test with the quarter quad data. There are several reasons that could account for these discrepancies and the model is probably much more accurate than the ROC test indicates.

The test data used for model evaluation were from a quarter quad survey of acreage estimates per quarter quad at the time of the survey. *E. esula* is still spreading in the state, and the model, based upon environmental surrogates, is one predicting the probability of occurrence. It could be that the areas where *E. esula* was predicted present in the model, but recorded absent in the survey, exist because (1) of incomplete data, because the spread of *E. esula* has not yet reached those sites, or (2) an invasion at the site was controlled and therefore not reported in the survey. Another test of the model that may provide better results would be to generate a modeled surface for an area such as Montana where *E. esula* has been spreading for a much longer time and evaluate the model for that area or to compare areas that *E. esula* invades in the next five years to the modeled surface for Colorado.

Additionally, as the survey consists of estimates at a very different scale, the comparability of the two datasets is questionable. In the case of an invasive species that is continuing its spread, the use of current distribution such as the survey might best be used to determine watch areas for the spread of an invasive species. The areas with modeled predicted presence and survey reported absences should be the areas of the state where managers focus early detection and prevention efforts for the spread of *E. esula*, especially those areas close to reported infestations.

The variables selected for use in the final model reflect the ecology of *E. esula*. For this species, the model would probably be improved by incorporating remotely sensed imagery taken during blooming, as it has a unique spectral signature (for example see Williams and Hunt 2002). This type of input would be useful to determine areas currently invaded as opposed to those with the potential of being invaded. Additional field data from areas not surveyed would also likely improve the model accuracy. If ancillary data were available at a higher resolution the model might also be improved, but that is unlikely in the near future for such a large extent. When creating models for such large extents, problems may arise as predictor variable's relationship to the species presence may change through space. For example, the elevation limit for a species would be different in Montana than New Mexico because of the change in latitude. More physiologically based variables may be better than those such as elevation, but these variables are hard to derive. Other variables that better capture processes related to *E. esula* distribution may also improve the model.

Again, this is an iterative process with the results of the model driving further field efforts, which are in turn used in the model. Future sampling efforts in the state should be guided by the error surface produced by the model. Data from these areas could be used to refine the model to generate a more accurate picture of *E. esula's* potential distribution.

Broad extent, moderate grain datasets such as this one are becoming available. Broad extent datasets with fine spatial resolution may be particularly important for predicting the occurrence of invasive species such as *E. esula* that have not reached their full spread potential. In these cases, dispersal is an important variable that may be missed or downplayed when the spatial extent considered is small. Potential source populations may be missed or areas with habitat potential may not be identified because the species has not yet established in areas covered by this small extent. In this paper, potential *E. esula* distribution in Colorado was described, although this species has not spread through much of the state. However, states to the North have had major problems and economic impacts from this species. The predicted model of occurrence provides land managers and private landowners with an idea of the potential problem in Colorado, and perhaps will aid in the prevention of its spread into some of these areas. With the synergy of large datasets as described in chapter 1, modeling the potential distribution of non-native, invasive species is possible. Modeling techniques such as those described above used in concert with synergy of datasets have promise as aids to land managers to detect, map, and control the spread of weeds such as *E. esula*.

3.5 Literature Cited

ArcGIS. 2002. ArcGIS. Environmental Systems Research Institute, Redlands, California.

Austin, M. P. 2002. Case studies of the use of environmental gradients in vegetation and fauna modeling: theory and practice in Australia and New Zealand. Pages 868 in J. M. Scott, P. J. Heglund, M. L. Morrison, J. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington.

Belcher, J. W., and S. D. Wilson. 1989. Leafy spurge and the species composition of a mixed-grass prairie. Journal of Range Management 42:172-175.

Campbell, J. B. 2002. Introduction to remote sensing, 3rd edition. Guilford Press, New York.

Cliff, A. D., and J. K. Ord. 1973. Spatial autocorrelation. Pion, London,.

Cyr, D. R., and J. D. Bewley. 1989. Carbon and Nitrogen Reserves of Leafy Spurge (Euphorbia-Esula) Roots as Related to Overwintering Strategy. Physiologia Plantarum 77:67-72.

Cyr, D. R., and J. D. Bewley. 1990. Seasonal variation in nitrogen storage reserves in the roots of leafy spurge (Euphorbia esula) and responses to decapitation and defoliation. Physiologia Plantarum 78:361 - 366.

Everitt, B. 1994. A handbook of statistical analyses using S-PLUS, 1st edition. Chapman & Hall, London ; New York.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38-49.

Gelbard, J. L., and J. Belnap. 2003. Roads as conduits for exotic plant invasions in a semiarid landscape. Conservation Biology 17:420-432.

Hanley, J. A., and B. J. Mcneil. 1982. The Meaning and Use of the Area under a Receiver Operating Characteristic (Roc) Curve. Radiology 143:29-36.

Lichstein, J. W., T. R. Simons, S. A. Shriner, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs 72:445-463.

Lym, R. G. 1998. The biology and integrated management of leafy spurge (Euphorbia esula) on North Dakota rangeland. Weed Technology 12:367-373.

Lym, R. G., and C. G. Messersmith. 1987. Carbohydrates in leafy spurge roots as influenced by environment. Journal of Range Management 40:139 - 144.

Mack, R. N., and W. M. Lonsdale. 2001. Humans as global plant dispersers: getting more than we bargained for. BioScience 51:95-102.

Mack, R. N., D. Simberloff, W. M. Lonsdale, H. Evans, M. Clout, and F. A. Bazzaz. 2000. Biotic invasions: Causes, epidemiology, global consequences, and control. Ecological Applications 10:689-710.

MathSoft. 1999. S-PLUS 2000 Guide to Statistics, Volume 1. Data Analysis Products Division. MathSoft, Inc., Seattle, Washington.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models, 2nd edition. Chapman and Hall, London ; New York.

Morrow, L. A. 1979. Studies on the reproductive biology of leafy spurge (Euphorbia esula). Weed Science 27:106 - 109.

Olson, B., and R. Wallander. 1999. Carbon allocation in Euphorbia esula and neighbors after defoliation. Canadian Journal of Botany-Revue Canadienne De Botanique 77:1641-1647.

Pearce, J., and S. Ferrier. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecological Modelling 128:127-147.

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

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.

Rejmanek, M. 2000. Invasive plants: approaches and predictions. Austral Ecology 25:497-506.

Ringwall, K., M. Biondini, and C. Grugiel. 2000. Effects of nitrogen fertilization in leafy spurge root architecture. Journal of Range Management 53:228-232.

Schimming, W. K., and C. G. Messersmith. 1988. Freezing resistance of overwintering buds of four perennial weeds. Weed Science 36:568-573.

Selbo, S., and J. Carmichael. 1999. Reproductive biology of leafy spurge (Euphorbia esula L.): breeding system analysis. Canadian Journal of Botany 77:1684-1688.

Selleck, G. W., C. Frankton, and R. T. Coupland. 1962. Leafy Spurge in Saskatchewan. Ecological Monographs 32:1-&.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry : the principles and practice of statistics in biological research, 3rd edition. W.H. Freeman, New York.

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

Swets, J. A. 1988. Measuring the Accuracy of Diagnostic Systems. Science 240:1285-1293.

Trammell, M. A., and J. L. Butler. 1995. Effects of exotic plants on native ungulate use of habitat. Journal of Wildlife Management 59:808-816.

Upadhyay, R. R., F. Bakhtavar, M. Ghaisarzadeh, and J. Tilabi. 1978. Cocarcinogenic and irritant factors of Ephorbia esula L. latex. Tumori 64:99-102.

USDA NRCS. 2004. The PLANTS Database, Version 3.5 (http://plants.usda.gov). National Plant Data Center, Baton Rouge, LA 70874-7790 USA.

Watkins, R. Z., J. Q. Chen, J. Pickens, and K. D. Brosofske. 2003. Effects of forest roads on understory plants in a managed hardwood landscape. Conservation Biology 17:411-419.

Williams, A. P., and E. R. Hunt. 2002. Estimation of leafy spurge cover from hyperspectral imagery using mixture tuned matched filtering. Remote Sensing of Environment 82:446-456.

Table 3-1. Summary statistics, including minimum and maximum values, mean and standard deviation per grid cell for the independent variables for the sample locations (n=2684).

Variable	Maximum	Minimum	Mean	Standard Deviation
Absolute aspect (°)	180	0	72.5	49.1
Slope (°)	60	0	8	7.7
Elevation (m)	3756	1082	2280.9	477.7
Sand fraction (%)	76	0	45.2	9.1
Silt fraction (%)	55	0	35.6	6.0
Clay fraction (%)	38	0	18.8	5.6
Average annual precipitation	10	1	6.4	1.5
Population density per zip code (people/km^2)	945	0	28.2	124.7
Distance to water (m)	2954	0	415.5	437.4
NDVI	352.5	-66	149.6	58.8
Tassel cap1	311	-49.6	91.9	43.9
Tassel cap 2	366	-39	158	50
Tassel cap 3	314	-56.6	162.5	49.8
Tassel cap 4	419	-245.7	127.8	45.4

Table 3-2. Coefficients for regression for probability of occurrence for leafy spurge in Colorado from the GLM.

Variable	Coefficient	t-value
Intercept	6.3	5.2
Absolute aspect	-0.0025	-1.6
Elevation	-0.0053	3.9
Slope	-0.021	-16.9
Clay Soil Fraction	0.055	-1.9
Distance from water bodies	-0.0025	-8.5
Tassel cap 1	0.031	7.5
Tassel cap 2	0.0075	2.8
Tassel cap 3	0.011	2.5
Tassel cap 4	0.014	7.2
NDVI	-0.017	-7.4

Table 3-3. Results from the GLM and spatial autocorrelation test of residuals.

	Null Deviance	Residual Deviance	Psuedo-R^2	Residual Degrees of Freedom	Moran's *I*
E. esula	3681.6	1216.1	0.67	2673	0.53 (p-value = 0)

Table 3-4. Semi-variogram values for the regression residuals for the *E. esula* GLM for three types of curves with the AICC values.

Regression Model	Nugget	Sill	Range	Semivariogram Model	AICC
E. esula	0.095	0.33	1748.7	Gaussian	133.8
E. esula	0.08	0.33	2188.2	Spherical	133.9
E. esula	0	0.34	960.2	Exponential	136.9

Table 3-5. Mann-Whitney test measure of discrimination for the GLM model and the spatial GLM model.

Statistic	GLM model	Spatial GLM Model
Mean	0.562	0.581
Maximum	0.626	0.554
Minimum	0.547	0.588
Sum	0.562	0.588

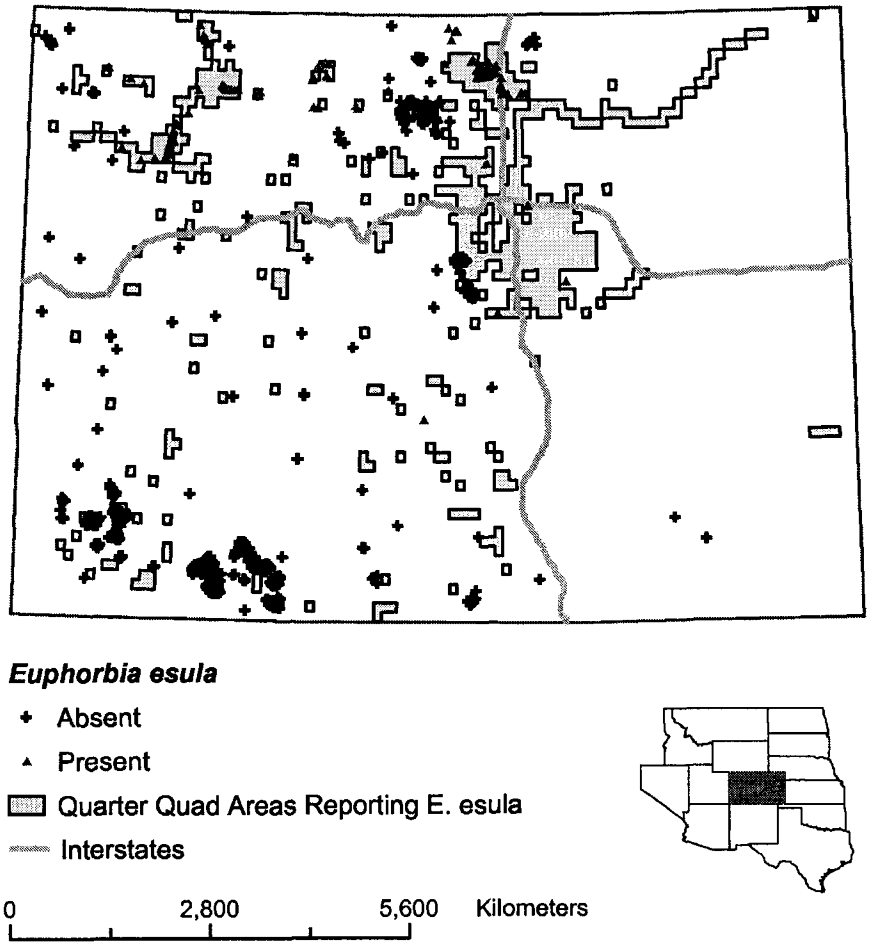
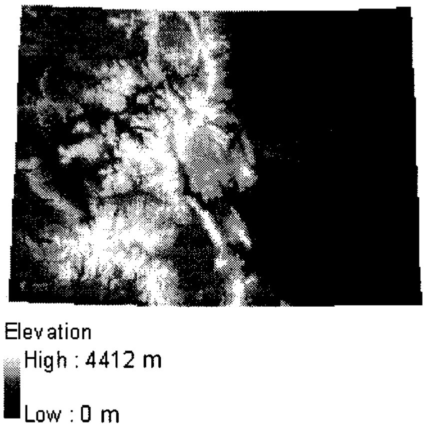
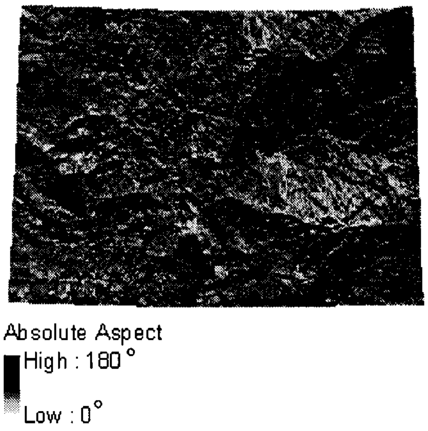


Figure 3-1. Field data locations of the presence and absence of *E. esula* in Colorado.

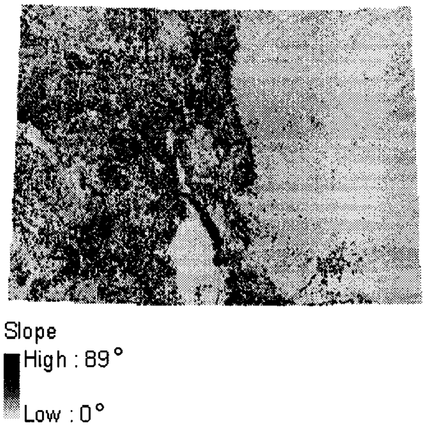
a)


Elevation
High : 4412 m
Low : 0 m

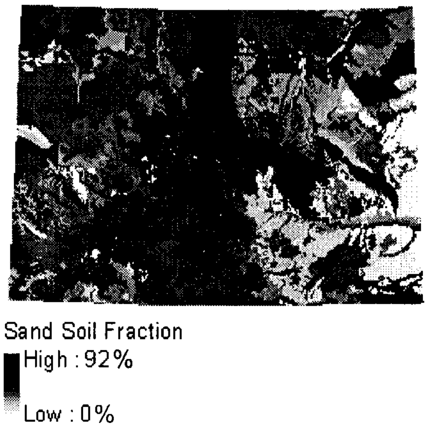
b)


Absolute Aspect
High : 180 °
Low : 0°

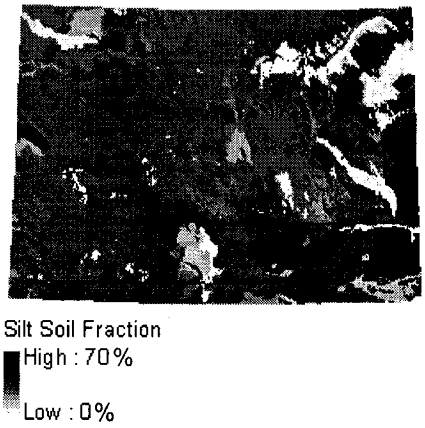
c)


Slope
High : 89 °
Low : 0 °

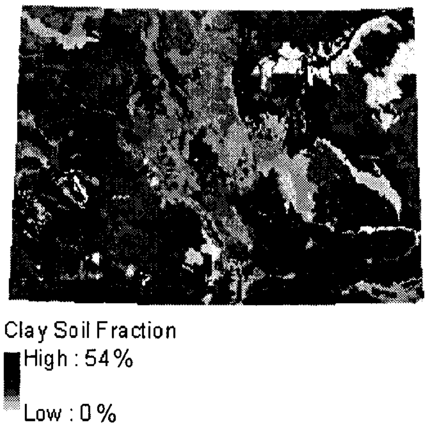
d)


Sand Soil Fraction
High : 92%
Low : 0%

e)


Silt Soil Fraction
High : 70%
Low : 0%

f)


Clay Soil Fraction
High : 54%
Low : 0 %

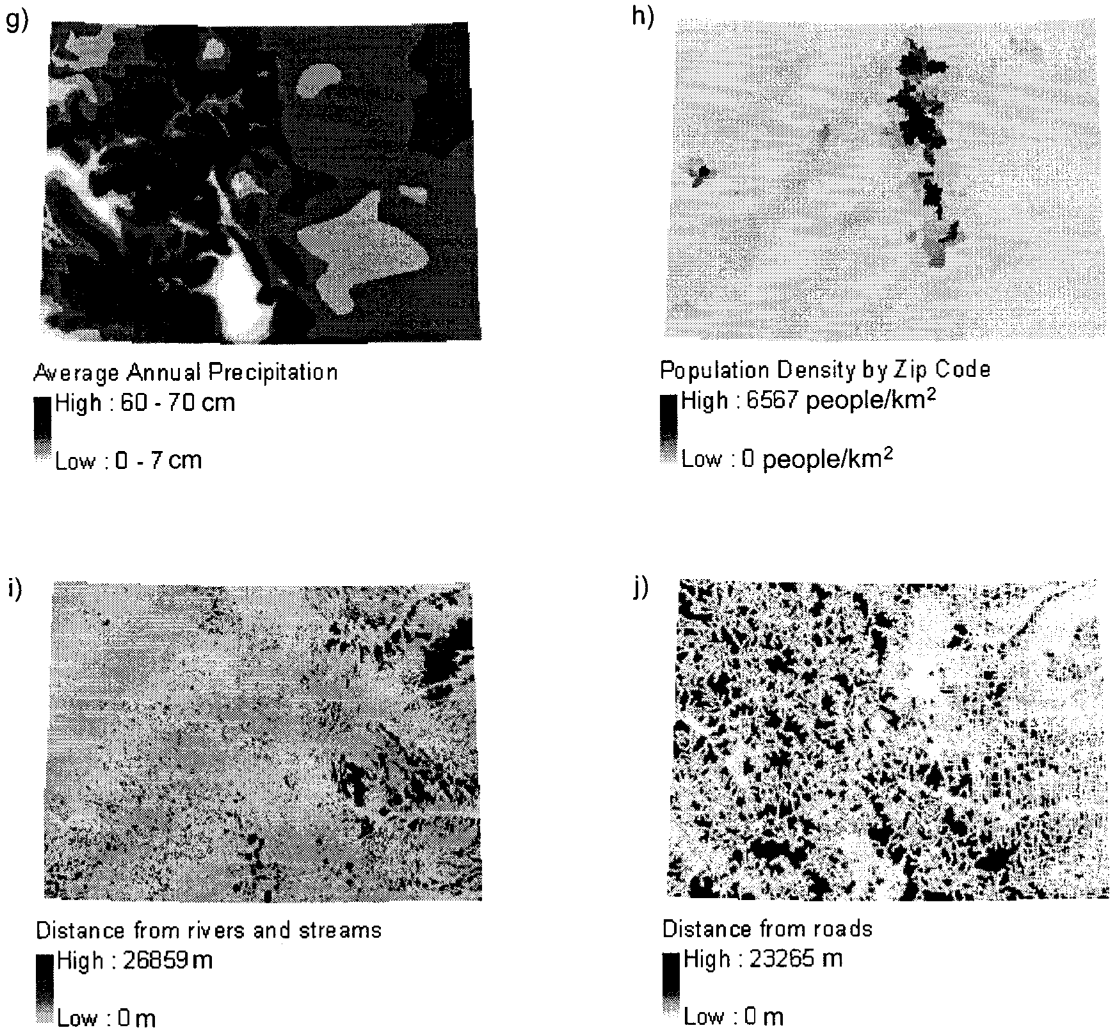
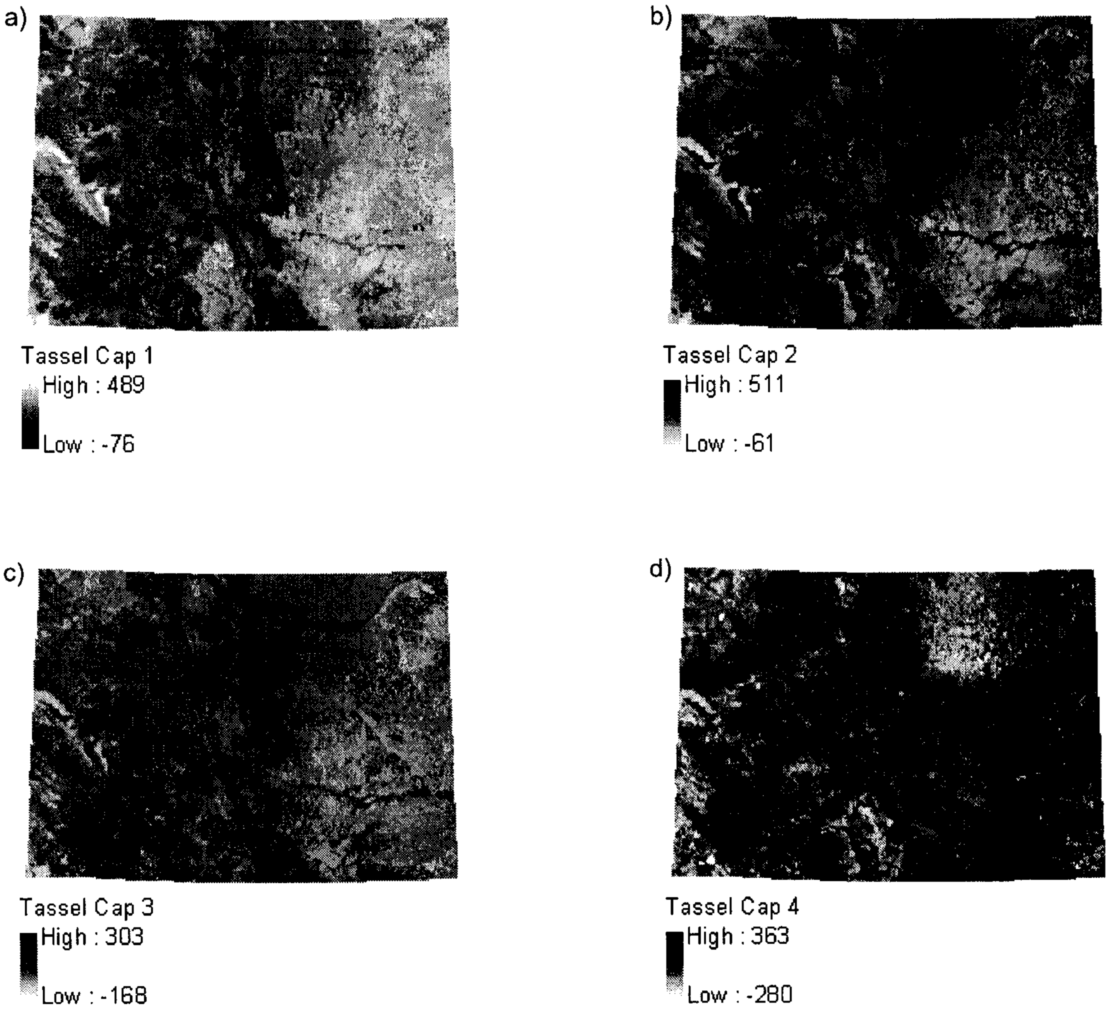


Figure 3-2. Topographic and environmental GIS layers for Colorado.



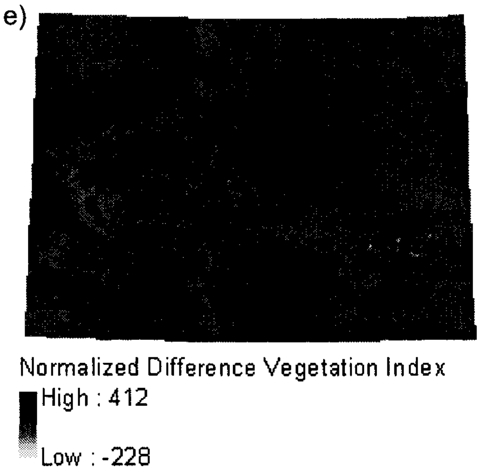


Figure 3-3. Remotely sensed imagery used as parameters in the model including tassel cap 1 to 4 (a through d) and normalized difference vegetation index (e).

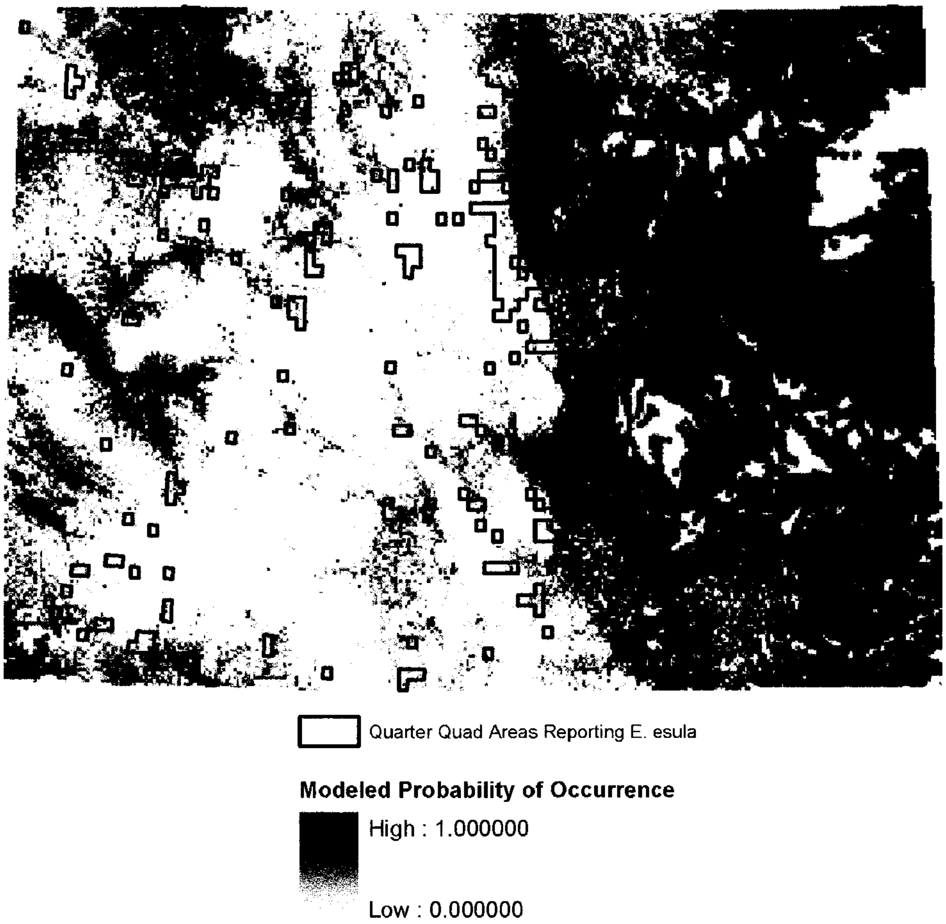


Figure 3-4. Predicted probability of occurrence of *E. esula* in Colorado from the general linear model and kriged residuals. Reported locations of *E. esula* from the state's quarter quad survey of county weed managers are also shown.

CONCLUSION

Non-native species are causing economic problems, threatening human health, and negatively impacting biodiversity, and these problems can be expected to increase. Additionally, collecting vast amounts of field data is costly, stressing already limited resources. Chapter 1 illustrated the benefits of creating large databases to incorporate existing datasets from disparate studies, individuals, and organizations. By sharing data, everyone gains more knowledge about species distributions and abundance. Analysis of the combined datasets can highlight areas that have been overlooked and understudied where future efforts should be focused.

Chapter 2 provides one example of what can be done with the combined databases. These species richness models are unique in the broad spatial extent they cover while maintaining a relatively fine spatial resolution. They are part of an iterative process, with the error surface and modeled output guiding future sampling efforts to areas where knowledge is poor or areas where invasion potential is great. Chapter 3 shows how the database can be used to create a model of potential distribution for a specific invasive species of interest at a broad spatial extent. This type of model can again be used to guide future field sampling, to aid early detection efforts and to better predict the spread of invasive species. These models can also be used to help land managers determine where to invest limited resources for early detection and control of non-native species.

Models across large spatial extents with fine resolution are important for non-native species management and control. Models for small areas may overlook potential habitat in an area because the species has not yet spread into the small area being considered. Coarse resolution models provide some insight into invasion patterns, but for a land manager can be like looking for a needle in a haystack to detect a single plant at an early stage of invasion in a large area.

Generating models with a 30-m^2 grid cell size posed some interesting theoretical and practical problems. It is difficult to find datasets at this spatial extent with a fine spatial grain. Knowing what scale might be appropriate for modeling purposes and what variables might be useful poses theoretical questions. Different variables may be appropriate at different scales, and knowing what variables might be important in predicting a variable of interest is a difficult theoretical question. The modeling techniques used alleviate some of those scale problems because they are designed to capture both broad and fine scale trends. Despite the theoretical issues associated with a model of this type, it was useful to determine trends across the state and help identify areas where groups in the state should focus control or mitigation efforts and future sampling efforts. The state was a politically useful extent to model, but a larger scale while maintaining a small grid size would not be recommended. The theoretical and practical problems would make it very difficult, and the utility could also be questioned.

Practical issues resulted from the file size of the rasters and the computational requirements for the kriging process. The rasters for the state each approached the two gigabyte file size limit from the Microsoft operating system. The kriging process at the scale crashed in some programs I tried to use, again due to the size of the files.

Given the above issues, I would not recommend using a large extent for most modeling purposes. In this dissertation, the 30m grid cell size was retained to make the models useful for managers. Different grains and extents, however, have different purposes and uses (Table 1). Narrower extents than the one used here while maintaining a comparable grain size would be easier to handle. However, for newly spreading species in an area, a broader extent is required to capture the initial invasion and spread. For a manager concerned with the way an established plant in their management unit will spread, a model covering an extent slightly broader than their unit would be sufficient. Thus, depending on the intended use of a model, and the area of interest, different extents and grains are appropriate.

The creation of large databases such as this holds many more opportunities for further research. Currently, data on species abundance are limited. Generating predictive models of individual species abundance and overall non-native species cover would be another useful tool for land managers. Additionally, with further data contributions a time series will be generated, and analysis of spread through time will be possible.

Another area for future work is in improving and automating the parameters used to generate the models. Most large extent datasets have coarse spatial resolution. Processes controlling the distribution of species can be very local and fine-scale. With technological advances in remote sensing, there is the potential to generate higher resolution input parameters. Different parameters may have effects at different scales, so exploring different scales might produce new insights into invasion patterns. The database has the information that is necessary to do this, but ancillary data to generate the models are currently unavailable at resolutions finer than 30-m.

The techniques described in this dissertation illustrate ways to capitalize on existing data to create models that may aid land managers in their task to control the spread of invasive, non-native species. Novel uses of existing datasets and models covering large geographic regions are becoming increasingly useful as we face declining budgets and increasing natural resource crises while attempting to locate invasive species in their initial stages of invasion. These techniques do not need to be limited to non-native species, but may be used to determine hotspots of biodiversity or potential distributions of rare and endangered species. They may therefore prove useful not only in combating the problems for biodiversity imposed by non-native species, but may have utility in other areas of conservation.

Table 1. Purpose and utility of different grains and extents.

	Small Grain	Large Grain
Small Extent	Local patterns of invasion for already established species in area of interest; useful to management	Local patterns of invasion of already established species in area of interest; difficult to locate species for management
Large Extent	Landscape patterns of invasion or of newly invading species; useful to management	Landscape patterns of invasion or of newly invading species; not particularly useful for management

Appendix A. Studies on vulnerability to invasion including the type of study, the scale the study was conducted, and the results applicable to patterns of invasion from the study.

Study	Type	Scale	Results
(Bergelson et al. 1993)	Experiment	Multiple	Rates of spread of a weed changes with different spatial patterns of disturbance
(Brown and Peet 2003)	Observational	Multiple	Comparison of five scales of invasion found switch in relationship between native and non-native species in riparian areas. Riparian areas were more invaded than upland areas.
(Burke and Grime 1996)	Experiment	Local	Limestone grassland in United Kingdom. Most invasible with high levels of disturbance. Even more invaded when disturbance associated with eutrification.
(Dukes 2001)	Experiment	Local	Grassland community microcosm's invisibility by *Centaurea solstitialis*. High functional diversity decreased success. Species diversity not correlated with vulnerability to invasion. Species diversity positively correlated with persistence of species after invasion.
(Fox and Fox 1986)	Review		Invasion occurs with at least minimal alteration of endogenous disturbance regimes
(Greenburg et al. 1997)	Observational	Local	Compared invasion along roadsides with native and non-native soils and clearcuts to hold disturbance constant in xeric Florida sand pine scrub. Modified soil type was more invaded than native, even with disturbance. Roadways facilitate non-native propagule transport.
(Harrison 1999)	Observational	Local	California serpentine and non-serpentine grasslands compared indicate that invasion is less in poor soils (but whether because poor soils are less invasible or because current non-natives are unsuited is unknown).
(Hobbs and Huenneke 1992)	Review		Exogenous disturbance such as fire, grazing, and soil disturbance outside of the natural cycle (either greater or less) leads to increased invisibility.
(Huenneke et al. 1990)	Experiment	Local	Applied plant nutrients (N and P) to plots in a Californian annual serpentine grassland. Grazing exclusion resulted in decreased native forb presence. Adding nutrients resulted in increased invasion (reduced total richness, increased non-native dominance).

(Larson et al. 2001)	Observational	Regional	Mixed grass prairie (THRO, both units, 11 plant communities). Vegetation type explained more variance in vulnerability to invasion than disturbance or park unit. Three way model with vegetation type, park unit, and disturbance best explains frequency and number of non-natives in aggregate- perhaps a reflection of nitrogen and water availability differences. Riparian zones are more invaded. Different species respond differently to disturbance and to vegetation type.
(Levine 2000)	Experiment	Local	California riparian area diversity is positively correlated with invisibility.
(Levine and D'Antonio 1999)	Review		Generally, species rich communities are more invasible than species-poor communities.
(Lonsdale 1999)	Observational	Global	Variances in non-native species number explained by number of native species, island or mainland status, and reserve or not reserve status. On the mainland degree of invasion increased with latitude. In reserves number of visitors also explained number of non-native species.
(MacDonald et al. 1989)	Observational	National	North American nature reserves are invaded.
(Planty-Tabacchi et al. 1996)	Observational	Watershed and patch	Studied Adour River in southwestern France, McKenzie River in Oregon, and the Dungeness and Hoh Rivers in Washington. The McKenzie and Adour similarities indicated that invasion along riparian areas can be predicted by species richness. At the patch scale, riparian zones are more invaded than upland zones.
(Rejmanek and Richardson 1996)	Review	Global	Island isolation from the mainland is correlated with their vulnerability to invasion. Species rich islands are more invaded, but could result because they are more suited to agriculture.
(Robinson et al. 1995)	Experiment	Community	California winter annual grassland with plots $2m^2$ to $32m^2$. Invaded plots typically had greater disturbance (trampled vegetation, bare soil), lowers levels of dominance, and higher species richness. Plot size was not significant. However, 70% of the variation was not explained.
(Smith and Knapp 1999)	Experiment	~60ha	Kanza Prairie, C_4 dominated grassland. Frequent fire reduced the number of non-natives found (e.g., vulnerability to invasion). Non-native richness negatively correlated with ANPP and aboveground grass production. Grazing treatments both showed a positive correlation between native and non-native species richness.
(Smith and Knapp 2001)	Experiment	Local	Northeastern Kansas Kanza Prairie grassland. Fire reduced vulnerability of habitats to invasion. Size of the local pool of non-native species was positively correlated with non-native species, and appears to counter the negative effect of fire on number of non-natives.
(Stohlgren et al. 1998b)	Observational	Multiple	Central grassland riparian versus upland areas with similar grazing intensities. $1m^2$ scale no patterns. $1000m^2$ scale showed riparian zones had higher native and non-native richness than upland sites.

(Stohlgren et al. 1999a)	Observational	Multiple	Rocky Mountains and central grasslands. At the 1m^2 scale, non-native species richness was negatively correlated with plant species richness and cover. At the 1000m^2 scale, non-native species richness correlated with soil % N and native cover. At the landscape scale, non-native species cover correlated to total foliar cover, mean soil % N, and total number of non-native species.
(Stohlgren et al. 1999b)	Observational	Multiple	Central grassland grazed and grazing exclosure areas. At 1m^2 and 1000m^2 scales across all areas, grazed and ungrazed plots did not differ significantly in non-native species richness. Elevation was strongly correlated with non-native diversity. Non-native species richness was positively correlated with soil % N.
(Tyser and Worley 1992)	Observational	Community	Glacier National Park invasion of roadsides and trails in the backcountry. Non-natives decrease with distance from roadsides. Trails exhibited a similar trend, though the effect extended less distance.

Literature Cited

Bergelson, J., J. A. Newman, and E. M. Floresrous. 1993. Rates of weed spread in spatially heterogeneous environments. Ecology 74:999-1011.

Brown, R. L., and R. K. Peet. 2003. Diversity and invasibility of southern Appalachian plant communities. Ecology 84:32-39.

Burke, M. J. W., and J. P. Grime. 1996. An experimental study of plant community invasibility. Ecology 77:776-790.

Dukes, J. S. 2001. Biodiversity and invasibility in grassland microcosms. Oecologia 126:563-568.

Fox, M. D., and B. J. Fox. 1986. The susceptibility of natural communities to invasion. Pages 57-66 in R. H. Groves and J. J. Burdon, editors. Ecology of biological invasions: and Australian perspective. Cambridge University Press, Cambridge.

Greenburg, 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.

Harrison, S. 1999. Local and regional diversity in a patchy landscape: native, alien, and endemic herbs on serpentine. Ecology 80:70-80.

Hobbs, R. J., and L. F. Huenneke. 1992. Disturbance, diversity, and invasion: implications for conservation. Conservation Biology 6:324-337.

Huenneke, L. F., S. P. Hamburg, R. Koide, H. A. Mooney, and P. M. Vitousek. 1990. Effects of soil resources on plant invasion and community structure in Californian serpentine grassland. Ecology 71:478-491.

Larson, D. L., P. J. Anderson, and W. Newton. 2001. Alien plant invasion in mixed-grass prairie: Effects of vegetation type and anthropogenic disturbance. Ecological Applications 11:128-141.

Levine, J. M. 2000. Species diversity and biological invasions: relating local process to community pattern. Science 288:852-854.

Levine, J. M., and C. D'Antonio. 1999. Elton revisited: a review of evidence linking diversity and invasibility. Oikos 87:15-26.

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

MacDonald, I. A., L. L. Loope, M. B. Usher, and O. Hamann. 1989. Wildlife conservation and the invasion of nature reserves by introduced species: a global perspective. Pages 215-255 in J. A. Drake, H. A. Mooney, F. di Castri, R. H. Groves, F. J. Kruger, M. Rejmanek, and M. Williamson, editors. Biological invasions: a global perspective. John Wiley and Sons, New York.

Planty-Tabacchi, A.-M., E. Tabacchi, R. J. Naiman, C. DeFerrari, and H. Decamps. 1996. Invasibility of species-rich communities in riparian zones. Conservation Biology 10:598-607.

Rejmanek, M., and D. M. Richardson. 1996. What attributes make some plant species more invasive? Ecology 77:1655-1661.

Robinson, G. R., J. F. Quinn, and M. L. Stanton. 1995. Invasibility of experimental habitat islands in a California winter annual grassland. Ecology 76:786-794.

Smith, M. D., and A. K. Knapp. 1999. Exotic plant species in a C-4-dominated grassland: invasibility, disturbance, and community structure. Oecologia 120:605-612.

Smith, M. D., and A. K. Knapp. 2001. Size of the local species pool determines invasibility of a C-4-dominated grassland. Oikos 92:55-61.

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

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

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.

Tyser, R. W., and C. A. Worley. 1992. Alien flora in grasslands adjacent to road and trail corridors in Glacier National Park, Montana (U.S.A.). Conservation Biology 6:253-262.

Appendix B. Datasets incorporated into the large database including the data collector, data type, sample size, and any citations resulting from the data.

Source	Data Type	Sample Size	Citation(s)
U.S. Geological Survey	Modified Whittaker Plots*	328	(Stohlgren et al. 1997, Stohlgren et al. 1998a, Stohlgren et al. 1999a, Stohlgren et al. 1999b, Chong 2002)
Researchers	Modified Whittaker Plots	52	(Manier 2002)
Graduate Students	Modified Whittaker Plots*	128	(Petterson 1999, Tobler 2000, Uowolo 2002, Barnett and Stohlgren 2003)
	GPS data	2 (621)	(Goslee et al. 2003)
	Forest Health Monitoring Plots*	51	(Geer 2002)
San Luis Valley GIS Authority	GPS data	6 (1592)	(Magee 2002)
USDA Forest Service	Forest Health Monitoring Plots	33	
	Modified Whittaker Plots	87	(Fornwalt et al. 2003)
The Nature Conservancy	GPS data	15 (413)	(Shultz 2002)
Colorado Natural Heritage Program	GPS data	5 (208)	(Lavender 2002)
Biota Of North America Program	County Species Lists*	63	(Kartesz 2002)
Weed Programs	GPS data	19 (7820)	(May 2002)
National Park Service	Park Species Lists*	8	(Information Center for the Environment 2001)
	GPS data	1 (799)	(National Park Service 2003)
Long Term Ecological Research Sites	Species Lists	2	(Niwot Ridge LTER 2002, Shortgrasse Steppe LTER 2002)
Bureau of Land Management	Daubenmire transects	1264	(Johnston and Huckaby 2001)
	GPS data	33 (3182)	(Vieira 2003)
Colorado State Parks	Species Lists	41	(Billerbeck 2003)
	GPS data	83 (9085)	(Billerbeck 2003)
Colorado Department of Transportation	GPS data	21 (6606)	(Colorado Department of Transportation 2002)

* Data will be served on-line at http://www.niiss.org. Other datasets may be provided dependent upon permission of data owners. Otherwise, contact individuals to receive datasets.

Literature Cited

Barnett, D. T., and T. J. Stohlgren. 2003. A nested-intensity design for surveying plant diversity. Biodiversity and Conservation 12:255-278.

Billerbeck, R. 2003. Colorado State Parks Weed Data. Colorado State Parks.

Chong, G. W. 2002. Multi-scale sampling of native and non-native plant diversity: examples of data analyses and applications. Ph.D. Colorado State University, Fort Collins.

Colorado Department of Transportation. 2002. Colorado Department of Transportation Geographic Data. Colorado Department of Transportation.

Fornwalt, P. J., M. R. Kaufmann, L. S. Huckaby, J. A. Stoker, and T. J. Stohlgren. 2003. Non-native plant invasions in managed and protected ponderosa pine Douglas-fir forests of the Colorado Front Range. Forest Ecology and Management 177:515-527.

Geer, S. 2002. Non-native plant invasions along roads in the Rocky National Park, CO and surrounding area. Masters. Colorado State University, Fort Collins, CO.

Goslee, S. C., K. G. Beck, and D. P. C. Peters. 2003. Distribution of known Russian knapweed in Colorado: Climate and environmental factors. Journal of Range Management 56:206-212.

Information Center for the Environment. 2001. Species in Parks: Flora and Fauna Databases. Information Center for the Environment.

Johnston, B., C, and L. Huckaby. 2001. Ecological types of the upper Gunnison Basin: vegetation-soil-landform-geology-climate-water land classes for natural resource management. Technical Report R2; RR-02001-01, U.S. Dept of Agriculture, Forest Service, Rocky Mountain Region, Denver, CO.

Kartesz, J. T. 2002. Biota of North America Program County Weed Database. Biota of North America Program.

Lavender, A. 2002. Colorado Natural Heritage Program Weed Data. Colorado Natural Heritage Program.

Magee, A. W. 2002. San Luis Valley GIS Weed Mapping Data. San Luis Valley GIS / GPS Authority.

Manier, D. 2002. Modified Whittaker Plot Data. Natural Resource Ecology Laboratory.

May, S. 2002. Larimer County GIS Weed Mapping Data. Larimer County Weed District.

National Park Service. 2003. Data and Information: Data Clearinghouse. National Park Service.

Niwot Ridge LTER. 2002. Niwot Ridge LTER Data Catalog and Search Page. Niwot Ridge LTER.

Petterson, E. S. 1999. Prescribed fire effects on plant communities in Rocky Mountain bighorn sheep habitat. Colorado State University, Fort Collins.

Shortgrasse Steppe LTER. 2002. Data Library. Shortgrasse Steppe LTER.

Shultz, T. 2002. The Nature Conservancy Weed Data. The Nature Conservancy.

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

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.

Tobler, M. A. 2000. Spatial heterogeneity in a protected ponderosa pine forest. Colorado State University, Fort Collins.

Uowolo, A. 2002. Successional development of riparian vegetation on the floodplains of the Green and Yampa Rivers. Masters Thesis. Colorado State University, Fort Collins.

Vieira, J. 2003. Royal Gorge Area Weed Mapping Data. Bureau of Land Management, Canon City Field Office, Canon City, Colorado.

Appendix C. GIS and remotely sensed imagery used as parameters in modeling, including the original scale of the data, any manipulations of the original data, summary statistics for the final raster used, and the source of the original, unmanipulated data.

Auxiliary Variable	Original Scale	Manipulation	Summary Statistics	Original Source
Absolute aspect	~90m grid	Calculated as the absolute value of 180 – aspect, with aspect created from DEM using aspect option in ArcMap (2002)	90.3 (0-180)	National Elevation Dataset (http://gisdata.usgs.net/NED/default.asp)
Slope	~90m grid	Created from DEM using calculate slope option in ArcMap (2002)	7.8 (0-89)	National Elevation Dataset (http://gisdata.usgs.net/NED/default.asp)
Elevation	1 arc-second (~ 30m grid)	Mosaic of DEM tiles using ERDAS-IMAGINE (1998)	2078 (1022 - 4412)	National Elevation Dataset (http://gisdata.usgs.net/NED/default.asp)
Soil fractions of sand, silt, clay	~1000m grid	Calculated average percent for top three standard soil layers (top 20cm) and resampled to 30m grid using SHAPEGRID in ArcINFO (2002)	46, 37, 17 (0-92, 0-70, 0-54)	Soil information for environmental modeling and ecosystem management (Miller and White 1998) (http://www.essc.psu.edu/soil_info/index.cgi?index.html)
Average Annual Precipitation	1:200,000	Resampled to 30m grid in ArcMap (2002)	13.5 (0-70)	Colorado Bureau of Land Management (http://www.co.blm.gov/metadata/cothemes.htm)
Population density per zip code	Zip Code	Area divided by population (people/km^2) converted to 30m grid in ArcMap (2002)	16 (0 – 6567)	ESRI zip code coverage and tiger 2000 census data for U.S. zip codes (http://factfinder.census.gov/servlet/ThematicMapFramesetServlet)
Distance from roads	All roads	Euclidean distance from 30m merged grid of roads using Map Calculator in ArcMap (2002)	1536 (<30 – 23265)	Highways, Major Roads, Local Roads, and Forest Service Roads – Colorado Department Of Transportation
Distance from stream or river	1:100000 Digital Line Graphs	Euclidean distance 30m resolution from lake or stream shapefile using Map Calculator in ArcMap (2002)	863 (<30 – 26859)	Colorado Department of Water Resources

Normalized Difference Vegetation Index	30m	NDVI transformation: (by NASA) $\frac{(NearInfrared - Visible)}{(NearInfrared + Visible)}$	46 (-228 – 412)	Landsat 7 TM mosaic from images by NASA Goddard
Tassel Cap 1	30m	Tassel Cap 1 transformation by NASA	72 (-76 – 490)	Landsat 7 TM mosaic from images by NASA Goddard
Tassel Cap 2	30m	Tassel Cap 2 transformation by NASA	111 (-61 – 511)	Landsat 7 TM mosaic from images by NASA Goddard
Tassel Cap 3	30m	Tassel Cap 3 transformation by NASA	128 (-168 – 303)	Landsat 7 TM mosaic from images by NASA Goddard

Literature Cited

Arc/Info. 2002. Arc/Info. Version 8.2. Environmental Systems Research Institute, Redlands, California.

ESRI. 2002. ArcGIS. Environmental Systems Research Institute, Redlands, California.

ERDAS-IMAGINE. 1998. ERDAS-IMAGINE. ERDAS, Inc, Atlanta, GA.

Miller, D. A., and R. A. White. 1998. A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling. Earth Interactions 2: [Available on-line at http://EarthInteractions.org].

Appendix D. The 48 vegetation classes for Colorado from Theobald et al. (2003) and their area with the associated mean, maximum, and standard deviation of non-native species richness for the model.

Vegetation Type	Area (ha)	Mean	Maximum	Standard Deviation
Residential - low	211019	19.0	32	11.9
Residential - high	1591260	13.0	32	11.2
Urban/ recreational grasses	215402	12.9	32	10.3
Commercial	120527	9.9	32	10.9
Orchards/ vineyards	12568	4.3	18	5.2
Wetlands - herbaceous	171261	2.3	18	1.0
Water	1147530	2.2	32	3.9
Wetlands - woody	32046	2.2	15	1.1
Sand dunes	111131	2.0	3	0.5
Pasture/ hay	8018350	1.8	32	1.8
Eastern plains shrubs	2163480	1.7	32	1.1
Salt desert shrub	7209180	1.7	18	1.1
Riparian shrub	3715240	1.6	32	1.0
Foothills grassland	28593100	1.5	32	2.0
Riparian grassland	8059140	1.5	32	1.7
Row crops	8418160	1.3	32	1.5
Shortgrass prairie	50733500	1.3	32	1.2
Sagebrush intermountain	11306500	1.3	7	0.9
Deciduous riparian	1130320	1.2	29	0.8
Deciduous front range shrub	5649040	1.2	15	0.8
Eastern sandhills sagebrush	48236	1.2	5	0.6
Transitional	220694	1.2	3	0.7
Ponderosa Pine foothills	6745200	1.1	30	1.0
Ponderosa Pine Montane	6586630	1.1	15	1.0
Deciduous intermountain	9144120	1.1	3	0.8
Mixed woodland	4777420	1.0	30	0.9
Ponderosa Pine/ Douglas Fir	3136010	1.0	15	1.0
Gambel oak	2019710	1.0	3	0.9
Small grains	12800800	0.9	32	0.9
Pinon/ Juniper	11509900	0.9	29	0.7
Lodgepole Pine	7917030	0.9	15	0.7
Transitional (quarries/ mines/ gravel)	48817	0.9	7	0.9
Fallow	5808290	0.8	32	1.5
Aspen	15952700	0.8	18	0.7
Mixed forest	1708790	0.8	18	0.7
Subalpine riparian shrub	52774	0.8	6	0.5
Spruce-Fir	5740630	0.8	6	0.7
Douglas Fir	5947470	0.7	14	0.7
Lodgepole Pine with Douglas Fir	3464480	0.7	7	0.5

Vegetation Type	Area (ha)	Mean	Maximum	Standard Deviation
Subalpine shrub	1357580	0.6	6	0.6
Limber pine	636915	0.6	3	0.6
Subalpine meadow	2832680	0.5	6	0.6
Transitional (bare rock/ sand/ clay)	2614800	0.4	32	1.1
Krummholz	487662	0.4	2	0.5
Mid/tall grass prairie	11144900	0.2	10	0.5
Alpine shrub	1529000	0.2	2	0.4
Alpine meadow	5068300	0.2	2	0.4
Snow/ ice	340286	0.1	3	0.2

Appendix E. Location of Modified Whitaker plots used to test the spatial and non-spatial GLM models for non-native species richness in Colorado including the number of non-native species recorded in the survey of the plot and the number predicted by the spatial GLM and by the non-spatial GLM.

Location	Survey Count	Spatial GLM	Non-Spatial GLM
Hayman Burn	1	6	6
Hayman Burn	5	6	6
Hayman Burn	2	6	6
Hayman Burn	2	6	6
Hayman Burn	2	6	6
Hayman Burn	7	5	5
Hayman Burn	3	6	5
Hayman Burn	4	5	5
Hayman Burn	8	5	5
Hayman Burn	10	5	5
Hayman Burn	4	6	6
Hayman Burn	3	6	6
Hayman Burn	6	5	6
Hayman Burn	6	5	5
Hayman Burn	7	6	6
Hayman Burn	6	6	6
Hayman Burn	1	6	5
Hayman Burn	3	6	6
Hayman Burn	3	6	6
Hayman Burn	3	6	6
Gunnison Basin	1	1	2
Gunnison Basin	1	1	1
Gunnison Basin	1	1	1
Gunnison Basin	3	1	1
Gunnison Basin	5	1	1
Gunnison Basin	5	1	1
Gunnison Basin	1	1	1
Gunnison Basin	2	1	1
Gunnison Basin	2	1	1
Gunnison Basin	1	0	1
Gunnison Basin	3	1	1
Gunnison Basin	2	1	1
Gunnison Basin	1	1	1
Gunnison Basin	3	1	1
Gunnison Basin	3	1	1
Gunnison Basin	3	1	2
Gunnison Basin	3	1	2
Gunnison Basin	2	1	2
Gunnison Basin	2	1	2

Location	Survey Count	Spatial GLM	Non-Spatial GLM
Gunnison Basin	3	1	2
Gunnison Basin	4	1	2
Gunnison Basin	2	1	2
Gunnison Basin	2	1	2
Gunnison Basin	3	1	2
Gunnison Basin	2	0	1
Gunnison Basin	3	6	1
Gunnison Basin	3	6	1
Gunnison Basin	4	6	1
Gunnison Basin	3	6	1
Gunnison Basin	1	6	1
Gunnison Basin	3	5	1
Gunnison Basin	4	6	2
Gunnison Basin	4	5	1
Gunnison Basin	2	6	1
Gunnison Basin	4	6	2
Gunnison Basin	3	6	2
Gunnison Basin	3	6	2
Front Range Foothills	5	6	1
Front Range Foothills	2	5	2
Front Range Foothills	1	6	1
Front Range Foothills	4	5	2
Front Range Foothills	1	5	2
Front Range Foothills	2	5	1
Front Range Foothills	1	6	2
Front Range Foothills	4	6	2
Front Range Foothills	3	5	2
Front Range Foothills	3	5	1
Front Range Foothills	2	6	1
Front Range Foothills	3	6	1
Front Range Foothills	13	6	1

Appendix F. S-plus functions used with example syntax in italics.

spwtdist: calculate the spatial weights matrix

```
sprich.wt<-spwtdist(sprich[,1],sprich[,2])
```

combscr: combinatorial screening for OLS variables

```
combscr(sprich[,3],sprich[,5-17])
```

ols: Run OLS regression using variables from combscr

```
sprich.ols<-ols(sprich[,3],sprich[,c(5,9,14,16,17,12,13)],w=sprich.wt)
```

glm: create full model glm

```
sprich.glm<-glm(NNATCNT ~ ABASP + SLP + DEM + SLT + CLY + PRECP + SND + POPDEN +
        RDDST + TC2 + TC3 + TC1 + NDVI, data = sprich, quasi=(link(log,variance= 'mu'))
eues.glm<-glm(NNATCNT ~ ABASP + SLP + DEM + SLT + CLY + PRECP +  SND + POPDEN +
        RDDST + TC2 + TC3 + TC1 + NDVI, data = eues, family = binomial)
```

step(glm): use stepwise to select the variables to use

```
step(sprich.glm)
```

hist: examine histogram of glm residuals

```
hist(residuals.glm(eues.glm),nclass = 20)
```

morani: check residuals for spatial autocorrelation

```
morani(residuals.glm(eues.glm),w=eues.wt)
```

variogrm: create variogram to obtain estimates of nugget, sill, and range

```
eues.var<-variogrm(eues[,1],eues[,2],residuals.glm(eues.glm),15,dmax=6000)
```

fitvar: fit the three types of variograms using estimate values from variogrm for nugget, sill, and range

```
eues.gau(or sph or exp)<-fitvar(eues.var,0.05,0.15,600,model="gau"(or sph or exp),wt=T)
```

lst: create a list of the spatial extent of the area of interest and the variable of interest

```
eues.lst<-list(x=eues[,1],y=eues[,2],z=residuals.glm(eues.glm),xl=140012,xu=763227,
        yl=409413,yu=4546663)
```

crossval: perform cross-validation using lst, the fitvar curve, and guess of number of nearest neighbors-redo with different numbers to maximize the R^2 and minimize the standard error

```
sprich.val<-crossval(sprich.lst,sprich.sph,5)
```

var(crossval$resid)^5: R^2 for cross-validation

1-var(crossval$resid)/var(variable of interest): standard error for cross-validation

Import the residuals to ArcGIS and krig them using the variogram type and values from fitvar and the number of nearest neighbors.

Appendix G. Cross-correlation results of native species richness and non-native species richness with variables from Table 2-1 not chosen as significant in the regression which if exhibiting significant cross-correlation could be used in co-kriging. The only significant variable at $\alpha = 0.01$ was non-native richness with tassel cap 1, but the cross-correlation statistic was less than 0.1.

	Variable	Cross-correlation Statistic	p-Value
Native Species Richness	Tassel Cap 1	-0.02	0.06
	Tassel Cap 2	0.01	0.32
	Tassel Cap 3	0.02	0.02
	NDVI	0.005	0.61
Non-native Species Richness	Sand	-0.02	0.06
	Silt	0.008	0.45
	Tassel Cap 1	-0.05	< 0.01