Sampling and modeling plant infestations : alternatives for identifying invasive plant distributions in rangeland environments
by Elizabeth Ann Roberts

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Land Resources and Environmental Sciences
Montana State University

Abstract:
Millions of hectares of North American rangelands are infested with invasive plant species (Lajeunesse et al. 1999). Consequently, the integrity of our natural systems and viability of our regional economy are threatened by the spread of these exotics. In order to successfully adapt our management strategies and improve our ability to halt the spread of invasive plants, the production of accurate, time-repeated infestation maps is critical. Traditional survey mapping is the standard for mapping invasive plant infestations. Survey mapping, however, is time consuming and often completed ineffectively. Time constraints limit the possibility of repeated mapping and the resulting maps are often not accurate; faster, more efficient mapping methods are necessary. In this study, alternatives to traditional survey mapping were examined. Presence/absence GPS-based infestation maps of Agroptilon repens L. and Centaurea maculosa Lam. from two different Montana rangeland sites were used to (1) test the usefulness of inverse distance weighting (EDW) for predicting invasive plant locations, and (2) determine whether available GIS layers could improve prediction success attained by IDW. An in-the-field accuracy assessment was completed for the GPS-based infestation maps. At both sites, map accuracies were high, and were considered reasonable representations of the invasive plant distributions. Samples were gathered from the GPS-based infestation maps through repeated computer-based sampling simulations. Three sampling methods and six sampling densities were tested. IDW was applied to each sampling strategy (sample method x sample density) to predict the presence or absence of the invasive plant species. The GPS-based infestation maps were used as references to determine the accuracy of EDW interpolation results. Some differences among the 18 sampling method x sampling density combinations were detected from the prediction accuracies using ANOVA and multiple comparison analysis. Sampling at a density of 0.25% (~1pt/ha) with a systematic sampling method was determined to be the preferred sampling strategy at our sites. This strategy resulted in overall accuracies near to and above 85%. Preliminary classification tree analysis was also conducted to test the relationship between readily available GIS data layers and invasive plant locations. Results indicated, of 8 GIS data layers, proximity to predicted nearest invasive plant was overwhelmingly the strongest predictor in determining invasive plant locations. Some correlations were found between 4 of the other 7 GIS layers and the invasive plant locations at each site. We concluded, of the variables tested (aside from invasive plant proximity), none offered enough predictive power for use in alternatives to traditional mapping at the two sites. IDW in combination with a systematic sampling at a density of ~lpt/ha, however, was recommended for predicting presence or absence of C. maculosa and A. repens distributions.

SAMPLING AND MODELING PLANT INFESTATIONS: ALTERNATIVES FOR IDENTIFYING INVASIVE PLANT DISTRIBUTIONS IN RANGELAND ENVIRONMENTS

by

Elizabeth Ann Roberts

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Land Resources and Environmental Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

December 2001

APPROVAL

Of a thesis submitted by

Elizabeth Ann Roberts

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic styles, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. Roger L. Sheley — Roger L. Sheley
Co-chairperson, Graduate Committee Date

Dr. Rick Lawrence — Rick Lawrence 12/6/01
Co-chairperson, Graduate Committee Date

Approved for the Department of Land Resources and Environmental Sciences

Dr. Jeffrey Jacobsen — Jeffrey S. Jacobsen 12/7/01
Head, Major Department Date

Approved for the College of Graduate Studies

Dr. Bruce R. McLeod — Bruce R. McLeod 12-10-01
Graduate Dean Date

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University-Bozeman, I agree that the Library shall make it available to borrowers under rules of the library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis is whole or in parts may be granted only by the copyright holder.

Signature Elizabeth A Roberts

Date 12/07/01

ACKNOWLEDGEMENTS

I would like to thank Roger Sheley for providing the inspiration, enthusiasm, guidance, and humor necessary to conduct this project. Thanks to Drs. Rick Lawrence and Richard Aspinall for their expertise in all things spatial. Thank you to Dr. Bruce Maxwell for being dedicated to practical land management issues and for doing the science right in a way that people can use. I have a great deal of admiration and owe many thanks to my GIS and invasive plant colleagues, and their dedication to answering questions involving practical land management questions and the stewardship of our natural landscapes. Thank you also to Rob Lloyd for being ever-present and ever-supportive.

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Millions of hectares of North American rangelands are infested with invasive plant species (Lajeunesse et al. 1999). Consequently, the integrity of our natural systems and viability of our regional economy are threatened by the spread of these exotics. In order to successfully adapt our management strategies and improve our ability to halt the spread of invasive plants, the production of accurate, time-repeated infestation maps is critical. Traditional survey mapping is the standard for mapping invasive plant infestations. Survey mapping, however, is time consuming and often completed ineffectively. Time constraints limit the possibility of repeated mapping and the resulting maps are often not accurate; faster, more efficient mapping methods are necessary. In this study, alternatives to traditional survey mapping were examined. Presence/absence GPS-based infestation maps of *Agroptilon repens* L. and *Centaurea maculosa* Lam. from two different Montana rangeland sites were used to (1) test the usefulness of inverse distance weighting (IDW) for predicting invasive plant locations, and (2) determine whether available GIS layers could improve prediction success attained by IDW. An in-the-field accuracy assessment was completed for the GPS-based infestation maps. At both sites, map accuracies were high, and were considered reasonable representations of the invasive plant distributions. Samples were gathered from the GPS-based infestation maps through repeated computer-based sampling simulations. Three sampling methods and six sampling densities were tested. IDW was applied to each sampling strategy (sample method x sample density) to predict the presence or absence of the invasive plant species. The GPS-based infestation maps were used as references to determine the accuracy of IDW interpolation results. Some differences among the 18 sampling method x sampling density combinations were detected from the prediction accuracies using ANOVA and multiple comparison analysis. Sampling at a density of 0.25% (~1pt/ha) with a systematic sampling method was determined to be the preferred sampling strategy at our sites. This strategy resulted in overall accuracies near to and above 85%. Preliminary classification tree analysis was also conducted to test the relationship between readily available GIS data layers and invasive plant locations. Results indicated, of 8 GIS data layers, proximity to predicted nearest invasive plant was overwhelmingly the strongest predictor in determining invasive plant locations. Some correlations were found between 4 of the other 7 GIS layers and the invasive plant locations at each site. We concluded, of the variables tested (aside from invasive plant proximity), none offered enough predictive power for use in alternatives to traditional mapping at the two sites. IDW in combination with a systematic sampling at a density of ~1pt/ha, however, was recommended for predicting presence or absence of *C. maculosa* and *A. repens* distributions.

INTRODUCTION

Invasive plant maps, created repeatedly over time, are critical to successful invasive plant management and eradication. In many cases, preliminary mapping is completed, management strategies are developed and implemented, but subsequent maps are not created. Consequently, management efforts are not properly assessed and management is not optimized; this wastes vital resources and leads to inefficient invasive plant control. The absence of repeated mapping is attributed to the time and cost of producing accurate survey maps (Austin 1998; Lees et al. 1991; Moore et al. 1991; Nicholls 1989). Therefore, faster, more efficient methods for producing accurate invasive plant distribution maps are needed.

Traditional survey mapping is the standard method for creating invasive plant maps. Traditional surveys combine hand or GPS mapping with on-the-ground or aerial surveying to delineate infestation boundaries (Johnson 1999). The high map accuracy necessary (recommended >80%) for evaluating management success is difficult to achieve and requires intensive mapping efforts (Cooksey et al. 1999). Successful mapping requires evaluation of the entire management area and only a few species mapped at a time. Alternative methods, using sampling strategies and interpolation or correlation modeling, might be able to replace traditional surveys by producing equivalent or higher accuracies and decreasing the time and money required for mapping, thus making repeated mapping more feasible.

Some research has been conducted using sampling and models for predicting plant distributions as an alternative to traditional mapping methods. The majority of the work, for invasive plant mapping, however, has focused on small areas dealing with simplified, agricultural monocultures (Aarts 1986; Donald 1994; Heisel et al. 1996). A few regional predictive mapping studies have been reported (Austin et al. 1990; Hill et al. 1997; Lees et al. 1991). Little research, however, has been devoted to invasive plant predictions on rangeland systems at the management scale (e.g. 200 to 3000ha). Given the many people and indigenous wildlife relying on viable and healthy rangelands, successful invasive plant management is essential in the rangeland environment.

Interpolation modeling holds a possibility for predicting infestation distributions. Only values at sample locations are required to build an interpolation model. Success is hinged on the level of spatial correlation among the sample points. Correlation methods might also be useful. Once a correlation model is developed, sampling might not be necessary (if a robust model can be developed). For a correlation model to be used for predictive mapping, strong correlations between predictor variables and invasive plant locations must be demonstrated.

The overall goals of this study were to (1) identify whether a simple interpolation method could be used to create presence/absence distribution maps at two sites for two invasive plant species; *Agroptilon repens* L. (Russian knapweed) and *Centaurea maculosa* Lam. (spotted knapweed) and (2) determine whether the correlations between readily available GIS data layers and the invasive plant locations, would have the potential to improve prediction success of the interpolation model. Chapter 2 is a review of existing mapping literature as it pertains to invasive plants and discusses the issues surrounding successful distribution models. Chapter 3 investigates whether inverse distance weighted (IDW) interpolation can act as an alternative to traditional survey mapping. Chapter 3 also discusses a preferred sampling method and sampling density for predicting *A. repens* and *C. maculosa* distributions. Chapter 4 focuses on whether correlations between invasive plant locations and 7 medium-resolution and 1 high-resolution GIS data layers could be used to improve IDW prediction results. The high-resolution data layer reviewed in Chapter 4 was derived from IDW results described in Chapter 3. The other GIS data layers were acquired from easily accessed Internet sources (data clearinghouses). For the methods common to both the IDW (Chapter 3) and the classification tree analysis (Chapter 4) portions of the study, see the methods and results sections in Chapter 3. Common methods include study site descriptions, details on obtaining the GPS-based infestation maps, and completion of accuracy assessments for the GPS-based infestation maps. Chapter 5 summarizes the findings of Chapter 3 and 4 and continues the discussion of alternative methods to traditional mapping as practical tools for land managers.

LITERATURE REVIEW

Introduction

Use of sampling and computer modeling to predict invasive plant distributions has become both more popular and more feasible since the early 1980s when technological advances in computing first allowed analysis of large data sets (Franklin 1995). Some approaches and methods to create invasive plant distribution maps have been tested (Donald 1994; Heisel et al. 1996). Issues critical to this work are considerations of scale (Bian et al. 1993; Stohlgren et al. 1997), determining an appropriate sampling method (Fortin et al. 1989; Mohler 1983; Stohlgren et al. 1998), and determining an appropriate sample density (Fortin et al. 1989).

Scale

Spatial scale can refer to either the geographic area of a study site (i.e., its extent) or the degree of detail the study attempts to describe (i.e., its resolution or grain) (Goodchild et al. 1997; Wiens 1989). In terms of invasive plant predictions, plant location patterns appear differently depending on the scale at which they are evaluated (Levin 1992). Therefore, research objectives must be clearly defined before an appropriate experimental scale can be determined (Franklin 1995; Kent et al. 1992). In terms of scale and invasive plant management, the experimental and management scales (both extent and resolution) should be equal. This approach is necessary because it avoids having to change from the experimental scale to management scale; scaling errors are prevented and better management results (Firbank 1993). Often, the importance of scale in prediction mapping is overlooked and is based on the convenience of the experimenter or logistical capabilities (Firbank 1993; Levin 1992). Some factors important for determining an appropriate experimental scale are: (1) what species are being considered, (2) how much variability needs to be detected, (3) at what resolution is the variability detectable, (4) whether the data collection at a certain scale is feasible for land managers, and (5) what is the appropriate computer modeling grain size (Weins et al. 1986).

Choice of species is fundamental to the identification of an appropriate study scale. A resolution providing a highly detailed description of one species might be unable to resolve information for a different species (Firbank 1993). When multiple species are being considered in a single model, therefore, their ecology and growth patterns should be similar. Land management objectives and the ecology of plant species drive the amount of variability needed for species detection. If land managers wish to detect rare or newly invading plant species, then the amount of variability required to determine their abundance would be much higher than if the managers needed to determine abundance of a fully established species. Under conditions where a high amount of variability detection is necessary, a finer resolution and greater amount of data collection are required (Goedickemeier et al. 1997). If less variability detection is needed, the scale (both the extent and resolution) might be broadened, but the ability to detect variability will decrease (Levin 1992).

Choosing the appropriate resolution for the predictive invasive plant model is also important. Researchers suggest matching the resolution of the model (e.g., grid cell size) with the minimum mapping unit (MMU). Selecting a modeling resolution below the MMU is inappropriate because elements below the grain size are indeterminable (Wiens 1989). Using a resolution greater than the MMU is also problematic due to the "new properties" that emerge when cells are aggregated to coarser resolutions (Bian 1997).

Determining Sampling Method

Testing sampling methods and densities is an essential preliminary step for all ecologically-based research (Green 1979). For management purposes and in plant mapping, there are general guidelines for choosing an appropriate sampling method. Sampling strategies should: (1) provide the most amount of information with the least amount of expense (cost efficiency) (Stohlgren et al. 1997), (2) be based on sound statistical designs meeting the requirements of the predictive model (Dale 1999), and (3) be able to capture the variation in the response variable (Fortin et al. 1989).

Commonly used sampling methods for prediction mapping are systematic sampling and simple random sampling. Systematic sampling is a straightforward sampling scheme that is easy to follow in the field. Systematic sampling produces interval data that is evenly distributed across the entire geographic extent of the study area. A drawback to the systematic sampling scheme is that all subsequent samples are determined by the location of the first sample (Bourdeau 1953). In simple random sampling, every location in the survey area has an equal probability of being sampled. Simple random sampling produces unbiased samples and is considered best suited for standard statistical tests (Goedickemeier et al. 1997). A drawback to this method is the potential production of a sample that is not representative of the response variable.

It is unclear which sampling method is best for plant mapping. A study evaluating the prediction of oak-hickory densities and basal areas found systematic sampling performed better than random sampling (Bourdeau 1953). Systematic sampling has also been preferred in cases of low response levels, because it requires smaller sample sizes are able to capture the same amount of variation as larger random samples (Wildi 1986). Systematic sampling methods might also be more efficient than random sampling at the same sample size if there is strong spatial autocorrelation (Moore et al. 1986). However, systematic sampling has been found to be unable to capture the response variability when the frequency and intervals of the samples are out of phase with plant patterns (Fortin et al. 1989).

Determining Sample Size

Determining an appropriate sampling size is also a necessary precursor to developing a model. Sample sizes often must be increased when species are rare or a large number of species are being predicted. Larger sample sizes increase the likelihood of capturing the spatial relationships or spatial autocorrelation of plant distributions. Since spatial autocorrelation is the basis of interpolation models, larger sample sizes have the ability to improve interpolation model accuracy. The sample size, however, must be at a feasible size to collect. In addition, in some cases, increasing sample size might not improve prediction. Interpolation accuracy results for predicting sugar maple densities indicated, while increasing the sample size did improve prediction accuracy, choosing the appropriate sampling method was more important than increasing the sample size in capturing the spatial structure of response (Fortin et al. 1989).

Interpolation Modeling

Using spatial interpolation methods to build species distribution maps requires spatially autocorrelated sample points. These interpolation models are typically applied to continuous data. Such models produce a range of values at unknown locations based on the distance and numerical relationships among known locations. In inverse distance weighting (IDW), the user defines an exponential distance weighting power. The weighting value is applied to the IDW equation and fit to a set number of closest points.

The higher the exponent, the greater weight nearby known values will have in prediction (Bowman et al. 1995).

In invasive plant mapping, interpolation models are typically used to predict density values (Donald 1994; Heisel et al. 1996). Interplolation results can be adapted to non-continuous data by reclassifying ranges of values into categories. Typically, interpolation models are considered limited in the information they can provide. Interpolation models are designed to predict response values. Unlike correlation models, interpolation modeling cannot explicitly convey, or use, relationships among other environmental factors that might influence plant distributions. The success of the interpolation model however, can provide additional information aside from distribution; it can identify how much a plant's location is spatially correlated to its neighbor.

Correlation Modeling

Correlation models use variables to predict response values. They can also identify relationships between the response values and variable combinations and/or values within each variable. Using correlation models allow researchers to glean information potentially related to the ecology driving plant distribution and not just the plant's physical locations. Decision trees are often used for predicting plant distributions (Austin 1998; Carpenter et al. 1993; Jeltsch et al. 1998; Lees et al. 1991; Lenihan et al. 1993; Moore et al. 1991). Decision trees are based on a hierarchical structure of rules and recursive partitioning (Breiman et al. 1984). They are often referred to as automated taxonomic keys (Moore et al. 1991). Prediction occurs in a series of binary splits, each based on the rule resulting in the greatest increase in "class purity." There are two types of decision trees, regression and classification. Regression trees use least squares estimations and, therefore, require continuous responses. Classification trees do not use a least squares error process and, therefore, have no limitations on the types of predictive data sets used, but require a categorical response.

By design, decision trees are relatively insensitive to outliers (Breiman et al. 1984; Mathsoft 1999). Evaluation of results is straightforward and pragmatic because the variables and values at each binary split are identified (Franklin 1995). For plant prediction, the description of the variables and variable value at each split helps determine whether the splits used make sense ecologically.

Correlation Predictor Variables

Successful correlation models require data layers with strong relationships to the response. For plant prediction mapping, these variables should be based on the ecology of the plant species being studied. In addition, the data layers must be trustworthy (accurate) and, for practical management purposes, the data layers must be easily obtained.

Direct gradients (temperature and pH) and resource gradients (light, water, nutrients, carbon dioxide, and oxygen) determine plant patterns in the natural landscape (Austin et al. 1984; Franklin 1995; Moore et al. 1991). Since data for direct gradients are difficult to gather, indirect gradients (climate, geology, and other vegetation) correlated with plant distributions are used. Climatic influences are often represented by temperature, precipitation, and elevation variables. Geology is another influential factor in predicting plant patterns (Lees et al. 1991; Moore et al. 1991). Soil type and slope are indicators of parent material and other direct gradients such as soil water capacity (Despain 1973; Lees et al. 1991; Moore et al. 1991).

Specific to invasive plant species, disturbance has been found to play a key role in distribution patterns. A grassland study in Glacier National Park, Montana found human and animal disturbance increased susceptibility to plant invasion (Tyser et al. 1988). Disturbance factors are numerous and can include human activities and wildlife, as well as abiotic factors, such as fire, floods, landslides, etc. Examples of disturbance indicators are road proximity, land management practices, such as grazing and logging, and proximity to hiking and wildlife trails. Riparian habitats and waterways are consistently disturbed by flooding and are prime locations for infestations (Baker 1986).

An additional predictive variable, with the potential for having a strong influence on invasive plant distributions, is proximity to same-species invasive plant locations. The influence of neighboring same-species locations is an indicator of spatial dependence among sample points and is often avoided in correlation modeling because of the confusion it creates in the statistical model (Franklin 1995). Proximity of same-species plant individuals, however, has the potential to be a source of information rather than noise. If an inherent interdependence of distance between individual invasive plants existed, a proximity to nearest invasive plant variable could be incorporated into a correlation model as a means of identifying areas with a high probability for infestation.

USING SAMPLING AND INVERSE DISTANCE WEIGHTED MODELING FOR MAPPING INVASIVE PLANTS

Introduction

Spatial modeling is an increasingly sought, time-saving alternative to traditional survey methods for generating invasive plant distribution maps (Donald 1994; Heisel et al. 1996). Interpolation models use samples and spatial relationships among these samples to predict values at unknown locations. Interpolation modeling is commonly used to predict continuous variables, such as density. Interpolation models can also be used for predicting categorical data (i.e., presence/absence) by binning ranges of values into separate groups. Of the many interpolation methods, an easy to use, highly accessible method is inverse distance weighting (IDW). Like other interpolation methods, IDW uses linear combinations of weights at known points to estimate unknown location values (Fig. 1).

$$\hat{Z}(s_o) = \sum_{i=1}^{n} \lambda_i Z(s_i)$$

Figure 1. Linear interpolation equation used in IDW.

$\hat{Z}(s_0)$ equals the values at unknown locations and is determined by the weighting value (λ_i) and values at known locations $Z(si)$. In the IDW equation, $d(s_i,s_o)$ is the Euclidean

$$\lambda_i = [d(s_i,s_o)]^{-p} / \sum_{i=1}^{n} [d(s_i,s_o)]^{-p}$$

Figure 2. IDW definition of weights.

distance between s_i and s_o (Fig. 2). P is the power value selected to control how fast the weights tend to zero as the distance from the location increases. The higher the exponent, the more influence nearby known values will have on predicted values (Bowman et al. 1995).

Numerous papers have been published identifying IDW as a valuable interpolation method (Bowman et al. 1995; Collins et al. 1996; Dirks et al. 1998; Skov 2000). For invasive plant prediction, however, kriging is typically used instead of IDW (Donald 1994; Heisel et al. 1996). Despite limited use of IDW in invasive plant mapping, research in other disciplines has found IDW can rival other interpolation methods (Bowman et al. 1995; Dirks et al. 1998; Gotway et al. 1996).

Sampling is required for interpolation modeling, consequently, choosing an appropriate sample method is key for successful model development. The appropriate sampling method for predictive mapping is dependant upon the management objectives and plant distribution patterns (Elzinga et al. 1999; Fortin et al. 1989).

Three sampling options are evaluated in this study; these are systematic sampling, random sampling, and a hybrid, systematic-random sampling method. Each method has benefits, as well as limitations. Systematic sampling is commonly used because generating sample locations and gathering field samples in a grid pattern is relatively simple. A drawback to systematic sampling is that it is constrained by the sampling interval relative to response distribution. If the sampling interval is out of phase with the response and response patch sizes are not twice the size of the sampling interval, chances of obtaining a representative sample are low. In simple random sampling, every location in the survey area has an equal probability of being sampled. Gathering a representative sample with random sampling can also be difficult because it often requires large sample sizes (Goedickemeier et al. 1997). Therefore, in cases where a large sample size is gathered, but patch sizes are small, random sampling might be superior to systematic. Systematic sampling might be more successful in cases of strong autocorrelation (Moore et al. 1986). Less review of systematic-random sampling exists in the literature. It was used in this study, however, as an attempt to take advantage of the benefits of both the systematic and random sampling methods.

In order for IDW to be a successful tool for invasive plant managers, the accuracies for predicting invasive plant locations must achieve a level to make sampling desirable over traditional mapping methods. High accuracies are recommended for managers to quantitatively assess their management strategies and improve their management efforts (Cooksey et al. 1999). Managers, however, might choose moderate levels of accuracy if mapping time is decreased sufficiently. What is desirable, therefore, are accurate maps, created from sample sizes small enough to save time and money.

The purpose of this study was to evaluate the success of three sampling methods (random, systematic, and systematic-random) and six sampling densities using IDW to predict *Agroptilon repens* L. (Russian knapweed) and *Centaurea maculosa* Lam. (spotted knapweed) distribution patterns. We predicted that the systematic sampling method would consistently produce higher accuracy values than other methods. In addition, we predicted accurate presence/absence maps could be produced using sample sizes practical (time-saving) for land managers.

Methods

A Geographic Information System (GIS) and spatial interpolation methods were used to examine the ability of varying sampling strategies to predict *A. repens* and *C. maculosa* distribution patterns within two Montana rangeland environments. Invasive plant distribution maps, with known accuracies, were collected in the field using Global Positioning Systems (GPS). Predicted maps were created from sampling the GPS-based infestation maps. Accuracy of predicted maps were determined by using the GPS-based infestation maps as a reference. The study was completed in 7 general steps: (1) gathering invasive plant GPS locations at two study sites, (2) conducting accuracy assessment of the GPS-based infestation maps at each site, (3) restructuring invasive plant data from cover categories into present and absent categories and converting maps into raster format, (4) simulating 18 different sampling strategies (3 sample methods x 6 sample densities, with 3 replications of each) using a GIS and the GPS-based infestation maps, (5) conducting inverse distance weighted interpolation calculations on all simulated sample data sets, (6) assessing accuracies of the resulting presence/absence distribution maps using the original GPS-based infestation maps as references, (7) evaluating differences in accuracies with ANOVA and multiple comparison analysis (MCA).

Study Sites

Prediction success was evaluated for invasive plant distributions at two sites. The *A. repens* site is a 600ha riparian zone along the Missouri River on the Charles M. Russell Wildlife Refuge in northcentral Montana (Extents: 47°41'30"N, 108°47'30"W and 47°38'N, 108°42'30"W—NAD27). Elevation at the site ranged from 600 to 900m. Average annual precipitation was 25 - 31cm. The study area was infested primarily with *A. repens*. *A. repens* is an aggressive perennial. It produces seeds, but spreads primarily by rhizomitous adventitious roots. *A. repens* is able to suppress growth of nearby plants because of its rhizomitous root system, allelopathic properties, and its primarily local spread. Based on these properties, *A. repens* tends to form dense stands in areas with shallow water tables or extra water from irrigation (Watson 1980). Native vegetation at the *A. repens* site included *Salix* spp. (willow), *Populus deltoids* Bartr. ex Marsh. (cottonwood), *Symphoricarpos albus* (L.) Blake (snowberry), *Sarcobatus vermiculitus* (Hook.) Torrey (greasewood), and *Chrysothamnus viscidiflorus* (Hook.) Nutt. (rabbitbrush). Other non-natives were *Cirsium arvense* L. Scop. (Canada thistle), *Eurphorbia esula* L. (leafy spurge), *Centaurea maculosa* Lam. (spotted knapweed), *Cardaria pubescens* (C.A. Mey.) Jarmolenko (whitetop), *Agropyron cristatum* (L.) Gaertn. (crested wheatgrass), and *Bromus inermis* Leyss. (smooth brome).

The *C. maculosa* site encompassed 1200ha of upland, mixed forest-rangeland on the Northern Cheyenne Indian Reservation in southeastern Montana (Extents: 45°45'N, 107°00'W and 45°37'30"N, 106°52'30"W—NAD27). Intermittent streams ran throughout this area and elevation ranged from 900 – 1500m. Average annual precipitation was 36 -

41cm. This area was primarily infested with *C. maculosa. C. maculosa* is a rapidly spreading exotic invading much of the northwestern United States (Sheley et al. 1999). It is a taprooted perennial and produces large numbers of seeds. Seeds are dispersed both locally and over long distances, with local extension of peripheral stands playing a large role in its spread (Watson et al. 1974). Local dispersal occurs when animals jar plants and loosen the seeds, typically 1 - 2m from parent plants. Long distance dispersal occurs when seeds become attached to passing people, animals, and/or vehicles. Seeds can be carried along watercourses and are transported with crop seeds and hay. Native vegetation included *Pinus ponderosa* Dougl. ex Laws. & Laws. (ponderosa pine), *Juniperus scopulorum* Sarg. (juniper), *Pseudorogneria spicatum* (L.) Gaertn. (bluebunch wheatgrass), and *Agropiron smithii* (Rydb.) Gould (western wheatgrass). An additional non-native at this site was *Bromus japonicus* Thunb. ex Murr. (Japanese brome).

GPS-based Infestation Maps

In order to test the success of different sampling strategies using IDW, accurate, complete invasive plant distribution maps were required. At the *A. repens* site, point, line, and area infestation data were collected in 1997 and 1998 by United States Fish and Wildlife Service personnel using intensive ground and helicopter GPS mapping methods. At the *C. Maculosa* site, point, line, and area infestation data were collected in 1999 using helicopter GPS mapping methods. Data at both sites were collected according to the Montana Noxious Weed Survey and Mapping System standards (Cooksey et al. 1998). Infestations less than 2.02ha (5 acres) were identified as points and attributed within one of three size ranges 0 – 0.04ha, 0.04 – 0.4ha, and 0.4 – 2.02ha. Infestations greater than 2.02ha (5 acres) were mapped as area features. Infestations following linear features were collected as lines with an associated buffer width value. The original GPS-based infestation maps were collected and attributed with either a low, moderate, or high cover classification. Data were later reclassified as present or absent. Present locations were given a value of 1 and absent areas were assigned a value of zero. After the infestation maps were collected, the point and line data were buffered to areas equal to infestation size identified by the GPS data collector.

Accuracy Assessment of GPS-based Infestation Maps

Accuracy assessments of the infestation data at both study sites were conducted in the fall of 1999. An 85% confidence in accuracy assessment results required at least eleven random points per cover category (Tortora 1978). These were randomly selected for each of the four categories mapped (Absent, Present-Low, Present-Moderate, Present-High). Using Rockwell GPS Pluggers with 5 – 15 m navigational accuracy, points were located and accuracy was assessed. A contingency matrix was calculated based on field assessments for only presence/absence, causing a greater number of evaluated presence locations, than absent. Due to access constraints, accuracies for only 10 of the 11 absent locations were assessed at the *A. repens* site.

Assessing Sampling Methods and Sampling Densities

The GPS-based infestation maps were converted from vector to raster (grid) format in the GIS. The cell size was determined by the need to match experimental scale with management scale (Firbank 1993) and was set to a resolution of 5m. For the *A. repens* site, the raster grid included 5,987,050 cells. For the *C. maculosa* site, the total number of grid cells was 13,517,075.

Sampling Strategies

Eighteen sampling strategies were evaluated. The sampling strategies were based on three sampling methods (systematic, random, and systematic-random) and six sampling densities (0.04, 0.06, 0.08, 0.11, 0.16, and 0.25%). The six sample densities were approximately 0.2, 0.3, 0.35, 0.45, 0.7, and 1.0 pts/ha, respectively.

Samples were gathered by applying each sample strategy (sample method x sample density) to the rasterized digital GPS-based infestation maps. In order to test for differences among sampling strategies, each strategy was replicated 3 times for the 18 method x density combinations for a total of 54 sample data sets at each site. Sampling of the GPS data sets was completed in Environmental Systems Research Institute's (ESRI) ArcViewGIS software. GIS-based computer code was written to automate the sampling process. For systematic sampling, the first sample point in the systematic sampling strategies was randomly shifted +/- 50m. Systematic-random samples were generated by randomizing sampling locations on the y-axis and setting an even sampling interval along the x-axis.

Analysis

At each site, the 54 data set combinations were analyzed using the IDW interpolation function in ArcView Spatial Analyst. User inputs to the IDW function were a power value (exponent) and *n* (the number of nearest sample points used in the interpolation of each cell). Distance power values used in IDW typically range between 1 and 3, with 2 being the most common (Gotway et al. 1996). In this project, a value of 2 was used. *N* can either be set as a fixed number of sample points or radius distance value. Researchers have used sample points values ranging from $6 \geq n \geq 24$ (Zimmerman et al. 1999). Twelve sample points, within range of recommended values for abruptly changing surfaces, was chosen (Declercq 1996). The IDW function was applied to each sample data set and produced a grid map with continuous predicted values ranging from 0 to 1. The resulting prediction grid was reclassified; values < 0.5 were identified as absent, and values ≥ 0.5 were identified as present.

Accuracy Assessment of Predicted Invasive Plant Distribution Maps

In addition to using the GPS-based infestation maps to simulate the sampling strategies, they were also used to determine the accuracy of the interpolation maps. The ability to use the GPS-based infestation maps as references was based on the high map accuracies at each of the sites. Confusion matrices of the predicted infestations were generated, and user's, producer's, and overall accuracies were calculated. These accuracy estimates provided a way to look at both how well the results could be trusted in the field (user's accuracy) and how well the model classified locations (producer's accuracy) (Fig. 3). Overall accuracy provided a generalized accuracy estimate, as it combined results from both the present and absent categories. Overall accuracy can mis-represent mapping success, however, when levels of one category in the geographic extents are much higher than others. Overall accuracy results, therefore, were considered more representative of prediction success at the *A. repens* site (presence/absence distributions were even) than at the *C. maculosa* site (presence was low relative to absence).

$$\text{User's accuracy} = \frac{\text{\# of correctly predicted locations in the category}}{\text{total \# of locations predicted in the category}} * 100$$

$$\text{Producer's accuracy} = \frac{\text{\# of correctly predicted locations in the category}}{\text{total \# of reference locations in the category}} * 100$$

$$\text{Overall accuracy} = \frac{\text{\# of locations predicted correct for all categories}}{\text{total \# of reference locations for all categories}} * 100$$

Figure 3. User's, producer's, and overall accuracy equations.

ANOVA was used to determine significant effects of sampling method and/or sampling density on user's and producer's accuracy for presence and overall accuracy. Three replications of the sample method x sample density combinations enabled a calculation of experiment-wide error protected means separations at each study site; this was done using multiple comparison analysis (MCA) with Tukey's test[1].

Results

GPS-based Infestation Map Accuracies

The number of GPS-based presence/absence locations evaluated at the *A. repens* site were 40 and 10, respectively (Table 1). At the *C. maculosa* site, 27 present locations and 18 absent locations were evaluated. Overall accuracies for the GPS-based presence/absence maps at the *A. repens* site was 94.0% and 80.0% at the *C. maculosa* site (Table 2). Although, at the *C. maculosa* site the accuracies for *C. maculosa* presence/absence distribution maps were lower, the maps at both sites were considered acceptable representations of the invasive plant patterns. Infestation levels were calculated to be 43.0% at the *A. repens* site and 12.5% at the *C. maculosa* site.

[1] Significant differences from MCA results are identified by a and bs in Figures 4, 5, & 6.

Table 1. Contingency matrix for accuracy assessment of GPS-based infestation maps at *A. repens* and *C. maculosa* sites.

	Truth-Present	Truth-Absent	*Row Total*
A. repens site			
Map-Present	38	2	*40*
Map-Absent	1	9	*10*
Col. Total	*39*	*11*	*50*
C.maculosa site			
Map-Present	23	4	*27*
Map-Absent	5	13	*18*
Col. Total	*28*	*17*	*45*

Table 2. Accuracy assessment of GPS-based infestation maps at *A. repens* and *C. maculosa* sites.

	User's	Producer's	Overall
A. repens site			
Present	95.0%	97.4%	
Absent	90.0%	81.8%	
			94.0%
C. maculosa site			
Present	85.2%	82.1%	
Absent	72.2%	76.5%	
			80.0%

Table 3. ANOVA *p*-values for sampling method and sampling density using user's and producer's presence accuracies and overall prediction accuracies for *A. repens* and *C. maculosa* sites.

		Accuracy Types		
		User's	Producer's	Overall
A.repens site	df		*p* -values	
Sample Method	2	0.0001	0.0020	0.5131
Sample Density	5	<0.0001	<0.0001	<0.0001
Sample Method x Sample Density	17	0.0108	0.9712	0.9189
C. Maculosa site				
Sample Method	2	0.0106	<0.0001	0.7605
Sample Density	5	<0.0001	<0.0001	<0.0001
Sample Method x Sample Density	17	0.0694	0.3689	0.9113

<u>IDW Predicted Map Accuracy Differences</u>

<u>User's Accuracy</u>. Effect of sample method and sample density on user's accuracy depended on the study site.

A. repens: At the *A. repens* site, interactions existed between sample method and sample density (Table 3). The effect of sample method on user's accuracy was, therefore, dependant upon the sample density. MCA indicated that the only differences among sample densities were where the systematic sample method produced higher accuracies than systematic-random; at 0.04% and 0.08% densities (Fig. 4).

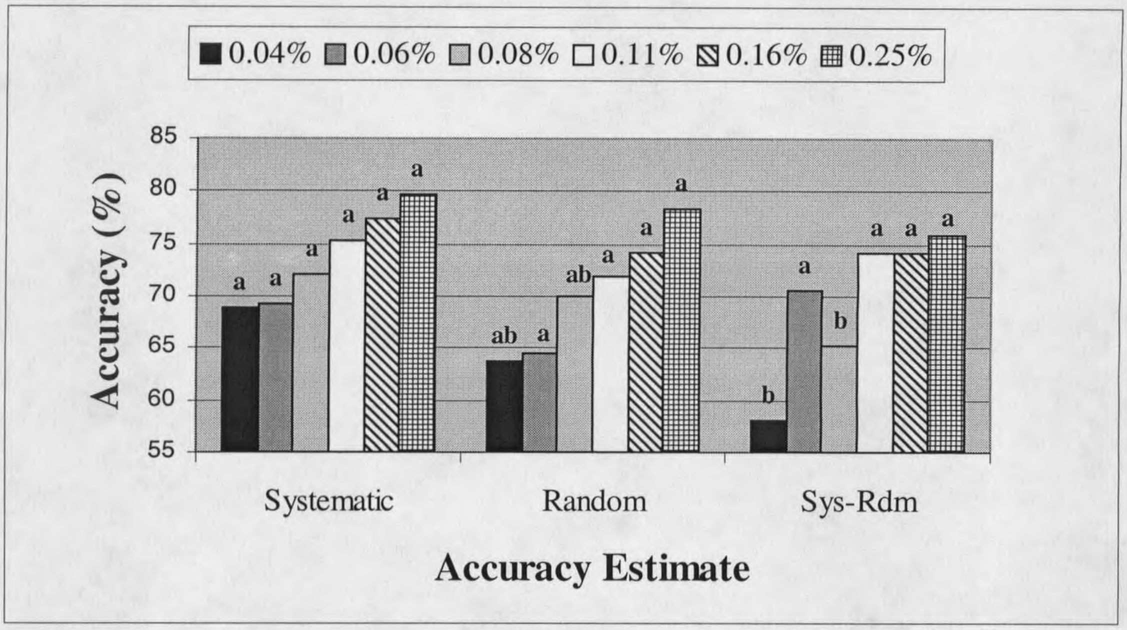

Figure 4. MCA of sample method x sample density presence prediction accuracies for systematic, random, and systematic-random sampling methods at *A. repens* site[2].

C. maculosa: At the *C. maculosa* site, there were no interactions between sample method and sample density (Table 3). Sample method and sample density main effects influenced user's accuracy. Systematic sampling produced higher user's accuracies (2.7%) than the systematic-random sample method, but not significantly higher than random sampling (Fig. 5). Systematic-random and random sample methods produced similar user's accuracies. *C. maculosa* user's accuracies for the systematic, random, and systematic-random sample methods were 81.0%, 79.1% and 78.3%, respectively.

The highest sample density (0.25%) produced the highest user's accuracies (Fig. 6). Increases occurred between 0.04% and 0.08% and 0.06% and 0.11%. User's accuracy for the *C. maculosa* site was the only time the highest two sample densities (0.16% and

[2] Significant differences among sample method x density combinations are identified by a and bs.