**MONTANA STATE UNIVERSITY**

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

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

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

($\lambda_i$) and values at known locations $Z(s_i)$. 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 $\geq 0.5$ 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 |

IDW Predicted Map Accuracy Differences

User's Accuracy. 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.

**a**



**b**



Figure 5. MCA of sample methods for user's and producer's accuracy for presence and overall prediction accuracies (a) *A. repens* site, (b) *C. maculosa* site[3].

---

[3] Significant differences among sample methods are identified by a and bs.
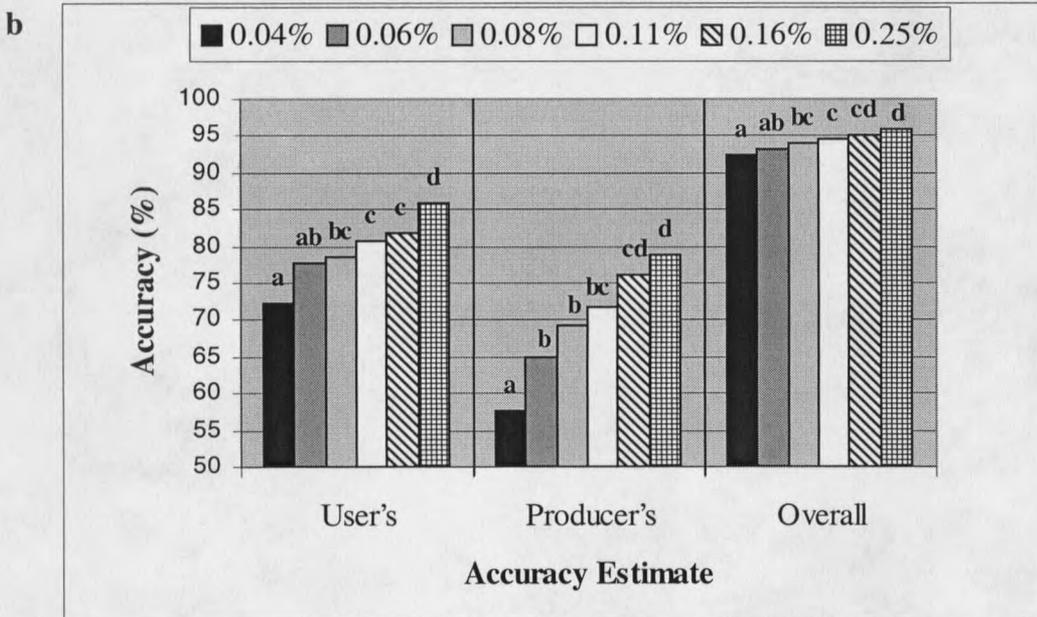
Figure 6. MCA of sample densities for user's, producer's, and overall accuracy for presence (a) *A. repens* site, (b) *C. maculosa* site[4].

---

[4] Significant differences among sample densities are identified by a and bs.

0.25%) were different from each other. Sample density user's accuracies ranged from 72.3% to 85.8% at the *C. maculosa* site.

Producer's Accuracy. The main effects of sample method and sample density influenced producer's presence accuracy at both sites (Table 3). No interaction between sample method and sample density was detected at either site.

**A. repens**: Systematic sampling produced a higher average accuracy (7.9%) than the systematic-random method (Fig. 5). There was no difference between the random and systematic-random sample methods. Producer's accuracies for the systematic, random, and systematic-random sample methods were 73.1%, 68.4% and 65.2%, respectively. Increasing sample density produced higher producer's accuracies (Fig. 6). The 0.25% sample density resulted in higher producer's accuracies than all other densities except for the 0.16% sample density. The producer's accuracy at the lowest sample density was 58.0% and 79.3% at the highest sample density.

**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 than the random sample method (7.0%) and systematic-random sampling (7.3%) (Fig. 5). Systematic-random and random sample methods produced similar accuracies. *C. maculosa* presence producer's accuracies for the systematic, random, and systematic-random sample methods were 74.5%, 67.5% and 67.2%, respectively. Higher producer's accuracies occurred when sample density was increased from 0.04% and 0.06% (Fig. 6).

The 0.16% sample density produced higher producer's accuracies than 0.08%. The 0.25% sample density produced higher accuracies than 0.11%. There was no difference in producer's accuracy at the two highest densities. Producer's accuracy was 57.7% at the lowest sample density and 78.7% at the highest sample density.

Overall Accuracy. Sample method did not affect overall accuracy at either site (Table 3). Sample density influenced overall accuracy at both sites.

**A. repens**: Mean overall accuracy across all sample methods was 82.0% at the *A. repens* site. Differences between 0.04% and 0.06%, 0.06% and 0.16%, and 0.11% and 0.25% sample densities were detected (Fig. 6). The average accuracy values ranged between 76.9% at the lowest sample density and 86.8% at the highest sample density.

**C. maculosa**: At the *C. maculosa* site, the mean overall accuracy across sample methods was 94.2%. Increases in some sample densities produced higher accuracies (Fig. 6). Accuracy differences occurred between 0.04% and 0.08%, 0.06% and 0.11%, and 0.11% and 0.25% sample densities. The overall accuracy ranged between 92.3% at the lowest sample density and 95.8% at the highest sample density.

Map Comparisons of Sample Methods

A visual comparison of the IDW maps, for one replication of the three sampling methods at the 0.25% sample density, showed for the *A. repens* site, systematic sampling

missed the fewest infestations (Fig 7). Systematic sampling, however, resulted in a larger

amount of over-prediction (predicted incorrect) of presence than random sampling. The

majority of the incorrectly classified locations using systematic sampling appear on the

edges of the large patches. Random sampling resulted in incorrect predictions in the

central portion of the study area. Systematic-random sampling incorrectly predicted more

areas present than random and missed more infestations than systematic. The maps also

indicated systematic sampling was able to capture a larger number of the smaller

infestations (<2.02ha) more successfully than either random or systematic-random

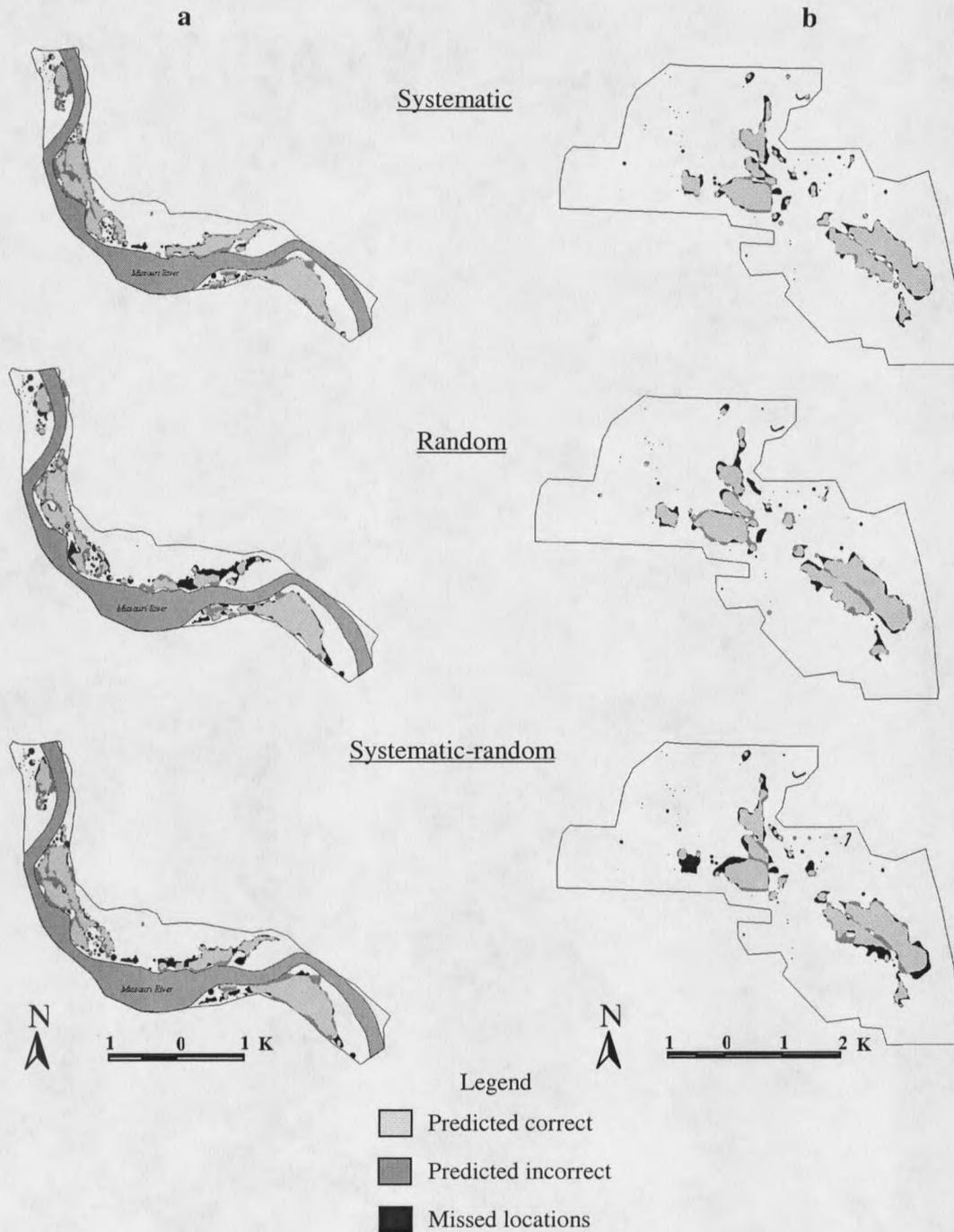sampling. Similar results occurred at the *C. maculosa* site.

Figure 7. Comparison of predicted infestation maps at 0.25% sampling density: predicted correct vs. predicted incorrect vs. missed locations at (a) *A. repens* site, (b) *C. maculosa* site.

Discussion


Sampling density had the greatest and most consistent effect on prediction

accuracies. At the 0.25% sample density, overall accuracies ranged from 78.0% to 86.8%

at the *A. repens* site and from 92.3% to 95.8% at the *C. maculosa* site. The highest values

meet the U.S. Geological Survey (USGS) 85% classification accuracy standard for

vegetation mapping (Anderson et al. 1976) and we believe are suitable accuracy levels

for invasive plant management. In our experience, traditional survey maps (except at the

most intensive level) rarely exceed these levels. Since sampling (even at the 0.25%

density) would take less time than traditional surveys, IDW is a desirable alternative to

traditional survey mapping. Our results also indicate that decreasing sample density

might cause significant decreases in prediction accuracies, except (in our case) at the

0.16% density level. This leads to the conclusion that optimum densities might fluctuate

and the most appropriate sample density for land managers will be based on management

conditions, such as site complexity, accuracy needs, and time constraints.

Sample method did not have as strong an influence on accuracy values as sample

density. However, at both our study sites, for some accuracy estimates, but not all,

systematic sampling performed better than the other sampling methods. Neither the

random nor the systematic-random sample methods performed better than systematic

sampling for any of the accuracy estimates either site.

There has been some concern expressed in the literature regarding errors caused

by systematic sampling. Shifts in systematic sample locations have been found to cause

drastically different results (Podani 1984). Other research has found a decrease in interpolation success might occur if the sampling interval is not synchronized with the plant distribution (Fortin et al. 1989). Our analyses do not support these concerns. Grid origins were randomly shifted +/- 50m in all the systematic sampling replications and only small variations in accuracy results at each density were evident; the average variation in accuracy was 3.1% at the *A. repens* sited and 2.2% at the *C. maculosa* site. One possible explanation for the success of the systematic sampling method is the relatively high infestation levels at the two sites. The *A. repens* site was 43.0% infested and the *C. maculosa* site was 12.5% infested. These infestation levels might be high enough to eliminate out-of-phase effects between a systematic sampling grid and invasive plant distribution.

IDW is based on the theory that nearer locations have more similar conditions than locations further away (spatial autocorrelation). For IDW to be as successful in predicting invasive plant locations for *A. repens* and *C. maculosa*, strong spatial autocorrelation among sample locations must have existed for the two study sites. This study has identified that the spatial autocorrelation is strong enough that prediction success can occur at infestation levels of 43.0% for *A. repens* and 12.1% for *C. maculosa*. For land-use managers, the question then becomes, what is the *minimum* infestation level where the spatial autocorrelation among sampled locations can be used with interpolation to produce high accuracy distribution maps for these and other invasive plant species. This question warrants additional research.

Based on study results, some limitations using sampling and IDW for producing invasive plant distribution maps exist. The first limitation is the inability of all methods, to predict presence/absence along the infestation patch edges. Some spatial uncertainty is inevitable with spatial data. Uncertainty is commonly estimated for digitizer error, has been applied to map classifications (Aspinall et al. 1995), and can be used with the sampling/interpolation results in this study. Using our data, when systematically sampling at 100m intervals (0.25% density), the mean distance of each unsampled location in each of the study areas was 38.2m. This distance value is an estimate of the positional uncertainty to expect from the predicted infestation boundaries for the systematic/0.25% sample strategy. Use of IDW to identify reduction or spread at the patch boundary when changes are less than ~40m, therefore, is inappropriate. Until spread has occurred at levels greater than the spatial uncertainty, our results indicate, at our study sites, land managers could calculate changes only in total infestation with confidence .

A second limitation is the missed infestation patches less than 2.02ha in size by all sample methods. According to sampling theory, when systematic sampling is used, the minimum detectable patch size is twice the area determined by the sampling interval (Theobald 1989). At the 0.25% sample density, the intersample distance was 100m, thus making 2ha the minimum detectable patch size. Most of the small infestations less than 2.02ha were not identified using the systematic sampling method. More sample points would be necessary to produce high accuracy maps if a greater number of small patches exist or if infestation levels are low. At the 0.25% density, however, the random and systematic-random sample methods were also unable to capture the small infestation

patches. In fact, as evident from the distribution maps, the systematic sampling method was more successful in capturing the small patches than the other methods at both study sites.

IDW has rarely been used in invasive plant prediction studies. Research has typically focused on other interpolation methods (Donald 1994; Heisel et al. 1996). While our study did not compare IDW versus other methods, our data shows IDW as a successful method for predicting presence/absence in landscapes with low variability and when the sample data amply represents the response. High prediction accuracy occurred at the *A. repens* and *C. maculosa* sites despite differences in infestation levels. Increased sample sizes might be required, however, when species are rare or a large number of species are being predicted (Wildi 1986).

This study has shown that sampling and IDW can produce high accuracy presence/absence distribution maps for two invasive plant species at two study sites. These prediction maps provide evidence that IDW might be able to serve as a substitute for traditional mapping. Appropriate map uses would be estimating total infestation area and determining invasive plant distribution patterns. Land managers are warned that using systematic sampling at ~1pt/ha interval might not be as successful when infestation levels are lower and patch sizes are small (<2.0ha) or when managers desire patch boundary change detection smaller than 40m (the predicted boundary uncertainty). If identifying the smaller patches and/or information regarding fine scale patch boundary change is necessary, a shorter sampling interval will be required. Research should focus on determining minimum infestation level requirements for predicting distributions for *A.*

*repens, C. maculosa*, and other invasive species. The high accuracies of the IDW maps are a fortuitous consequence of the high degree of spatial autocorrelation among the invasive plant sample locations. Because of this spatial autocorrelation, using a systematic sample method at a 0.25% (~1pt/ha) sample density might provide an improvement over typical traditional map accuracies and save survey time and money.

# USING CLASSIFICATION TREE ANALYSIS FOR IMPROVING PREDICTIONS OF INVASIVE PLANT LOCATIONS

## Introduction

Correlation models identify statistical relationships between combinations of predictor variables to a particular response value. In invasive plant modeling, correlation models might be used to test hypotheses of invasive plant distribution occurrence and/or to predict their distribution patterns. Correlation modeling, for both explaining and predicting invasive plant distributions, can be useful as a land management tool if biophysical variables to develop the models are readily available and the relationship between the variables and the invasive plant locations is strong.

Data summarized in Chapter 3 showed, at one site each, inverse distance weighting (IDW) interpolation modeling could be used to accurately predict presence/absence of *Agroptilon repens* L. (Russian knapweed) and *Centaurea maculosa* Lam. (spotted knapweed) species using a systematic/0.25% (~1pt/ha) sampling strategy. Due to the strong spatial autocorrelation of invasive plant locations and the high predictive power of IDW interpolation, a same-species invasive plant proximity variable could be developed from IDW prediction maps. An invasive plant proximity variable (derived from Chapter 3 results), in conjunction with other readily available GIS data layers representing biophysical variables, could be used to predict invasive plant spatial

distributions using correlation modeling methods and possibly increase prediction accuracies over IDW alone. In addition, if relationships were highly correlated, then there is future promise that GIS data layers could function as indicators of dispersal mechanisms, which could lead to predicting patterns of invasive plant spread.

In this portion of the study, the predictive power of an invasive plant proximity predictive variable was tested using classification tree analysis. The correlation between invasive plant locations and other biophysical variables, as represented by GIS data layers, was also evaluated. The other biophysical variables used were aspect, elevation, slope, principal soil component, land cover, distance to water, and distance to road. These variables were chosen based on their general relationships with plant distribution patterns and their accessibility to land managers for use.

Direct gradients (temperature, pH) and resource gradients (light, water, nutrients, carbon dioxide, and oxygen) are correlated with plant patterns in natural landscapes (Austin et al. 1984; Franklin 1995; Moore et al. 1991). Since data for direct gradients are difficult to gather, indirect gradients correlated with landscape patterns often are used. Examples of indirect gradients are climate, vegetation, and geology. Climatic variables are often represented by temperature, precipitation, and elevation and are indicators of invasive plant distributions (Chicoine et al. 1985; Forcella et al. 1983). Elevation was found to be the strongest predictor for plant patterns in the Northern Rockies (Anderson 1990). Based on its relationship with nutrient and water availability and space, presence of "other vegetation" is a strong indicator of invasive plant distribution patterns. Other vegetation, as related to canopy cover and light availability, was found influential in

predicting *C. maculosa* distribution patterns (Marcus et al. 1998). Geology is often used

to indicate parent material and other direct gradients such as soil water capacity (Despain

1973; Lees et al. 1991; Moore et al. 1991). Particular to invasive plant, disturbance has

been found to play a strong role in distribution patterns. A grassland study in Glacier

National Park, Montana found human and animal disturbance increases susceptibility to

*C. maculosa* invasion (Tyser et al. 1988).

Classification tree analysis (CTA) often is chosen to evaluate correlations

between variables and a response because it is easy to evaluate and a combination of

continuous and categorical data can be used. Additional advantages are that the variables

and values at each binary split are identified and can be examined to determine whether

they make ecological sense (Franklin 1995). Classification trees are also considered well

suited to represent the complexity of interactions among plant communities and diverse

topographic and edaphic variables that determine their environment (Lawrence et al.

2000; Moore et al. 1991).

Classification tree analysis is used when the response and some or all of the

predictor variables are categorical. CTA is based on a hierarchical structure of rules and

recursive partitioning (Breiman et al. 1984). The trees are also referred to as automated

taxonomic keys (Moore et al. 1991). Prediction occurs in a series of binary splits based

on rules resulting from the greatest increase in class purity.

A critical component of classification tree analysis is pruning. The pruning

process consolidates the full-grown tree by merging subsequent nodes, using a cost

complexity criterion (Mathsoft 1999). The criterion considers the properties of the entire

model and reduces tree complexity, while minimizing increases in classification error (Moore et al. 1991). Misclassification error is based on a comparison between predicted response values and the actual value at known locations. Pruning is a common practice in decision tree model development because it simplifies and increases the efficiency of the model and reduces overfitting (Moore et al. 1991).

This study is the second portion of an on-going project testing interpolation and correlation modeling as alternatives to traditional survey mapping. Our objectives were to examine how available GIS data layers representing biophysical variables correlated with *A. repens* and *C. maculosa* presence/absence. If strong correlations existed, then the variables could be used in conjunction with IDW to further develop a distribution model for predicting invasive plant distributions.

## Methods

GPS-based infestation maps, with known accuracies, a Geographic Information System (GIS), and classification tree analysis (CTA) were used to examine the correlation between GIS data layers and *A. repens* and *C. maculosa* distribution patterns within two Montana rangeland environments. Accuracy of distribution maps were determined by comparing predicted results to the GPS-based infestation maps. The study consisted of 5 general steps: (1) delineating invasive plant boundaries using GPS at two study sites, (2) accuracy assessment of GPS-based maps at each site, (3) restructuring of invasive plant data from cover categories into present and absent categories and

converting maps into raster format (for detailed description of methods 1 – 3, see Methods section of Chapter 3), (4) formatting of 8 predictor variables, and (5) evaluating classification tree analysis to identify relationships between environmental factors and infestation locations.

Samples for the CTA were generated from the GPS-based infestation maps using a 0.25% sampling density and systematic sampling at each of the study sites. Based on an accuracy assessment, the GPS-based infestation maps were determined to be a reasonable approximation of the invasive plant distribution at the two sites (Chapter 3, Tables 1&2).

Predictor Variables

A database was developed to determine variables correlated with *A. repens* and *C. maculosa* distributions. The response and eight biophysical variables were included. Biophysical variables were derived from spatial datasets and were selected based on ease of derivation and availability. All of the variables, except invasive plant proximity (IPP) had a resolution of approximately 30m (Table 4). The high resolution IPP variable was calculated from IDW presence/absence prediction maps (Chapter 3). Since a goal of the study was to determine whether readily available GIS data layers could account for invasive plant variability and potentially improve the prediction success of invasive plant locations over IDW, the predictor variables were resampled to match the 5 m resolution of the infestation locations.

Resampling a GIS data set to a resolution higher than its native cell size poses analysis difficulties. This is because, when GIS data sets are rescaled to a higher

resolution, no additional information can be created; the data sets are limited to detect information no finer than their native cell size (Jensen 1996; Star et al. 1990). Justification of resampling to a smaller cell size, for this study, existed in our need to evaluate the usefulness of readily available data; higher resolution GIS data layers that matched the scale of invasive plant location data, did not yet exist. A consequence of the rescaling is that only variability occurring on the 30m level could be detected by the coarser data sets (equal to the resolution of the GIS data layers' native cell size). Even though the classification tree analysis was conducted at 5m, evaluating correlations at 5m was useful, due to the possibility that some of the invasive plant distribution variation might occur at the level of the coarser data sets. It is also important to note that some invasive plant distribution variation was likely occur at the higher resolution (5m) and therefore, be undetectable by the GIS data layers.

Each of the eight predictive variables fell into one of three categories, indirect gradients, dispersal, or disturbance (Table 4). Indirect gradients represented 5 of the 8 variables. These were aspect, elevation, slope, principal soil component, and land cover. Elevation and aspect were used to represent climatic influences. Slope and principal soil component represented the geologic influences. Proximity to nearest invasive plant and distance to water represented dispersal mechanisms. Distance to road represented disturbance influences. Additional predictor variables known to influence invasive plant distributions, such as proximity to wildlife trails, non-maintained road systems, intermittent streams, and periodic disturbance (such as logging and grazing), were not used in the analysis because they were not readily available.

Classification Tree Analysis

The variables were compiled into a GIS, georeferenced, and converted to a raster format. Continuous variables (such as elevation) were re-sampled using cubic-convoluted methods and categorical variables (such as land cover) were resampled using nearest neighbor. Cubic convoluted resampling methods are appropriate for continuous variables, however, in terms of elevation, the averaging used might have further reduced the accuracy of some values. All variables were exported to an S-PLUS data frame for statistical analysis. Classification trees were derived for both study sites. Trees were grown until observations at each node were either all homogenous or there were fewer than 5 observations.

Classification trees were developed with and without the invasive plant proximity (IPP) predictor variable. This was done to determine the predictive ability of the biophysical variables when invasive plant proximity was unknown. *A. repens* and *C. maculosa* were well established at their respective sites. Consequently, proximity to same-species invasive plant was found to have a strong influence on invasive plant distribution and explained much of the variability when CTA was conducted. Statistically, the strong correlations of the IPP variable, masked the correlations of the other biophysical variables. The IPP variable was, therefore, removed from analysis, allowing the evaluation of the amount of variability explained by the other biophysical variables. Principal soil component data were available for only the southern half of the *C. maculosa* site. CTA was, therefore, first run on the southern half of the study site,

which included the principal soil component data layer. CTA was then applied to the

entire *C. maculosa* site without principal soil component.

The classification trees were pruned based on an S-PLUS cross-validation

technique (Venables et al. 1997). Cross-validation randomly separated the sample data

into 10 equal sub-samples. For each sample data set, nine classification trees were grown

and tested against the tenth. Deviance reduction, size (number of terminal nodes), and

cost complexity criterion (k) were calculated. Trees were pruned at the terminal node size

where the last large reduction in deviance occurred. Using the pruned classification trees,

important biophysical variables were assessed.

Table 4. Response and biophysical variables used in CTA for *A. repens* and *C. maculosa* sites[5,6,7].

| Variables | Abbrev. | Type | Source | Source scale |
|---|---|---|---|---|
| **Response** | | | | |
| Infestation Locations | | Cat | On-the-ground, aerial GPS surveys | 5 |
| **Predictor variables** | | | | |
| *Indirect gradients* | | | | |
| Aspect | Asp | Cat | Digital Elevation Model (DEM) | 30 |
| Elevation | Elev | Cont | DEM | 30 |
| Slope | Slope | Cont | DEM | 30 |
| Principal soil component | Psc | Cat | NRCS SSURGO soils data | 1:24k |
| Land cover | Landcov | Cat | GAP Land Cover using Landsat | 30 |
| *Dispersal* | | | | |
| Distance to water | Dtwr | Cont | DRG Topographic Map | 1:24k |
| Invasive plant proximity | Ipp | Cont | Inverse Distance Weighting (IDW) | 5 |
| *Disturbance* | | | | |
| Distance to road | Dtrd | Cont | DRG Topographic Map | 1:24k |

[5] Soils data only covers southern half of *C. Maculosa* site.
[6] In the Type column, the abbreviation Cat refers to categorical data and Cont refers to continuous.
[7] In the Source column, the abbreviation DRG refers to digital raster graphic files.

## Results

Classification Tree Analysis *A. repens* site

<u>With Invasive Plant Proximity (IPP) variable</u>. After pruning, IPP was the only important splitting variable, explaining 86.0% of the deviance (Table 5). The locations with the highest probabilities for *A. repens* presence were where the nearest invasive plant neighbor was less than 36m (Fig. 8).

**Invasive Plant**
**Proximity** $^1$< 36m$^2$

| | |
|---|---|
| 181 | 412$^3$ |
| 98.0% | 1.0%$^4$ |

$^1$- splitting variable
$^2$- splitting criteria
$^3$- number of sites in split
$^4$- present classification probability
$^5$- classification of terminal node.
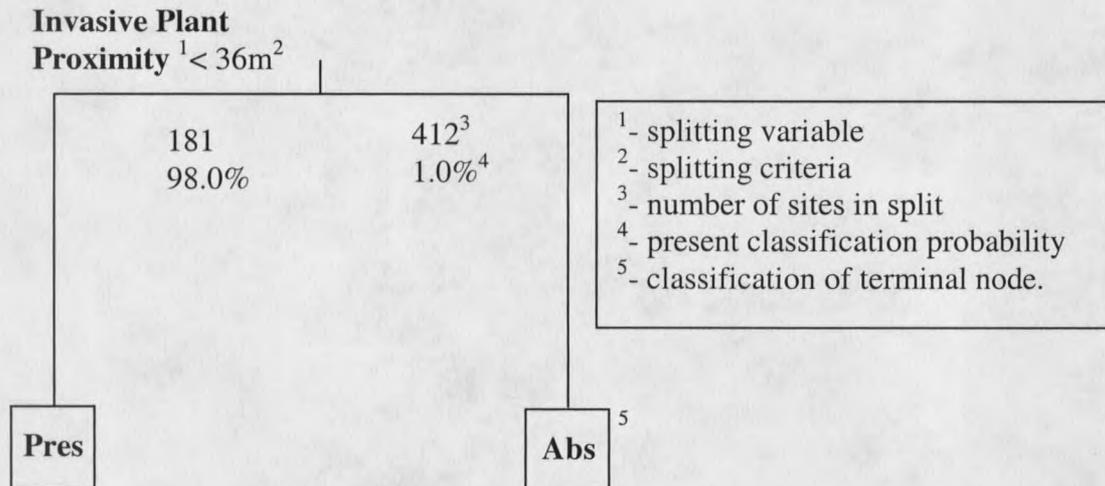
**Pres**

**Abs** $^5$

Figure 8. CTA splitting rule with invasive plant proximity (IPP) predictor variable at *A. repens* site.

<u>Without Invasive Plant Proximity (IPP) variable.</u> When all predictor variables except IPP were included in the CTA at the *A. repens* site, land cover, elevation, distance to water, and principal soil component were the splitting variables (Fig. 9, Table 5). The primary

Land cover

| Altered Herbaceous | Low Cover Grassland |
|---|---|
| Mod./High Cover Grassland | Silver Sage |
| Mixed Mesic Shrub | Greasewood |
| Mixed Broadleaf Forest | Wyoming Big Sage |
| Ponderosa Pine | Greasewood/Big Sage |
| 45.0% | Rocky Mtn Juniper |
| 365 | Water |
| | Broadleaf Domin. Riparian |
| | Badlands |

**Elevation** < 700m

| | | Shrub Badlands |
|---|---|---|
| 57.0% | 12.0% | Missouri Breaks |
| 287 | 78 | 8.0% |
| | | 228 |

**Abs**

**Distance
to water** < 45m

| 0.0% | 57.0% |
|---|---|
| 19 | 68 |

**Abs**

**Principal Soil Component**

| Clay | Loam |
|---|---|
| 68.0% Silt-Clay | 41.0% |
| 161 | 107 |

**Abs**

**Pres**

**Land cover**

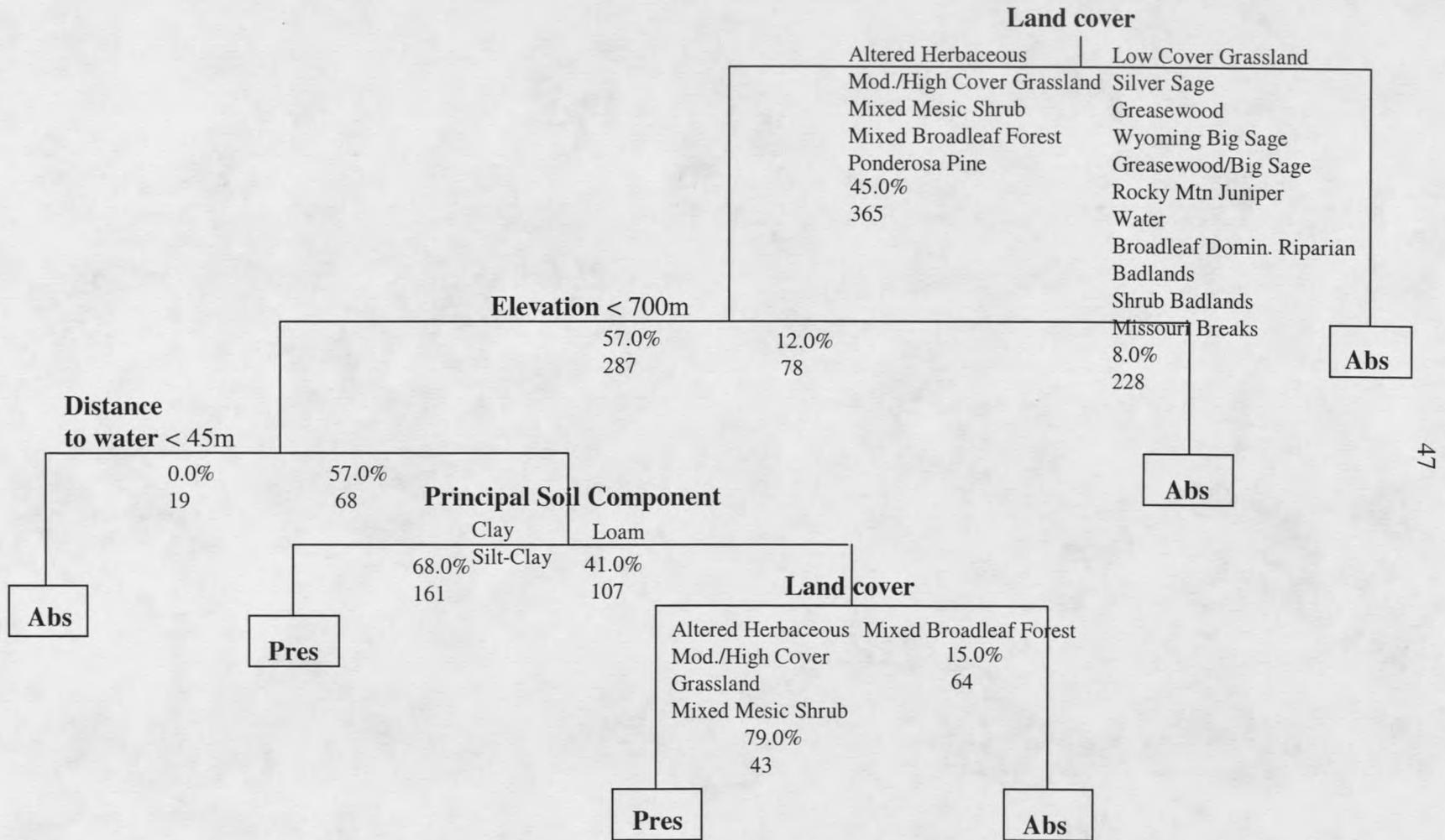| Altered Herbaceous | Mixed Broadleaf Forest |
|---|---|
| Mod./High Cover | 15.0% |
| Grassland | 64 |
| Mixed Mesic Shrub | |
| 79.0% | |
| 43 | |

**Pres**

**Abs**

47

Figure 9. CTA splitting rules without invasive plant proximity (IPP) predictor variable at the *A. repens* site.

variable was land cover, which explained 21.0% of the deviance. The other 3 splitting

variables, elevation, distance to water, and principal soil component, together

explained 10.1% of the deviance. Locations with either altered herbaceous, moderate/high

cover grasslands, mixed mesic shrub, mixed broadleaf forest, or Ponderosa pine, at elevations

less than 700m had a high probability of A. repens presence (Fig. 9). Locations with clay and

silt-clay soil types increased the probability of A. repens presence. Other high

probability locations were where loam was the principal soil component and the land cover

was either altered herbaceous, moderate/high cover grassland, and mixed mesic shrub.

Table 5. CTA results with and without the invasive plant proximity (IPP) predictor variable for A. repens and C. maculosa sites.

|  | A. repens site | | C. maculosa site | |
|---|---|---|---|---|
|  | w/ IPP | w/o IPP | w/ IPP | w/o IPP |
| # Terminal Nodes | 2 | 6 | 2 | 7 |
| Total % Reduction of Deviance | 86.0% | 31.1% | 97.1% | 5.6% |
| Variables Used | IPP | Landcov | IPP | Elev |
|  |  | Elev |  | Dtrd |
|  |  | Dtwr |  | Landcov |
|  |  | Psc |  | Slope |

Classification Tree Analysis *C. maculosa* site

With Invasive Plant Proximity variable. After pruning, IPP was the only important splitting

variable, explaining 97.1% of the deviance (Table 5). The locations with the highest

probabilities for *C. maculosa* presence were where the IPP was less than 20m (Fig. 10).
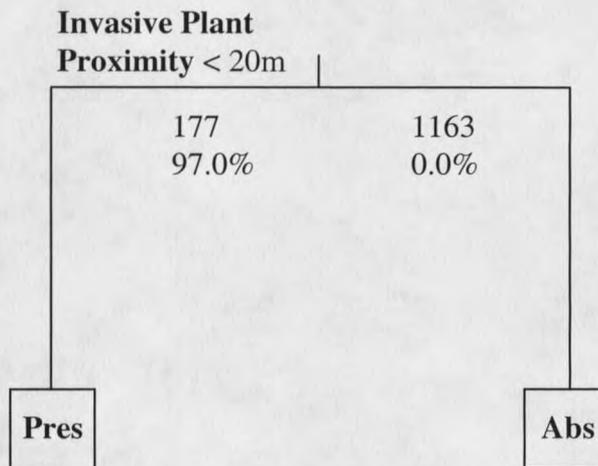
**Invasive Plant
Proximity** < 20m

| 177 | 1163 |
| 97.0% | 0.0% |

| Pres | | Abs |

Figure 10. CTA splitting rule with invasive plant proximity (IPP) predictor variable at *C.
maculosa* site.

Without Invasive Plant Proximity (IPP) variable. The CTA run on the southern half

of the *C. maculosa* site found the principal soil component was not an important variable

for predicting locations of *C. maculosa* presence. All predictor variables except IPP and

principal soil component were included in the CTA for the entire *C. maculosa* site.
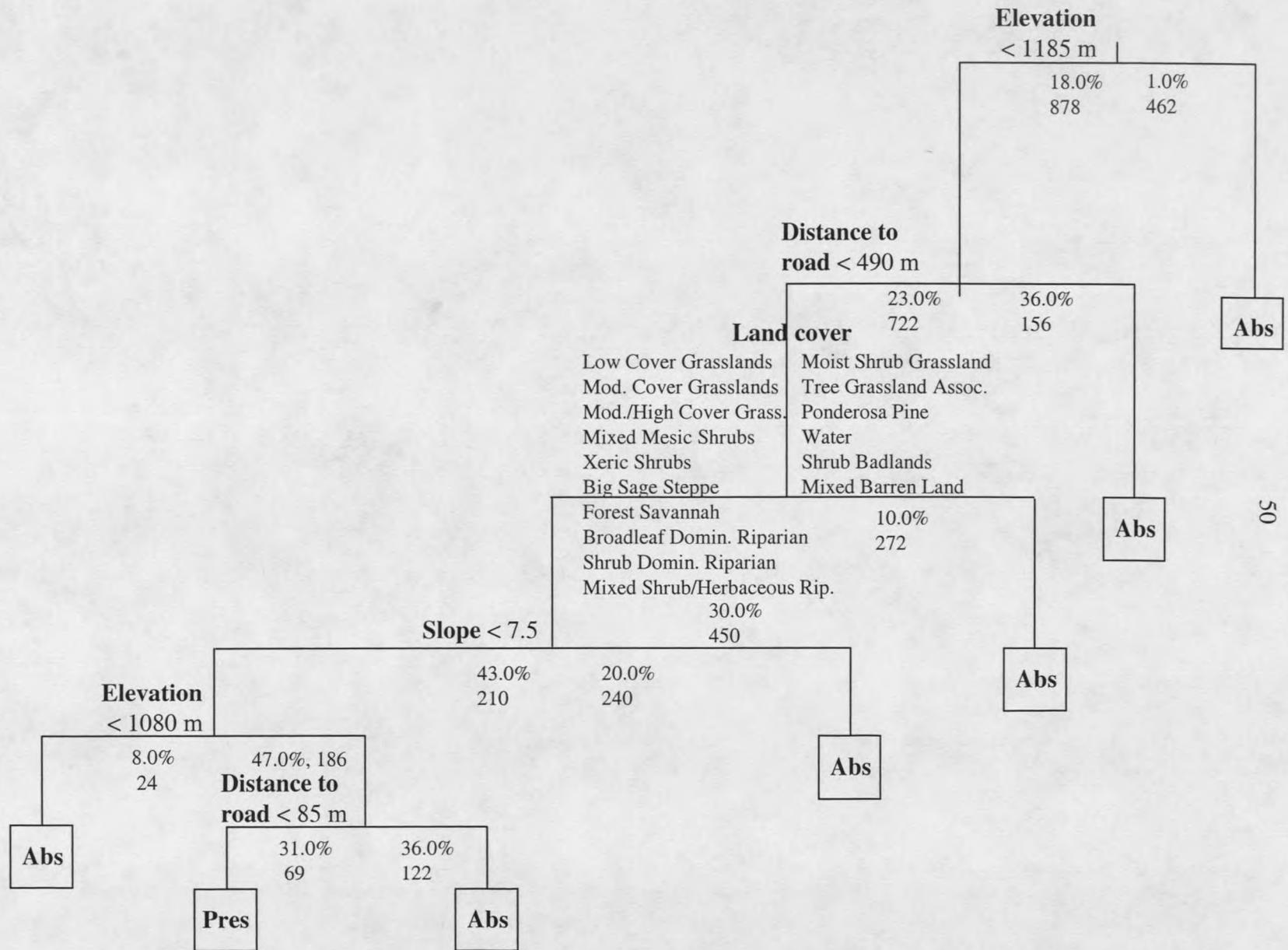
**Elevation**
**< 1185 m** |

18.0%          1.0%
878            462

**Distance to**
**road** < 490 m

23.0%          36.0%
**Land cover**   722            156

Low Cover Grasslands   Moist Shrub Grassland
Mod. Cover Grasslands  Tree Grassland Assoc.
Mod./High Cover Grass. Ponderosa Pine
Mixed Mesic Shrubs     Water
Xeric Shrubs           Shrub Badlands
Big Sage Steppe        Mixed Barren Land

Forest Savannah              10.0%
Broadleaf Domin. Riparian    272
Shrub Domin. Riparian
Mixed Shrub/Herbaceous Rip.
                      30.0%
**Slope** < 7.5       450

43.0%          20.0%
**Elevation**  210            240
**< 1080 m**

8.0%           47.0%, 186
24             **Distance to**
               **road** < 85 m

**Abs**        31.0%          36.0%
               69             122

               **Pres**        **Abs**

               **Abs**

               **Abs**

       **Abs**

       **Abs**

**Abs**

50

Figure 11. CTA splitting rules without distance to invasive plant proximity (IPP) predictor variable at the *C. maculosa* site.

Elevation, distance to road, land cover, and slope were identified as the important splitting variables (Table 5, Fig. 11). The primary splitting variable was elevation, which explained 2.4% of the deviance. The other splitting rules combined reduced deviance by 3.2%.

All *C. maculosa* present locations were at elevations between 1080 and 1185m, land cover was either low, moderate or moderate/high cover grasslands, mixed mesic or xeric shrubs, big sage steppe, forest savannah, broadleaf dominated riparian, shrub dominated riparian, or mixed shrub/herbaceous riparian areas, slope was less than 7.5 degrees, and distance to road was less than 85m. (Fig. 11).

## Discussion

Based on the classification tree correlation analysis results, the relationship between same-species invasive plant proximity and *C. maculosa* and *A. repens* locations was high at our two study sites. Important correlations existed between 4 of the other biophysical variables and the response at each site. These correlations, however, were much weaker than invasive plant proximity and considered not strong enough for predicting invasive plant distributions.

The accuracies of the IDW prediction maps creating the invasive plant proximity (IPP) variable for this portion of the study were high. It was expected, therefore, that CTA would identify IPP as a strong predictor variable. Not expected was the amount that the IPP variable would be able to reduce the deviance (97.1% at the *A. repens* site and by

86.0% at the *C. maculosa* site). When IPP was included in the CTA, the high levels of

deviance reduction by the IPP variable masked the influence of the other predictor

variables, causing them to be eliminated during the pruning process.

When CTA was run without the IPP variable, correlations between the other

biophysical variables and the response were uncovered. None of the other variables,

however, were able to reduce the deviance near to IPP levels. Due to the low amounts of

variability explained by the 7 other biophysical variables, they were not considered suited

for predicting invasive plant distributions. Some value might exist, however, at newly

invading sites and deserves examination.

At the *A. repens* site, presence was associated with land cover types requiring

greater amount of soil moisture (herbaceous, grasslands, mesic shrubs, and broadleaf

forests). This is consistent with the ecological characteristics of *A. repens*. *A. repens* has

been found to be most competitive in wetter soils (Watson 1980). In land cover types

associated with drier soils (low cover grassland, sage and greasewood, junipers, badlands,

etc.), the probability of presence was low. Other land cover associated with low

probabilities of *A. repens* were riparian areas dominated by broadleaf land cover. This

lack of association between broadleaf forest and *A. repens* is more difficult to explain, but

might be associated with the dense willow stands growing along the Missouri river

shoreline, which might prohibit *A. repens* establishment. Another splitting rule identified

with *A. repens* presence was elevations <700m. These lower elevation locations are in the

bottomlands of the study site, nearer the river, possibly where the soil is wetter. At

locations where the distance to water was <45m, there was a zero probability of *A. repens*

presence. The predicted absence of *A. repens* at these distances might also be explained by the dense willow stands growing along portions of the riverbanks. Interestingly, in sites where clay is the principal soil component, there was a 68.0% probability *A. repens* would be present. Research conducted in Colorado, has found relationships between clay soils and existence of *A. repens* (Beck, unpublished). *A. repens* had the highest probability of being present, however, in the loam soils where land cover types were altered herbaceous, moderate/high cover grasslands, and mixed mesic shrubs.

Explanations for splitting criteria were less straightforward for the *C. maculosa* site than the *A. repens* site. Thresholds existed for elevation, distance to roads, and slope. At elevations higher than 1185 and less than 1080m, *C. maculosa* had a low probability of being present. A distance to road threshold of 490m existed, beyond which, the CTA results indicated no *C. maculosa* locations. The distance to nearest road <85m was the only splitting point that indicated *C. maculosa* presence.

Even though correlations in the CTA without the IPP predictor variable were too low, at the two study sites, to be considered suitable for predicting invasive plant distributions, the biophysical variables CTA identified might be more suitable for predicting presence locations in areas with lower infestation levels. Biophysical variables, such as slope, disturbance levels, and proximity to land cover types, have been shown to be influential in determining plant establishment (Lawrence et al. 2000). The relationship between biophysical variables and invasive plant establishment, therefore, might be useful in assessing areas at risk to new invasion and permanent invasive plant establishment, and should be evaluated.

Due to the design of the correlation model, inferences based on the results are limited. The major limiting factor is the difference in resolutions between the response and the non-IPP variables. As stated in the methods section, all the variables, aside from IPP, used in the classification tree analysis were derived from data sets coarser than the response. The approximate resolution for all of the variables was 30m. The CTA analysis was conducted at the 5m resolution. The CTA results indicated, therefore, how well medium-resolution GIS data layers (at 30m) compared with response variability at 5m. In order to be confident with inferences made in this study, additional evaluation should be conducted with the response set to the 30m resolution of the GIS data layers.

An additional limitation is based on the source of the samples used for the classification tree analysis. The samples were gathered from the GPS-based infestation maps, rather than collected directly from the field. The infestation maps were developed from point, line, and area data. Both the points and lines were later buffered in the GIS. Although the accuracy assessment indicated accuracies of the maps were high, there could be some error in correlation results due to the inaccuracies of the samples gathered from these procedures.

The overwhelmingly strong correlation between the invasive plant proximity variable and invasive plant locations, at both sites, was increasing evidence that using interpolation modeling (IDW) would be most influential in assisting land managers in determining their invasive plant distributions. Results showed, biophysical variables, at the time of the study, had little potential to improve IDW predictions. The IPP variable derived from the IDW results explained nearly all of the variation at the two sites. The

other GIS data layers were masked by the IPP variable when all variables were included in the analysis and not highly correlated when evaluated independently. The other biophysical variables, therefore, were determined unsuitable for predictive mapping of invasive plant distributions, even as a supplement to the IPP variable.

This work was conducted on only two sites, both of which had relatively high levels of establishment. The biophysical variables tested in this study might be more correlated (causing a stronger potential for predictive ability) in situations where infestations are newly established. Correlation analysis should, therefore, be examined at other sites with lower infestation levels. In addition, due to the disparity in resolutions between the coarser GIS data layers and the invasive plant response, and the problems caused by resampling data sets to finer resolutions than their native cell size, analysis should be re-run, using the same data as used in this study, at the 30m resolution. Finally, in order to improve our understanding of the correlations between invasive plant locations and biophysical variables, and to make stronger conclusion as to whether biophysical variables have the potential to improve invasive plant prediction accuracies (as a supplement to, or instead of IDW), predictive data sets need to exist at the 5m scale; equal to that of the scale needed for invasive plant management.

# SUMMARY

In this project, IDW as an alternative method to traditional methods for invasive plant mapping was examined and preliminary work evaluating the correlation between available GIS data layers and invasive plant locations was conducted. Results identified sampling strategies in conjunction with IDW, could be useful for accurately predicting invasive plant distributions. Interpolation modeling was determined to be useful for predicting invasive plant distributions, while readily available predictive GIS data layers were not.

IDW prediction results indicated the preferred sampling strategy for predicting *C. maculosa* and *A. repens* distributions was systematic sampling at a density of ~1pt/ha. The accuracies resulting from this sampling strategy produced overall accuracies of near to and above 85.0%. Invasive plant proximity (as derived from IDW results), when evaluated in classification tree analysis, was also identified as highly related to invasive plant locations. The other GIS data layers chosen by CTA were not influential enough to assist in predicting plant locations, but they might be used to determine areas susceptible to new invasions and determine where these invasions might occur. At newly invading sites, the spatial dependence among individual plants might be weaker than at established sites, allowing the biophysical variables to play a more important role in predicting invasive plant locations.

This research has provided evidence that IDW might be an appropriate substitute for traditional mapping at sites where *C. maculosa* and *A. repens* are firmly established. This study offers the exciting possibility that minimal sampling efforts will be able to produce distribution maps rivaling traditional mapping methods. Using sampling and IDW, invasive plant locations were accurately predicted and distribution patterns determined at two sites. Additional testing of invasive plant species at other sites is necessary to determine the strength of IDW as an alternative mapping tool in invasive plant management in rangeland environments. The influence of biophysical variables on invasive plant distributions needs to be reassessed, as higher resolution data sets become available.

REFERENCES CITED

Aarts, H.F.M. 1986. A computerized model for predicting changes in a population of *Galium aparine. Proceedings of the EWRS Symposium 1986, Economic Weed Control*:277-284.

Anderson, J. 1990. Vegetation and Vegetation Dynamics in the Greater Yellowstone Ecosystem: An Introduction: Manuscript prepared for Greater Yellowstone Coalition Science Council.

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witner. 1976. A land use and land cover classification system for use with remote sensor data: USGS.

Aspinall, R.J., and D.M. Pearson. 1995. Describing and managing uncertainty of categorical maps in GIS. In *Innovations in GIS 2: Selected papers from the second national conference on GIS research UK*, edited by Fisher, P. London: Taylor & Francis.

Austin, M.P. 1998. An ecological perspective on biodiversity investigations: Examples from Australian Eucalypt forests. *Annals of the Missouri Botanical Garden* 85 (1):2-17.

Austin, M.P., R.B. Cunningham, and P.M. Fleming. 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* 55:11-27.

Austin, M.P., A.O. Nicholls, and C.R. Margules. 1990. Measurement of the realized niche: Environmental niches of five Eucalyptus species. *Ecological Monographs* 60:161-177.

Baker, H.G. 1986. Patterns of plant invasion in North America. In *Ecology of Biological Invasions of North America and Hawaii*, edited by Mooney, H. A. and Drake, J. A. New York: Springer-Verlag.

Bian, L. 1997. Multiscale nature of spatial data in scaling up environmental models. In *Scale in Remote Sensing in GIS*, edited by Quattrochi, D. A. and Goodchild, M. F. Boca Raton: Lewis Publishers.

Bian, L., and S.J. Walsh. 1993. Scale dependencies of vegetation and topography in a mountainous environment of Montana. *Professional Geographer* 45:1-11.

Bourdeau, P.F. 1953. A test of random versus systematic ecological sampling. *Ecology* 34:499-512.

Bowman, G.K., F.J. Molz, and O. Guven. 1995. An evaluation of interpolation methodologies for generating three-dimensional hydraulic property distributions from measured data. *Ground Water* 33 (2):247-258.

Breiman, L., J.H. Friedman, R.A. Olshon, and C.J. Stone. 1984. *Classification and Regression Trees*. Belmont: Wadsworth International Group.

Carpenter, G., A.N. Gillision, and J. Winter. 1993. DOMAIN: A flexible modeling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation* 2:667-680.

Chicoine, T.K., P.K. Fay, and G.A. Nielsen. 1985. Predicting weed migration from soil and climate maps. *Weed Science* 34:57-61.

Collins, F.C., and P.V. Bolstad. 1996. A Comparison of Spatial Interpolation Techniques in Temperature Estimation. Doctor of Philosophy in Forestry, College of Forestry and Wildlife, Virginia Polytechic Institute and State University, Blacksburg.

Cooksey, D., E.A. Roberts, and R.L. Sheley. 1999. *Montana Noxious Weed Survey and Mapping System: Weed mapping handbook* (v. 2.0) [Handbook]. MSU-Bozeman, October 17, 2000 [cited April, 10 2001]. Available from http://www.montana.edu/places/mtweeds/pubs.html.

Cooksey, D., and R.L. Sheley. 2000. *Mapping Noxious Weeds in Montana. EB 148* [Extension Bulletin]. MSU Publications, 1996 1998 [cited October 2000]. Available from http://www.montana.edu/places/mtweeds/pubs.html.

Dale, M.R.T. 1999. *Spatial Pattern Analysis in Plant Ecology*. Cambridge: Cambridge University Press.

Declercq, F.A.N. 1996. Interpolation methods for scattered sample data: Accuracy, spatial patterns, processing time. *Cartography and Geographic Information Systems* 23 (3):128-144.

Despain, D.G. 1973. Vegetation in the Bighorn Mountains, Wyoming, in relation to substrate and climate. *Ecological Monographs* 43:329-355.

Dirks, K.N., J.E. Hay, C.D. Stow, and D. Harris. 1998. High-resolution studies of rainfall on Norfolk Island; Part II, Interpolation of rainfall data. *Journal of Hydrology* 208 (3-4):187-193.

Donald, W.W. 1994. Geostatistics for mapping weeds, with a Canada Thistle (*Cirsium arvense*) patch as a case study. *Weed Science* 42:648-657.

Elzinga, C.L., D.W. Salzer, and J.W. Willoughby. 1999. *Measuring & Monitoring Plant Populations*. Denver: Bureau of Land Management.

Firbank, L.G. 1993. Implications of scale on the ecology and management of weeds. In *Landscape ecology and agroecosystems*, edited by Bunce, R. G. H., Ryszkowski, L. and Paoletti, M. G. Boca Raton: Lewis Publishers.

Forcella, F., and S.J. Harvey. 1983. Eurasian weed infestation in western Montana in relation to vegetation and disturbance. *Madrono* 30:102-109.

Fortin, M.-J., P. Drapeau, and P. Legendre. 1989. Spatial autocorrelation and sampling design in plant ecology. *Vegetatio* 83:209-222.

Franklin, J. 1995. Predictive vegetation mapping: Geographic modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19:474-499.

Goedickemeier, I., O. Wildi, and F. Kienast. 1997. Sampling for vegetation survey: Some properties of a GIS-based stratification compared to other statistical sampling methods. *Coenoses* 12 (1):43-50.

Goodchild, M.F., and D.A. Quattrochi. 1997. Scale, multiscaling, remote sensing, and GIS. In *Scale in Remote Sensing and GIS*, edited by Goodchild, M. F. and Quattrochi, D. A. Boca Raton: Lewis.

Gotway, C.A., R.B. Ferguson, G.W. Hergert, and T.A. Peterson. 1996. Division S-8 nutrient management & soil & plant Analysis: Comparision of kriging and inverse-distance methods for mapping soil parameters. *Soil Science Society American Journal (must check journal title)* 60:1237-1247.

Green, R.H. 1979. *Sampling Design and Statistical Methods for Environmental Biologists*. New York: Wiley.

Heisel, T., C. Andreasen, and A.K. Ersboll. 1996. Annual weed distributions can be mapped with kriging. *Weed Research* 36:325-337.

Hill, M.J., R.J. Aspinall, and W.D. Williams. 1997. Knowledge-based and inductive modelling of rough fescue (*Festuca altaica, F. campestris* and *F. hallii*) distribution in Alberta, Canada. *Ecological Modelling* 103:135-150.

Jeltsch, F., S.J. Milton, W.R.J. Dean, N. Van Rooyen, and K.A. Moloney. 1998. Modeling the impact of small-scale heterogeneities on tree-grass coexistence in semi-arid savannas. *Ecology* 79:780-794.

Jensen, J.R. 1996. *Introductory Digital Image Processing: A remote sensing perspective.* 2nd ed. Upper Saddle River: Prentice Hall.

Johnson, D.E. 1999. Surveying, mapping, and monitoring noxious weeds on rangelands. In *Biology and Management of Noxious Rangeland Weeds*, edited by Sheley, R. L. and Petroff, J. K. Corvallis: Oregon State University Press.

Kent, M., and P. Coker. 1992. *Vegetation Description and Analysis: A Practical Approach.* London: Belhaven Press.

Lajeunesse, S., R. Sheley, C. Duncan, and R. Lym. 1999. Leafy spurge. In *Biology and Management of Noxious Rangeland Weeds*, edited by Sheley, R. L. Corvallis: Oregon State University Press.

Lawrence, R.L., and W.J. Ripple. 2000. Fifteen years of revegetation of Mount St. Helens: A landscape-scale analysis. *Ecology* 81 (10):2742-2752.

Lees, B.G., and K. Ritman. 1991. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environmental Management* 15 (6):823-831.

Lenihan, J.M., and R.P. Neilson. 1993. A rule-based vegetation formation model for Canada. *Journal of Biogeography* 20:615-628.

Levin, S.A. 1992. The problem of pattern and scale in ecology. *Ecology* 73:1943-1967.

Marcus, W.A., G. Milner, and B. Maxwell. 1998. Spotted knapweed distribution in stock camps and trails of the Selway-Bitterroot Wilderness. *Great Basin Naturalist* 58:156-166.

Mathsoft. 1999. *S-PLUS 2000: Guide to Statistics.* 2 vols. Vol. 1. Seattle: MathSoft, Inc.

Mohler, C.L. 1983. Effect of sampling pattern on estimation of species distributions along gradients. *Vegetatio* 54:97-102.

Moore, D.M., B.G. Lees, and S.M. Davey. 1991. A new method for predicting vegetation distributions using decision tree analysis in a geographic information systems. *Environmental Management* 15 (1):59-71.

Moore, P.D., and S.B. Chapman. 1986. *Methods in Plant Ecology*. 2nd ed. Oxford: Blackwell.

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

Podani, J. 1984. Analysis of mapped and simulated vegetation patterns by means of computerized sampling techniques. *Acta Bot. Hung.* 30:403-425.

Sheley, R.L., J.S. Jacobs, and M.L. Carpinelli. 1999. Spotted knapweed. In *Biology and Management of Noxious Rangeland Weeds*, edited by Sheley, R. L. and Petroff, J. K. Corvallis: Oregon State University Press.

Skov, F. 2000. Potential plant distribution mapping based on climatic similarity. *Taxon* 49:503-515.

Star, J., and J. Estes. 1990. *Geographic Information Systems: An introduction*. Englewood Cliffs: Prentice Hall.

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

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

Theobald, D.M. 1989. Accuracy and bias issues in surface representation. In *Accuracy of Spatial Databases*, edited by Goodchild, M. F. and Gopal, S. London: Taylor & Francis.

Tortora, R.D. 1978. A note on sample size estimation for multinomial populations. *The American Statistician* 32 (3):100-103.

Tyser, R.W., and C.H. Key. 1988. Spotted knapweed in natural area fescue grasslands: An ecological assessment. *Northwest Science* 62 (4):151-160.

Venables, W.N., and B.D. Ripley. 1997. *Modern Applied Statistics with S-PLUS*. 2 ed. New York: Springer-Verlag.

Watson, A.K. 1980. The biology of Canadian weeds (Centaurea repens (L.)). *Canadian Journal of Plant Science* 60:993-1004.

Watson, A.K., and A.J. Renney. 1974. The biology of Canadian weeds. 6. *Centaurea diffusa and C. maculosa. Canadian Journal of Plant Sciences* 54:687-701.

Weins, J.A., J.F. Addicott, T.J. Case, and J. Diamond. 1986. Overview: The importance of spatial and temporal scale in ecological investigrations. In *Community Ecology*, edited by Diamond, J. and Cade, T. J. New York: Harper & Row.

Wiens, J.A. 1989. Spatial scaling in ecology. *Functional Ecology* 3:385-397.

Wildi, O. 1986. Neue Methoden der Vegetationskartierung. *Vierteljahresschr. Nat. forsch. Ges. Zur.* 131 (3):183-189.

Zimmerman, D., C. Pavlik, A. Ruggles, and M.P. Armstrong. 1999. An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Mathematical Geology* 31 (4):375-390.

APPENDICES

# APPENDIX A

## *A. REPENS* AND *C. MACULOSA* STUDY SITES

**Montana**

a

**A. repens Study Site**

**A. repens**

Present
Absent
No Data

Missouri River

N

1000     0     1000   Meters

b

**C. Maculosa Study Site**

**C. maculosa**
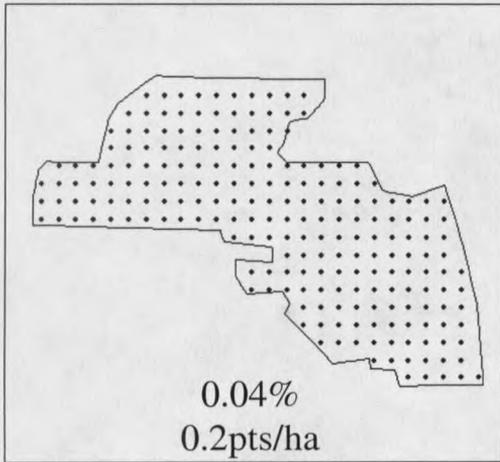
Present
Absent
No Data

Streams

N

1000     0     1000   Meters

APPENDIX B

ILLUSTRATIONS OF SAMPLE METHODS

Illustrations of the three different sampling methods used (a) systematic, (b) random, (c) systematic-random.

APPENDIX C

ILLUSTRATIONS OF SAMPLE DENSITIES

0.04%
0.2pts/ha

0.11%
0.45pts/ha

0.06%
0.3pts/ha

0.16%
0.7pts/ha

0.08%
0.35pts/ha

0.25%
1pts/ha

Appearance of sample point densities are greater than they truly are and are for illustrative purposes only. Sample densities were approximated; they differed slightly depending on the study site.

APPENDIX D

ARCVIEWGIS SAMPLING SCRIPT

```
' Name:  Sampling script for Invasive Plant Distribution Prediction Project (using IDW)

' Description: creates a ArcView shapefile with point locations derived from one of the following methods
'    1.  Simple random sampling
'    2.  Systematic - regular grid
'    3.  Systematic random - random in a grid

' This script is specific to Charles M. Russell wildlife refuge data located in central Montana.
' This is because in the Systematic sampling portion, an extents origin is hardwired
' into the script. To make the script universal, you would have to change it to a user entry and delete
' the hardwired origin numbers.
'
' Elizabeth Roberts, Richard Aspinall and Gretchen Burton
' Summer 2000
'
'****** Setup Info
theTitle = "Select Random Locations"

' Check if Spatial Analyst is loaded
test = Extension.Find("Spatial Analyst")
if (test = NIL) then
   MsgBox.Error("Spatial Analyst is not loaded.  Install Spatial Analyst and run again.",theTitle)
   return(NIL)
end

' Set the Work Directory, Project, get Active View and set Projection
theProject = av.GetProject
theWorkDir = theProject.GetWorkDir
theWorkDir.SetCWD
theView = av.GetActiveDoc
viewUnits = theView.GetDisplay.GetUnits
thePrj = theView.GetProjection

'    Create a List of GRID Themes in the View
theView = av.GetActiveDoc
ThemeList = List.Make
for each t in theView.GetThemes
  if (t.Is(GTHEME)) then
    ThemeList.Add(t)
  end
end

' Choose method of sampling
methods = {"Simple Random", "Systematic", "Systematic Grid with Random"}
method = MsgBox.ListAsString(methods,"Choose method of sampling",theTitle)


'******* Simple Random Sampling
if (method = "Simple Random") then
   ' Set the Mapextent, Cellsize and any Mask for the Analysis using the AnalysisPropertiesDialog
   ae = AnalysisPropertiesDialog.Show(theView, FALSE, theTitle)
   if (ae = NIL) then return NIL end
   ae.Activate

   cellSize = 1
   resoln = ae.GetCellSize(cellSize)        ' get cell size to use
```

```
' MsgBox.Info("Resolution" ++ cellsize.AsString," ')

' Make a random GRID with specified cellsize, map extent etc
randGrid = Grid.MakeRandom


' Find number of rows and columns in the GRID
rc = randGrid.GetNumRowsAndCols          ' get rows and columns
rows = rc.Get(0)                  ' rows
cols = rc.Get(1)                  ' columns
' MsgBox.Info("Rows:"++rows.AsString ++ " columns" ++ cols.AsString, theTitle)

Cells = rows * cols               ' Number of cells in GRID
' MsgBox.Info("Cells" ++ Cells.AsString, theTitle)
' Select proportion of area or number of samples to select
type = {"Proportion of area","Number of samples"}
samplesize = MsgBox.ListAsString(type,"Choose method for sample size specification",theTitle)

if (samplesize = "Proportion of area") then
  prop = MsgBox.Input("Enter proportion as a percentage",theTitle,"10")
  ' MsgBox.Info(prop,"Proportion")
  s = Cells * (prop.AsNumber / 100)
elseif (samplesize = "Number of samples") then
  size = MsgBox.Input("Enter sample size as count",theTitle,"100")
  ' MsgBox.Info(size,"Sample size")
  s = size.AsNumber
  prop = s * 100 / Cells
else
end

' Report number of cells and proportion
msgBox.Info("Cells:" ++ s.AsString ++ "Proportion:" ++ prop.AsString ++ "%","theTitle")

' Rescale the randGrid to give n sets of samples of size 's'
sClasses = (Cells / s)
' MsgBox.Info(sClasses.AsString,theTitle)
randGrid = (randGrid * sClasses).Int

' Select one of the values from this GRID to be the sample sites
randGrid = (randGrid = 1).Con(1.AsGrid,0.AsGrid)



' ******* Systematic Sampling (with randomly shifted origin)
' Uses the SHIFT command to change the start of the bottom left origin
' If creating new resolution grid, must have grid to sample active

elseif (method = "Systematic") then

' Title of dialog boxes
 theShiftTitle= "Shift Grid"

' Set Analysis Properties
  ae = AnalysisPropertiesDialog.Show(theView,TRUE,theTitle)
    if (ae = NIL) then return NIL end
    ae.Activate
    cellSize = 1
    resoln = ae.GetCellSize(cellSize)       ' get cell size to use
```

```
'MsgBox.Info("Resolution" ++ cellsize.AsString," ")

    theShiftTitle= "Shift Grid"

'Ask if Grid is already made, if no create new resolution grid
 if (MsgBox.YesNo("Is Grid to shift already created?",theShiftTitle, true).Not) then

   rsmplGrid = Grid.MakeRandom

   theGTheme = GTheme.Make(rsmplgrid)    'create a theme
   theView.AddTheme(theGTheme)      ' add theme to the view

 end

'Choose grid to shift
 ThemeList = List.Make
 for each t in theView.GetThemes
   if (t.Is(GTHEME)) then
     ThemeList.Add(t)
   end
   end
 a = MsgBox.List(ThemeList, "Choose the GRID to shift", theShiftTitle)
 tmpgrid= a.getgrid

 'Cell resolution and x,y origin for shifting Grid
 oldoriginx= 665568.2235
 oldoriginy= 5277876.1414

'Tell User x,y coord origin using for calculation
origin= "(665568.2235, 5277876.1414)"

 if (MsgBox.YesNo("Is the Grid origin?:"++ origin,theShiftTitle, true).Not) then
 exit
 end

clist={"100","125","150","175","200","250"}
cellchoice= MsgBox.ListAsString(clist, "Choose the Cellsize:", theShiftTitle)
 if (cellchoice = nil) then
 exit end
cellsize2 =cellchoice.asnumber

 no = number
 Rdm = no.MakeRandom (-50, 50)

MsgBox.Info("Random #:"++Rdm.AsString,"Randomly Selected # Meters to Shift Grid Origin:")
shiftvalX = Rdm
shiftvalY=shiftvalX

neworiginX = oldoriginX+shiftvalx
neworiginY = oldoriginY+shiftvalY

'Shift selected grid
randGrid = tmpGrid.SHIFT(point.make(neworiginX,neworiginY),(cellsize2))


' ******* Systematic-random sampling
'Systematic Portion (resamples grid, to set resolution)
```

```
elseif (method = "Systematic Grid with Random") then

    ae = AnalysisPropertiesDialog.Show(theView,TRUE,theTitle)
     if (ae = NIL) then return NIL end
    ae.Activate
    cellSize = 1
    resoln = ae.GetCellSize(cellSize)        ' get cell size to use
    MsgBox.Info("Resolution" ++ cellsize.AsString," ")

    randGrid = Grid.MakeRandom

    theGTheme=GTheme.Make(randGrid)    'create a theme
    theGTheme.SetName("Sys Grid")       ' set name of theme
    theView.AddTheme(theGTheme)       ' add theme to the view

    'Random portion of systematic-random sample

  ' Find number of rows and columns in the GRID
    rc = randGrid.GetNumRowsAndCols         ' get rows and columns
    rows = rc.Get(0)                  ' rows
    cols = rc.Get(1)                  ' columns
    ' MsgBox.Info("Rows:"++rows.AsString ++ " columns" ++ cols.AsString, theTitle)

    Cells = rows * cols              ' Number of cells in GRID
    ' MsgBox.Info("Cells" ++ Cells.AsString, theTitle)
    ' Select proportion of area or number of samples to select
    type = {"Proportion of area","Number of samples"}
    samplesize = MsgBox.ListAsString(type,"Choose method for sample size specification",theTitle)

    if (samplesize = "Proportion of area") then
      prop = MsgBox.Input("Enter proportion as a percentage",theTitle,"10")
      ' MsgBox.Info(prop,"Proportion")
      s = Cells * (prop.AsNumber / 100)
    elseif (samplesize = "Number of samples") then
      size = MsgBox.Input("Enter sample size as count",theTitle,"100")
      ' MsgBox.Info(size,"Sample size")
      s = size.AsNumber
      prop = s * 100 / Cells
    else
    end

    ' Report number of cells and proportion
    msgBox.Info("Cells:" ++ s.AsString ++ "Proportion:" ++ prop.AsString ++ "%","theTitle")

    ' Rescale the randGrid to give n sets of samples of size 's'
    sClasses = (Cells / s)
    ' MsgBox.Info(sClasses.AsString,theTitle)
    randGrid = (randGrid * sClasses).Int

    ' Select one of the values from this GRID to be the sample sites
    randGrid = (randGrid = 1).Con(1.AsGrid,0.AsGrid)

    randGrid = (randGrid <= 0).SetNull(randGrid)
    randGrid = (randGrid <= 0).SetNull(randGrid)

else
  return nil
```

```
end

randGrid = (randGrid <= 0).SetNull(randGrid)

' Output Sample locations
theGTheme1 = GTheme.Make(randGrid)    ' Create a Theme
theGTheme1.SetName("Samples")        ' Set Name of Theme
theView.AddTheme(theGTheme1)         ' Add Theme to the View

' Make a point file from the Grid
thefilename = FileDialog.Put("NewFile.shp".asfilename,"*.shp","Enter the new output filename.")
if (thefilename = nil) then
  exit
end

anFN = thefilename.asstring
position = anFN.IndexOf(".")
if (position = -1) then
  thefilename = (anFN + ".shp").asfilename
else
  anFN = anFN.Left(position -1)
  thefilename = (anFN + ".shp").asfilename
end

if (thePrj.isnull.not) then
  keepprj = msgbox.YesNo("Do you want the new shape file to remain in the current map projection? If you choose yes
then it will be saved in the current projection. If you choose No then it will be converted to decimal
degrees","Projection",true)
else
  keepprj = true
end

'for each t in theView.GetActiveThemes
'  if (t.Is(GTHEME)) then
'    ' get grid and display source name
'    theGrid = randGrid
'    if (keepprj = true) then
'       lineResult = theGrid.AsPointFTab(thefilename, prj.makenull)
'    else
'       lineResult = theGrid.AsPoint(thefilename,thePrj)
'    end
'  end
'end

if (MsgBox.YesNo("Add shapefile as theme to a view?",theTitle, true).Not) then
  exit
end

' Get a list of all Views
viewList = { }
for each d in av.GetProject.GetDocs
  if (d.Is(View)) then
    viewList.Add( d )
  end
end

' Add the shapefile to a View - choose the View
```

```
viewList.Add("<New View>")
addToView = MsgBox.ListAsString( viewList,"Add Theme to:",theTitle)
if (addToView <> nil) then
 if (addToView = "<New View>") then
  addToView = View.Make
  addToView.GetWin.Open
 end
 theFTheme = FTheme.Make(lineResult)
 addToView.AddTheme(theFtheme)
 addToView.GetWin.Activate
end
```

APPENDIX E
ABSENCE ACCURACY RESULTS

| | | Absence Accuracy Estimates (%) | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | Producer's | | | User's | | |
| | | Sample Method | | | | | |
| | | Systematic | Random | Sys-Rdm | Systematic | Random | Sys-Rdm |
| | Sample Density | | | | | | |
| *A. repens* site | 0.04% | 87.3% | 85.4% | 82.8% | 84.3% | 82.3% | 80.5% |
| | 0.06% | 86.1% | 82.9% | 88.5% | 87.1% | 86.4% | 84.5% |
| | 0.08% | 87.6% | 87.7% | 85.7% | 87.9% | 85.0% | 83.4% |
| | 0.11% | 89.3% | 88.4% | 90.0% | 88.6% | 86.1% | 85.0% |
| | 0.16% | 90.0% | 88.6% | 89.2% | 90.0% | 88.5% | 87.1% |
| | 0.25% | 90.8% | 90.6% | 88.9% | 91.8% | 90.0% | 90.7% |
| *C. maculosa* site | | | | | | | |
| | 0.04% | 97.7% | 96.7% | 96.9% | 94.7% | 94.7% | 93.7% |
| | 0.06% | 96.8% | 97.8% | 97.1% | 95.9% | 94.6% | 95.4% |
| | 0.08% | 97.5% | 97.2% | 97.5% | 96.6% | 95.4% | 95.6% |
| | 0.11% | 97.4% | 97.9% | 97.5% | 97.0% | 95.5% | 96.0% |
| | 0.16% | 98.0% | 97.5% | 97.6% | 97.3% | 96.6% | 96.3% |
| | 0.25% | 98.3% | 98.1% | 98.3% | 97.7% | 97.0% | 96.7% |

APPENDIX F

DETAILS OF RESPONSE AND GIS DATA LAYERS

**Response**

At the *A. repens* site, invasive plant data were collected in 1997 and 1998 by

United States Fish and Wildlife Service personnel using ground and helicopter GPS

mapping methods to collect point, line and area coverclass data using the Montana

Noxious Weed Survey and Mapping System standards. Infestations less than 2.02ha (5

acres) were identified as points. Infestations greater than 2.02ha (5 acres) were mapped as

polygons. Infestations following linear features were collected as lines with an associated

buffer width value.

At the *C. Maculosa* site, infestation maps were created in 1999 using helicopter

GPS mapping methods using the Montana Noxious Weed Survey and Mapping System

standards.

**Aspect, Slope, Elevation**

At both sites, a USGS digital elevation model (DEM) was used to calculate the

**aspect, slope,** and **elevation** at both the *A. repens* and *C. maculosa* sites. The DEM had a

30-meter grid resolution and were either of good or fair quality. **Aspect** was calculated

using the *Derive Aspect* command in ESRI's ArcView GIS. The *Derive Aspect* command

identifies the steepest down-slope direction from each cell to its neighbors. The values of

the **Aspect** grid were converted to 8 directional categories, north, northeast, east,

southeast, south, southwest, west, and northwest. **Slope** was calculated using the *Derive*

*Slope* command in ArcView. *Derive Slope* identifies the maximum rate of change, from

each grid cell to its neighbors. **Elevation** units were in meters above sea level based on DEM values.

**Principal soil component**

For both study sites, the soils data was derived from the Natural Resource and Conservation Service (NRCS) Soil Survey Geographic (SSURGO) database. SSURGO data were not entirely complete at either site. Soils data for the *C. maculosa* site were available for only the southern portion of the study area and were obtained from the Bureau of Indian Affairs and linked to the soils attribute database. For the *A. repens* site, complete and certified SSURGO data were available for the western portion of the study area. Soils data for the eastern half was completed by georegistering in-process polygon files and linking them to SSURGO soil attribute tables. **Principal soil component** was identified by the soil texture of the highest percentage soil type.

**Land Cover**

**Land cover** at both sites was derived from the Montana Land Cover Database from the University of Montana. Land cover was derived from Landsat Thematic Mapper images collected in 1993. Grid cell resolution was 30m. Vegetation classification was based on the cover-type assigned in supervised classification.

## Distance to Water

**Distance to water** was based on the Euclidean distance from the water locations to all other cells. At the *A. repens* site, the data layer was generated from on-screen digitized river and steam data from 1:24,000 USGS digital raster graphics (DRGs); the Missouri River and perennial streams were identified. At the *C. maculosa* site, all determinable water sources were on-screen digitized from 1:12,000 USGS digital orthophoto quarter quads (DOQQs).

## Invasive Plant Proximity

**Invasive Plant Proximity** was calculated using interpolated presence/absence invasive plant prediction maps from Chapter 3. The invasive plant maps were calculated the *Inverse Distance Weighting (IDW)* command in ArcView. The *IDW* command interpolated values using a point sample data set, where values of 1 represented invasive plant presence and 0 represented invasive plant absence. The Euclidean distance of all cells identified as absent from cells identified as present was calculated.

## Distance to Road

**Distance to road** at the *A. repens* site was calculated from road ArcView line files digitized from 1:24,000 USGS digital raster graphic (DRG) files. Roads included unimproved off-trail roads maintained by the Fish and Wildlife Service. At the *C. maculosa* site, roads included unmaintained roads and jeep trails. Roads were on-screen

digitized from 1:12,000 USGS digital orthophoto quarter quads (DOQQs). The Euclidean

distance from the roads to all cells was calculated.

APPENDIX G

MAPS OF RESPONSE AND SIGNIFICANT
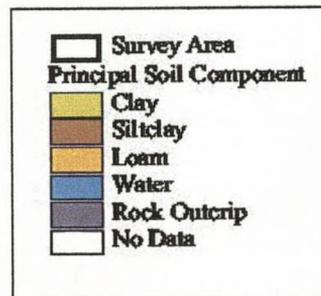GIS DATA LAYERS

**A. repens (response)**

Survey Area
A. repens
Present
Absent
No Data

**Invasive Plant Proximity**

Survey Area
Invasive Plant Proximity (m)
0 – 63
64 – 137
138 – 216
217 – 295
296 – 383
384 – 495
496 – 639
640 – 859
No Data

**Land Cover**

Survey Area

Vegetation
Mixed Broadleaf Forest
Altered Herbaceous
Broadleaf Dominated Riparian
Very Low Cover Grassland
Low Cover Grassland
Moderate/Higher Cover Grassland
Shrub Dominated Riparian
Missouri Breaks
Shrub Badlands

Badlands
Wyoming Big Sage / Steppe
Mixed Mesic Shrub
Ponderosa Pine
Greasewood / Big Sagebrush
Rocky Mountain Juniper
Greasewood
Silver Sage
Water
No Data

**Elevation**

Survey Area
Elevation (m)
- 675 - 685
- 686 - 692
- 693 - 696
- 697 - 700
- 701 - 704
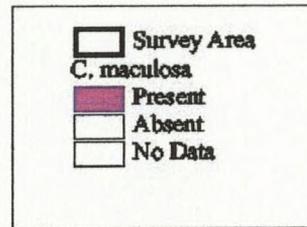- 705 - 708
- 709 - 713
- 714 - 719
- 720 - 734
- No Data

**Distance to Water**

Survey Area
Distance to Water (m)
- 5 - 76
- 77 - 155
- 156 - 237
- 238 - 322
- 323 - 409
- 410 - 500
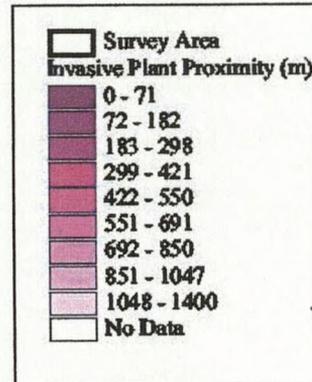- 501 - 603
- 604 - 731
- 732 - 909
- No Data

**Principal Soil Component**

Survey Area
Principal Soil Component
- Clay
- Siltclay
- Loam
- Water
- Rock Outcrip
- No Data

## *C. maculosa* (response)



Survey Area
C. maculosa
Present
Absent
No Data

500    0    500   Meters

## Invasive Plant Proximity



Survey Area
Invasive Plant Proximity (m)
0 - 71
72 - 182
183 - 298
299 - 421
422 - 550
551 - 691
692 - 850
851 - 1047
1048 - 1400
No Data

500    0    500   Meters

## Land Cover



Survey Area

Vegetation Types
Mixed Riparian
Herbaceous Riparian
Shrub Domin. Riparian
Mixed Broadleaf Forest
Mixed Conifer/ Broad. Forest
Ponderosa Pine
Forest Savannah
Very Low Cover Forest
Trees / Grassland

Moist Shrub Grassland
Upland Grassland
Low Cover Grassland
Mod./High Cover Grassland
Mixed Mesic Shrub
Dry Shrub-land
Shrub Badlands
Mixed Barren Land
Water
No Data

500    0    500   Meters

# Elevation

Survey Area

**Elevation (m)**

- 1062 - 1086
- 1087 - 1110
- 1111 - 1135
- 1136 - 1159
- 1160 - 1183
- 1184 - 1208
- 1209 - 1232
- 1233 - 1256
- 1257 - 1281
- No Data

500  0  500  Meters

N

# Distance to Roads

Survey Area

**Distance to Roads (m)**

- 0 - 88
- 89 - 176
- 177 - 264
- 265 - 352
- 353 - 440
- 441 - 528
- 529 - 616
- 617 - 704
- 705 - 792
- 793 - 880
- 881 - 969
- No Data

500  0  500  Meters

N

# Slope

Survey Area

**Slope (%)**

- 0 - 5
- 6 - 11
- 11 - 17
- 17 - 23
- 23 - 29
- 29 - 34
- 34 - 40
- 40 - 45
- 45 - 55
- No Data

500  0  500  Meters

N

APPENDIX H

CROSS-VALIDATION OF CLASSIFICATION TREES

a

b

c

d

Cross-validation Results[8]

---

[8] Deviance reduction versus size (number of terminal nodes before pruning) and k (cost-complexity criterion), O - indicates node size where tree was pruned, (a) *A. repens* site, w/o IPP, (b) *A. repens* site w/IPP, (c) *C. maculosa* site, w/o IPP, (d) *C. maculosa*, w/ IPP.