THESIS

MODELING THE EFFECTS OF HABITAT QUALITY ON BLACK-TAILED PRAIRIE DOG HABITAT OCCUPANCY USING SPATIALLY CORRELATED DATA

Submitted by

Jon R. Belak

Department of Fishery and Wildlife Biology

In partial fulfillment of the requirements

for the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2001

COLORADO STATE UNIVERSITY

November 2001

WE HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER OUR SUPERVISION BY JON R. BELAK ENTITLED MODELING THE EFFECTS OF HABITAT QUALITY ON BLACK-TAILED PRAIRIE DOG HABITAT OCCUPANCY USING SPATIALLY CORRELATED DATA BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

Committee on Graduate Work

Adviser

Department Head

ABSTRACT OF THESIS

MODELING THE EFFECTS OF HABITAT QUALITY ON BLACK-TAILED PRAIRIE DOG HABITAT OCCUPANCY USING SPATIALLY CORRELATED DATA

The black-tailed prairie dog congregates in clusters of distinct patches across the relatively homogeneous grasslands in which it occurs. This patchy distribution is in part a response to differences in habitat quality between sites, but there is also a social component to the habitat selection of this colonial species. The approach described in this paper divides the landscape into 30 m square pixels, using measures of habitat quality quantified at this scale to model the relationships between presence of prairie dogs and habitat attributes at the pixel level. The inadequacy of non-spatial analyses using these data is demonstrated as justification for a global model that accounts for spatial autocorrelation in habitat occupancy, indirectly modeling the social component of habitat selection and providing more precise and unbiased estimates of the effects of habitat variables.

Jon R. Belak
Fishery and Wildlife Biology
Colorado State University
Fort Collins, Colorado 80523
Fall 2001

ACKNOWLEDGMENTS

This research was supported in part by funds provided by the Rocky Mountain Research Station, Forest Service, U.S. Department of Agriculture (Contract # RMRS-99217-RJVA), and was also made possible by the help of many people, some of which are listed below.

I thank my advisor, Barry Noon, for allowing me to be his student. I have learned many things from Barry during the course of my graduate career, and he has been instrumental in providing me with the guidance, assistance, and insight I needed to complete my research. John Wiens was responsible for stimulating my interest in Landscape Ecology, and has been an inspiring teacher and helpful committee member. Julie Savidge provided insight into the biological reality behind my research; I appreciate her knowledge of prairie ecosystems and her contribution as a member of my graduate committee.

I am particularly grateful to Rudy King of the U.S. Forest Service Rocky Mountain Research Station Statistics Unit for generously providing many hours of statistical assistance, access to Forest Service computing resources, detailed reviews of my manuscripts and presentations, and a large amount of moral support. This project would not have been possible without his help. Robin McGee of the College of Natural Resources Advanced Technology Lab, Michelle Clifford and Jeff Greene of the University of Colorado's Information Technology Services Microsystems Group, and Dennis Mastin of Sun Microsystems also furnished technical assistance and the computing power necessary to complete the analysis. Denis Dean and Roger Hoffer of the CSU department of Forest Sciences and Mike Morrison, formerly of the USFS Pueblo Integrated Resource Inventory Center, gave useful advice about GIS and remote sensing questions on many occasions.

My contacts at the United States Forest Service were also indispensable. Dan Uresk, John Sidle, and Greg Schenbeck were helpful at many points throughout the research process, from providing the initial data and funding, to giving comments on methods and results. Glen Moravek and Julie Reedy Wheeler of Fort Pierre National Grassland and Doug Sargent of Buffalo Gap National Grassland provided tabular data and advice specific to each grassland, and John Hof from the Rocky Mountain Research Station provided previously digitized coverages of prairie dog colonies for the Conata Basin study area.

Dave Wolf and William Belton furnished Landsat TM imagery from Forest Service archives, and James Lacy reproduced imagery from the CALMIT archives at University of Nebraska at Lincoln for this study.

My deepest appreciation goes to my wife Amy. Without her support I would be in a completely different place. I thank my family for nurturing me and giving me the desire to learn and experience life, and our animals, Buzzy, Roland, and Malcolm for their unconditional love.

This research is dedicated to Joshua, Eli, Nina, Jessica, Brandon, Sophia, and all my other young friends and relatives. You belong to the generation that must change the way humans live on this planet. I have the greatest confidence in your abilities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

CONSERVATION CONTEXT

The black-tailed prairie dog (hereafter referred to as "prairie dog") is a diurnal rodent native to the prairies of Western North America. The colonial nature of this species, probably its most distinctive characteristic, results in large groups of individuals aggregated in distinct populations across the landscape. Before extensive human modification of the prairie, prairie dogs likely had equilibrium metapopulation structure over large portions of their range (Gilpin 1999), with mosaics of spatially distinct populations connected by dispersal, a dynamic spatial pattern but relatively stable proportion of habitat occupancy over time, and a high degree of overall population persistence.

Prairie dogs are still broadly distributed throughout their historic range, but habitat fragmentation, introduced disease, and direct human persecution of the species have shifted the patch-level population structure toward smaller and more isolated clusters of populations with little connectivity between clusters. It is likely that this spatial subdivision decreases overall persistence probability for the species. According to the United States Fish and Wildlife Service, prairie dogs occupy only 5% of the habitat they occupied at the turn of the century, and 1% of the area occupied before modification of their habitat by humans (USDI, 1999). With continued habitat loss, it is possible that the species will be unable to survive as small clusters of populations sparsely distributed across its former range.

It is well known that prairie dogs modify the habitat they occupy and serve as prey for other species that are known or suspected to be in decline. Associated species include the Black-footed Ferret (*Mustela nigripes*), Mountain Plover (*Charadrius montanus*), Burrowing Owl (*Athene cunicularia*), Golden Eagle (*Aquila chrysaetos*), Swift Fox (*Vulpes velox*), and Ferruginous Hawk (*Buteo regalis*). Prairie dog colonies have been shown to support greater densities of small mammals and greater density and species richness of birds than adjacent mixed-grass sites (Agnew 1983, Agnew et al. 1986). Prairie dogs affect grassland plant-species composition, enhancing habitat for Bison (*Bison bison*) and Pronghorn Antelope (*Antilocapra americana*), and the species also improves forage quality and net N-mineralization in surface soils (Krueger 1986, Detling 1998). In general, the patchy and shifting nature of prairie dog disturbance over time is thought to enhance range and soil quality over broad temporal and spatial scales (Whicker and Detling 1988). These effects have led some researchers to suggest that prairie dogs may be a keystone species, with ecological importance disproportionate to their biomass (Kotliar et al. 1999, but also see Stapp 1998). Loss of the prairie dog could have large, cascading effects on the prairie ecosystem as a whole, and this provides additional incentive for conservation of the species.

In 1994, concern over long-term viability led to a petition to declare the prairie dog a candidate for listing under the Endangered Species Act (USDI, 1999). This petition was declined after a 90-day review. Two new petitions were submitted in 1998, leading to a 12-month status review. On February 3, 2000, the USFWS issued a 12 month finding that threatened status was warranted, but precluded by the existence of other species with higher listing priorities (USDI, 2000). This decision put the responsibility for management of the species largely in the hands of individual states, Indian nations, and private landowners.

A cooperative prairie dog conservation effort was initiated in 1998 that involved 11 states and several Indian nations. The initial agreement established an interstate conservation team with representatives from each state and Indian nation, and each state created its own working group of stakeholders to focus on the development of Candidate Conservation Agreements with Assurances (CCAAs). These are agreements made with the US Fish and Wildlife Service in which the property holder agrees to a certain level of conservation for a species of concern in return for assurances that additional federal regulation will not be imposed in that area should that species be listed as endangered. Although Indian nations have recently withdrawn from the official CCAA process (Mike Fox, Native American Fish and Wildlife Society, Pers. Comm.), they still intend to submit prairie dog management plans by the deadline that are similar to those that states will submit as part of a CCAA.

The first step in the CCAA negotiation process is the development of umbrella CCAA's between each participant and the federal government that set target acreage objectives for prairie dog habitat and set management guidelines for these areas. Each of these agreements must be approved by all parties in the group. These are scheduled to be in place by October 2001, and it is expected they will support de-listing of the species during the 2002 Candidate Assessment. CCAA's involving private landholders are expected to follow. These types of management agreements are unprecedented conservation tools that have tremendous potential, both for species preservation and for misuse. They are to be tested first on the conservation of the prairie dog.

Given the current situation, evaluating the conservation potential of areas where the species is already protected is a logical first step. The Forest Service has issued a moratorium on control efforts for the National Grasslands, and has proposed managing for larger populations. Although these proposed management plans have not received final approval at this point, it makes sense to maximize conservation of prairie dogs on federal land if federally mandated conservation of the prairie dog conservation in other areas is to be politically viable.

Identifying areas of suitable habitat is necessary for effective management to maintain and augment existing prairie dog colony networks. This requires understanding relationships between habitat covariates and observed patterns of prairie dog habitat use. The research described below quantifies these relationships by estimating habitat covariates from existing data sources and exploring their relationship with known habitat occupancy patterns using synthetic statistical models that predict habitat suitability based on a small set of habitat covariates.

PREVIOUS BLACK-TAILED PRAIRIE DOG HABITAT USE RESEARCH

Past attempts to study prairie dog habitat use focused mainly on relationships with other grazing animals, dietary preferences, characteristics of plant and animal communities associated with prairie dog colonies, and the social structure that generates habitat occupancy patterns. These studies are useful to help define the micro-site characteristics related to prairie dog habitat occupancy.

Early prairie dog research focused on examining the degree of dietary overlap (i.e. competition) between the prairie dog and domestic cattle. The simplest way to make such a comparison is to examine the average biomass intake of an individual of each species. Merriam (1901) estimated that 256 prairie dogs consume the same amount of forage in a year as a single cow, but this statistic had no empirical basis. When studies measuring biomass consumption of each species (USDA 1968, Hansen and Cavender 1973) are compared, the biomass consumption of a single cow or cow-calf unit is roughly equal to that of 532 or 389 prairie dogs, respectively. This type of analysis assumes 100% dietary overlap and no loss of herbage due to prairie dog digging and clipping (or cattle trampling). Prairie dog digging and clipping may be incorporated by introducing prairie dogs into fenced enclosures and measuring the reduction in forage production that results (Taylor and Loftfield 1924), but the artificial introduction of individuals into an enclosed area where they did not previously occur has unknown effects on the results.

More importantly, neither type of comparison takes into account the nutritional content of the forage being consumed. Subsequent studies (O'Meilia et al. 1982, Uresk and Bjugstad 1983) have not detected any negative effects on cattle due to prairie dog competition, and examinations of nutritional content of plants found on and off prairie dog colonies have found that though prairie dogs reduce the overall quantity of biomass within a colony, the quality of the remaining forage is often higher than that found off towns (Coppock 1981). These studies are not conclusive, however, and further research is necessary to determine the extent of competition between these two species given differences in plant nutritional content within and outside prairie dog colonies.

Studies of cattle-prairie dog interactions clearly show a synergistic relationship between cattle grazing and prairie dog habitat occupancy. In 1949, Osborn and Allen credited regenerating tall vegetation in a fallow pasture for the elimination of a small prairie dog colony located in tall-grass prairie. Snell and Hlavachick (1980) found that elimination of cattle grazing during the growing season within and around a large colony in Kansas and the resulting vegetation growth was associated with a 90% decrease in the size of the colony. Knowles (1982), studying prairie dog distribution in eastern Montana, concluded that prairie dog distribution was mainly influenced by heavy livestock grazing pressure and other land disturbances created by humans. Uresk et al. (1982) found that the burrow densities within Conata Basin in Buffalo Gap National Grassland increased twice as fast in areas grazed by cattle vs. ungrazed areas. Cincotta (1985) reported that prairie dog colony expansion on the adjacent Badlands National Park was greater in heavily grazed areas and areas previously disturbed by homesteading activity. This was corroborated by Langer in 1998 using a GIS analysis based on comparison of old homesteading records with known prairie dog distributions. Cincotta et al. (1988) built a linear regression model to predict prairie dog establishment adjacent to existing colonies. They found that population density, visibility through vegetation, and the interaction of these two terms were significantly associated with colony expansion ($p < 0.10$). These studies indicate that prairie dogs prefer areas with reduced vegetation density created by grazing pressure.

Dietary preferences of prairie dogs as reconstructed through examination of their stomach contents also reveal information about habitat use. Analysis of prairie dog food preferences is problematic, however, because they are "selective opportunists" (Clippinger 1989). Prairie dogs select different plants and parts of plants during the growing season, but are capable of surviving on a wide range of forage with no measurable effects on their health, and dietary preferences within and between colonies vary significantly even when overall species composition is similar (Fagerstone et al. 1977). Summers and Linder (1978) examined the diets of prairie dogs living within two separate colonies and four distinct vegetation types in South Dakota. There were significant differences in vegetation composition when burrows within a single vegetation type, between types in the same colony, and between colonies were compared. Consumption patterns between colonies were significantly different. Fagerstone et al. (1981) studied variation in the diet of 158 prairie dogs collected from 12 colonies in Buffalo Gap National Grassland over six times of year. Vegetation composition varied significantly between (but not within) colonies. Consumption of brome (*Bromus* spp.) and buffalograss (*Buchloe dactyloides*), both preferred forages, varied significantly within colonies, while between colonies there were significant differences in consumption of red three-awn (*Aristida longiseta*) and prairie dogweed (*Dyssodia papposa*), which are generally not preferred.

These studies illustrate the difficulty in making generalizations about the prairie dog diet between areas and predicting habitat suitability on the basis of plant species composition. The fact that species composition and availability are altered over time in and around colonies further confounds vegetation use versus availability relationships. Prairie dogs decrease the standing crop, cover, and frequency of perennial grasses while increasing these measures for annual species (Lerwick 1974). Ratios of forb versus graminoid species as well as overall plant diversity are significantly higher in the core areas of established colonies (Coppock 1981), and peak live biomass of graminoid vegetation in areas occupied for 1-2, 3-8, and over 25 years, was 39%, 61%, and 97% lower than in adjacent unoccupied areas (Coppock et al. 1980). Bonham and Hannan (1978) reported a twofold decrease in blue grama (*Bouteloua gracilis*) clump size within prairie dog towns compared to adjacent unoccupied prairie. The fact that prairie dogs alter species composition over time makes the use of vegetation type as a predictor of prairie dog occupancy problematic since differences in vegetation could be a result of continued habitat occupancy as well as the reason an area is initially occupied. In addition, categorical vegetation predictors can only be incorporated within the analysis framework as dummy variables, which have inherently less predictive power and may not be averaged over time like continuous variables. The relationship between prairie dog habitat occupancy and vegetation density is more clear cut, though problems in attributing causality still exist. Regardless of the vegetation types present in an area, it is unlikely to be occupied by prairie dogs if vegetation is too dense (Merriam 1901, Osborn and Allen 1949, King 1955, Koford 1958, Hoogland 1981, Clippinger 1989).

Prairie dog habitat occupancy results from the interaction between selection at the level of the individual and the colonial nature of the species. Both of these processes structure habitat use patterns at multiple scales. As Hoogland (1995) notes, he has never seen an individual prairie dog living apart from others; they are an obligate colonial species. Even so, resource scarcity and inbreeding avoidance require that some individuals, usually males, migrate from their natal *coteries*, the basic matrilineal social unit of a colony. Groups of coteries are subdivided into *wards* by intervening non-habitat, and clusters of wards constitute a *colony*. Even though individuals from different coteries rarely physically interact unless they are dispersing or defending territory, within a town all individuals share in the responsibilities of providing alarm calls to protect the colony as a whole (Hoogland 1995). These three hierarchical levels of structure generate the clumped pattern of prairie dog habitat occupancy that is characteristic of the species' habitat use patterns.

The main conclusions relevant to this study that may be drawn from the above research are as follows:

- Extreme values for some habitat variables (e.g. vegetation density) apparently constrain patterns of prairie dog habitat use (Cincotta et al. 1988).
- Within the occupied ranges of these constraining variables, little is known about prairie dog preference and whether these variables have any direct cause-effect interpretation.
- Prairie dogs are ecosystem engineers (Jones et al. 1994). This behavior complicates habitat modeling since one is measuring habitat attributes that change as a result of the animal's behavior as well as more static attributes that may have stimulated the initial habitat selection.

A REVIEW OF EXISTING PRAIRIE DOG HABITAT SUITABILITY MODELS

The first attempt to quantify prairie dog habitat suitability in terms of measurable habitat variables was Clippinger's (1989) Habitat Suitability Index (HSI) model. This is one of the series of USFWS HSI models which relate the habitat use of a wildlife species to a set of habitat attributes that presumably affect food, cover, and reproduction. Variation in the values of individual habitat attributes results in variation in the HSI (habitat suitability index). This number is scaled from zero to one, with one being most suitable and 0 being unsuitable. HSI's for each attribute are then combined using a specific algorithm. The resulting index ranges between 0 and 1 and is intended to express the suitability of the area based on all relevant characteristics of habitat quality.

The Clippinger HSI model defined functions to express habitat suitability based on percent herbaceous cover, percent slope, average vegetation height, and soil type (Figure 1). The herbaceous cover function was defined by searching the literature for the maximum and minimum values observed on active prairie dog colonies. The minimum observed % herbaceous cover observed on active colonies is 25% (Fagerstone et al. 1977), and this is the lowest cover value with an HSI of one. Below this point, habitat suitability declines abruptly in a linear fashion until it reaches zero at a percent cover value of 15%. The HSI value at the maximum observed herbaceous cover of 91% (Uresk 1984) approaches zero; as cover decreases the HSI abruptly increases, reaching an HSI of one at 80% cover. Slope has an HSI value of one until 10% slope is reached, after which it declines in a linear fashion, reaching an HSI of zero at 20% slope. Average vegetation height has an HSI of one from 0-25 cm, after which the index value drops sharply, decreasing its slope at 40 cm. and becoming zero at 80 cm. Relative forb cover was originally included in the model, but was removed during the model verification process. Soil type, a categorical variable, is represented as a histogram. Loam and sandy loam have reduced HSI's (0.8 and 0.6), sand has an HSI of zero, and all other types have an HSI of one.

Model validation consisted of randomly selecting 21 1-ha plots within occupied areas of Rocky Mountain Arsenal near Denver, CO. Visual counts of prairie dogs in each plot were used to estimate population density, and these figures were compared to HSI values calculated for each plot based on the values of the habitat attributes. Model performance was fair ($r = 0.49$, $p = 0.02$), but the author noted that there were few index values between 0 to 0.5 and hypothesized that the colonial nature of prairie dogs requires a level of habitat quality sufficient to support a minimum of two coteries or 15 individuals per ha if an area is to remain occupied over time. This may be true, but it is also likely that HSI values were inflated by the small size of test plots and the use of categorical habitat quality attributes with a relatively few number of possible values. This model is based upon previous research from studies conducted throughout the species range, not a coherent set of habitat observations from one location, and the area chosen for model validation may not be representative of conditions throughout the species' range.

A second habitat model was produced by the United States Forest Service for the National Grasslands (unpublished). This model used data from previous studies to define the habitat attributes important for prairie dog habitat occupancy and the levels of each that represent preferred, marginal, or unsuitable habitat. The habitat attributes were used in a geographic information systems (GIS) intersection analysis to classify the public areas of all National Grasslands managed for prairie dogs. This is a consequence model; that is, it explores the results of a set of a priori assumptions. Table 1 shows the habitat traits used to define suitability. For all of these habitat attributes, eliminating areas classified as unsuitable or marginal removes only a small fraction of the area under consideration, and there are only a few possible values for each attribute used as an independent variable in the model. Like the HSI model, variables are categorical or summarized as categories. Consequently, the model is unable to eliminate many areas from consideration.

For each grassland, the output of the model was overlaid with a coverage showing all areas known to be occupied by prairie dogs. After the classification criteria for some predictors were modified, 90-95% of all occupied areas fell within preferred or marginal habitat. Visual comparison of model output to observed habitat occupancy for Fort Pierre National Grassland, however, reveals that most areas classified as preferred were never occupied, many areas designated as marginal were occupied consistently, and less than one-third of the area initially considered by the model was eliminated (Figs. 2 & 3). In addition, visual examination of reclassified model predictions (suitable or unsuitable instead of preferred, marginal, and unsuitable) within a portion of the Conata Basin study area (Fig. 4) reveals that the model predicts small areas of unsuitable habitat in a seemingly random, "salt and pepper" fashion across the landscape. Since the analysis is based upon intersecting a series of individual complex polygon coverages for each attribute, creating new coverages with millions of smaller polygons that have all the attributes of their parent coverages, then applying suitability rules based on all habitat attributes to each small polygon, it is inherently susceptible to scattered error patterns like those shown in Figure 4. The model seems to make reasonable predictions on the left side of the figure, but the overall tendency toward this type of error pattern decreases the usefulness of model output.

These results also suggest that too little is known about the habitat associations of prairie dogs for effective parameterization of a consequence model. While it is possible to define broad habitat tolerances within which the species can survive and to produce maps of their expected distribution, the results of these analyses tell us little about habitat selection. A better understanding of habitat use patterns requires exploration of fine-grained associations between habitat occupancy and quantitative habitat variables using statistical models.

Proctor (1998) performed the first such analysis, relating patterns of habitat occupancy (occupied/unoccupied) at the scale of a 30 m Landsat TM pixel to vegetation type, slope, soil texture, and soil depth quantified at the same scale. The final model, based on a classification and regression tree (CART) analysis (Morgan and Messenger 1973), used vegetation, slope, and soil texture to predict four levels of suitable habitat. This model has been used by Montana Fish, Wildlife, and Parks to identify 32 million acres of prairie dog habitat across eastern Montana, including 2.3 million acres of preferred habitat.

This model, unlike those described above, was based on associations between observed habitat use and measures of habitat quality within the area modeled rather than values taken from various published studies using data from different areas. Also, areas known to be unsuitable such as open water, forested areas, and wetlands were eliminated from the data, allowing the modeling process to focus on habitat relationships of interest and to exclude others that may introduce noise. Even though including these areas may have added to the apparent predictive power of a model, this would add little to our knowledge of prairie dog habitat relationships. By focusing more closely on areas of interest, the model was able to make relatively precise and accurate predictions.

When Proctor's final CART model was used to make predictions based on the data used to create it, Category 1 (unsuitable) habitat made up 82.2% of the study areas, but contained only 14.8% of the total area occupied by prairie dogs. All other habitat categories were to some extent suitable, and 85.3% of these areas were occupied even though they made up only 17.8% of the total area (Table 2). These results indicate that the model was able to discriminate suitable from unsuitable areas with reasonable accuracy. The relative contribution of different attributes and the relationship of each with habitat occupancy is could not be estimated, however, since habitat was divided into classes based on limited combinations of categorical predictors.

In addition, observations (e.g. adjacent pixels) were treated as statistically independent even though they were not. This can lead to spurious significant relationships between habitat predictors and occupancy (see below). Also, predictors improve in CART models simply as a function of sample size (Breiman et al. 1984). Since CART works by repeatedly subsetting the data based on all possible values for different categorical predictors, as long as sample sizes are large it is fairly likely that particular values for each predictor will produce a classification much better than one that could have developed by chance. Proctor's data sets had 22382 and 26108 observations.

The authors observe that occupancy of areas not classified as highest quality (category 5: 0-4% slopes, preferred vegetation) usually occurred when these areas were adjacent to high quality occupied habitat. This discrepancy in occupancy patterns is likely due to spatial factors that are important to prairie dog habitat occupancy but were not incorporated in the modeling process.

The research summarized above indicates that there is no obvious set of habitat attributes that clearly define prairie dog habitat. That is, the species is widely, but patchily distributed throughout prairie ecosystems. Other than differences in gross habitat attributes (e.g. slope, elevation), areas occupied by prairie dogs within this ecosystem seem similar to unoccupied areas. As a consequence, it is difficult to map, a priori, the distribution of suitable habitat in the absence of concurrent prairie dog survey data.

Our inability to map the distribution of suitable prairie dog habitat has 3 possible explanations: (1) we are unaware of, or incapable of measuring and mapping the habitat attributes that prairie dogs are actually evaluating when selecting habitat; (2) the spatial distribution of prairie dogs is largely random with respect to prairie ecosystems, and an animal's settling response is driven more by the presence of conspecifics than by habitat; or (3) there is both a social and a habitat component to selection. In this case, habitat associations shown by the species may be an expression of necessary but not sufficient conditions.

That is, true habitat relations exist, but the strength of the stimulus to settle based upon these habitat attributes is strongly modulated by presence/absence of conspecifics.

The models described below are based on the assumption that the latter hypothesis is true, and model spatial correlations in the dependent variables to account for the social component of prairie dog habitat selection. The sections that follow explain how spatial statistics can be used to minimize social effects that confound the study of species/habitat relationships, allowing habitat quality factors that are related to patterns of habitat use to be more reliably identified.

FITTING SPATIAL MODELS TO BIOLOGICAL PATTERNS

What is Spatial Autocorrelation and why is it important?

Spatial Auto-Correlation (SAC) is the tendency for things that are close to one another to be more similar (or different) than those farther from one another. In this case, prairie dog habitat occupancy is thought to be positively correlated in space, with less suitable areas being occupied mainly because they are adjacent to more suitable occupied areas. As a simplified case, imagine a map of a small fragment of prairie dog habitat with occupancy and habitat attributes resampled into 30 m square cells (Fig. 5). Darker green cells have identical values for all predictors. These areas are highly suitable for prairie dogs and are consistently occupied. Light green cells are also identical to one another, but are marginal in quality and sporadically occupied. Red cells are known to be unsuitable (forest, open water), and are excluded from the analysis. Occupied areas are cross-hatched. All 25 cells of optimal habitat are occupied, but 24 cells in adjacent marginal habitat are occupied as well.

The non-spatial regression analyses previously discussed treat all occupied areas as equivalent; each occupied cell in marginal habitat has the same influence on the final model as occupied cells in optimal habitat. The influence of marginal cells that are occupied only because they are adjacent to occupied, highly suitable cells leads to models that cannot discriminate between different levels of habitat quality. Predictors that incorporate habitat and spatial factors simultaneously may be used in these non-spatial models in an attempt to incorporate missing spatial relationships, but the biological implications of these indirect predictors are often ambiguous, and their use does not necessarily lead to a model that is easy to interpret (Guisan and Zimmermann 2000). When spatial autocorrelation is modeled directly, effects due to correlations in the response variable may be partitioned from those due to relationships with habitat quality, leading to a clearer interpretation of model results and their implications for habitat management.

Statistical significance and the failure to incorporate SAC

In any non-spatial logistic regression analyses on GIS raster data, adjacent 30 m squares of land are represented by independent observations in the dataset, even though the distance between samples is much smaller than the distance over which the dependent and independent variables are correlated in space (Legendre and Fortin 1989). As Figure 6 illustrates, assuming independence between non-independent observations overestimates degrees of freedom and biases estimates of regression coefficients and their standard errors, resulting in coefficients being declared significantly different from zero when they are not. In this way, habitat predictors that are not strongly associated with habitat occupancy may be highly significant in non-spatial regression analyses, but far from significant when spatial autocorrelations in the data are accounted for.

Analytical techniques for Modeling Spatial Autocorrelation

Spatial correlation can be directly incorporated into logistic regression models in two different ways, and each offers advantages and disadvantages. Autologistic models condition the response of each individual pixel on the response of cells within some predefined neighborhood surrounding it. A spatial covariate predictor that is a distance-weighted average of the number of occupied pixels in the neighborhood of each observation is estimated, closely modeling a particular spatial configuration. The value for this predictor is observation-specific, and is used directly in the main effects model. This method has obvious advantages in terms of increased accuracy, and it has performed well when used to model species habitat relationships (Preisler and Mitchell 1993, Augustin et al. 1996, Huffer and Wu 1998). The procedure is computationally intensive, however. The final models used in this study (see results) used a neighborhood 30 cells square. Using the same neighborhood for an autologistic analysis would require that each pixel be assigned a weighted average based on values in ca. 900 neighboring pixels. The data sets used in this study would require roughly 186,500 and 438,000 such calculations to prepare the data, a task that is computationally prohibitive at this time. More fundamentally, these models could also be said to suffer from the "glorification of the particular" (Roughgarden 1983); they are so closely tied to the area modeled that they lack generality and cannot be readily applied to other areas.

Generalized linear mixed models (GLMM's) that incorporate spatial autocorrelation build spatial covariance models based on prediction error from the main effects model rather than including a spatial predictor in the main model itself. Spatial GLMM's can be further divided into two groups, those with a marginal specification that model spatial covariance separately for each observation (McCullagh and Nelder 1989, Liang and Zeger 1986, Gotway and Stroup 1997) and conditional models that assume that the data are conditionally independent with constant variance and use a global spatial covariance model (Gotway and Wolfinger *in press*). The former approach attempts to directly specify the marginal moments for each observation, while the latter uses a conditional specification based on an underlying spatial process, and is a model for the mean response rather than the individual response. Marginal models cannot be used at this point on data sets as large as the ones considered here using current applications and computing technology (Marcia Gumpertz, pers. comm.). In addition, they do not provide estimates of the standard error of regression coefficients derived by maximum likelihood.

Conditional spatial GLMM models, on the other hand, provide maximum likelihood-based estimates of standard error and can be implemented using readily available software and estimation methods that are stable and robust on large data sets (Littell et al. 1996). The spatial extent of the study areas considered is still too large to be modeled as a single entity, but the analysis allows the area to be split into blocks that are used as subjects for global spatial and main effects models. In this way, the generality of the model may be further improved.

The conditional spatial correlation model is used to alter the error surface minimized in the main effects model, removing the effects of spatial correlation from the main model rather than incorporating them as autologistic models do. In this way, the spatial particulars of the study area are not "glorified," but rather treated as a nuisance and accounted for so that species/habitat relationships that are not confounded by spatial correlation are revealed. This allows the model to be more transportable; even though it is still based on a particular set of conditions, the fact that spatial relationships have been accounted for and at least partially removed allows models to be compared and extrapolated between areas with more reliability.

A spatially adjusted regression model takes the prediction errors from the main model, which will initially predict occupancy in far too many marginal cells, and estimates separate components for spatial and model-related error. This allows the habitat quality effect of occupied marginal cells to be distinguished from the effect due to close proximity to occupied, more suitable cells. When the main model is adjusted based on the fitted spatial covariance model, the effects of SAC are removed, and the influence of sub-optimal cells on model specification is indirectly minimized. The final model estimates the coefficients associated with habitat variables more accurately and therefore defines habitat more precisely than a non-spatial model could. Spatially adjusted models also provide more realistic estimates of model and coefficient error.

It is important to note that there will be error in the output of any model that predicts the space use of a colonial species based only on habitat attributes and associated habitat occupancy patterns. In prairie dogs, space use is influenced not only by habitat quality and social behavior, but also by the historic legacy of past occupancy. The models used in this research incorporate measures of habitat quality directly, but social characteristics and historical factors are incorporated only indirectly. Social characteristics are modeled through spatial correlation models, and the historical factors of past prairie dog poisoning and shooting through using the proxy variable distance to private land. Needless to say, these are crude approximations of an intricate reality, but the point of this study is to explore whether these approaches are adequate to improve model performance and increase our knowledge of prairie dog/habitat relationships. In the absence of detailed, long term field data to model habitat use and selection, these models represent the best way to incorporate these patterns.

Spatial Modeling and Scaling Issues

The first step in creating a spatial model is deciding the scale(s) at which to study the phenomenon of interest. This is driven by theoretical and practical considerations, but must ultimately make biological sense if the model is to be useful. Related to this is the choice of whether to employ a vector or raster-based analysis. Vector analyses are based on the attributes of a collection of polygons that are created based on features of the mapped area (e.g. dominant vegetation), while raster data are created by artificially dividing the map into individual cells of equal size and assigning habitat values to individual cells. Vector-based analyses are intuitively appealing since they categorize a map feature uniformly within the boundaries of the polygon. Raster-based analyses, on the other hand, offer easier data collection and quantitative analysis. The choice between the two is often dictated by available data but should also be based on the phenomenon being modeled.

The scale of observation used to study many organisms that occupy habitat in a non-random fashion is that of an individual unit, or patch, of habitat. This unit can be defined independent of knowledge about organism occupancy for many species, assuming an a priori knowledge of the biotic or abiotic characteristics of a species' habitat. Some of these species, such as butterflies that occupy meadows, frogs that occupy ponds, and owls that occupy remnant patches of old forest, have been successfully modeled at the habitat patch scale using spatial models (Noon and McKelvey 1992, Hanski 1994, Hanski et al. 1996, Sjogren-Gulve and Ray 1996). Prairie dogs, in contrast, occupy a relatively homogeneous area with a high proportion of areas that could be occupied. Even though habitat quality is likely quite important for habitat occupancy, relevant habitat traits vary along gradients, and the habitat patch is nowhere to be found. Habitat suitability for this species cannot be defined independent of habitat occupancy, and this makes the habitat patch an invalid grain size at which to gather and process data.

Another option is to let the organism define the scale, quantifying habitat covariates and occupancy within individual polygons defined by the organisms themselves. This choice is intuitively appealing, but tremendously complicates data collection and analysis. For example, the coterie would seem to be an appropriate scale of observation, but coteries are irregular in shape, change unpredictably over time and space, and do not necessarily have a coherent identity over time. The same could be said of colonies, which have even less identity as a spatial whole. Prairie dog colonies move around the landscape over time, expanding, contracting, changing configuration, dying, connecting, and separating. In many cases, it is impossible to say where one colony begins and another ends. Also, habitat attributes such as slope are scale dependent, and would likely lose most or all of their information content through averaging

Modifications Inspired by Initial Model Evaluation

Spatial model error adjacent to private inholdings within Conata Basin was high, and this introduced doubts about the way distance to private land had been quantified for use in initial models. As previously mentioned, this area is managed as a black-footed ferret reintroduction area. Since prairie dogs are the obligate prey species of the black-footed ferret, shooting has been prohibited in this area since 1994, and prairie dog poisoning is limited mostly to areas on the edges of the basin that are adjacent to private land. Even though there are a few private inholdings, these areas received very little poisoning, were generally not fenced off from publicly owned areas, and were managed in a manner similar to publicly-owned areas. Given this, it made little sense to use distance from these areas as a predictor of suitability. Similarly, prairie dog control did not occur on the grasslands adjacent to Badlands National Park, where prairie dogs are protected. These areas were removed as sources for distance calculations, new distance data were created, and estimation was repeated using the new distance data along with the predictors from model 3 and the predictors from model 4. These models were called models 5 and 6, respectively.

Model predictions improved around private inholdings as a result of these changes , but error still occurred in areas adjacent to private land at the edges of the study area. Figure 28 shows results for model 5, which had the best AIC of the two (model 5 AIC = 400351, model 6 AIC = 400386). Based on further examination of errors around the edges of the grassland, it was unclear whether removing the distance to private land predictor would improve or degrade overall performance. It was suspected that, like precipitation, the distance to private land variable had been retained in the model because it increased predictions of unsuitable habitat in unoccupied areas at the expense of increasing false negative errors. Spatially, the predictor seemed to be creating a halo of error around the edges of the study area, with a high proportion of unsuitable habitat predicted in these areas and a low proportion in interior areas.

Biologically, it was clear that many areas adjacent to private land were suitable since they were occupied at some point; otherwise, they would not have been poisoned. Poisoning no longer occurs on the grasslands, and studies have shown that prairie dog populations are capable of recovering within a few years in areas where they have been exterminated (Uresk and Schenbeck, 1987). Areas are not poisoned unless they are occupied, and poisoning is obviously closely related with habitat occupancy, since only trend in the response and improving negative predictions at the expense of positive ones. Doubts as to the biological validity of this predictor as well as its suspected link to region-specific loss of model sensitivity warranted its removal. Model 3 is estimated without precipitation, and Model 4 also removes high mean clay content of all soil horizons.

All four models were estimated using both 900 m and 1200 m spatial block sizes. Comparison of spatial models 1-4 using AIC (Table 12) revealed that when the same model was calculated using different spatial block sizes, those 900 m spatial blocks explained the data much better. Within this block size, model 3 was the best spatial model of the four. This model was selected for comparison with a) the model that would have been chosen using non-spatial analyses and b) a non-spatial model using the predictors in model 3. Note that this last model is a by-product of the spatial analysis, since the spatial analysis was necessary to select the variables used.

Comparison of AIC values from these three models (Table 13) demonstrate a dramatic improvement in AIC value when the spatial model is used. The probability (w_r) that the spatial model is the best fit to the data is close to 1.00. Regression coefficients associated with the various predictors also changed dramatically between the final spatial model and its non-spatial counterpart, indicating that modeling spatial covariance has a large effect on overall model results. Maximum likelihood estimates for spatial model 3 are shown in Table 14. The next step was to determine how these differences between models translated into differences in predictions on modeled data.

Initial Model Evaluation

The false negative prediction error rate of the full non-spatial model seemed to have little relationship with the spatial context of colonies, whereas those of spatial models often occurred at the edges of colonies and showed specific spatial patterns (Figs 26 and 27) . These patterns motivated further modifications, as described below.

suitable and occupied areas) is encouraging. Plotting model predictions on the map shows that the spatial model only roughly approximates patterns of observed prairie dog habitat use, however (Fig. 25). The model has a highly negative intercept, requiring that the cut point be adjusted unusually low (0.008) in order to maintain 95% sensitivity. This results in poor ability to discriminate unsuitable areas. The relatively poor performance is not surprising given the final set of predictors used. Two of the three (distance to private and estimated annual precipitation) describe habitat quality quite indirectly if at all. The improvement in fit using a simple spatial model instead of a more saturated non-spatial model was encouraging, but model results indicated that there was a need to develop further predictor variables and improve the accuracy of existing ones if possible. Variables were modified, replaced, and augmented as described previously in an attempt to improve model performance on the Conata Basin dataset.

Conata Basin

The general approach to analysis of Conata Basin data was the same as that used for Fort Pierre except for the elimination of preliminary stepwise regression models. As before, all candidate predictors were highly significant in the non-spatial logistic regression analyses; in the spatial models that followed, predictors were removed in steps. Table 11 shows the predictors highly significant in the non-spatial models and those included as candidates for the initial set of spatial models. The first spatial model (Model 1) had half the number of predictors used in the non-spatial models. Model 2 is the same as 1 except for the removal of % silt and clay of the top soil horizons.

All the Conata Basin models described above used precipitation data, and all had fairly poor sensitivity in the eastern parts of the grassland. Further examination revealed that habitat occupancy and precipitation data had a similar overall trend of decrease from west to east, a trend that was reflected in an increase in false negatives from west to east The precipitation trend was similar to that observed in the Fort Pierre data, but the regression coefficient associated with precipitation changed sign between study areas, an indication that it had no clear relationship with habitat suitability. Even though precipitation had an obvious negative effect on prediction of suitable areas, it was likely significant due to the fact that it identified much of the eastern sections of the grassland that were unoccupied as unsuitable, mimicking a

SPATIAL ANALYSES

Fort Pierre

Spatial analyses on the Fort Pierre data initially included all highly significant predictors from the non-spatial logistic regressions. A subsequent correlation analysis on selected predictors revealed that the three predictors that describe clay content of all soil horizons (none of which were significant) were so highly correlated that they could not be used together. Regression analyses on highly correlated predictors introduce error into the model by distorting estimates of regression coefficients and standard errors (Neter et al. 1989). It was necessary to select one of these predictors to include in the model and remove the other two. The predictor that describes the highest clay content of all soil horizons in each map unit was retained based on its p-value, but it remained insignificant when a reduced spatial model was calculated.

Many of the predictors that were highly significant in the Fort Pierre non-spatial models became insignificant when spatial correlation in model errors was used to adjust the main effects model. When these terms were dropped from the model, only three variables remained: mean clay content of the top soil horizon, estimated annual precipitation, and distance to private land. As previously mentioned, none of the models calculated with different block sizes using Fort Pierre data can be directly compared using AIC. Spatial models calculated using 900 m spatial blocks were used as the final models for comparison purposes because this block size produced the best AIC values on the second study area, where performance could be quantified. AIC values for the spatial model with significant predictors, the non-spatial model with the same variables, and the full non-spatial model are shown in Table 9. Based on AIC values, the estimated probability w_r (Burnham and Anderson 1998) that the reduced spatial model is better than the others exceeds 99%.

Accuracy statistics produced through resubstitution analysis on the modeled data (Table 10) show that when correct prediction of occupied areas (sensitivity) is held at 95%, the spatial model has roughly double the rate of false positive errors, but one third the amount of false negative errors. False positive errors cannot be verified, as discussed below, but the lower rate of false negatives (i.e. failing to detect intensity/pasture area and the interaction between distance to private land and poisoning probability, but all other terms were highly significant ($p < 0.01$) and were retained in the model.

This model with interactions was then compared with one that included only the significant main effects (Table 6). The larger model did not appear to have greater explanatory power based on the estimates for percent concordance (86.5% vs. 86.6%), the percentage of response pairs properly identified if all possible outcomes from the model are paired and compared to what was observed. Interaction terms in the full model also had unreasonably large coefficients and standard errors, an indication of model overfitting (Hosmer and Lemeshow 2000); when they were included, the standard errors associated with main effects were larger. For these reasons, interaction terms were removed.

Further output for the main effects model (Table 7) shows that the Wald and profile likelihood confidence intervals for the odds ratio estimates do not include 1, so there is a 95% probability that there is a difference between the predicted response as the level of a each predictor changes. The three tests of the global null hypothesis also show that at least one coefficient is significantly greater or less than zero; the model predicts with greater accuracy than one could by guessing. The model is not sound, however, and the statistics are misleading. A concordance rate like that observed could be produced by a model that predicted no habitat occupancy correctly. The Hosmer and Lemeshow goodness-of-fit test shows a significant amount of variation in the model that remains unexplained, and the classification table (Table 8) demonstrates a high degree of false predictions and lack of discrimination. Confidence intervals are actually much wider for reasons mentioned earlier. The use of an inadequate analysis framework obscures information that might be available if spatial correlations in model error were incorporated.

Variogram analysis of model residuals (Fig. 10) showed error was spatially correlated to a distance of roughly 700 m. This distance was used as an initial range estimate for subsequent spatial models.

RESULTS

NON-SPATIAL ANALYSES

Linear regression models of proportional occupancy using all habitat predictors and no interactions performed poorly. Even though all terms except one interaction (slope * mean %clay) remained after stepwise selection and were highly significant, the linear model explained only 55% of the variability in the response, and the standard errors of the parameter estimates were large (Table 5). Boxplots of residuals from this model at all levels of observed occupancy (Fig 12) show that the model predicts unoccupied areas fairly accurately, but shows an increasing trend of over-prediction as observed occupancy increases.

Boxplots summarizing the range of values of each predictor at various levels of the response (Fig 13-24) further demonstrated that the data did not support modeling occupancy as a proportional response: none of the terms in the model demonstrate a linear response with occupancy. Examination of the boxplots with a binary response in mind, however, revealed strong patterns for some predictors. Precipitation and poisoning both have a strong positive relationship with the response, while occupancy and the clay content variables were clearly negatively related. Livestock stocking levels (recommended and actual) and slope had a weak negative correlation with occupancy, and standard deviation of mean stocking intensity and distance to water and private land had little apparent relationship with occupancy at all. Based on these results, the use of a binary response variable and logistic regression model for additional analyses was justified.

Stepwise logistic regression using all the interactions and main effects in Table 3 was used to eliminate any extraneous predictors. In the initial model run, the program detected a quasi-complete separation of data points and failed to converge to a solution. This error was caused by the poisoning predictor, which by definition was zero for all unoccupied areas (an area had to be occupied to be poisoned). With no overlap between the distribution of the covariates between two outcome groups, maximum likelihood estimation cannot proceed (Hosmer and Lemeshow 2000), so it was necessary to remove poisoning from the model. Stepwise selection also removed the standard deviation of stocking was removed, and several variables that became insignificant in the process were also removed. Details are given in the results section.

95% *sensitivity* (correct prediction of occupied areas). *Commission error*, predicting occupancy for unoccupied areas, is higher in all models as a result, but this type of error is expected due to the patchy but temporally dynamic distribution of prairie dog colonies. Given this dynamic spatial distribution, only a fraction of suitable areas are expected to be occupied at any given time, and unoccupied areas are not necessarily unsuitable.

Accuracy assessment statistics calculated after data resubstitution and cutpoint selection can reveal important details about model performance, but some are misleading. For example, results of resubstitution analysis on the spatial and non-spatial Fort Pierre models (Table 9) indicate that non-spatial models generally have higher sensitivity when specificity is held constant, which in this case means they performed better at classifying areas that are unoccupied and possibly unsuitable. Higher specificity ratings may reflect an ability to identify unsuitable habitat, but they also may be a result of incorrectly identifying suitable but unoccupied habitat as unsuitable. There is no way to verify whether or not areas are unsuitable when suitability is only known in relation to observed prairie dog habitat occupancy, especially in a species that is known to occupy only a fraction of available habitat at any given time. Similarly, statistics based on overall accuracy (e.g. Kappa) are not useful when accuracy in one component of a binary response is a priority and accuracy in the other component is also important, but impossible to verify.

False negatives are easily confirmed, however, since by definition occupied areas are suitable. They are also the most important type of error to minimize in a model that attempts to define suitable habitat for the black-tailed prairie dog on those public lands essential for the conservation of the species. This species has lost at least 98% of the habitat it occupied before the arrival of white settlers in North America, and most of that habitat now provides irreplaceable food resources and probably will never be reclaimed. The top priority in this situation is to identify suitable areas on public lands that remain so that they all may be included in habitat assessments. Minimizing false positive errors at a set level of sensitivity (95%) was therefore the focus of quantitative model evaluation.

Map-based evaluation focused on comparing the spatial pattern of model errors at individually determined cut points. Patterns in model error prompted the elimination and modification of several additional predictors. Estimated annual precipitation was removed, distance to private land was modified to exclude National Park areas and inholdings as sources for distance calculation, distance to private land square) while maintaining identical output data sets. These spatial models may be directly compared with each other using AIC. Non-spatial models were subjected to the same 900 m spatial blocking code and analyzed using a non-spatial version of the same routine as the final spatial models, so these results are directly comparable as well.

Evaluating Model Output

In general, model performance may be evaluated by predicting the response for the full data set used to build the model (*resubstitution* or *verification*), for portions of the original dataset (*crossvalidation*), or for independent data sets (*validation*). Model validation between study areas could not be supported due to previously mentioned differences in the input data. *Crossvalidation* requires repeatedly excluding different portions of the input dataset, recalculating the model, and predicting for the data excluded. This approach was impractical since performing the number of model runs necessary to complete a single crossvalidation analysis would take months or years with the computing resources available. Validation of the final Conata Basin models was possible only for areas immediately adjacent to Conata Basin where prairie dog habitat occupancy and habitat covariate data were developed.

Given all these restrictions, resubstitution using the modeled dataset was the main method used to evaluate model performance. Predictions using the logistic regression equation are first transformed through the logit link function

$$\frac{1}{1 + e^{-\eta}}$$

where η is the linear predictor produced by the logistic regression equation. The results of this transformation range from 0 to 1, and must be reclassified back to a binary response to compare them with the original data.

The cutpoint used to classify model predictions has a strong effect on performance, and can only be optimized for one type of error at a time. In this study, it was most important to minimize *omission error*, the failure to predict suitability for areas that were occupied, so cutpoints were optimized to maintain blocks). Models using block sizes of 1500 m on a side were attempted, but these required excessive computing time and could not be estimated with available computing resources.

Spatial blocks for each model run were defined within the SAS code using the map coordinates associated with each observation (UTM Easting and Northing). This code uses minimum and maximum coordinate values along with user-specified values for the number of cells allowed in each dimension, assigning a block number to each observation using mathematical formulas and effectively dividing the area into symmetrical blocks. The nature of this blocking code made it impossible in some cases to compare models quantitatively using Akaike's Information Criterion (AIC) or similar measures because the data sets used to calculate the models sometimes varied with spatial block size.

The problem is illustrated in Figure 11

using a small section of Fort Pierre grassland. The blocks contain various numbers of observations based upon the amount of area within the block that was included in the analysis: if most of a block has been masked out because it is composed of non-habitat, that block will have few observations within it. Blocks with less than 20 observations must be excluded from the analysis, or convergence problems may result. When different block sizes are used to divide up the same area, different pieces can be included or excluded. Figure 10 shows occupied habitat in red, unoccupied habitat in white, and non-habitat excluded from the analysis in black. Spatial blocks are shown outlined in black. In the left diagram, all observations within the 750 m square block adjacent to the number 3 would be removed from the data, even though cells occupied by prairie dogs are included in this block. The 1500 m block size on the right includes the same area but easily meets the 20 observation minimum criteria. In the dataset that uses this block size, all of these observations are included in the analysis.

In general, more observations are excluded when smaller spatial block sizes are used, and almost none are excluded at larger block sizes. Unfortunately, it is virtually impossible to control where these deletions occur. This problem was most evident in the Fort Pierre study area, which is highly dissected with areas of non-habitat. In this area, the data sets produced by the spatial blocking code varied by a few thousand observations between block sizes. Even though these differences amounted to less than 1% of the total sample size, the models cannot be directly compared. In Conata Basin there is only one major private inholding, and it was possible to implement the blocking code for two block sizes (ca. 900 m and 1200 m

Given these computational limitations, the only way to make the modeling process viable was to divide the study area into blocks and use these as subjects for the spatial models. This extension of repeated measures design incorporates variation in two spatial dimensions instead of one temporal dimension, analyzing the repeated measures to produce global models for spatially correlated errors. These errors are quantified iteratively for each block in the dataset, and the global spatial model produced at the end is used to adjust a main effects model based on the entire dataset (ignoring the spatial blocking structure). The spatial model used for adjustment on the first iteration is created with user-specified distance parameters estimated with variogram analysis as described above. After the main effects model is calculated, the process repeats, using the error matrix from the first global iteration to refine the spatial covariance model and then the main effects model. In effect, there are two sets of iterations involved. The first set is used to calculate a global spatial model to adjust the error surface associated with the main effects equation. These iterations are nested within a second set that repeats the entire process, refining both the spatial and main effects models (as described above in the theory section) until convergence. The estimation procedure is described in Table 4.

This design, inspired by necessity, may actually lead to more general results. Previous spatial analyses using Proc Mixed have indicated that a spatial extent of roughly double the range is usually sufficient to capture spatial variation (Rudy King, pers. comm.). Extending the side dimensions of the spatial blocks past this range would not be expected to add explanatory power to the model, and could even detract from model performance by including too much spatial information at scales irrelevant to prairie dog habitat occupancy and "glorifying" the particular spatial pattern found in a dataset composed of a few, very large spatial blocks. In this situation, generalizing multiple small snapshots of spatial covariance into a single model rather than using one vast snapshot may be more robust.

Altering the size of the blocks used to calculate the spatially-adjusted models changes the extent of the spatial analysis window, dividing the study area into greater or fewer "snapshots" of data for analysis as individual subjects. Changing block size would be expected to affect calculations of spatial covariance, but the potential effects on predictor variables and their coefficients are not as clear. In order to explore how variation in spatial block size affected model output, separate spatial models were estimated using blocks of approximately 750, 900, and 1200 m on a side (these will be referred to as 750, 900, and 1200 m spatial

The spatially adjusted regression analysis described above represents the simplest way to account for the spatial correlations inherent in the dataset. Accounting for these correlations is necessary because the data violate the assumptions required by non-spatial linear and logistic regression analyses, and produce inaccurate and invalid results when these methods are used. More complex spatial analyses than the ones currently considered are theoretically possible, but not currently practical on large data sets like the ones used in this study. In addition, software to implement these analyses is not readily available. The spatial analyses provided by Proc Mixed with the Glimmix macro (SAS 1999) represent the only existing commercial software capable of a landscape level logistic regression analysis using a spatially adjusted model.

Spatial modeling is always a compromise between fitting the data at hand as accurately as possible and attempting to extract general models from specific situations. The aim of this model building exercise tends toward the latter; the purpose is not to predict the response of prairie dogs extremely accurately on any given 30 m square of the grasslands, but to discover the relative importance of known habitat attributes in order to create general models of how habitat attributes and patterns of prairie dog habitat occupancy are related. Modeling spatial covariance as a repeated measure in space could be an effective way to obtain that generality.

Implementation

Spatial analysis using Proc Mixed/Glimmix is an iterative procedure with a large number of computations; each analysis can take weeks to complete. The procedure also requires large amounts of random access memory (RAM) since an $N X N$ matrix (where N is the number of sites) must be held within memory for the duration of the calculation. Practical considerations related to these processing time and RAM requirements limit the extent of the geographic sampling frame that can be used for the spatial models to sizes much smaller than the study areas. Randomly subsampling 50% and 75% of the data from the full dataset was investigated as a way to increase the extent of the area used to build spatial models and maintain reasonable sample sizes, but repeated analyses using different samples drawn from the same data showed unacceptable variation in regression coefficients and their standard errors between samples.

concluded that the impact of the specific function used in the spatial model is typically negligible as long as a reasonable model with appropriate parameter estimates is used. The analyses performed for this research all use a spherical model, which may be expressed as

$$f(d_{ij}) = [\, 1 - 1.5\,(d_{ij}/\rho) + 0.5\,(d_{ij}/\rho)^3]\; 1\,(d_{ij} < \rho)$$

where ρ is the range of spatial correlation and the function $1\,(d_{ij} < \rho)$ equals 1 when $d_{ij} < \rho$ and 0 otherwise.

The spatial correlation model $f(d_{ij})$ is estimated using generalized least-squares (GLS), which minimizes the expression

$$(\mathbf{y} - \mathbf{X}\beta)'\, \mathbf{V}^{-1}\, (\mathbf{y} - \mathbf{X}\beta).$$

The matrix **V** in the above equation is the variance of y, and is equal to

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$$

Knowledge of **V** thus requires knowledge of **G** and **R,** which are unknown but may be estimated using likelihood-based methods that exploit the assumption that **e** and **u** are normally distributed. The software used to implement the analysis (SAS Proc Mixed) creates an objective function associated with maximum likelihood (ML) or restricted/residual maximum likelihood (REML), maximizing it over all unknown parameters to estimate **G** and **R**. Estimates of β and **u** are obtained by solving the standard mixed model equations. When the GLIMMIX macro is used with Proc Mixed, the variance function $\mathbf{R}_{\mu}$, which is based upon a spatial model, is used as a weight and the linear predictor y* replaces y in the mixed model equations. Proc Mixed is called iteratively and implements the two estimation methods described above until the relative deviation of the variance/covariance parameter estimates indicate convergence. This estimation method is called pseudo-likelihood (PL) or restricted pseudo-likelihood (REPL), and is described in detail in Wolfinger and O'Connell (1993) and Littell et al. (1996).

where e_i is the error associated with y_i, the i^{th} observation of y. The spatial component is defined by letting

$$\mathrm{Var}(e_i) = \sigma_i^2 \text{ and}$$

$$\mathrm{Cov}(e_i, e_j) = \sigma^2 [f(d_{ij})]$$

where d_{ij} is the distance between site i and site j. The spatial function $f(d_{ij})$ is the same for all pairs of equally distant locations (i.e. it assumes stationarity), but may incorporate the influence of direction (isotropy) if desired.

The spatial covariance models used by Proc Mixed rely on semivariance statistics originally created for predicting ore deposits in mining operations (Matheron 1963, Journel and Huijbregts 1978, David 1977). These are calculated by taking the variance between each possible pair of observations in a study area and sorting them according to distance. Variograms expressing the sum of squared differences between all observations (y) within each range of distance (x) are used in this analysis to provide the initial estimates of σ^2 and ρ required for the spatial models. These correspond to the sill and range of a variogram, respectively (Fig. 10). The sill of a variogram is the average semivariance observed between observations that are at or beyond the range of spatial correlation. The range of the variogram is the distance to which habitat occupancy is correlated; beyond this distance there ceases to be a relationship between distance and semivariance. This quantity is used to define the minimum block size needed to capture spatial variability. The nugget is the variance between observations that are adjacent, and represents variability in habitat occupancy due to process or measurement error. High nugget variance suggest that the pattern varies at a finer scale in space than the sampling scheme or that the pattern is inherently heterogeneous.

The shape of the variogram produced by the data is used to define the general function used to describe spatial dependence. Proc Mixed is capable of using several forms of spatial models, including spherical, exponential, gaussian, linear, linear log, and power functions. Littell et al. (1996) duplicated spatially adjusted regression analyses with Proc Mixed using different models on the same data, and design matrix, **u** is a vector of unknown random-effects parameters, and **e** is an unobserved vector of errors. These models are referred to as "mixed" because both fixed and random effects can be used, though random effects are not employed in the current analysis. Spatial covariance is modeled through analysis of correlations in the error term **e** using the **R** matrix, the matrix that defines the error expected due to SAC. The vectors **u** and **e** are assumed to be normally distributed with

$$E[\mathbf{u}] = 0 \text{ and } \mathrm{Var}[\mathbf{u}] = \mathbf{G}$$

$$E[\mathbf{e}] = 0 \text{ and } \mathrm{Var}[\mathbf{e}] = \mathbf{R}$$

R is calculated by defining an $N X N$ matrix **F**, where N is the number of pixels and each ij^{th} element $f(d_{ij})$ expresses covariance for each site as a function of distance from every other site. **R** is given one of two forms based on the observed variation in the response at the smallest lag distance (30 m). If the observed variance at close range is low, the model with no nugget variance (see below)

$$\mathbf{R} = \sigma^2 \mathbf{F}$$

is appropriate, otherwise

$$\mathbf{R} = \mathbf{I}\, \sigma_1^2 + \sigma^2 \mathbf{F}$$

is used, adding the nugget variance to each element in **R.** In these equations, σ^2 is the deviance between observations that are far enough apart to be unaffected by SAC. The calculation of **R** thus involves multiplying each element of the matrix **F** by the estimated variance between observations too distant from one another to be spatially correlated (deviance).

The basic form of the spatial correlation model is

$$y_i = \mu + e_i$$

independent variables in a model that fails to account for spatial autocorrelation can be highly significant yet have little predictive power. The results of these non-spatial linear and logistic regression analyses justified the use of a spatially adjusted GLMM analysis.

Spatial Models

Theory

Model errors like those described above are expected when the model fails to incorporate spatial autocorrelation in the response. In this case, it is suspected that habitat of lesser quality may be occupied more frequently than would be expected when it is located next to high quality, occupied habitat. Conversely, high quality habitat may be unoccupied if it is far from occupied areas. If this is not accounted for in habitat suitability analyses, occupied low quality habitat is treated as equivalent to high quality habitat, producing bias in estimates of the coefficients associated with habitat attributes and reducing the precision of estimates of the response. As the number of predictors in a model increases, the potential importance of accounting for spatial variability in the data also increases (Littell et al. 1996). In this context, models that adjust for spatially correlated error produce more accurate and unbiased estimates of the relationship between habitat attributes and habitat occupancy.

The spatial models used in this analysis divide the study area into blocks and use these blocks as subjects for a global spatial covariance model. The analysis uses a formulation of the generalized linear mixed model (GLMM). Mixed models relax the assumption of independent and identically distributed Gaussian random errors made in the general linear model, allowing for heterogeneous variances and correlated errors. The mixed model is written as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

where y is a data vector transformed using a differentiable monotonic link function, **X** is a known matrix of explanatory variables, β is a vector of unknown fixed-effects parameters, **Z** is a known random effects

DATA ANALYSIS

Non-spatial Models

The Fort Pierre data were initially processed for analysis by creating grids with nonzero values for occupied areas and zeros for unoccupied areas, with nonzero values in each grid corresponding to the length of time between the current and subsequent census. When added together and divided by the total number of years, the proportion of time a cell was occupied during the time series is estimated. This operation is not completely realistic because it assumes no changes between censuses, but it was thought to have the potential to provide more information than a binary response if supported by the data. Using the Fort Pierre data, preliminary statistical models were fit to this proportional response in order to identify weak predictors and establish the need for more complex models. These data were suitable for analysis with ordinary linear regression and spatial autoregressive models, and it was necessary to establish that these methods were not sufficient before pursuing more complex analyses. The results of these preliminary analyses proved that analysis with a proportional response was not supported by the data (see results) and justified the use of a binary response variable and logistic regression. The analysis process used on the Fort Pierre data set is presented in Figure 9.

Classifying occupancy as binary rather than proportional eliminates the need to model temporal autocorrelation because frequency of occupancy is no longer considered. It also allows the analysis to be readily implemented using available software that produces goodness of fit statistics for model evaluation. Perhaps most importantly, it eliminates the need for the troublesome assumption that there have been no changes in occupancy between censuses, since it is only necessary to know which pixels have been occupied at any time during the time period studied. These practical considerations also made logistic regression more appropriate as an analysis tool.

Simple logistic regression first was used to check the performance of a non-spatial model and screen variables for use in the initial spatial models. All predictors tested were highly significant, and model concordance statistics were good. Classification table results show that the model performs poorly when it comes to predicting suitable habitat, however. As previously discussed, Figure 6 shows how

Two images (August 6, 1991 and August 22, 1997) were selected for Conata Basin based on these considerations. These acquisition dates fall near the two endpoints of the occupancy time series, and may not be representative of intervening conditions, but they represented the only high-quality data available. The red and near-infrared bands of these images (bands 3 & 4) were converted into ArcInfo grids, and NDVI values were calculated for both using the Map Calculator in Arcview (ESRI, 1999). The average of values from both years was used as a predictor in the models.

Converting Coverages to Input Data

All data layers were ultimately converted into co-registered grid coverages. Areas to be excluded from the analysis were delineated by creating a mask grid coverage. The following areas were removed using the mask grid:

- all private land and all pastures containing any private land
- all wetlands, river corridors, and water features
- all wildlife areas, exclosures, and waterlots
- all areas with missing soils or other data layers
- a buffer of 15 m on each side of all primary roads and railways
- a buffer of 30 m on each side of all highways

Data from each of the co-registered grids were extracted using the sample command in ArcInfo, producing delimited text files containing an easting and northing corresponding to the centroid of each grid cell followed by the value for that cell from each grid sampled. These files were imported into Arcview and converted to Dbase format for editing. Data formats were standardized, pixels with missing observations were removed, and the final data sets were imported for statistical analysis.

Slope

Slope data were created from digital elevation models (DEM's) produced by the USGS (U.S.Department of the Interior, U.S. Geological Survey, 1992). DEM's are created from a sampled array of elevations from some number of ground positions at regularly spaced intervals. The slope data used in this study were created with a 7.5-Minute DEM's created using a data spacing of 30- x 30-m and elevation values determined by manual profiling using photogrammetric stereo-models. Limitations in the spatial resolution of these DEM data have implications for the insignificance of this predictor in the final models, as discussed below. DEM data were converted to floating point slope grids using Arctools-grid for use in the models.

NDVI

The Normalized Difference Vegetation Index (NDVI) is a ratio of the red and near-infrared bands of a satellite image that is designed to detect green vegetation. Due to data availability, this index could be used on the Conata Basin analysis only, as discussed above. Landsat Thematic Mapper (TM) images of this study area were selected for use in the analysis based on several criteria.

- Images with any cloud cover obscuring the study area were rejected.
- Discrimination of grassland vegetation types is clearest for images acquired in late summer or early autumn (Mike Morrison, pers. comm.); only images from these time periods were considered.
- Imagery used to compare or average index values between years should be radiometrically corrected to adjust for atmospheric influences and acquired as close to the same date as possible in order to minimize the confounding influences of atmospheric and phenological factors (Roger Hoffer, pers. comm.). Uncorrected images or images not acquired within three weeks of the same calendar date were rejected.

ArcInfo (ESRI, 1998). This command creates grids of distance values using an input grid with non-zero real numbers for all sources and no data for areas for which distance values are desired. It was necessary to repeat this analysis for each pasture to find distance to water within a pasture, using Forest Service coverages of livestock water developments and permanent water features from the National Wetlands Inventory as sources instead of private lands. The repetitive nature of the water distance analyses warranted writing a macro to automate the process (see appendix).

Use of Weather Data

Weather station data (HPRCC, 2001) for stations adjacent to each study area were used to estimate precipitation values across the grasslands. Average annual values for each station were first calculated for the time periods studied, and inverse distance-weighted interpolation was used to estimate values for each pixel in the dataset. Although precipitation data does include moisture due to snowfall, the efficiency of snow capture by measuring devices is quite variable (Doesken et al. 1996). Snowpack data available for the study areas had too many missing observations, however, and Snow Water Equivalent (SWE) data needed to convert snow to moisture data were not available.

Incorporating Actual Use Grazing Data

National Grasslands grazing records were used to calculate animal unit months (AUM's) per unit area for all pastures. An animal unit month is the amount of forage consumed by a 450 kg beef cow (or bull) with or without a nursing calf in one month, which is estimated at 355 kg of dry forage per month or ca. 11.8 kg. per day. Grazing records showed the number of cows and bulls grazed in each pasture and the dates during which this occurred. These data were converted to AUM's and appended to the GIS coverage of pastures. Dividing the AUM's by the area of each pasture produces the predictor variable used in the analyses.

pixels, so this predictor varies at a fairly coarse spatial scale. The Conata Basin analyses used prodnorm directly, dividing this quantity by soil survey polygon (map unit) area. Average size of a map unit in Conata Basin is ca. 20 ha or 222 pixels, and this predictor captures variation at a finer spatial scale.

The predictors for estimated percent sand, silt and clay in the top layer of soil were generated using a lookup table created by Dr. Jan Cipra in the CSU Soil Sciences GIS lab. This table uses the texture field in the database to estimate percentages for these three variables based on relationships observed in soil survey data taken throughout the United States. It is unknown how accurately the lookup table estimates these percentages for the soils in the two areas examined in this study, however. Also, this table became available after the Fort Pierre data analysis was completed, and was used only for the Conata Basin data.

Creating Distance Related Predictors

Studies of cattle distribution in relation to livestock watering points on arid rangelands indicate that cattle graze areas close to watering points preferentially (Andrew 1988, Owens et al. 1991). Prairie dogs are also known to respond positively to increased cattle grazing pressure (Uresk et al. 1981). Distance to water was used as a proxy for cattle grazing intensity differences related to distance from water sources. When pastures are fairly small and watering points are distributed evenly, livestock position within a pasture may be unimportant for water availability, and the relationship between this predictor and habitat occupancy would likely be weak. If distance from water strongly influenced the cattle grazing patterns, however, and if prairie dogs responded to these differences in grazing intensity, the predictor could be highly significant as a predictor of prairie dog habitat suitability.

Distance to private land was included mainly to reflect the probability of poisoning in colonies on the National Grasslands adjacent to private land. Poisoning in the area has occurred in response to the complaints and legal actions of adjacent private landowners. This predictor may also incorporate shooting effects, or the decreased emigration that might occur when an area is adjacent to a hostile matrix where there are few colonies or none.

Predictors that describe distances were created through analysis of existing data. Distance to private land was easy to calculate in one step by using the eucdistance command in the GRID module of same soil types present in the same proportions, and are distinct in this way from all other map units. The relationship between a map unit and soil polygons delineated within a GIS is a one to many relationship, with one to several hundred map units explaining soil patterns for entire survey areas (individual counties).

The procedures used to perform a soil survey are not easily described, since they vary according to the particular soils and related survey logistics of each area, and must be flexible in order to capture the variation present in each particular situation. Even though the basic objective of soil surveys is the same for all kinds of land, the number of mapping units, their composition, and the level of detail needed in mapping vary with the complexity of soil patterns as well as specific needs of users. Thus the specific methods used and the soil survey that results are matched to the soils and the soil-related problems of the area (USDA 1998).

In general, soil samples are taken along a series of transects, with deep soil pits dug when necessary for more detailed analysis. Multiple soil samples within a map unit are designated using *sequences.* During a soil survey, map units on soil maps are constantly revised to reflect new information gained through the study of soil sequences. The end result is designed to reduce the bewildering number of soil types to a more manageable system of map units that can be practically used to assess soil properties for a wide range of purposes.

Clay content is recorded in the database using the lowest value observed for each horizon in each sequence as well as the highest. Predictors for the model were generated using these data by averaging low, mean, and high clay content estimates for each soil horizon present in the map unit. Soil clay content predictors for the Buffalo Gap dataset were calculated by weighting the contribution of each sequence to the statistic based on its estimated proportion within the map unit. These data are more accurate than those used at Fort Pierre, which used a simple average of all sequences within the map unit and did not weight them by proportional area.

Prodnorm is a field within the soil database that estimates the number of pounds of forage that could be produced annually within a map unit given normal growing conditions. This field was used by the NRCS to derive recommended animal unit months permitted within each grazing allotment. In the Fort Pierre dataset, NRCS recommended AUM's divided by allotment area was used as an index of soil productivity and a potential predictor of habitat occupancy. Mean allotment size is ca. 178 ha or 1976 impossible to incorporate poisoning data, so distance to private land was used instead as a proxy variable for the probability of being exposed to poison.

Modification of Soil Data

The USFS soil data layer lacked detailed information about soils, characterizing areas based on range-sites. This classification system, used extensively in range management, combines information about the upper soil horizon, its climax vegetation, and land use. These data were inadequate for several reasons: they confounded soil and vegetation characteristics, ignored subsurface soil features, could not be translated into quantitative soil traits, and as categorical predictors had weak predictive power. It was necessary to restore the links to the NRCS SSURGO soil survey database from which this coverage was generated (USDA, 1997) in order to estimate more quantitative variables to use in model development. In order to link the two data sources, a field within the coverage originally created by appending two fields in the soil database needed to be created in the original soils data. This made it possible to relate map units in the GIS coverage to soil attributes quantified by soil surveys.

Soil survey data provide a wide range of information for many different applications. Estimated clay content and soil productivity were the only variables that could be taken directly from the soils database that seemed relevant to prairie dog habitat suitability. Other soil variables used in the analysis were generated using lookup tables (see below). Clay content was summarized based on averaging values for all soil layers and for the top horizon only.

Explanation of further differences between the clay content predictors requires a brief description of the SSURGO database and the soil survey methodology behind it. This database is created using soil maps that delineate the spatial distribution of different soil types. Each soil type has a unique set of interrelated properties and is a product of the parent material from which the soil formed, its environment, and its history. Soils are identified by names that serve as references to a national system of soil taxonomy, and are linked to geographic areas indirectly by using soil survey map units, which are collections of areas defined and named alike because they are similar in terms of their soil components or miscellaneous areas or both (USDA 1998). Individual polygons that are designated as being in the same map unit have the

DATA PREPARATION

Preparing the Habitat Occupancy Data

Fort Pierre prairie dog colony survey data for the years modeled, recorded by drawing polygons on custom paper maps of the grassland (ca. 1:62,000 scale), were digitized using a digitizing table and georeferenced to produce GIS coverages for the years 1975, 1977, 1980, 1981, and 1983. Conata Basin occupancy data for the years 1990 and 1994 were recorded on several 1:24000 scale USGS 7.5 minute quadrangle maps or mylar sheets registered to quad maps. These were scanned at high resolution and converted into geo-referenced image files so that more accurate on-screen digitizing techniques could be used. After digitizing, the individual quads were edge-matched and converted into a single GIS coverage for each year. Survey data from 1996/1997, previously digitized for another modeling study (Bevers et al. 1997) were supplied by John Hof. All survey data were converted into ArcInfo polygon coverages, correcting spatial errors as needed. Further analysis of these coverages is described below.

Quantifying Known Poisoning Events

Grasslands personnel documented poisoning events on Fort Pierre during the time period modeled by referencing a colony number from the colony survey map of the same year and estimating the colony's total area as well as the area poisoned. These tabular data documenting poisoning events were used to estimate the poisoning probability within each colony for each poisoning event by dividing the estimated area poisoned in that colony by its total area. Converting these vector coverages into grids creates identical values in each cell that is within a colony boundary. The cell values correspond to the estimated probability that the cell of interest was poisoned in the year in question. When grids from a time series of years are added together and divided by the number of poisoning events, the cell-specific poisoning probabilities apply to the entire time period. These data were included in the model in order to account for changes that were known to have affected prairie dog habitat occupancy, but the fact that unoccupied areas were never poisoned made these data was incompatible with the estimation process. This made it

METHODS

Although they are within the same region and state, the two study areas are dramatically different in terms of their environment, management, and the data available to model prairie dog habitat suitability. The data used for the Conata Basin analyses were of higher quality for several reasons. Prairie dog habitat occupancy data for this area were recorded using tiled 1:24,000 USGS 7.5 minute topographic quadrangles, while at Fort Pierre lower resolution maps of the entire grassland were used as a base layer. Remotely-sensed data could not be used in the Fort Pierre analysis because the Landsat Multi-spectral Scanner (MSS) imagery acquired concurrently with the relatively old (1975-1983) Fort Pierre habitat occupancy data had insufficient resolution. Conata Basin habitat occupancy data, on the other hand, were acquired more recently (1990-1997), allowing the use of Landsat Thematic Mapper (TM) imagery with a 30 m resolution.

In addition, Fort Pierre was the first study area modeled; the Conata Basin data set was prepared and analyzed with the benefit of that experience. Some predictors used at Conata Basin (% sand, silt and clay) were developed after the Fort Pierre data were finalized, and some were calculated in a more accurate way (clay content soil variables) than they had been at Fort Pierre. Finally, it is likely that the configuration of prairie dog habitat and habitat occupancy in Conata Basin lends itself to statistical analysis more easily than that of Fort Pierre, as discussed below. In the sections that follow, distinctions between the two data sets and their analysis are made as necessary; if no distinction is made, it may be assumed that the methods for each dataset are identical. The procedures used at Fort Pierre are presented in Figure 9.

DATA ACQUISITION

GIS data required for the models were obtained from various sources. Some data layers were created by modifying existing National Grasslands GIS coverages and converting them directly into grids. Tabular data in paper or electronic form were appended to existing GIS attribute tables to generate data for other predictors. Values for some predictors were estimated using interpolation of point estimates or GIS-based macros, and occupancy data were created from existing maps. Details about the development of different data layers for each predictor and the biological rationale behind each are given in Table 3.

the combination of similar management histories within study areas that are fairly different in their ecology and physical structure was thought to provide an assay of the generality of model results, even though models cannot be directly compared between study areas due to differences in input data.

INFORMATION ABOUT THE STUDY AREAS

The Fort Pierre National Grassland (Fig. 7) is located in central South Dakota near the eastern limit of the Black-tailed prairie dog's range. The grassland is on the western side of the transition from mixed-grass to tall-grass prairie, and precipitation as well as productivity are relatively high compared to other areas where prairie dogs are found. Managed primarily for grazing, the grassland has a low concentration of prairie dogs, and occupancy is thought to be limited to areas grazed by cattle. A privately owned biological reserve within which prairie dogs are protected is adjacent to the northwest corner of Fort Pierre, and the Lower Brule Sioux Indian reservation is adjacent to the northeast corner. Other than this, the grassland is surrounded by and dissected with private property where prairie dogs are actively controlled, and these areas are considered non-habitat for the purposes of this study. A time series of prairie dog colony locations mapped from 1975 to 1983 was available for this area.

Conata Basin (outlined in red in Fig. 8) is an area within Buffalo Gap Grasslands in southwest South Dakota that is partially surrounded by the rock formations of Badlands National Park. This mixed-grass prairie is less productive than Fort Pierre and has more changes in topography due to badlands formations. The area has been a black-footed ferret reintroduction site since early 1994, and prairie dogs receive protection as the primary prey of this endangered species. The species is also protected in areas of Badlands National Park to the north. Prairie dog populations in Conata Basin probably have some degree of population exchange with colonies in adjacent populated areas, but this is not considered in this study. Conata Basin, in contrast to Fort Pierre, has only one major private inholding, and habitat is far more contiguous in this area. Colony location data in Conata Basin collected from 1990-1997 were used in the analyses.

The history of the two areas is similar in terms of land use and prairie dog management. Prairie dog colonies were subject to poisoning and shooting before and during the time periods modeled in both areas, and each has a history of homesteading activity in the early 1900's , followed by a period of overgrazing. Grazing intensity in the past few decades has been lowered, and the management focus has recently shifted away from maximizing cattle stocking rates and toward sustainable use. The effects of this past land-use on prairie dog habitat occupancy during the time periods modeled is unknown. In general, values over space. These polygon definition ambiguities and averaging problems would likely introduce so much noise into the data that they would not be useful, and the modeling framework that could incorporate these ambiguities would be intractable. In fact, there may be no way to reduce the complexity of the spatial pattern of prairie dog habitat use to a clearly defined set of polygons at any spatial scale. For these reasons, a cell-based, or raster analysis was chosen. Cell-based analyses reduce a continuous reality to one that can be modeled in a consistent manner, and define the unit of observation independent of current habitat occupancy patterns, avoiding the ambiguities inherent in observations quantified at the patch or coterie level.

The cell size used in a raster analysis defines the spatial scale at which information is averaged into individual observations. The choice of cell size should be supported by objective biological criteria. Prairie dog feeding and affects vegetation in a radius of approximately 10-25 m from an individual burrow (Gold 1976). This could be considered the minimum area of habitat that can be colonized by an expanding coterie, though average coterie size and estimated minimum colony size are roughly three times larger (Clippinger, 1989, Hoogland 1995).

A grain size roughly equivalent to the average area around burrow openings that is affected by prairie dog grazing makes sense biologically and coincidentally, matches up well with constraints imposed by available data. This study used USGS DEM's and Landsat TM satellite imagery with a minimum resolution of 30 m. It is possible to resample these data to a coarser resolution, but this subsumes spatial variance within samples and averages it out of the analysis (Wiens 1989). Also, resampling inevitably increases classification error in the data, especially if raster cells are increased or decreased by a fraction of their former size (Heuvelink 1998). Increasing spatial resolution without using new data is, of course, impossible. In general, it is best to use these GIS data at their original resolution. Fortunately, this scale corresponded roughly with the area around a single burrow for which measurable grazing pressure occurs. Given the above considerations, the 30 m square pixel was the most practical choice for analysis.

areas that are occupied by prairie dogs are poisoned. Distance to private land was created originally for models with a proportional response, and made some sense in that context, but it was possible that including it in binary response models was detrimental. These considerations made it necessary to estimate additional models using the same predictors as in models 3 and 4, but omitting distance to private land to explore the effects of removing this predictor. These models are called models 7 and 8.

Of these two, model 7 had the lowest AIC value (Model 7 AIC=394837, Model 8 AIC=394852), and was chosen as the final spatial model. A non-spatial version of this model was generated for comparison purposes (AIC=822532). Resubstitution results for these non-spatial and spatial models are shown in Figures 29 and 30. Even though the accuracy statistics are similar for the non-spatial and spatial model using the final set of predictors (Table 15), the non-spatial model is able to exclude less non-habitat from consideration at the same rates of error. Both versions of model 7 have a similar rate of false negative error, roughly 1/3 that of the full non-spatial model. Both reduced models have sensitivity and specificity similar to the full non-spatial model, but the rate of false positive errors of these models is roughly double that of the full model. It is unclear whether these errors represent bad predictions or merely suitable but unoccupied habitat; in terms of the types of model error that are verifiable, spatial model 7 is clearly superior to the others, and non-spatial model 7 is superior to the non-spatial full model. The regression coefficients and standard errors of model 7 are shown in Table 16.

Validations of the final spatial model

Performance of the final spatial model was validated by substituting occupancy data from the maximum recorded period of habitat occupancy in the area (1983) and visually examining model fit in modeled (Fig. 31) and unmodeled (Fig. 32) areas. Observed model fit was good, with a low rate of false negative error and good specificity. Model error tended to occur at the edges of colonies, and tracked colony boundaries fairly closely, unlike the errors of non-spatial models, which often seemed to have little relationship with habitat context.

DISCUSSION

"To do science is to search for repeated patterns, not simply to accumulate facts."

Robert MacArthur

The modeling process described above cannot be used to provide direct inference to biological factors associated with prairie dog habitat suitability. Such inference would require the use of predictor variables that have a direct functional relationship with demographic parameters. Although true suitability models like this have been developed for a few, well-studied "flagship" species, the current approach is rooted in the pragmatic realization that data do not exist to model habitat suitability in such detail for most species, and they probably never will. This research attempts to use modern statistical theory, GIS data, and computing technology to circumvent the need for the detailed habitat and demographic data in order to address the conservation problem at hand .

Science is a system of shorthand that allows us to understand what is unknown based upon what is known. When it is most effective, science allows important characteristics of a system to be understood and predicted based upon the values of a few easily measured variables. Even though it may be necessary to explore a situation fairly thoroughly to extract this information, the goal of predictive science is never the exhaustive quantification of a particular scenario (the previously mentioned "glorification of the particular"), but the summarization of important relationships that may be generalized to other situations. Wel (1975) illustrates this in his attempt to model an idealized drawing of an elephant composed of 36 points connected by lines. Although the drawing could be reproduced exactly using a model with 36 terms, the essential features can be approximated by fitting a least squares Fourier sine series function with 30 terms. The author notes that the resulting approximation might not satisfy a third grade art teacher, but could carry a chemical engineer to an initial design. Science allows us to eliminate details not needed to describe the essential components of a system under study, in this case a rather rough picture of an elephant.

These goals of science are well suited to modern species conservation, which must be based upon existing or easily acquired data if it is to be proactive. Funding for ecological research is usually available to study only a limited number of species; most of these are of direct value to humans or are already imperiled. This leads to a situation in which research resources are allocated using a triage paradigm. Important details critical to the survival of a given species are often not known until that species is on the brink of extinction, and funding is often not available for detailed study of ongoing conservation problems until they are in their final stages. These realities are conditioned not by science, but by human culture and politics. In the case of the prairie dog, the funding situation is due to the structure imposed by federal and state laws related to species conservation in the United States, but the basic issues are universal and driven by global economic considerations. They can be expected in any conservation situation in any area of the world.

In this context it is expected by society that the scientific method will allow inference to the important characteristics of ecological systems. The integrative properties of science are needed to filter available data, deduce the essential relationships to be explored, and obtain a basic understanding with a minimum of additional study. This is the social reality within which conservation biology must operate if it is to be effective. As Lubchenco notes (1998), the role of science is slowly shifting from a focus on the refinement of resource extraction methods to one of evaluating the health of the environment and determining what can be done to safeguard or improve environmental health. These problems are infinitely more complex, and require that all available tools and data be incorporated in order to achieve the most complete understanding. The evolving fields of GIS, spatial statistics and landscape ecology provide a framework to explore existing or readily-developed data sources, and the exploration of these methods is mandatory to develop the theory and practice of a modern ecology that attempts to address the conservation problems that we as a species have created.

This is not to say that direct inference through focused field research is not possible; in fact, the most valuable information yielded by the modeling process is often that which indicates what additional data should be gathered. The integration of data processing and analysis techniques with field data collected over large spatial and temporal scales, and the refinement of both using insights gained over time is the key to further progress in ecology This study was an attempt to explore some of the analysis tools that have become available for this type of ecological research.

SPATIAL MODELS AND PREDICTOR EFFECTIVENESS

The most striking differences between the spatial and non-spatial models is the way predictors that are highly significant in non-spatial models become insignificant when spatial correlation is modeled directly. There are three main reasons why this could occur:

1. The predictor is quantified at the wrong scale, and has too little information content to be effective.
2. The manner in which this predictor is correlated over space matches almost perfectly with prairie dog habitat occupancy, the spatial model accounts for it well, and it is unnecessary to model it directly.
3. The predictor becomes ineffective when spatial correlations in habitat occupancy are directly accounted for because the influence of marginal habitat is indirectly minimized.

It is possible that some predictors would be more significant if they were quantified at an appropriate spatial scale. This first point is illustrated in Figure 33, where the unit of observation, a 30 m square pixel, is represented by the two green squares. When variation in a predictor occurs within a different "domain of scale" than the one within which it is sampled, correlations that are strong can disappear or change their relationship with the response. In general, when the scale of observation is too coarse, mechanistic relationships and the patterns they produce are averaged out of the data (Wiens 1989). The slope of the terrain and recommended livestock stocking intensity data used in this study provide examples of two ways input data can be too coarsely quantified.

Slope has variation at the sub-pixel level that is not captured at the scale of observation. This interpretation is supported by the relatively high nugget variance in slope variograms from both study areas. Since input data were based upon a 30 m USGS DEM, spatial resolution is fundamentally limited to the 30 m spacing of sampling points used to create the DEM. The numbers in the first panel of Figure 33 represent estimated % slope within each quarter of the unit of a pixel. The average value of slope for this pixel is only 2.2%, even though the lower left quarter of the pixel has an extreme value of slope for this area (15%). If this area was unoccupied due to extreme slope values, it is highly unlikely that this relationship would be captured in the model, since the average slope of this pixel is only 5.5%. Shortridge (2001) examined the effects of resampling DEM raster data to coarser resolutions, and found that average and maximum slope values consistently decrease regardless of interpolation method. The same is true of unmeasured variation between sample points in the original data. When extreme values within 30 m pixels are averaged away during data collection, a previously strong relationship between habitat occupancy and slope may be obscured.

Some habitat data are only available at spatial scales much broader than the scale at which observations are quantified. Observed grazing intensity, for example, is defined by the number of cattle assigned to each pasture and the length of time they are allowed to remain. Since all pixels within a pasture share the same value for this variable, variation within the range of spatial covariance (represented by the circle around the pixel in Figure 33) is likely to be small. Unless grazing intensity between pastures has a very large effect on habitat occupancy, the predictor is unlikely to be significant. If this habitat quality attribute could be quantified at a finer spatial scale it might be more effective, but this would involve monitoring the activities of individual cattle or drastically reducing the size of pastures. Neither of these options were possible, and observed grazing intensity was not significantly associated with observed occupancy. NRCS recommended grazing intensity was also quantified at the pasture level in the Fort Pierre analysis and was insignificant, but when these data were replaced by the finer scale soil productivity data used to set these grazing levels, it had a strong relationhip with habitat occupancy.

Some predictors, on the other hand, might track the spatial pattern of model prediction error so closely that this variation is effectively captured by spatial correlation models. These predictors would become insignificant even though they could be biologically meaningful. The likelihood that global, isotropic spatial models would be capable of capturing spatial variability in a predictor to this extent is very low, but must be mentioned as a possibility.

Accounting for spatial correlations in the response variable resulted in the removal of several predictor variables considered highly significant in non-spatial models. Ver Hoef et al. (2001) demonstrate how non-spatial linear regression models result in too many variables declared significant by comparing linear spatial and non-spatial models estimated using simulated data sets with positive spatial autocorrelation. In a series of Monte Carlo simulations, the non-spatial model was allowed to have all the variables used to produce the dataset, while the spatial models were estimated with an incomplete set of predictors. Results of the reduced spatial models were consistently superior to the non-spatial "true" model, indicating that the spatial models also absorb the effects of unmodeled variables to give more valid and precise prediction and estimation. The authors followed this analysis with another that used actual field data, and found that mean-squared prediction errors were also much smaller for spatial models even though they used fewer parameters.

DIFFERENCES BETWEEN SPATIAL AND NON-SPATIAL MODEL OUTPUT

What are the specific consequences of failing to incorporate spatial correlations into the modeling process? It was thought that the effect of modeling occupied marginal cells adjacent to highly suitable, occupied cells without adjusting for SAC would be to inflate the suitability of similar unoccupied cells and result in too many areas being declared suitable (commission errors). Comparing spatial and non-spatial model 7, this appears to be true. When the full non-spatial and reduced spatial models are compared, however, the primary consequence is an increase in the rate of false negative predictions. When spatial relationships are not included in the models, and the model that would have been arrived at through non-spatial methods is used, the ability to correctly distinguish suitable areas from those that are unsuitable is reduced. These over-saturated non-spatial models also predict unoccupied areas as unsuitable with greater frequency, but without long-term occupancy data, the accuracy of this type of prediction cannot be verified.

Spatial models also allow weak or redundant predictors to be detected and removed. Comparing AIC values of the full non-spatial model to that of the non-spatial model with the same predictors as the spatial model (Table 14), AIC values of the reduced model are much greater. The AIC statistic is calculated using

(-2 * (log likelihood of the model | data)) - (2 * # parameters in the model)

Although the second term in the equation penalizes models with too many parameters, the difference in AIC values between all models was large enough to be unaffected by this subtraction; for large sample sizes, the parameter penalty has a negligible effect. The same ranking of models resulted when the first term in the above equation is the only one used. This means that even though the full model may have too many parameters, the reduced non-spatial model is superior to the full non-spatial model because has a better fit, regardless of the number of parameters. Recall that all terms in the full non-spatial model were highly significant, and since high p-values were the criteria used for eliminating variables, there would have been no way to arrive at the reduced model if only non-spatial methods were used. Even if one had no intention of using spatially adjusted regression for a final analysis, these methods produce a non-spatial model that is superior to the one arrived at through non-spatial methods.

FUNCTIONAL RELATIONSHIPS BETWEEN PREDICTORS AND HABITAT SUITABILITY

The predictors used to model habitat suitability vary in their functional relationship to prairie dog habitat occupancy, and this has implications for the utility of the final models. Ideally, the ecological gradients that define independent variables for a habitat suitability regression analysis should be quantified with very little error and have an explicit functional relationship with the response variable (Guisan and Zimmermann 2000). This is not always the case, of course. Predictor variables are measured along three types of ecological gradients: resource gradients, direct gradients, and indirect gradients (Austin 1980, Austin 1985, Austin et al. 1984, Austin and Smith 1989). Resource gradients are associated with the energy and matter consumed by an organism, direct gradients are environmental attributes that have physiological importance but are not consumed, and indirect gradients have no direct physiological relevance to an organism but indirectly represent a combination of resources and direct gradients (Guisan and Zimmermann 2000).

Indirect predictors are often quite powerful in that they incorporate multiple potential influences on a species' distribution at once, but they have the disadvantage that their relationship with these influences may be area-specific. In a different area with different variables associated with habitat selection, the same value of an indirect variable may be indicative of a different combination of direct and resource gradients. This is known as the "law of relative site constancy" (Walter and Walter 1953), and its implications are that models that are parameterized using predictors estimated from direct and resource gradients can be more easily generalized between areas. Unfortunately, indirect resource gradient data are often the most readily available for such analyses. If they are used to model a species' response to habitat variation, it is important that they represent the same direct and resource gradients between study areas if generality is desired.

Using the classification described above, the predictors used in the models may be categorized as follows:

Resource: AUMEAN, PROD, NDVI
Direct: ALL SOIL VARIABLES, PPT, WATER
Indirect: PRIVATE, SLOPE

The indirect predictors shown above may incorporate different combinations of direct and resource gradients when used in different areas. Although slope would be expected to have the same biological significance between sites, there is more variation in this predictor in the Conata Basin study area. Both areas are relatively flat, but Conata Basin has badlands rock formations that introduce more heterogeneity in slopes across space and reduce the range at which individual pixel values for slope are strongly correlated. Badlands formations and the areas around them, often barren non-habitat, are associated with changes in slope in Conata Basin and may help create stronger relationships with soil, vegetation, and other gradients than slope at Fort Pierre. Slopes were more significant as a predictor of occupancy in Conata Basin than they were at Fort Pierre, even though they were not significant enough to be included in a final model. If a greater geographic extent were modeled in the Conata Basin study area, badlands formations with even more variation in slope would be included, and the variable might be more significant.

The relationship between average annual precipitation and habitat occupancy illustrates a related problem with interpreting the biological significance of habitat suitability models. When trends in a predictor match trends in the response even though there may be little functional linkage between the two, false conclusions can be drawn that impair the generality and biological realism of results. Even when the models that result have better statistical performance, the loss of biological meaning is unacceptable and counterproductive to obtaining generality and insight.

Distance-based variables representing resource and indirect gradients performed poorly in terms of explaining variation in the response variable. Distance to water was a surrogate for cattle grazing intensity, and was based on the assumption that grazing intensity is higher near water sources. Pastures in both study areas were likely too small for such effects to be detected, and the relationship with observed occupancy was insignificant. Distance to private land, on the other hand, was highly significant but not useful for correctly predicting habitat occupancy. This predictor was significant because it increased predictions of unsuitable areas around the edges of the grassland (some of which were unoccupied), but it degraded overall performance by essentially removing all specificity in the center of the study area as well as increasing the false negative error rate. It is possible that distance-based predictors are a poor choice in general when using a spatial model so strongly based on distance. In addition to the potential for introducing redundant information, the interpretation of these predictors can be ambiguous since they can indirectly represent multiple environmental attributes.

All of the habitat attributes significant in the final spatial models measure gradients associated with the energy and matter consumed by prairie dogs (direct) or with environmental attributes that have physiological importance but are not consumed (resource). Predictors associated with direct gradients were most strongly related to habitat occupancy, and indicated that prairie dogs preferentially occupy areas of high soil productivity, and possibly increase soil productivity over time in occupied areas.

BIOLOGICAL SIGNIFICANCE OF SPATIAL MODEL OUTPUT

With the exception of distance to private land, most of the variables included in the final spatial models are related to soil or vegetation characteristics. This is logical for a species that lives inside burrows and subsists on the vegetation that surrounds them. The specific relationships between the various predictors and patterns of site occupancy provide additional information. Areas with lower mean clay content in the top soil horizon are associated with occupancy. This is likely a response to the fact that these areas are more fertile and produce preferred forage (grasses) in greater abundance. Similarly, estimated soil productivity is positively associated with occupancy, as is NDVI, an index of plant greenness and vigor. Although it is common wisdom that prairie dogs are attracted to areas that have poor fertility and further degrade the fertility of these areas over time, the results of this study demonstrate otherwise. Prairie dogs may be attracted to areas that are overgrazed and have low vegetation, but in the areas studied they preferentially occupy sites that are relatively fertile. There is evidence that soil fertility increases with prairie dog presence as well; vegetation growing in areas occupied by prairie dogs consistently has higher nitrogen per unit of above-ground net primary production (Whicker and Detling 1988). This makes it impossible to separate selection of fertile areas from fertility effects through continuing occupancy, and this caveat also applies to NDVI. The significance of estimated soil productivity, however, supports the idea that prairie dogs may have the ability to identify areas that are intrinsically fertile and occupy these areas preferentially.

High mean clay content of all soil horizons may be related to the climax vegetation community associated with different soil types, but characteristics of the soil as a burrow material are probably also important. Burrows in areas with too low of an overall clay content (sandy) have poor structural integrity, and are avoided, whereas areas with silty loam clay soils are thought to have the best burrow durability (Clippinger 1989). This predictor had a positive relationship with habitat occupancy, and though it was only marginally significant in the final spatial model (p = 0.1734), its contribution to decreasing AIC justified keeping it in the model. The Conata Basin study area is dissected with badlands rock formations and the barren soils that accompany them. These soils do not support much vegetation, are likely hard to excavate, and have a lower clay content than the more productive areas that surround them. Given this context, it is not surprising that this variable was significant. It is uncertain whether it would be useful in other areas, however.

The significance of distance to private land as a predictor of habitat suitability was most likely due to the poisoning that has occurred on the grasslands in the past. These control actions occurred primarily in response to landowner complaints, and examination of records shows that most poisoning occurred adjacent to boundaries with private land. Increased shooting pressure may be associated with proximity to private land in some areas, but shooting activity would need to be fairly intense to have a significant effect on colony area. It is also possible that proximity to non-habitat, privately-owned areas where prairie dogs are actively exterminated, has negative effects on colonies because they are not a source of immigrating individuals and result in mortality for those who disperse into it. This prevents the immigration and emigration between colonies thought to be necessary to maintain genetic variability and enhance metapopulation persistence of the species (Gilpin, 1999). These metapopulation effects may be significant, but are impossible to evaluate without extensive study, and do not appear to be supported by the observed pattern of habitat occupancy.

Since poisoning and shooting are no longer permitted in either study area, distances from private land are now of questionable relevance to estimates of habitat suitability. Long term data are not sufficient to understand the extent of historical influences on current occupancy patterns. It is not known, for example, how long it would take for a large area to be reoccupied if all prairie dogs were completely removed. When occupied areas still exist nearby, prairie dog populations rebound within a few years after poisoning. In Conata Basin, for example, colonies require treatment every three years to stop expansion (Schenbeck 1985, Uresk and Schenbeck 1987). When colonies are only partially treated, population recovery can occur within two years (Knowles 1982). It is reasonable to assume that they respond even more quickly after shooting activities stop, since shooting has the effect of creating small vacancies that are quickly filled rather than removing prairie dogs from large areas (Gilpin 1999). Even though past control activities have had some influence on current distribution, this relationship would not be expected to continue for long periods after control is discontinued.

The effect of removing the distance to private land predictor from the models was to increase predictions of unsuitable habitat within the study area, decrease error in predictions of suitable habitat at the margins of the study area, and produce more general models not tied to the specific history of the area modeled. All model performance statistics increased with the removal of this predictor, and comparisons of the spatial predictions of models that included distance to private land with those that did not indicated that a slight improvement in specificity at the edges of the study area (which may or may not be accurate) was the only positive effect on prediction.

The spatial component of the models also has biological implications. The criterion for defining a "complex" of prairie dog towns in most current management plans is that colonies should be separated by straight line distances of no more than 5 km to be included in a complex. This figure is loosely based upon the estimated maximum dispersal ability of the species. Radiotelemetry of dispersing prairie dogs, however, has given an average straight-line dispersal distance of only 2.4 km (Garrett and Franklin 1988). Observed mortality of dispersing animals in this study was 56%, significantly higher than that of prairie dogs that did not disperse (10%). Dispersers took a meandering route, usually following drainages and swales. Genetic research (Roach, 1999) has shown that the genetic distance between colonies is related to distance along these dispersal corridors, but not to straight line distance between colonies. Given that drainages are important for dispersal success and population interchange, straight line distance between colonies is not adequate as a measure of connectivity between colonies. Creating distance matrices that quantify distances between sites on the basis of drainage distance is not a trivial task, however, and must be done for each area an analysis is applied to. The estimated range of spatial correlation was consistent around 725 m for all spatial analyses and between data sets, and is straightforward to quantify using variogram analysis (refining estimates requires Mixed/Glimmix analysis). This measure of correlation could be used to provide a more quantitative and site specific measure of connectivity than 5 km rule currently used, which may not be adequate to ensure species persistence.

CONCLUSION

This research describes an analysis framework that allows reliable quantification of relationships between spatially correlated occupancy data and a set of habitat covariates. Results show that it is possible to predict prairie dog habitat occupancy more precisely and accurately than before using a set of four predictors from two commonly available data sources, a time series of habitat occupancy data, and spatially adjusted regression analysis. The analyses in this study used data similar to that in previous modeling work, but quantified observed habitat use more precisely by modeling spatially correlated prediction error to remove it from the estimation process for a main effects model.

Although this type of analysis is computationally intensive, it is conceptually simple and based on refinements of standard statistical theory. The analytical techniques used here have existed for years, but the software necessary to implement them and the hardware that makes it practical at the landscape scale have only become available recently. The models explored in this thesis describe spatial variation in a very rudimentary way, but represent one of the only analysis frameworks that is currently feasible at the landscape level.

One of the most exciting possibilities of this model structure is the potential to extend the analysis to the state or regional level by randomly sampling spatial blocks from a regional population of blocks. Spatial blocks in this study formed a contiguous mass of habitat, but since they are used as repeated measures to build the model, spatial blocks are assumed to be statistically independent. If selected randomly from a larger area, blocks could be used to create general models to define habitat quality/habitat occupancy relationships that could be applied to extensive areas. Unlike other spatial models (e.g. spatial autoregressive and autologistic models), a valid spatially adjusted regression model may be applied to an unmodeled area without the need to examine or analyze location-specific data. Application of spatially adjusted models is as straightforward as the application of non-spatial logistic models because the adjustment is already built into the equation.

The analyses used commonly available statistical software and readily available habitat data. Similar data exist to model many other species, and the analysis is relatively flexible in terms of incorporating different data types (presence/absence or count data) and different types of models for positive spatial autocorrelation (spherical, exponential, gaussian, linear, linear log, and power). These tools are available for use in further analyses of prairie dog/habitat relationships as well as those of other species.

A major advantage of these models is the ability to exclude predictors that have spurious significance in non-spatial analyses, yet become insignificant when spatial correlations in the response variable are accounted for. In the above analysis, over-parameterized, non-spatial models predicted that too many occupied areas were unsuitable. Spatial models with fewer parameters had a false negative error rate about one third that of the non-spatial full model, yet had similar statistics for correct predictions. Incorporation of spatial correlations thus allowed models to be simplified and the most important (and only verifiable) source of error to be reduced.

Using a non-spatial model with the same parameters as the spatial model gave accuracy statistics that are comparable with the spatial model, but with a far lower level of specificity than either of the other two models. In other words, using the same predictors without the spatial adjustment results in the same proportion of correct and incorrect predictions, but since a smaller area is being declared unsuitable, these standards are easier to meet, and the model is unable to remove many areas from consideration. This is the result anticipated in the introduction: a non-spatial model that is based upon the equal contribution of marginal and optimal quality habitat would be expected to have a reduced ability to idenfity unsuitable areas compared to a spatial model.

Using readily available data, the spatial models were able to quantify relationships between habitat quality and habitat occupancy that could not have been detected using other methods. The components necessary for application of this framework to a range-wide habitat suitability analysis of the Black-tailed prairie dog already exist, as they do for many other species. Ecology and science in general are at a stage where technology and theory have provided a realm of possibilities that remain to be investigated, and many tools that could be used to understand ecological relationships have not been tested using field data. Spatially adjusted regression models are among these tools, and represent a way to filter and simplify a complex reality, define essential relationships, and guide further research into species habitat relationships for prairie dogs and many other animals.

LITERATURE CITED

Agnew, W. 1983. Flora and fauna associated with prairie dog ecosystems. M.S. Thesis. Colorado State University, Fort Collins. 47 pp.

Agnew, W., D. W. Uresk, and R. M. Hansen. 1986. Flora and fauna associated with prairie dog colonies and adjacent ungrazed mixed grass prairie in western South Dakota. Journal of Range Management 39:135-139.

Andrew, M. H. 1988. Grazing impacts in relation to livestock watering points. Trends in ecology and evolution 3:336-339.

Augustin, N.H., M.A. Mugglestone, and S.T. Buckland. 1996. An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology 33:339-347.

Bonham, C.D., and J. Hannan. 1978. Blue grama and buffalograss patterns in and near a prairie dog town. Journal of Range Management 31:63-65.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. Classification and Regression Trees. Belmont, CA: Wadsworth. 358p.

Burnham, K.P. and D.R. Anderson. 1998. Model Selection and Inference: A Practical Information-Theoretic Approach. New York: Springer Verlag. 353 pp.

Cincotta, R. P. 1985. Habitat and dispersal of black-tailed prairie dogs in Badlands National Park. Ph.D. Dissertation, Colorado State University, Fort Collins. 52 pp.

Cincotta, R.P., D.W. Uresk, and R. M. Hansen. 1988. A statistical model of expansion in a colony of black-tailed prairie dogs. Pages 30-33 *in* Proceedings of the Eighth Great Plains Wildlife Damage Control Workshop. April 28-30, 1987. Rapid City, S.D. U.S. Dept. of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station General Technical Report RM-154.

Clippinger, N.W. 1989. Habitat suitability index models : Black-tailed prairie dog. U.S. Fish and Wildlife Service. Biological Report 82(10.156) 21 pp.

Coppock, D.L., J.K. Detling, and M.I. Dyer. 1980. Interactions among bison, prairie dogs, and vegetation in Wind Cave National Park. Final report to the National Park Service, Wind Cave National Park, South Dakota. 195 pp.

Coppock, D.L. 1981. Impacts of Black-tailed prairie dogs on vegetation in Wind Cave National Park. M.S. Thesis. Colorado State University, Fort Collins. 86 pp.

Detling, J.K. 1998. Mammalian herbivores: ecosystem level effects in two grassland national parks. Wildlife Society bulletin 26:438-448.

Doesken, N. J., and J. Arthur. 1996. The snow booklet: a guide to the science, climatology, and measurement of snow in the United States. Fort Collins, Colo., Colorado State University, Dept. of Atmospheric Science, Colorado Climate Center. 84 pp.

Environmental Systems Research Institute, 1998. Arc Info version 7.2.1. Redlands, CA.

Environmental Systems Research Institute, 1999. ArcView version 3.1. Redlands, CA

Fagerstone, K.A., H.P. Tietjen, and G.K. Lavoie. 1977. Effects of range treatment with 2, 4-D on prairie dog diet. Journal of Range Management 30:57-60.

Fagerstone, K.A., H.P. Tietjen, and O. Williams. 1981. Seasonal variation in the diet of black-tailed prairie dogs. Journal of Mammalogy 62:820-824.

Gilpin, M. 1999. An inquiry into the population viability of the prairie dog. Available: http://gemini.msu.montana.edu/~mgilpin/prairie_dog.html.

Gold, K.I. 1976. Effects of blacktailed prairie dog mounds on shortgrass vegetation. M.S. Thesis. Colorado State University, Fort Collins. 40 pp.

Gotway, C.A., and Stroup, W.W. 1997. A generalized linear model approach to spatial data analysis and prediction. Journal of Agricultural, Biological, and Environmental Statistics 2:157-178.

Gotway, C.A., and Wolfinger, R.D. 2001. Spatial prediction of counts and rates. *In press.*

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

Hanski, I. 1994. A practical model of metapopulation dynamics. Journal of Animal Ecology 63:151-162.

Hanski, I., A. Moilanen, T. Pakkala, and M. Kuussaari. 1996. The quantitative IF model and persistence of an endangered butterfly population. Conservation Biology 10:578-590.

Heuvelink, G.B.M. 1998. Error Propagation in Environmental Modeling with GIS. London: Taylor and Francis. 127 pp.

High Plains Regional Climate Center. 2001. Historical Data Summaries. Available: http://hpccsun.unl.edu/products.html

Hoogland, J.L. 1981. The evolution of coloniality in white-tailed and black-tailed prairie dogs (Sciuridae: Cynomys luecurus and C. ludovicianus). Ecology 62:252-272.

Hoogland 1995. The black-tailed prairie dog: social life of a burrowing mammal. Chicago : University of Chicago Press.

Hosmer, D.W., and S. Lemeshow. 2000. Applied logistic regression. 2nd edition. New York: John Wiley and Sons, Inc.

Huffer, F.W., and Wu, H. 1998. Markov chain monte carlo for autologistic regression models with application to the distribution of plant species. Biometrics 54:509-524.

Jones, C. G., Lawton, J. H., and Shachak, M. 1994. Organisms as ecosystem engineers. Oikos 69:373-386.

Journel, A.G., and C.J. Huijbregts. 1978. Mining Geostatistics. Academic Press, London. 600 pp.

King, J.A. 1955. Social behavior, social organization, and population dynamics in a black-tailed prairie dog town in the Black Hills of South Dakota. Contributions from the Laboratory of Vertebrate Biology No. 67. University of Michigan, Ann Arbor. 123 pp.

Knowles, C.J. 1982. Habitat affinity, populations, and control of black-tailed prairie dogs on the Charles M. Russell National Wildlife Refuge. PhD Dissertation, University of Montana, Missoula. 171 pp.

Koford, C.B. 1958. Prairie dogs, whitefaces, and blue grama. Wildlife Monographs 3. 78 pp.

Kotliar, N. B., B.W. Baker, A.D. Whicker, and G. Plumb. 1999. A critical review of assumptions about the prairie dog as a keystone species. Environmental Management 24:177-192.

Langer, T.J. 1998. Black-tailed prairie dogs as indicators of human-caused changes in landscape at Badlands National Park, South Dakota. M.S. Thesis, North Carolina State University, Raleigh. 94 pp.

Lerwick, A.C. 1974. The effects of the black-tailed prairie dog on vegetative composition and their diet in relation to cattle. M.S. Thesis. Colorado State University, Fort Collins. 106 pp.

Liang, K.Y. and Zeger, S.L. 1986. Longitudinal data analysis using generalized linear models. Biometrika 73:13-22.

Littell, R.C., G.A. Milliken, W.W. Stroup, and R.D. Wolfinger. 1996. SAS System for Mixed Models. Cary, N.C.: SAS Institute Inc. 633 pp.

Lubchenco, J. 1998. Entering the century of the environment: a new social contract for science. Science 279:491-497.

MacArthur, R.H. 1972. Geographical Ecology: Patterns in the Distribution of Species. New York: Harper and Row. 269 pp.

Matheron, G. 1963. Principles of geostatistics. Economic Geology 58:1246-1266.

McCullagh, P., and J.A. Nelder. 1989. Generalized Linear Models, Second Edition. New York: Chapman and Hall. 261 pp.

Merriam, C.H. 1901. The prairie dog of the Great Plains. U.S. Department of Agriculture Yearbook. 1901:257-270.

Morgan, J.N. and R.C. Messenger. 1973. THAID: A Sequential Search Program for the Analysis of Nominal Scale-dependent Variables. Ann Arbor: Institute for Social Research, University of Michigan.

Neter, J., W. Wasserman, and M.H. Kutner. 1989. Applied Linear Regression Models, Second Edition. Homewood, IL: Irwin. 667p.

Noon, B.R. and K.S. McKelvey. 1992. Stability properties of the spotted owl metapopulation in Southern California. Pages 187-206 *in* The California Spotted Owl: A Technical Assessment of its Current Status. U.S. Forest Service, Pacific Southwest Research Station General Technical Report GTR-133. Berkeley, CA. 285 pp.

O'Meilia, M.E., F.L. Knopf, and J.C. Lewis. 1982. Some consequences of competition between prairie dogs and beef cattle. Journal of Range Management 35:580-585.

Osborn, B., and P.F. Allan. 1949. Vegetation of an abandoned prairie dog town in tall-grass prairie. Ecology 30:322-332.

Owens, M. K., K. L. Launchbaugh, and J. W. Holloway. 1991. Pasture characteristics affecting spatial distribution of utilization by cattle in mixed brush communities. Journal of Range Management 44:118-123.

Preisler, H. K., and R. G. Mitchell. 1993. Colonization patterns of the mountain pine beetle in thinned and unthinned lodgepole stands. Forest Science 39:528-545.

Proctor, J. 1998. A GIS model for identifying potential black-tailed prairie dog habitat in the northern Great Plains shortgrass prairie. M.S. Thesis. University of Montana, Missoula. 56 pp.

Roach, J.L. 1999. Genetic analysis of a black-tailed prairie dog (*Cynomys ludovicianus*) metapopulation in shortgrass steppe. M.S. Thesis. Colorado State University, Fort Collins. 90 pp.

Roughgarden, J. 1983. Competition and theory in community ecology. American Naturalist 122:583-601.

SAS Institute. 1999. The SAS System, version 8.1. Cary, N.C.

Sjogren-Gulve, P., and C. Ray. 1996. Using logistic regression to model metapopulation dynamics: large-scale forestry extirpates the pool frog. Pages 111-138 *in* D.R. McCullough, ed. Metapopulations and Wildlife Conservation and Management. Island Press, New York.

Snell, G.P., and B.D. Hlavachick. 1980. Control of prairie dogs--the easy way. Rangelands 2:239-240.

Stapp, P. 1998. A reevaluation of the role of prairie dogs in great plains grasslands. Conservation Biology 12:1253-1259.

Summers, C.A., and R.L. Linder. 1978. Food habits of the black-tailed prairie dog in western South Dakota. Journal of Range Management 31:134-136.

Uresk, D.W., J.G. MacCracken, and A. J. Bjugstad. 1982. Prairie dog density and cattle grazing relationships. Pages 199-201 *in* Proceedings, 5th Great Plains Wildlife Damage Control Workshop, October 13-15, 1981. Nebraska Center for Continuing Education, Lincoln.

Uresk, D.W., and A.J. Bjugstad. 1983. Prairie dogs as ecosystem regulators on the northern high plains. Pages 57-67 *in* Proceedings, 7th North American Prairie Conference, August 4-6, 1980. Southwest Missouri State University, Springfield.

Uresk, D.W. 1984. Black-tailed prairie dog food habits and forage relationships in western South Dakota. Journal of Range Management 37:325-329.

Uresk, D.W., and G.L. Schenbeck. 1987. Effect of zinc phosphide rodenticide on prairie dog colony expansion as determined from aerial photography. Prairie Naturalist 19:57-61.

U. S. Department of Agriculture, Forest Service. 1968. Range and Environmental Analysis Handbook. Forest Service Handbook 2209.21. 244 pp.

U. S. Department of Agriculture, Natural Resources Conservation Service. 1997. National Map Unit Interpretation Records (MUIR) Database. Available: http://www.statlab.iastate.edu:80/soils/nssc/.

U. S. Department of Agriculture, Natural Resources Conservation Service. 1998. Keys to soil Taxonomy. Available: http://www.statlab.iastate.edu:80/soils/nssc/

U.S. Department of the Interior, U.S. Fish and Wildlife Service, 1999. Black-tailed prairie dog fact sheet. Available: http://www.r6.fws.gov/btprairiedog/facts.htm

U.S. Department of the Interior, U.S. Fish and Wildlife Service, 2000. Questions and answers: black-tailed prairie dog (12-month status review). Available: http://www.r6.fws.gov/btprairiedog/qanda.htm

U.S.Department of the Interior, U.S. Geological Survey, Digital Elevation Models -- Data Users Guide; Reston, VA. Available: ftp://mapping.usgs.gov/pub/ti/DEM

U.S. Department of the Interior, U.S. Geological Survey, 1992, Standards for digital elevation models: Reston, VA. Available: ftp://mapping.usgs.gov/pub/ti/DEM

Ver Hoef, J.M., N. Cressie, R.N. Fisher, and T.J. Case. 2001. Uncertainty and spatial linear models for ecological data. Pages 214-237 in Spatial Uncertainty in Ecology. Springer, New York.

Wel, J. 1975. Least squares fitting of an elephant. Chemtech Feb. 128-129.

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

Whicker, A.D., and J.K. Detling. 1988. Ecological consequences of prairie dog disturbances. BioScience 38:778-785.

Wolfinger, R., and M. O.Connell. 1993. Generalized linear mixed models: a pseudo-likelihood approach. J. Statist. Comput. Simul., 48: 233-243.

Table 1. Predictors used in the USFS Prairie Dog Habitat Capability Model

	Preferred	**Suitable**	**Unsuitable**
Range site	moderately clayey silt loams	dense clays, sandy and silty	badlands, very sandy soils
Vegetation	grassland	sagebrush and greasewood	coniferous and deciduous
Grazing	primary rangeland (water developments)	secondary rangeland (no water developments)	ungrazed
Slope	< 10%	10% - 30%	over 30%
Water/wetland	upland	upland	water/wetland

Table 2. Comparison of the Proctor et al. model predictions with observed data

Category	% of area w/ P.D.	% of total P.D.	% total area
1. Slope 4-25%, all veg. Types except low density grassland and barren sites	0.8	14.8	82.2
2. Slope 0-4%, veg. as above, soil texture clay and loam	10.7	25.4	10.9
3. Slope 0-4%, veg. as above, soil texture clay/loam	26.6	21.7	3.7
4. Slope 4-25%, low density grasslands and barren sites	10.2	1.7	0.7
5. Slope 0-4%, low density grasslands and barren sites	67.9	36.5	2.5
Total	100	100	100

Table 3. Candidate predictors used in initial Fort Pierre models

Predictor	Data Source & Processing	Potential Biological Rationale
Slope	USGS DEM converted to slope grid using Arc Tools	Small slopes selected by species
Mean %clay of top soil horizon in map unit	Map units in USFS soil survey layer linked with NRCS SSURGO soil database	top horizons with low clay content have lower veg density, are less prone to erosion, and may have superior structural char.
Mean % clay of all soil horizons in map unit		soils with lower mean clay content in all horizons may have superior structural char.
High mean % clay of all soil layers in map unit		if the highest mean clay content in a mapunit is above some threshold, it is unsuitable
Low mean % clay of all soil layers in map unit		if the lowest mean clay content is not above some threshold, it is unsuitable
Average annual precipitation	Data from surrounding weather stations used to interpolate values across grassland	More arid areas are selected by the species due to their decreased productivity
Distance to private land	Land Ownership layer & ArcInfo AML	Areas closer to private land are more likely to be poisoned or shot at.
Distance to water within a pasture	Water source coverages & Arc Info AML	Outside of some trampling radius, increased grazing pressure improves HS within some distance of water
Proportion of years poisoned	Paper maps and tables summarizing poisoning efforts transcribed and appended to occupancy coverages.	Poisoning events that were known to occur likely affected occupancy probability for all pixels within some range of the poisoning
NRCS stocking level ÷ pasture area	Tabular data transcribed and appended to allotment coverages	species favors less productive areas
Actual use stocking level mean	Tabular data transcribed and appended to pastures in allotment coverages	Species prefers areas that are intensely grazed
Actual use stocking level S.D.	Tabular data transcribed and appended to pastures in allotment coverages	Species prefers areas with high stocking level variance.
Mean %clayof top horizon * Slope	Manipulation of data done within SAS	Clay accumulates in closed depressions and other areas of low slope
Mean % clay of all horizons * Slope	Manipulation of data done within SAS	Clay accumulates in closed depressions and other areas of low slope
High mean % clay of all horizons * Slope	Manipulation of data done within SAS	Clay accumulates in closed depressions and other areas of low slope
Proportion of years poisoned * Distance to private land	Manipulation of data done within SAS	Poisoning is more likely for towns adjacent to private land.
NRCS stocking level * Actual use stocking level	Manipulation of data done within SAS	Actual use of grazing lands is influenced by the recommendations of the NRCS

Table 4. Description of the Glimmix estimation process

1)
    a) Estimate main effects model using entire dataset.
    b) Make predictions based on the initial main effects model.
    c) Use user-defined parameters to remove spatially correlated errors from the error surface, leaving model-based error, error due to unmodeled variation, and spatially generated error not accounted for by the initial spatial model.
    d) Transfer "pseudodata' from 1c to next iteration.
2)
    a) Estimate main effects model using same data as in 1a, but minimize error based. on error surface from 1c.
    b) Make predictions based on the revised main effects model.
    c) Use the errors from 2b to estimate a new spatial correlation model.
    d) Remove spatially correlated errors from error surface using the revised spatial model, leaving model-based error, error due to unmodeled variation, and spatially generated error not accounted for by the initial spatial model.
    e) Transfer pseudodata from 2d to the next iteration.
3) Repeat process until convergence.

Table 5. Output from initial Fort Pierre multiple linear regression analysis

Root MSE	0.21741	R-Square	0.5492
Dependent Mean	0.24469	Adj R-Sq	0.5490
Coeff Var	88.84917		

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
INTERCEPT	1	-2.98899	0.32158	-9.29	<.0001
AUMN	1	0.11707	0.03441	3.40	0.0007
AUS	1	0.76401	0.12774	5.98	<.0001
CLAYMEAN	1	-10.04077	0.48388	-20.75	<.0001
TOPCLAYM	1	-0.23523	0.02750	-8.55	<.0001
CLAYMH	1	3.02436	0.24073	12.56	<.0001
CLAYML	1	6.70459	0.27938	24.00	<.0001
PPOISON	1	4.17689	0.05662	73.77	<.0001
PRECIP	1	0.09206	0.00267	34.45	<.0001
NRCSM	1	0.18810	0.03976	4.73	<.0001
PRIVD	1	0.03031	0.00256	11.84	<.0001
SLOPES	1	-7.13047	0.43239	-16.49	<.0001
WATD	1	-0.06203	0.00241	-25.79	<.0001
SLOPE*TCM	1	-4.41916	0.50179	-8.81	<.0001
SLOPE*CMH	1	30.38309	1.75466	17.32	<.0001
SLOPE*CML	1	-19.10414	1.49495	-12.78	<.0001
PPOIS*PRIVD	1	-0.59867	0.07395	-8.10	<.0001
NRCS*AUMN	1	-0.18436	0.03194	-5.77	<.0001

Table 6. Comparison of initial non-spatial logistic regression models

MODEL WITH SIGNIFICANT INTERACTIONS

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-12.99	0.7689	285.4368	<.0001
AUMN	1	-4.81	0.4740	102.9890	<.0001
CLAYMEAN	1	-157.90	6.6001	572.4070	<.0001
TOPCLAYM	1	-7.22	0.3530	418.2051	<.0001
CLAYMH	1	63.73	3.5143	328.8566	<.0001
CLAYML	1	95.08	3.5260	727.0520	<.0001
PRECIP	1	1.38	0.0291	2251.7791	<.0001
NRCSM	1	-3.67	0.5568	43.3287	<.0001
PRIVD	1	0.34	0.0295	131.0657	<.0001
SLOPES	1	-46.26	5.4609	71.7494	<.0001
CLAYMEAN*SLOPES	1	-201.80	41.6861	23.4424	<.0001
TOPCLAYM*SLOPES	1	-41.47	6.2765	43.6482	<.0001
CLAYMH*SLOPES	1	280.00	43.1774	42.0481	<.0001
AUMN*NRCSM	1	1.78	0.4473	15.9032	<.0001

Association of Predicted Probabilities and Observed Responses

Percent Concordant	86.5	Somers' D	0.731
Percent Discordant	13.4	Gamma	0.732
Percent Tied	0.1	Tau-a	0.364
Pairs	384843536	c	0.865

MAIN EFFECTS MODEL

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-17.22	0.5731	903.3060	<.0001
AUMN	1	-2.96	0.0983	904.3561	<.0001
CLAYMEAN	1	-167.50	6.5763	648.7117	<.0001
TOPCLAYM	1	-9.74	0.2199	1964.7128	<.0001
CLAYMH	1	77.91	3.1935	595.2484	<.0001
CLAYML	1	95.00	3.5771	705.2745	<.0001
PRECIP	1	1.37	0.0281	2379.0446	<.0001
NRCSM	1	-1.40	0.0585	569.7598	<.0001
PRIVD	1	0.39	0.0298	175.5437	<.0001
SLOPES	1	-9.53	0.3683	670.0004	<.0001
WATD	1	0.08	0.0289	7.8803	0.0050

Association of Predicted Probabilities and Observed Responses

Percent Concordant	86.6	Somers' D	0.734
Percent Discordant	13.2	Gamma	0.735
Percent Tied	0.1	Tau-a	0.365
Pairs	89389077	c	0.867

Table 7. Model fit diagnostics for main effects non-spatial logistic model

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	18750.8948	10	<.0001
Score	15778.2372	10	<.0001
Wald	10880.0484	10	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
AUMN	0.052	0.043	0.063
CLAYMEAN	<0.001	<0.001	<0.001
TOPCLAYM	<0.001	<0.001	<0.001
CLAYMH	>999.999	>999.999	>999.999
CLAYML	>999.999	>999.999	>999.999
PRECIP	3.939	3.728	4.163
NRCSM	0.247	0.221	0.277
PRIVD	1.484	1.400	1.573
SLOPES	<0.001	<0.001	<0.001
WATD	1.084	1.025	1.148

Table 8. Classification Table for main effects non-spatial logistic model

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.000	18403	0	21135	0	46.5	100.0	0.0	53.5	.
0.100	18119	4726	16409	284	57.8	98.5	22.4	47.5	5.7
0.200	17101	10107	11028	1302	68.8	92.9	47.8	39.2	11.4
0.300	15985	14159	6976	2418	76.2	86.9	67.0	30.4	14.6
0.400	15045	16595	4540	3358	80.0	81.8	78.5	23.2	16.8
0.440	14529	17216	3919	3874	80.3	78.9	81.5	21.2	18.4
0.500	13577	18019	3116	4826	79.9	73.8	85.3	18.7	21.1
0.600	12392	18832	2303	6011	79.0	67.3	89.1	15.7	24.2
0.700	11073	19366	1769	7330	77.0	60.2	91.6	13.8	27.5
0.800	9391	19919	1216	9012	74.1	51.0	94.2	11.5	31.1
0.900	4500	20653	482	13903	63.6	24.5	97.7	9.7	40.2
1.000	0	21135	0	18403	53.5	0.0	100.0	.	46.5

Table 9. Comparison of AIC values for final Fort Pierre Models

Model	# Parm.	AIC	wr
Non-spatial full	11	3142004	0
Non-spatial reduced	4	2874122	0
Spatial reduced	6	2037564	1

Table 10. Comparison of accuracy assessment statistics for Fort Pierre final models

Model	Cutpoint	Sensitivity	Specificity	False Pos.	False Neg.
Non-spatial full	0.300	95.5	27.9	44.5	13.3
Non-spatial reduced	0.280	95.0	29.4	44.1	13.8
Spatial reduced	0.399	95.10	15.80	84.20	4.90

Table 11. Candidate predictors used in initial Conata Basin spatial models

Predictor	Non-spat.	Spat. 1	Spat. 2	Spat. 3	Spat. 4
Actual use stocking mean	x				
Low mean clay of all horizons	x				
Average mean clay of all horizons	x				
High mean clay of all horizons	x	x	x	x	
Low mean clay of top soil horizon	x				
Average mean clay of top soil horizon	x	x	x	x	x
High mean clay of top soil horizon	x				
Estimated average annual precipitation	x	x	x		
Estimated productivity/mapunit area	x	x	x	x	x
Distance to private land	x	x	x	x	x
Slope	x				
Distance to water in pasture	x				
Estimated % sand of top soil horizon	x				
Estimated % silt of top soil horizon	x	x			
Estimated % clay of top soil horizon	x	x			
Mean NDVI	x	x	x	x	x

Table 12. AIC statistics for initial Conata Basin spatial models

	# parms	900 meter spatial block		1200 meter spatial block	
		AIC	w_r	AIC	w_r
Model 1	11	404992	0.0000000000	405336	0.0000000000
Model 2	9	404758	0.0000000000	405120	0.0000000000
Model 3	8	399255	0.9999999995	--	--
Model 4	7	399298	0.0000000004	--	--

Table 13. Comparison of full and reduced non-spatial models with spatial model 3

	# parms	AIC	w_r
Full non-spatial model	13	834962	0.0000000000
Non-spatial model 3	6	821465	0.0000000000
Spatial model 3	8	399255	0.9999999999

Table 14. Maximum likelihood estimates for spatial model 3

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-1.4460	0.1697	276	-8.52	<.0001
MCH	0.2075	0.1453	19E4	1.43	0.1532
TCM	-0.3160	0.1480	19E4	-2.14	0.0327
PROD	0.1384	0.0213	19E4	6.49	<.0001
PRIVATE	0.2339	0.0669	19E4	3.50	0.0005
NDVI1	0.7491	0.0998	19E4	7.51	<.0001

Table 15. Comparison of accuracy assessment statistics for final models

	Sensitivity	Specificity	False +	False -
Full non-spatial model	95.0	29.4	44.1	13.8
Non-spatial model 7	95.6	21.8	78.2	4.4
Spatial model 7	95.2	25.2	74.8	4.8

Table 16. Regression coefficients and standard errors of the final spatial model

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-1.1032	0.1388	276	-7.95	<.0001
MCH	0.1954	0.1435	19E4	1.36	0.1734
TCM	-0.3059	0.1461	19E4	-2.09	0.0363
PROD	0.1355	0.02106	19E4	6.44	<.0001
NDVI1	0.7350	0.09862	19E4	7.45	<.0001

Figure 1. Clippinger's Habitat Suitability Index Functions

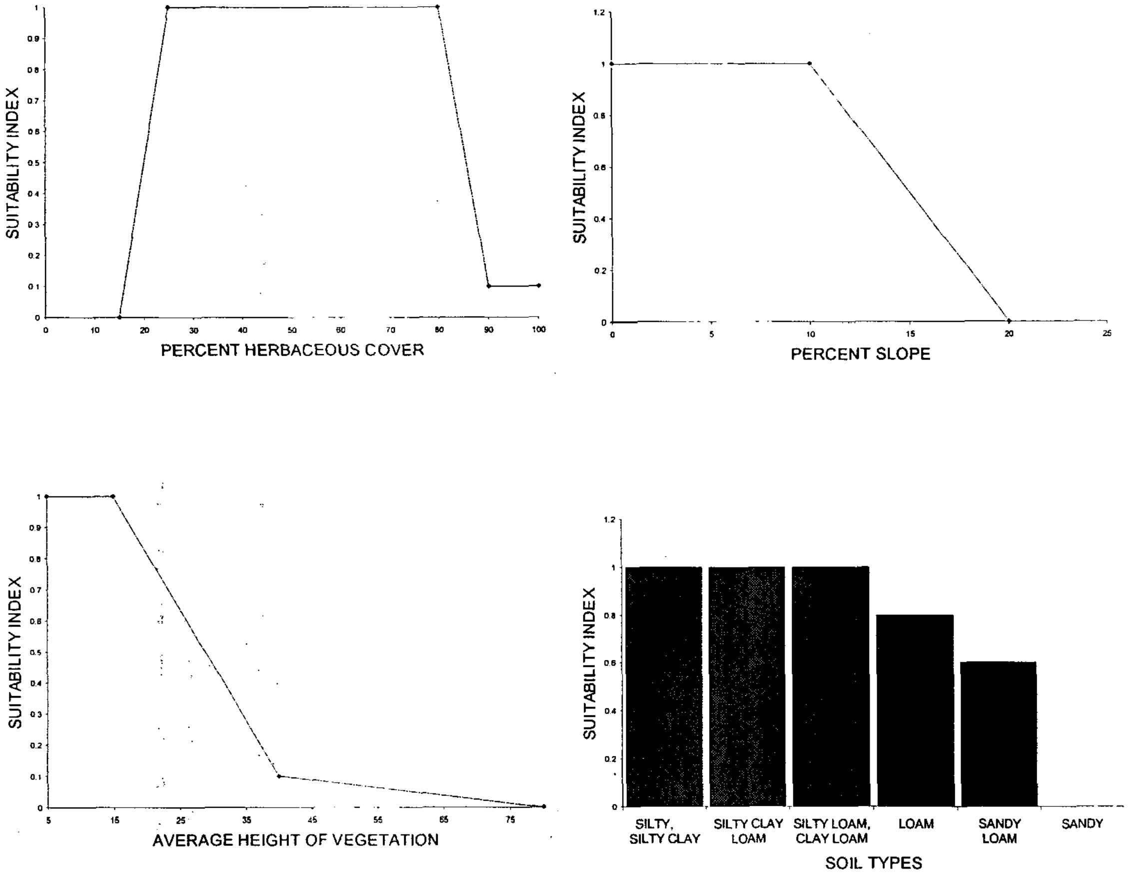


Figure 2. Comparison of Habitat Capability Model output and observed occupancy for Fort Pierre National Grassland

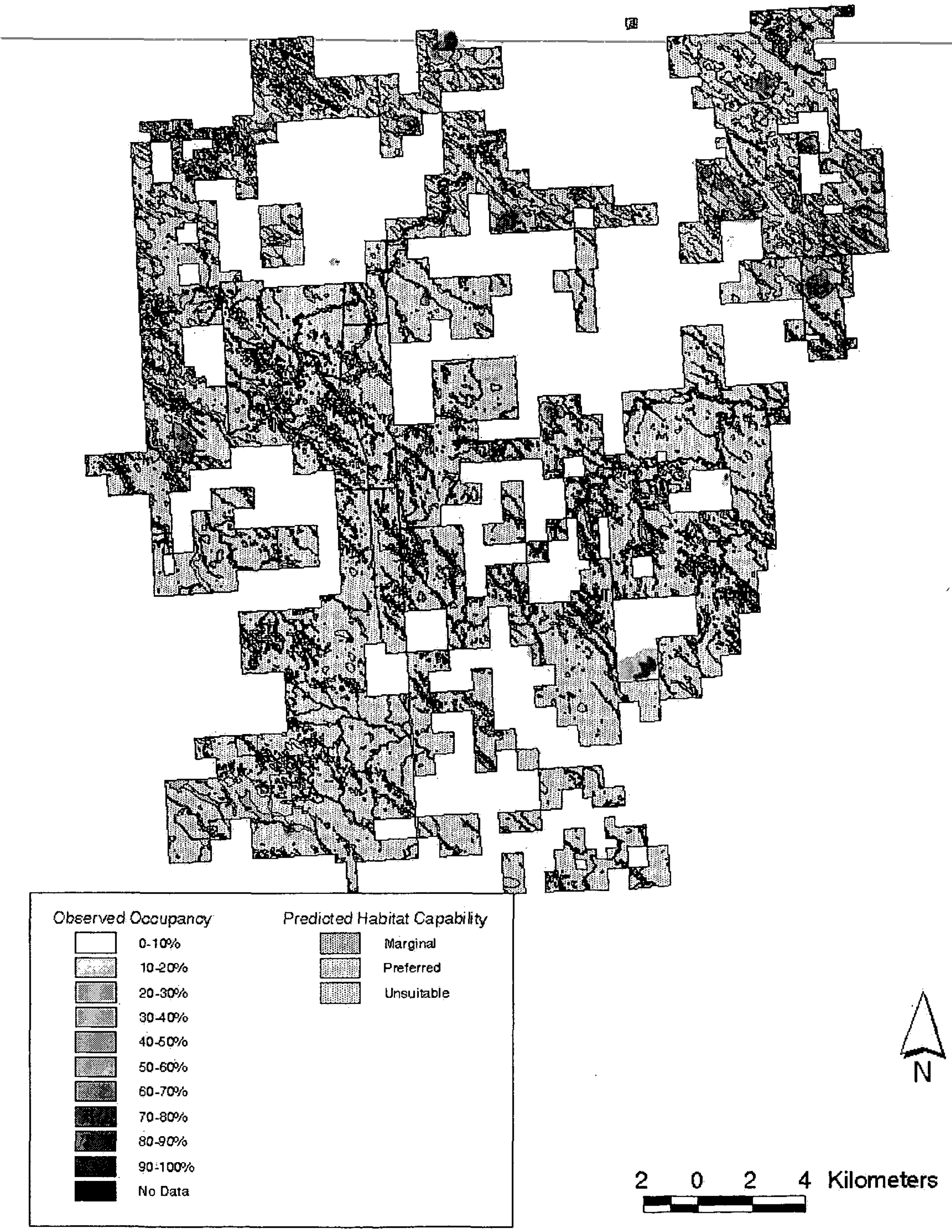


Figure 3. Comparison of Habitat Capability Model output and observed occupancy in the northeast corner of Fort Pierre National Grassland (close-up of Figure 2).

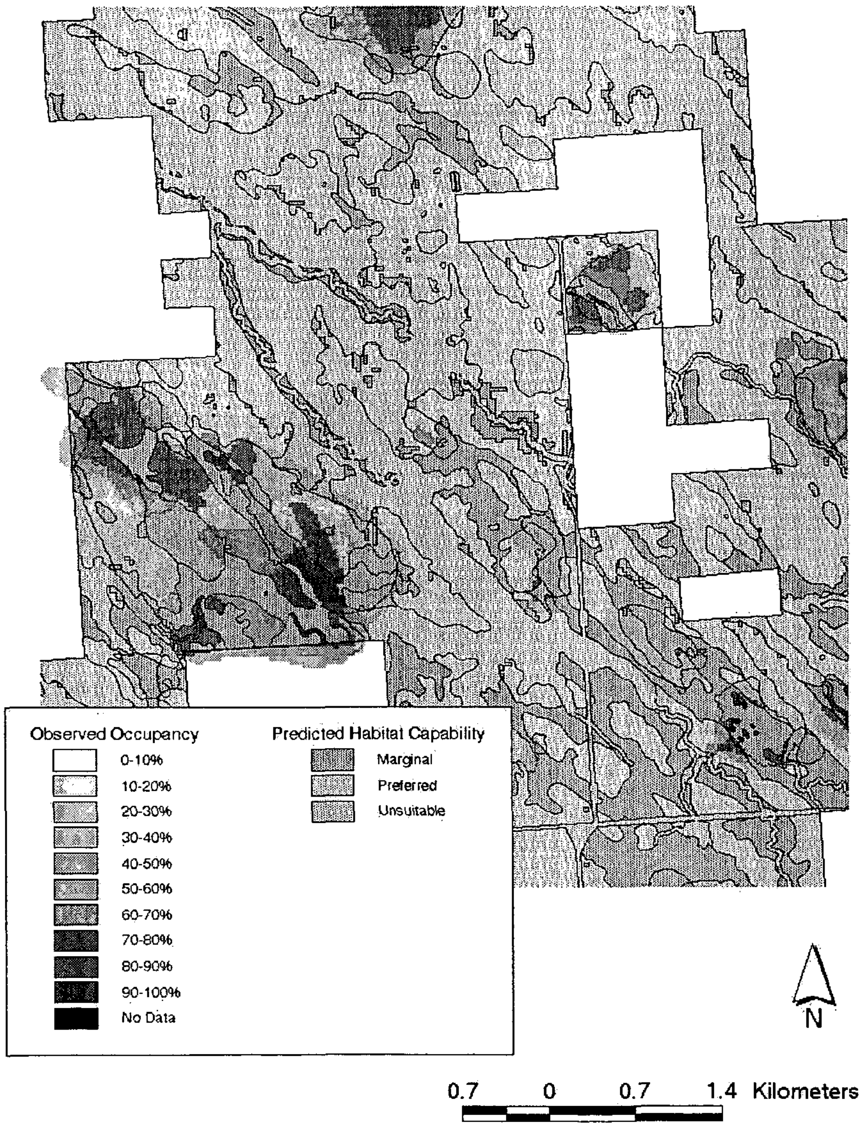


Figure 4. Habitat capability model performance on a portion of the Conata Basin study area

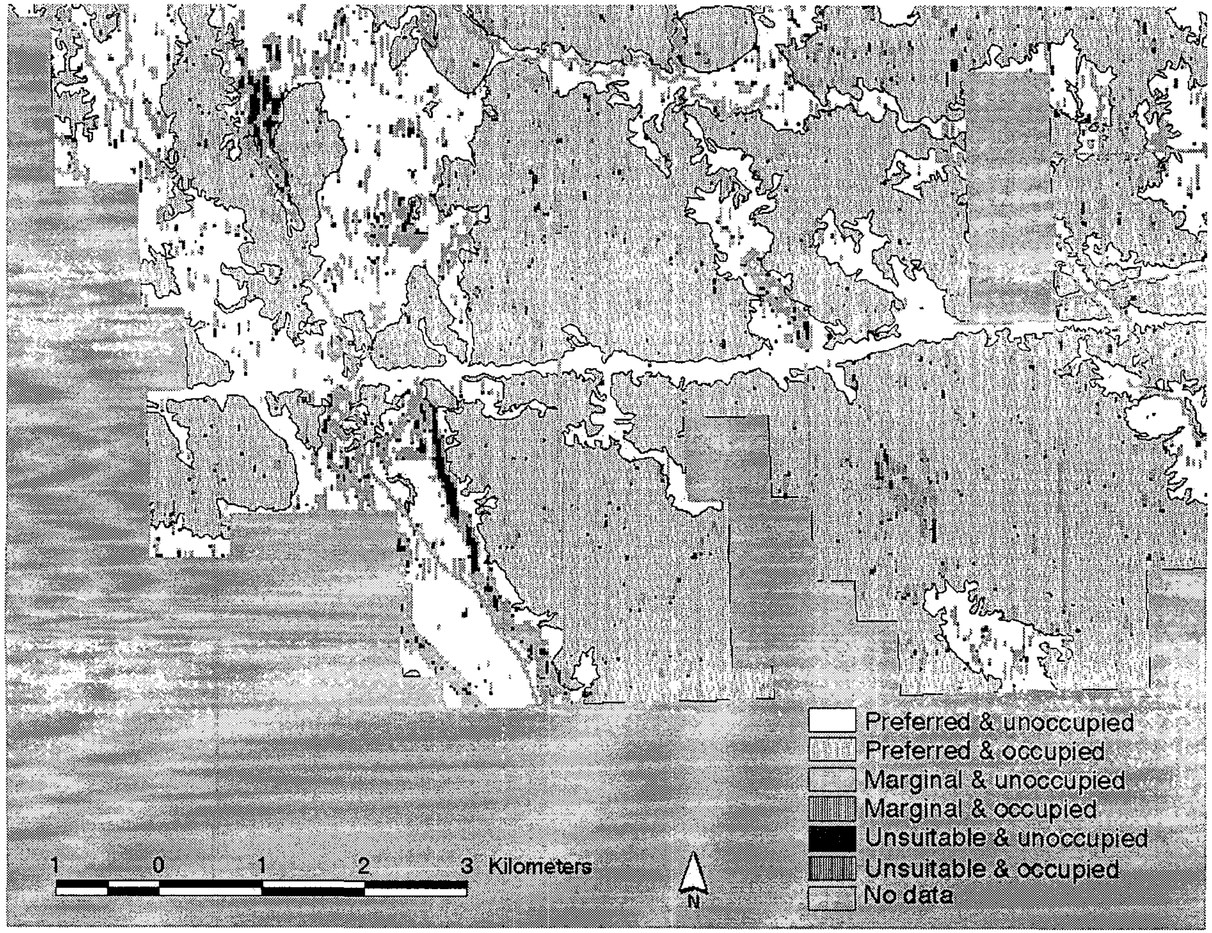


Figure 5. Spatial autocorrelation introduces error when non-spatial methods are used to model binary habitat occupancy as a function of habitat quality

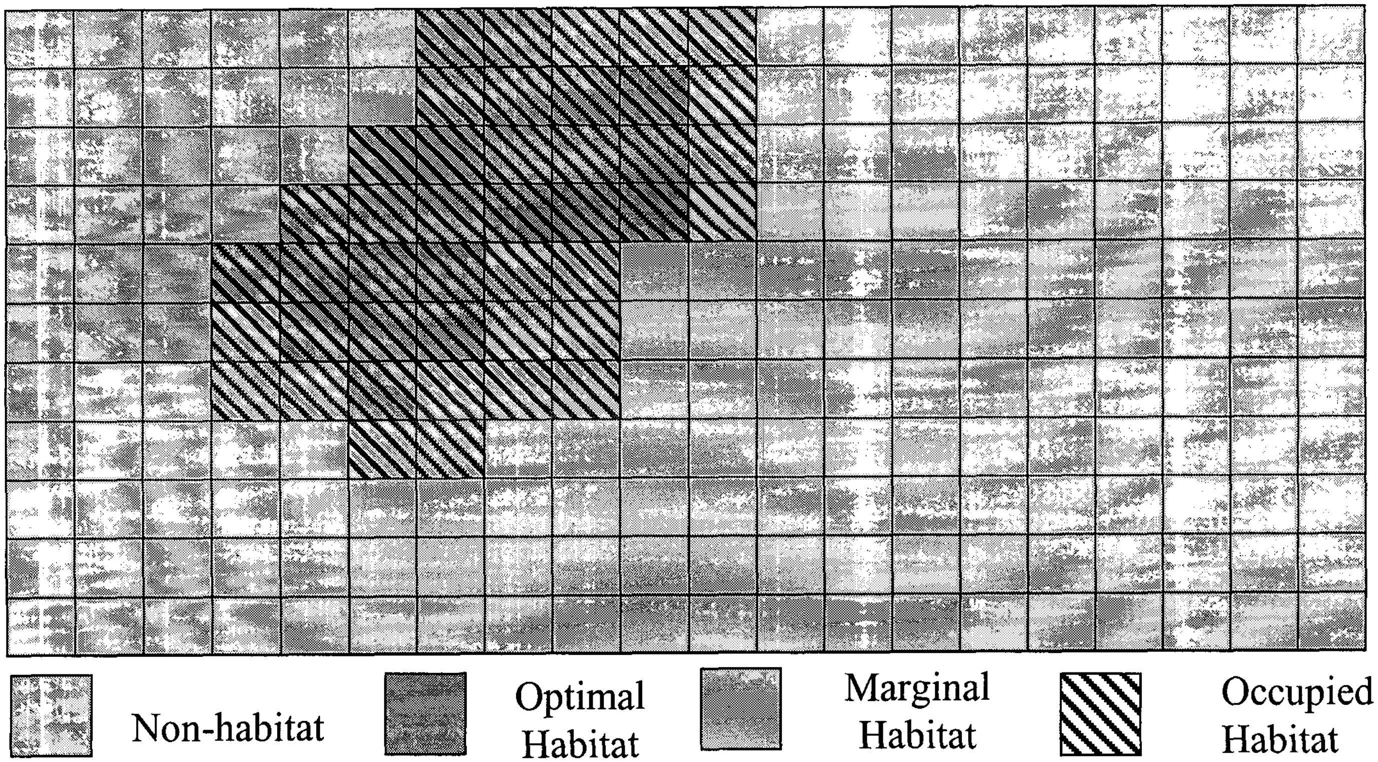


Figure 6. Effects of spatial autocorrelation on estimates of regression coefficients and standard errors

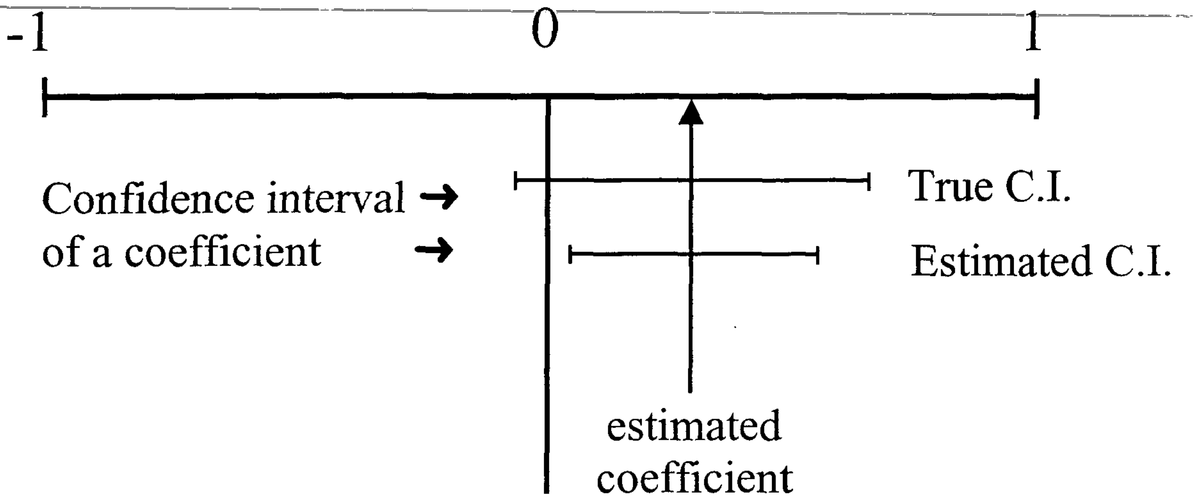


Figure 7. Pastures on public land within Fort Pierre National Grassland

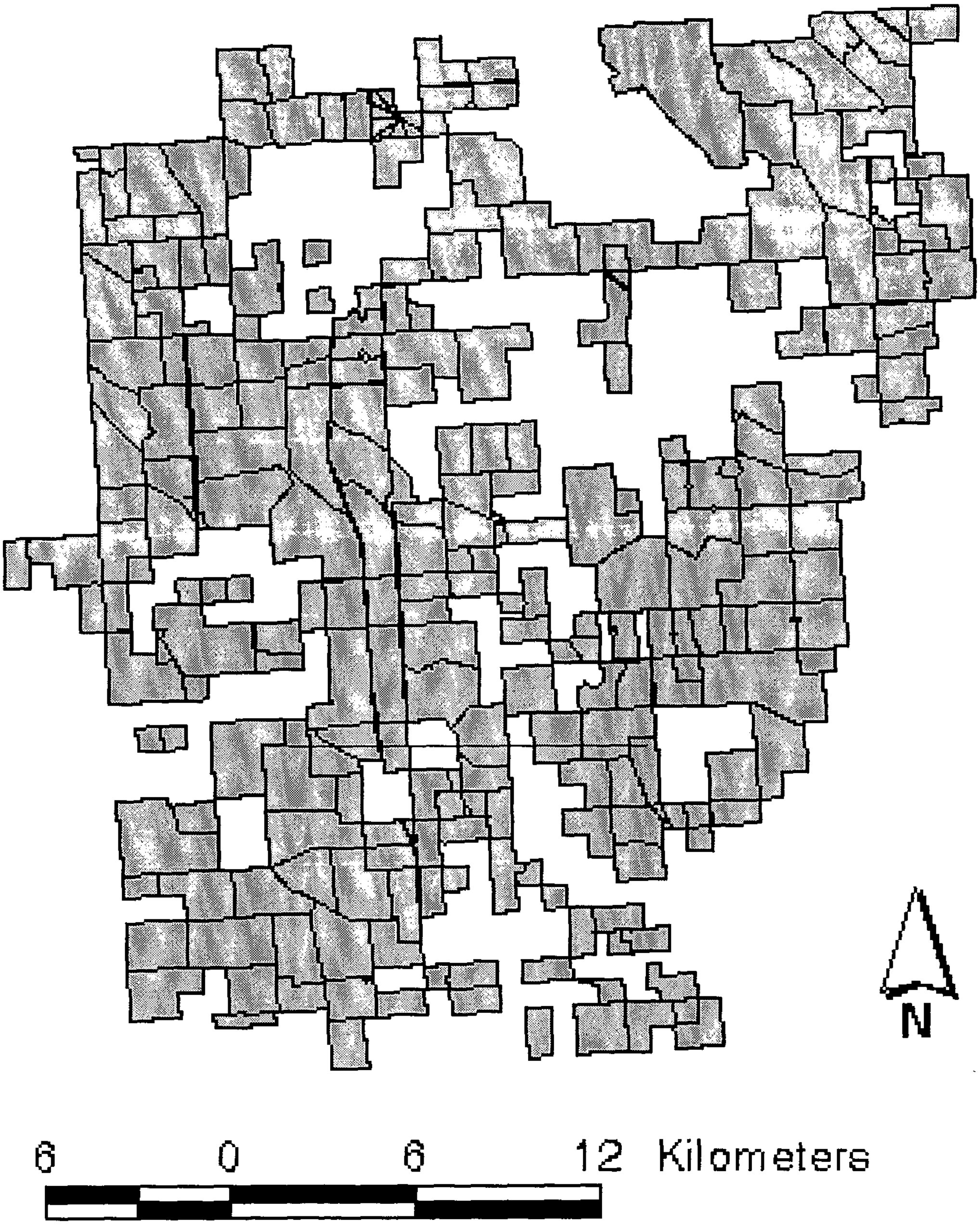
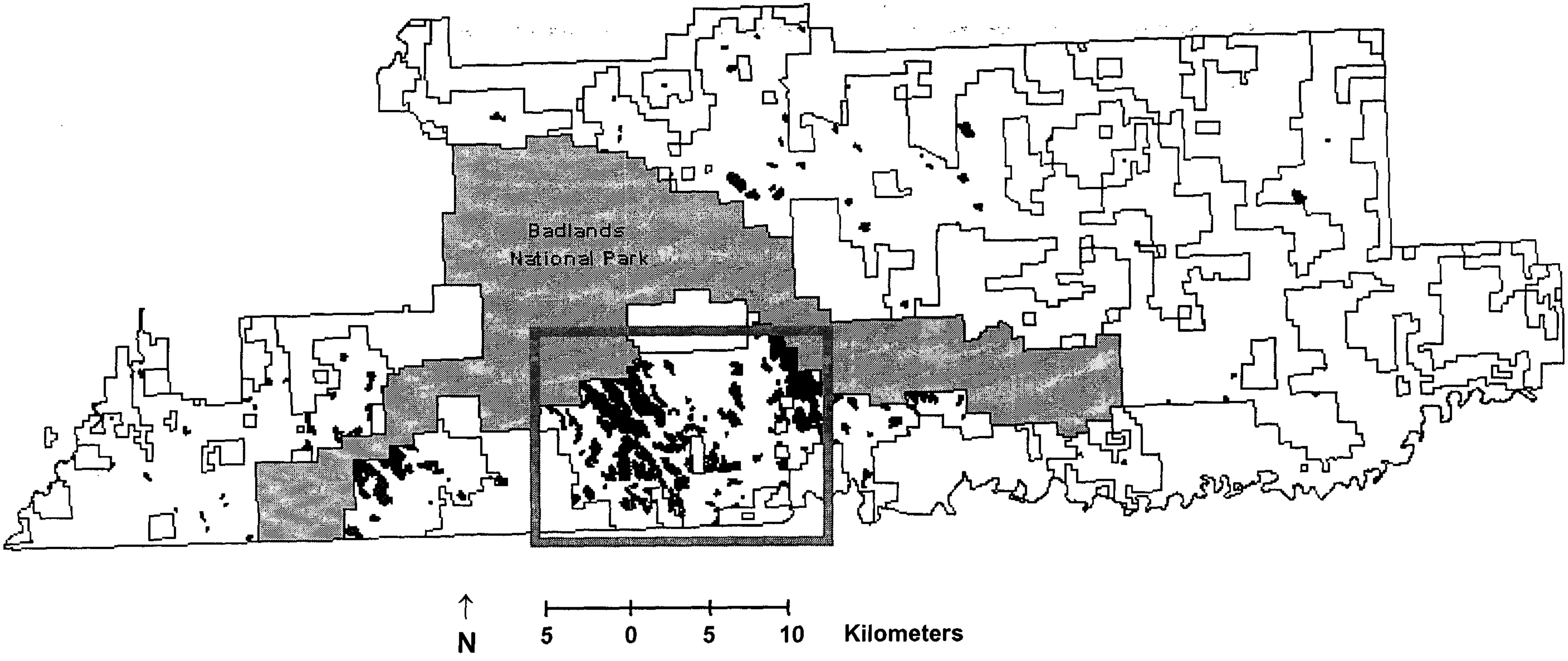


Figure 8. Conata Basin (shown outlined in red) in Buffalo Gap National Grassland



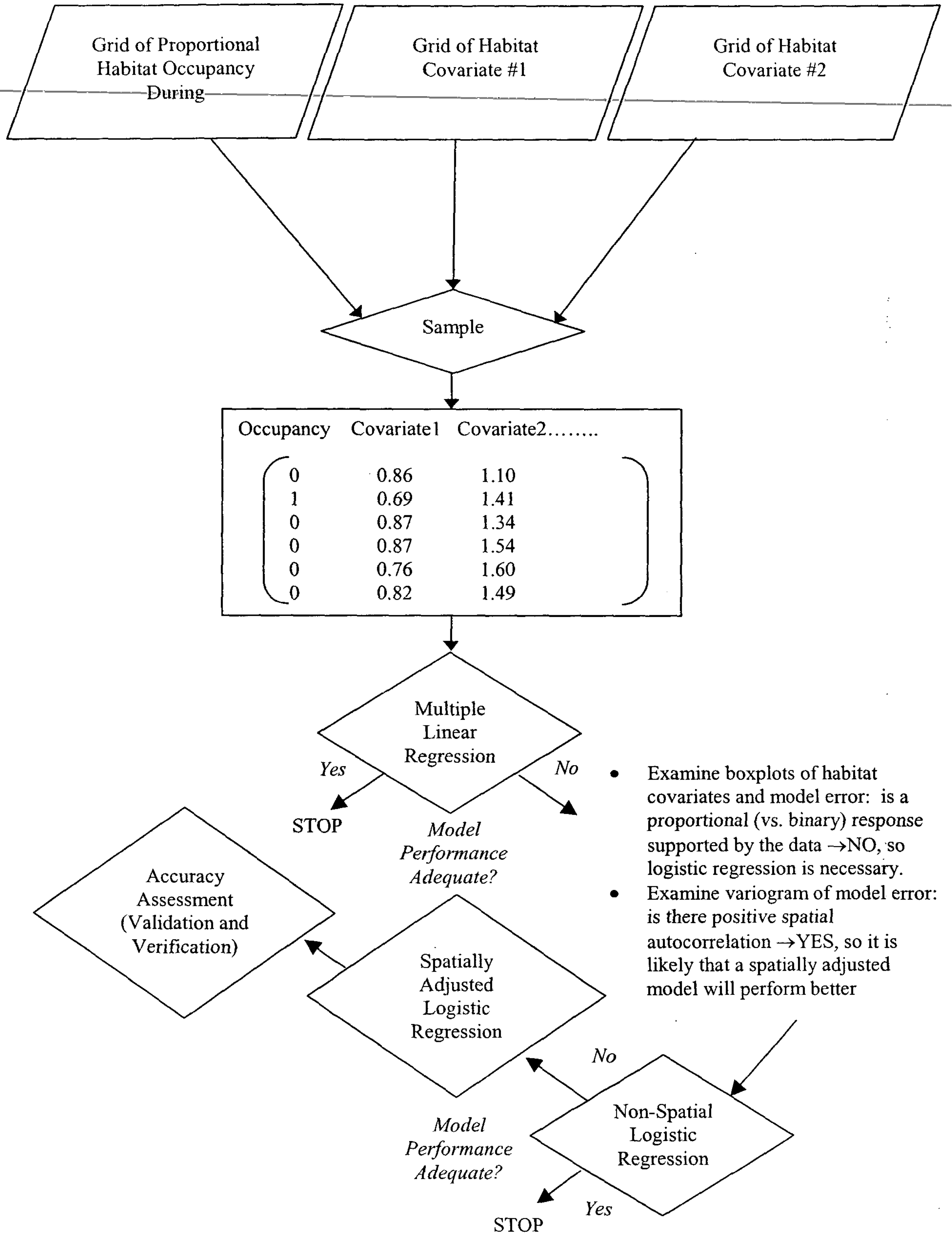


Figure 9. Flowchart of Analysis Process for Fort Pierre

Occupancy	Covariate1	Covariate2……..
0	0.86	1.10
1	0.69	1.41
0	0.87	1.34
0	0.87	1.54
0	0.76	1.60
0	0.82	1.49

- Examine boxplots of habitat covariates and model error: is a proportional (vs. binary) response supported by the data →NO, so logistic regression is necessary.
- Examine variogram of model error: is there positive spatial autocorrelation →YES, so it is likely that a spatially adjusted model will perform better

Figure 10. Variogram of residuals from Fort Pierre non-spatial linear model

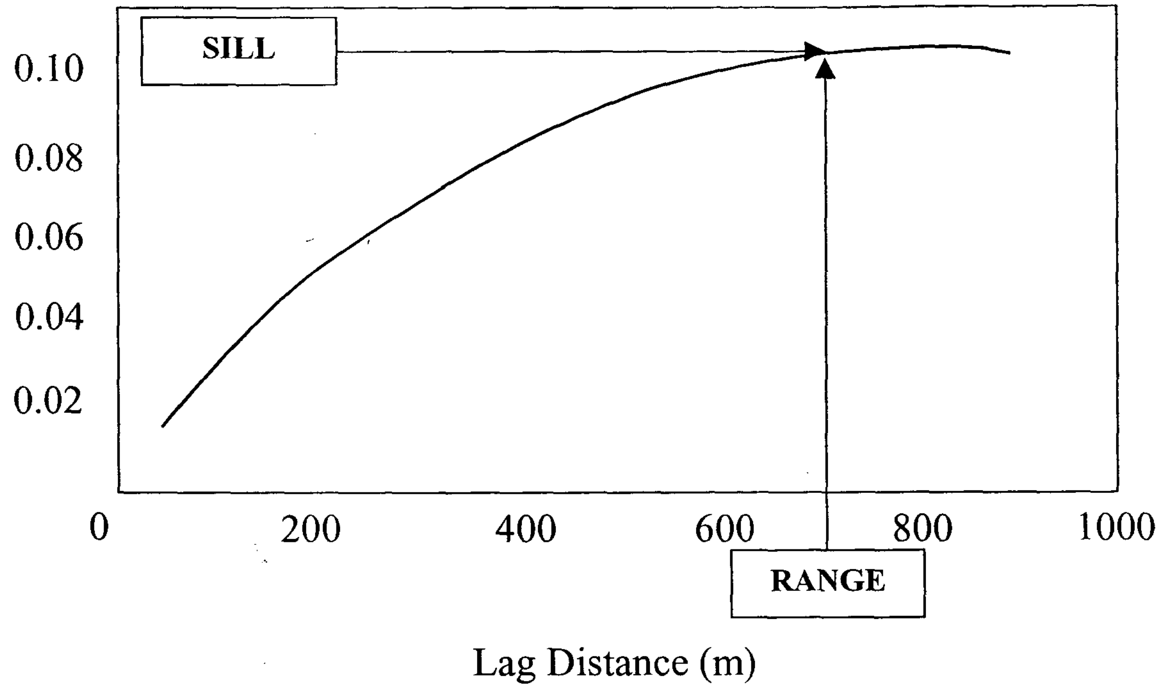


Figure 11. Data preparation using the smaller size spatial block (left panel) results in a block with less than 20 observations (outlined in blue) that must be removed, and this creates differences between data sets that prevent comparison of models using AIC.

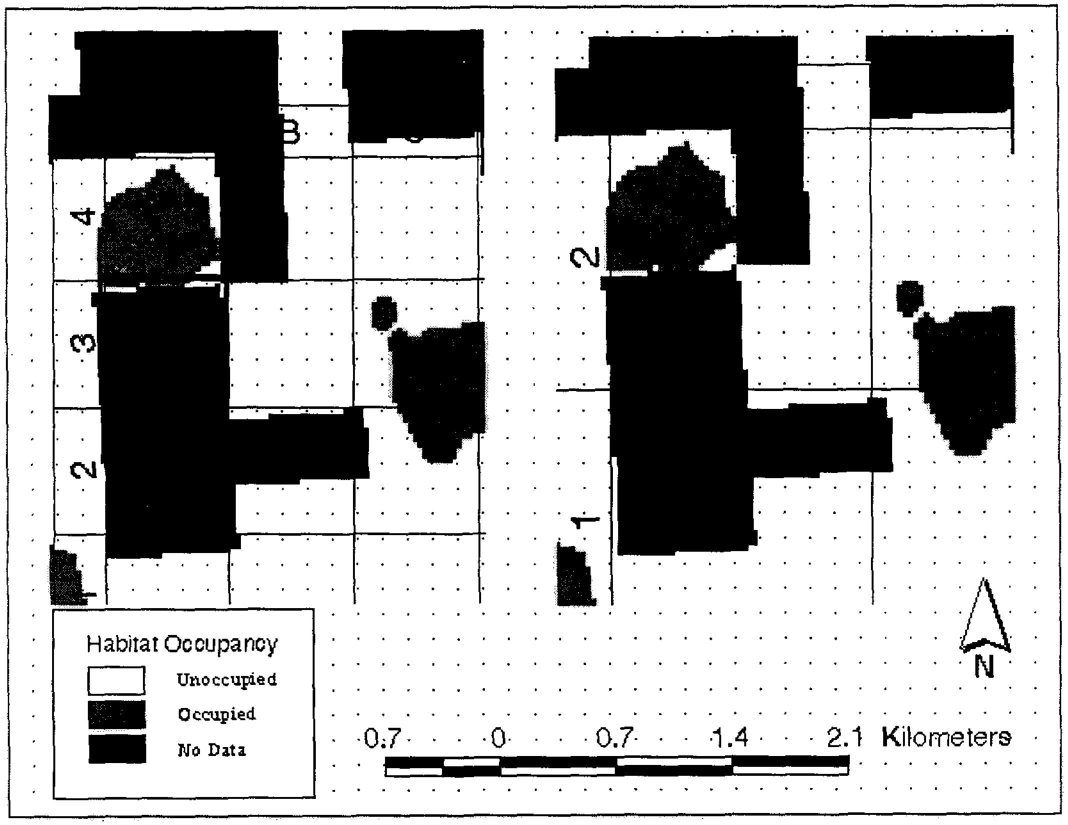


Figure 12. Error from the multiple linear regression model and its relationship to proportional occupancy

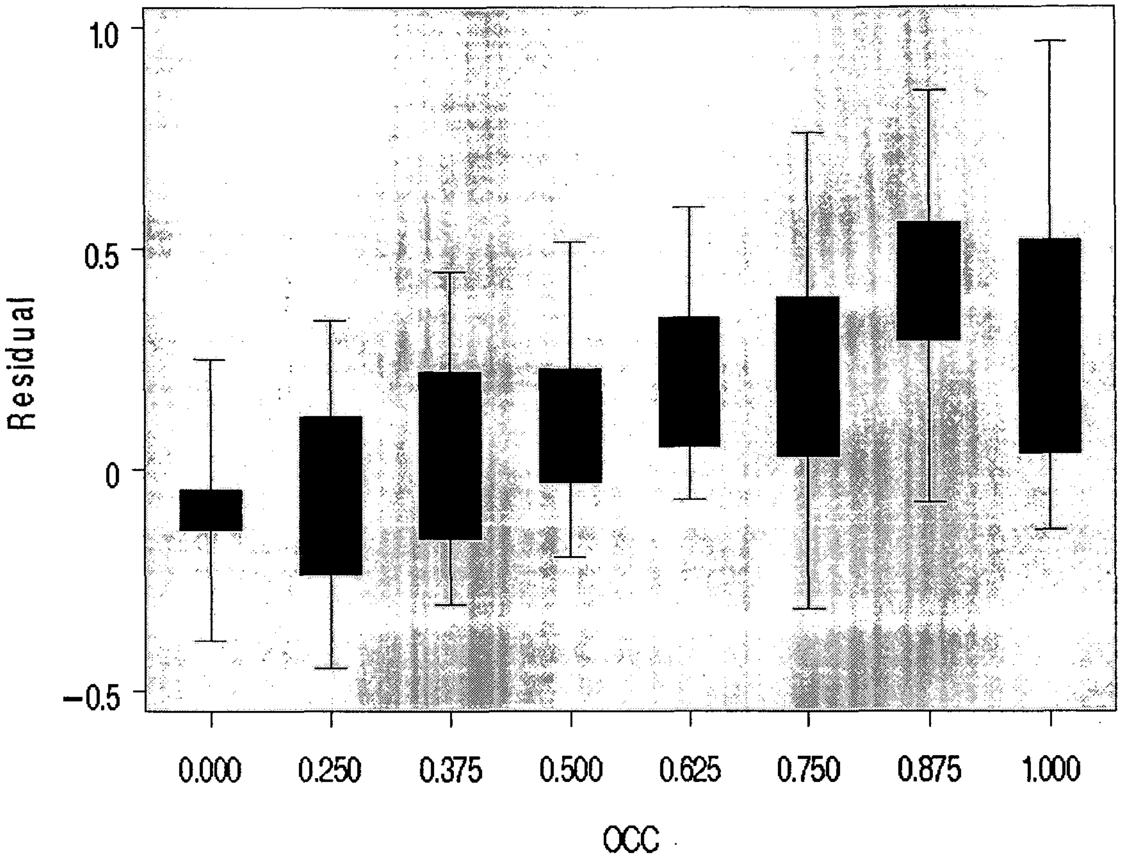


Figure 13. Mean actual use stocking levels vs. proportional occupancy

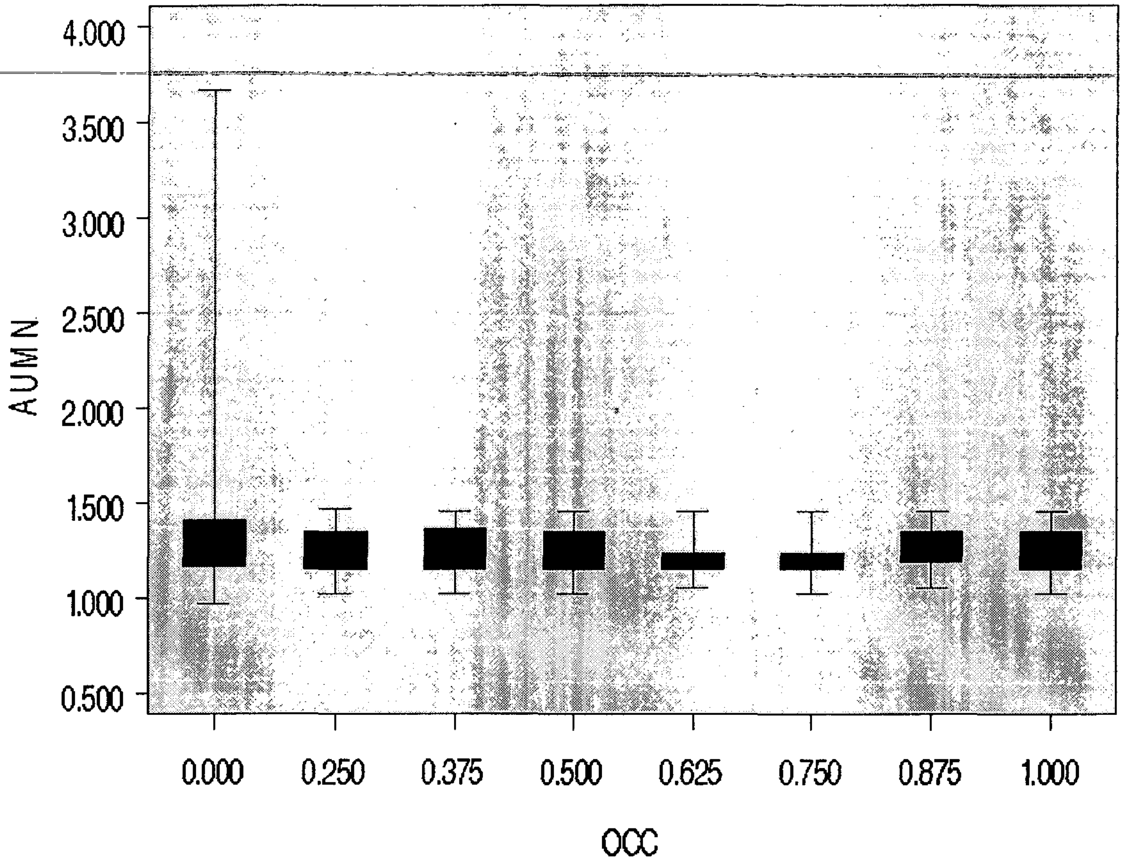


Figure 14. SD of actual use stocking levels vs. proportional occupancy

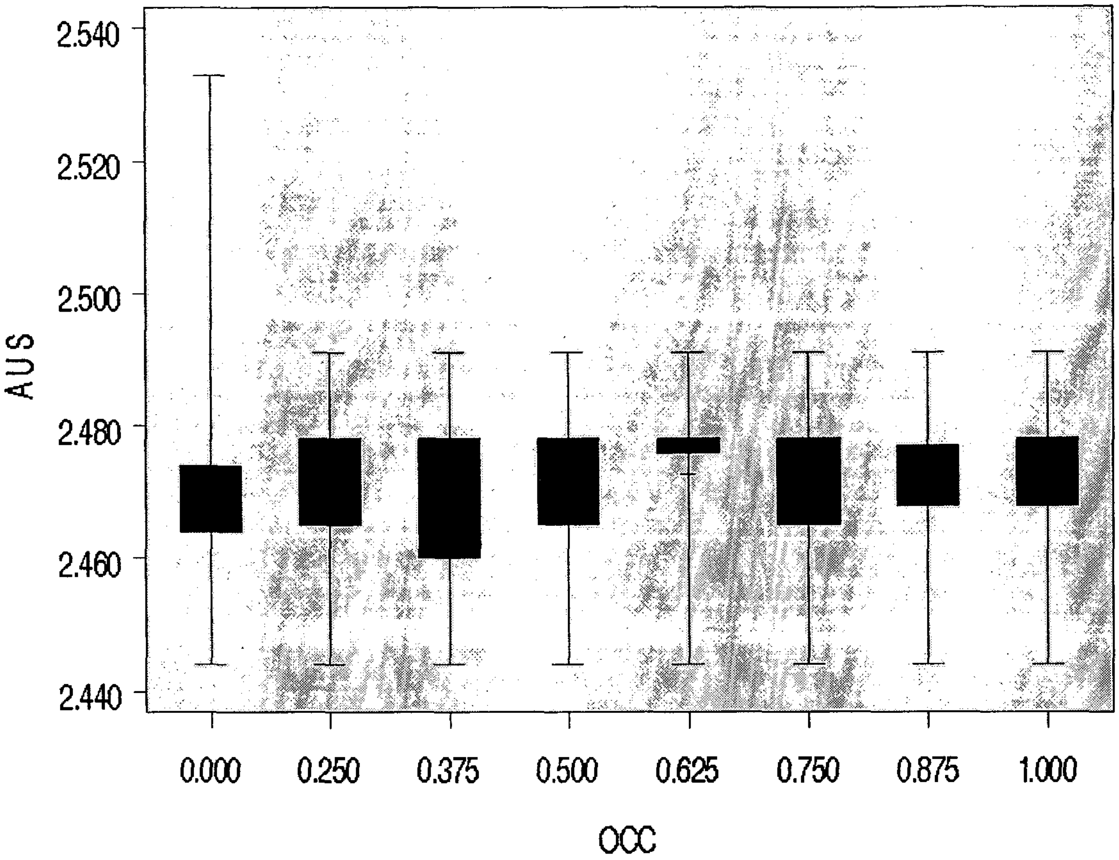


Figure 15. Mean clay content of all soil horizons vs. proportional occupancy

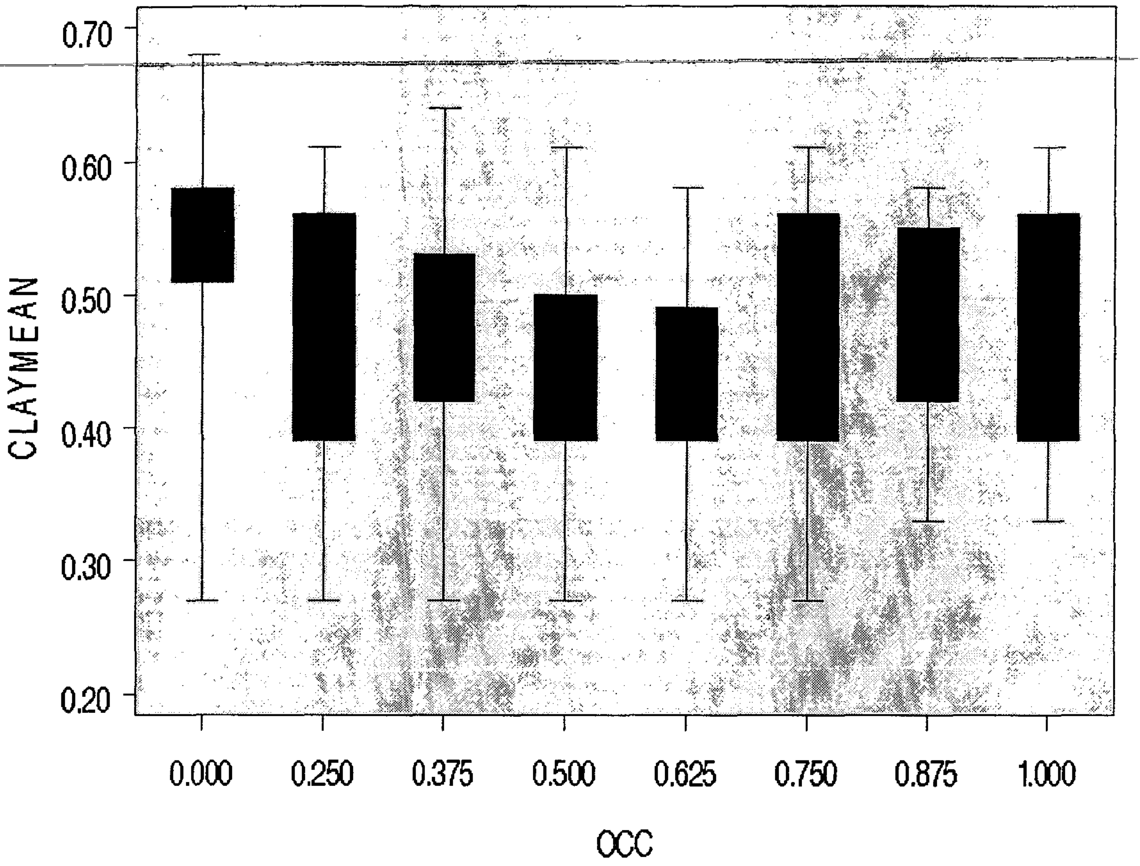


Figure 16. High mean clay content of all soil horizons vs. proportional occupancy

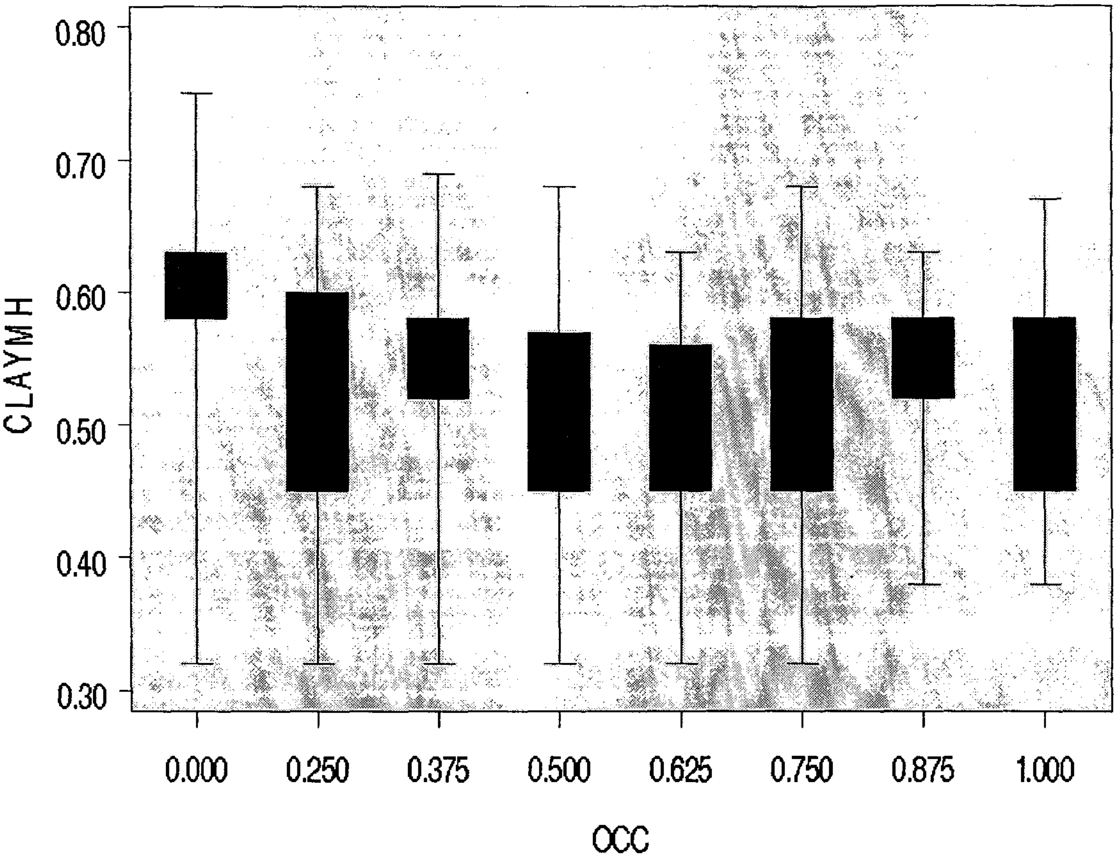


Figure 17. Low mean clay content of all soil horizons vs. proportional occupancy

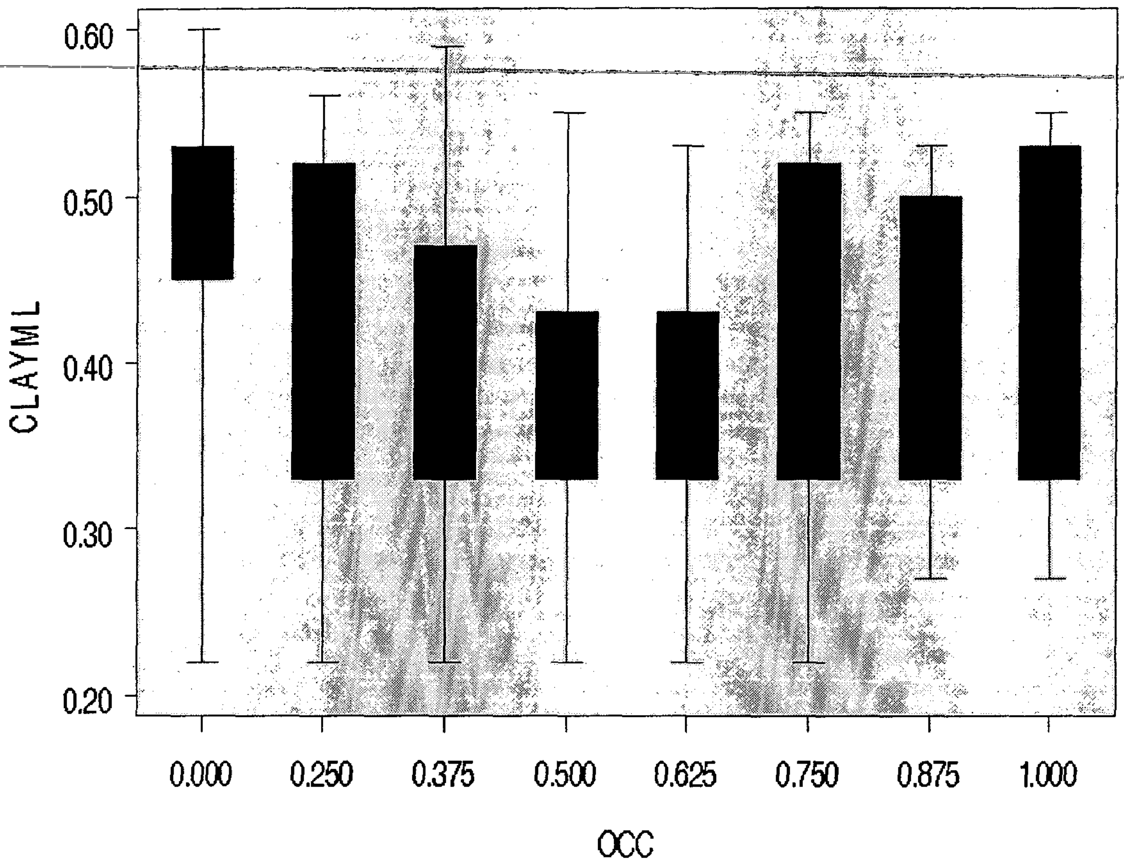


Figure 18. Mean clay content of top soil horizon vs. proportional occupancy

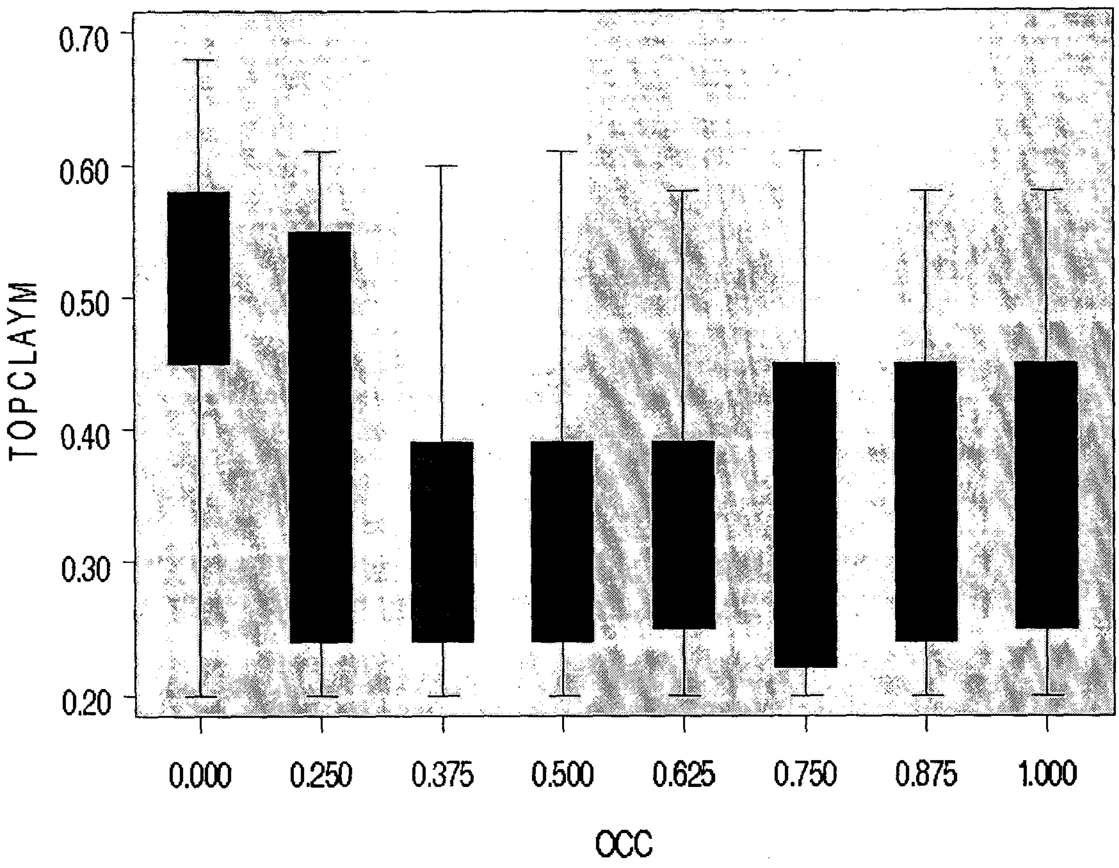


Figure 19. NRCS recommended stocking level vs. proportional occupancy

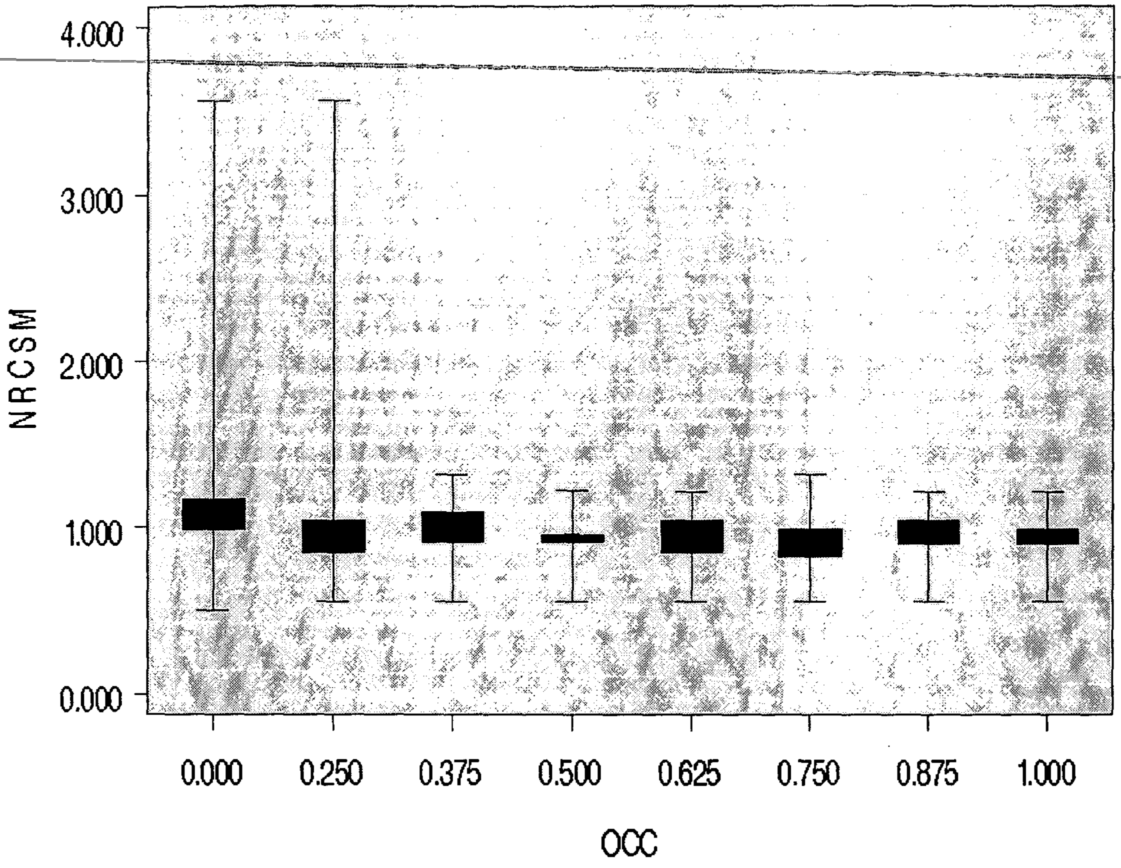


Figure 20. Poisoning probability vs. proportional occupancy

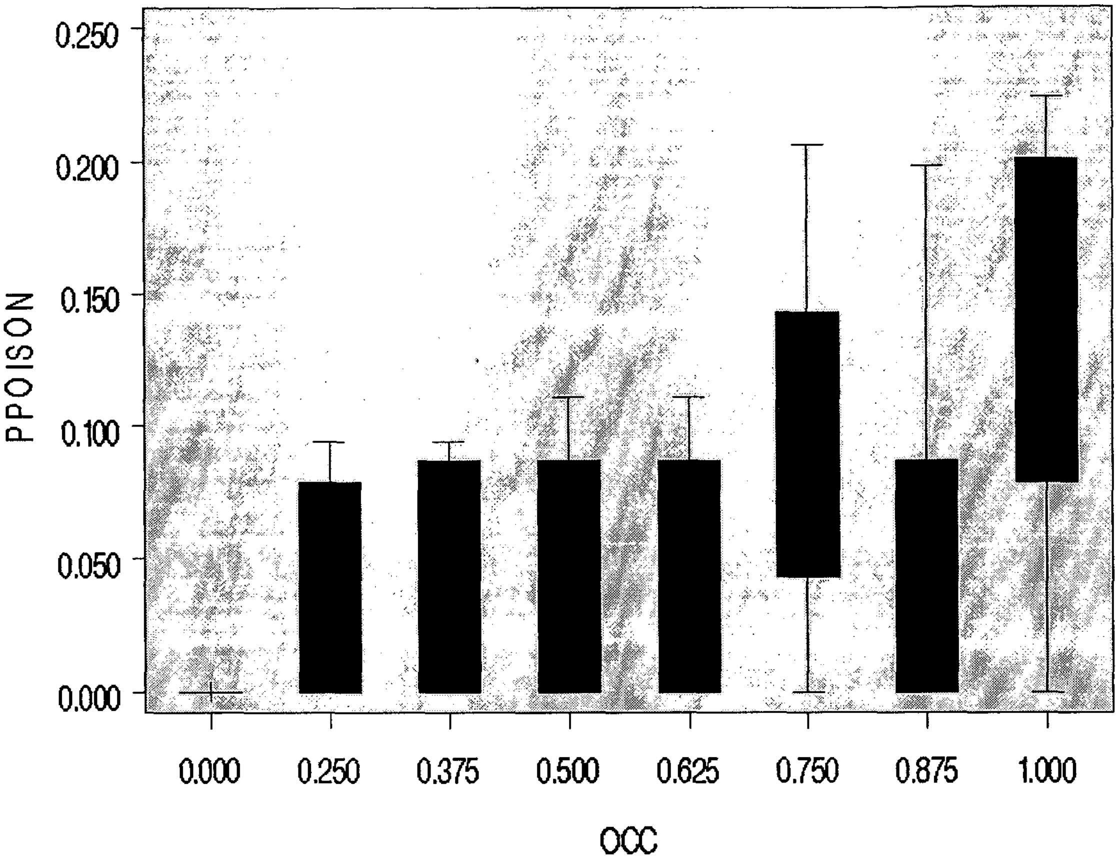


Figure 21. Mean annual precipitation vs. proportional occupancy

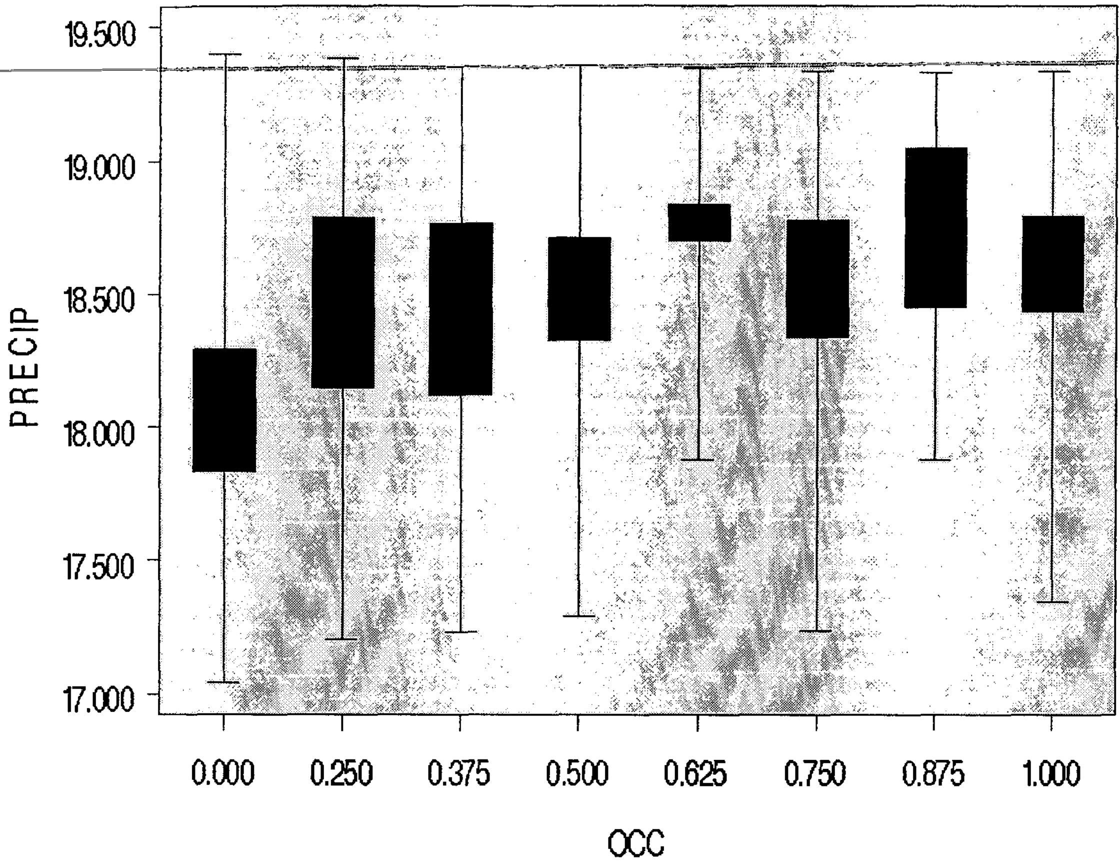


Figure 22. Distance to private land vs. proportional occupancy

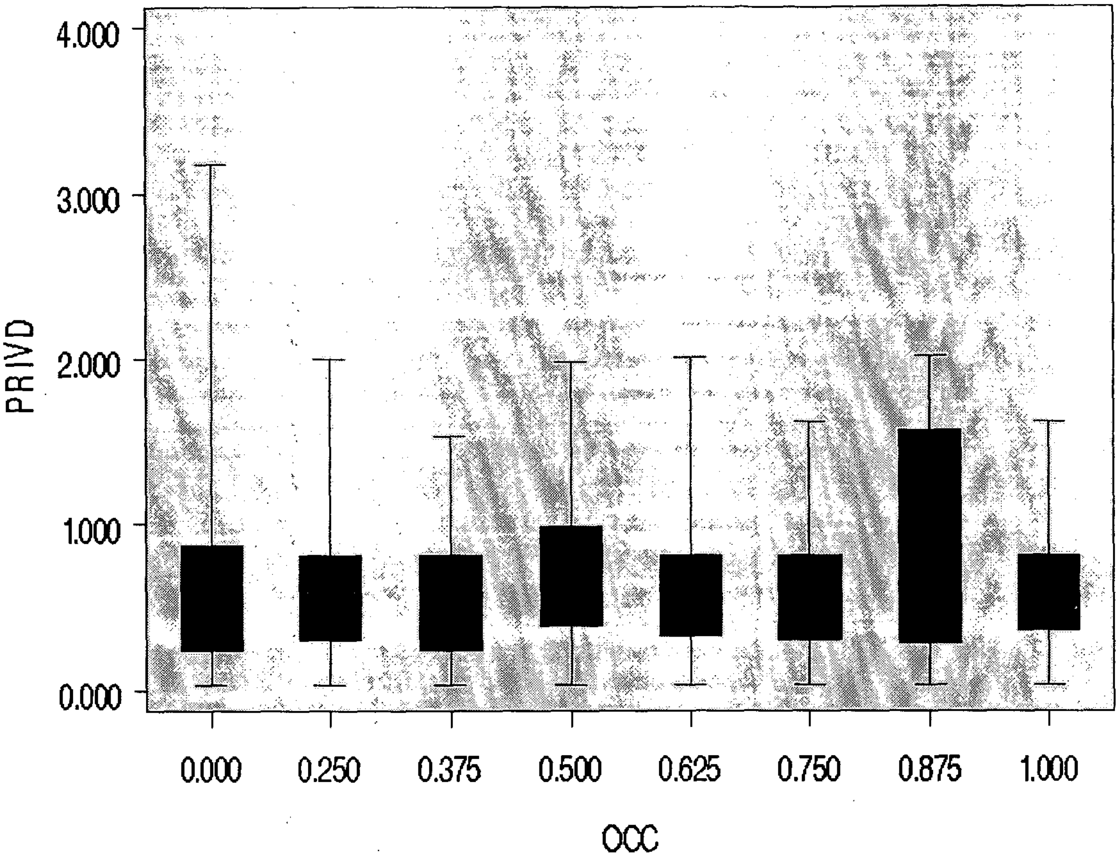


Figure 23. Slopes vs. proportional occupancy

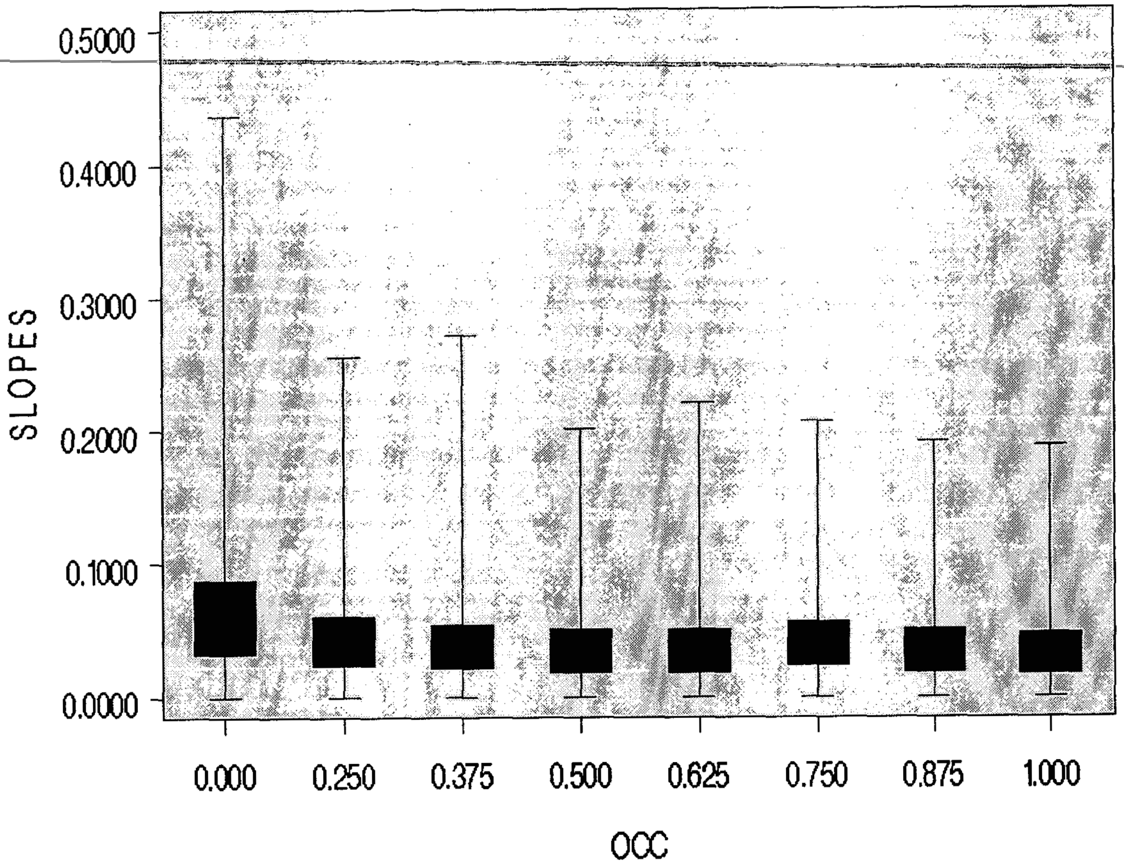


Figure 24. Distance to water vs. proportional occupancy

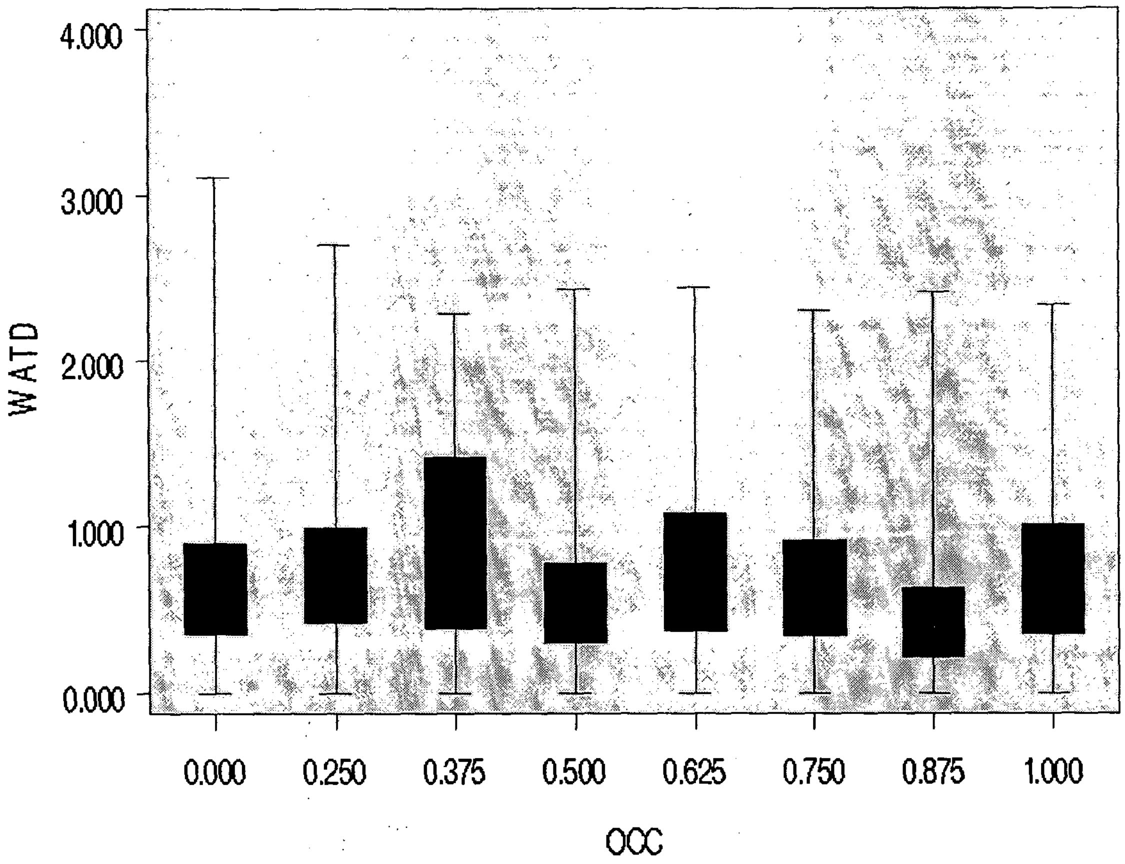


Figure 25. Fort Pierre spatial model resubstitution results with 95% sensitivity

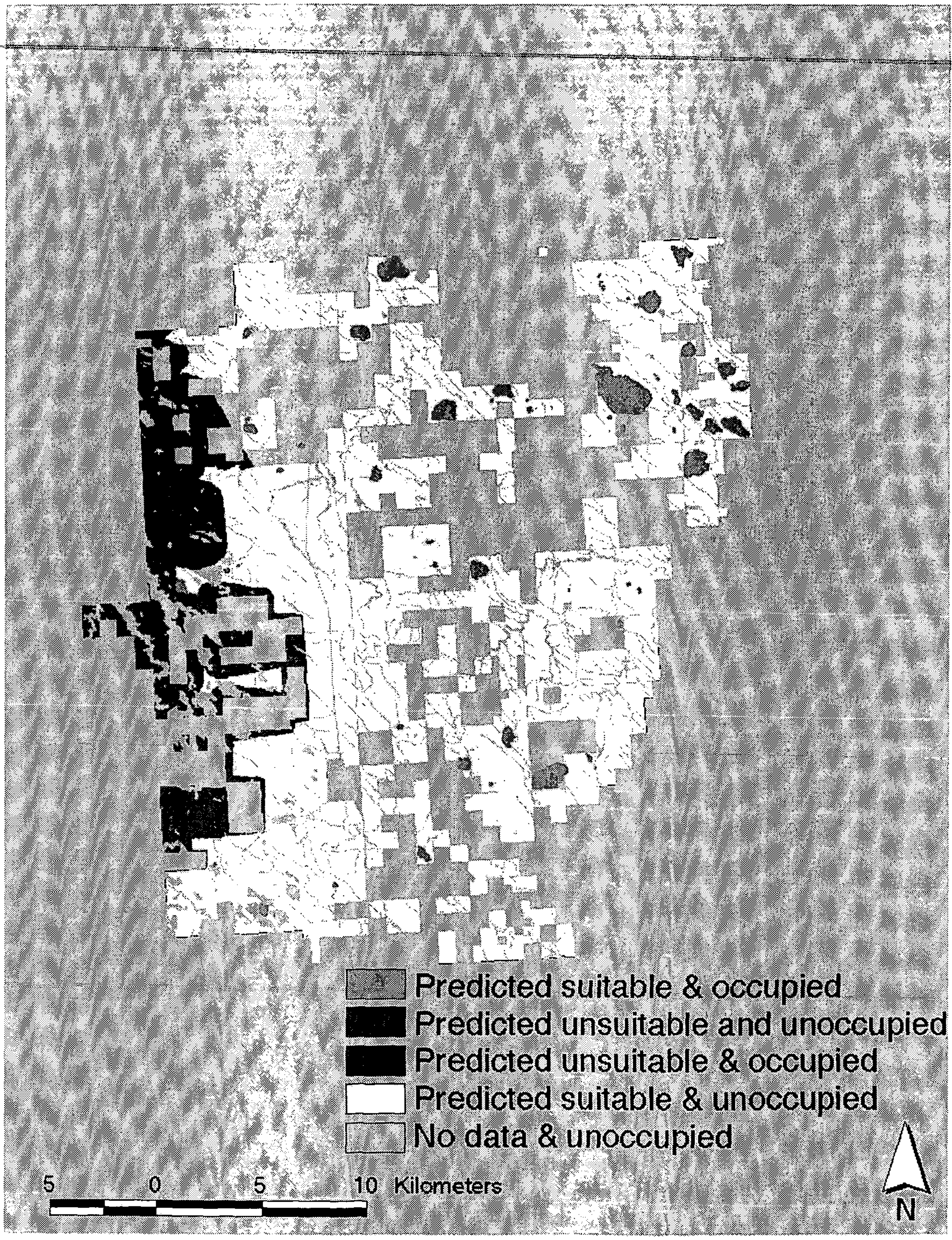


Figure 26. Resubstitution results for full non-spatial model

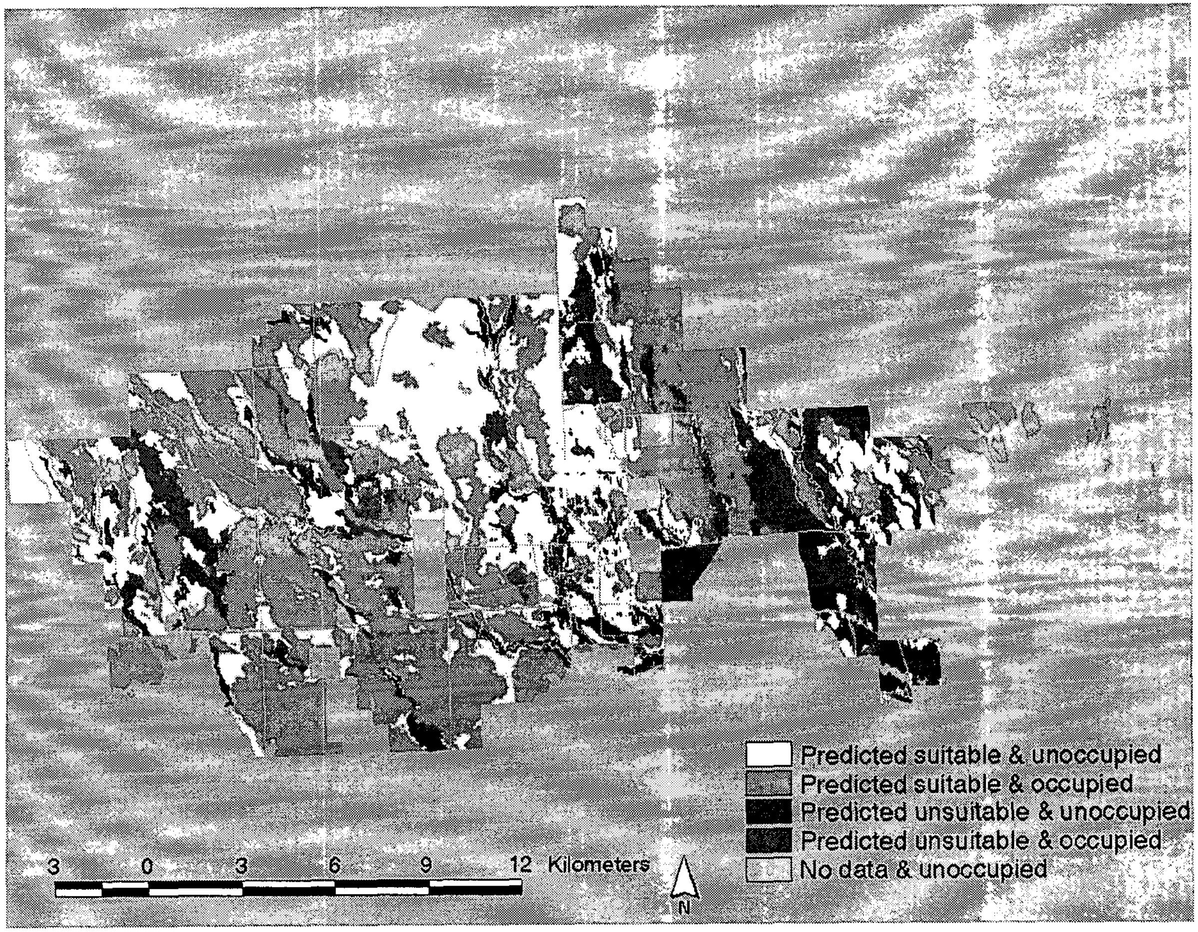


Figure 27. Resubstitution predictions for spatial model 3

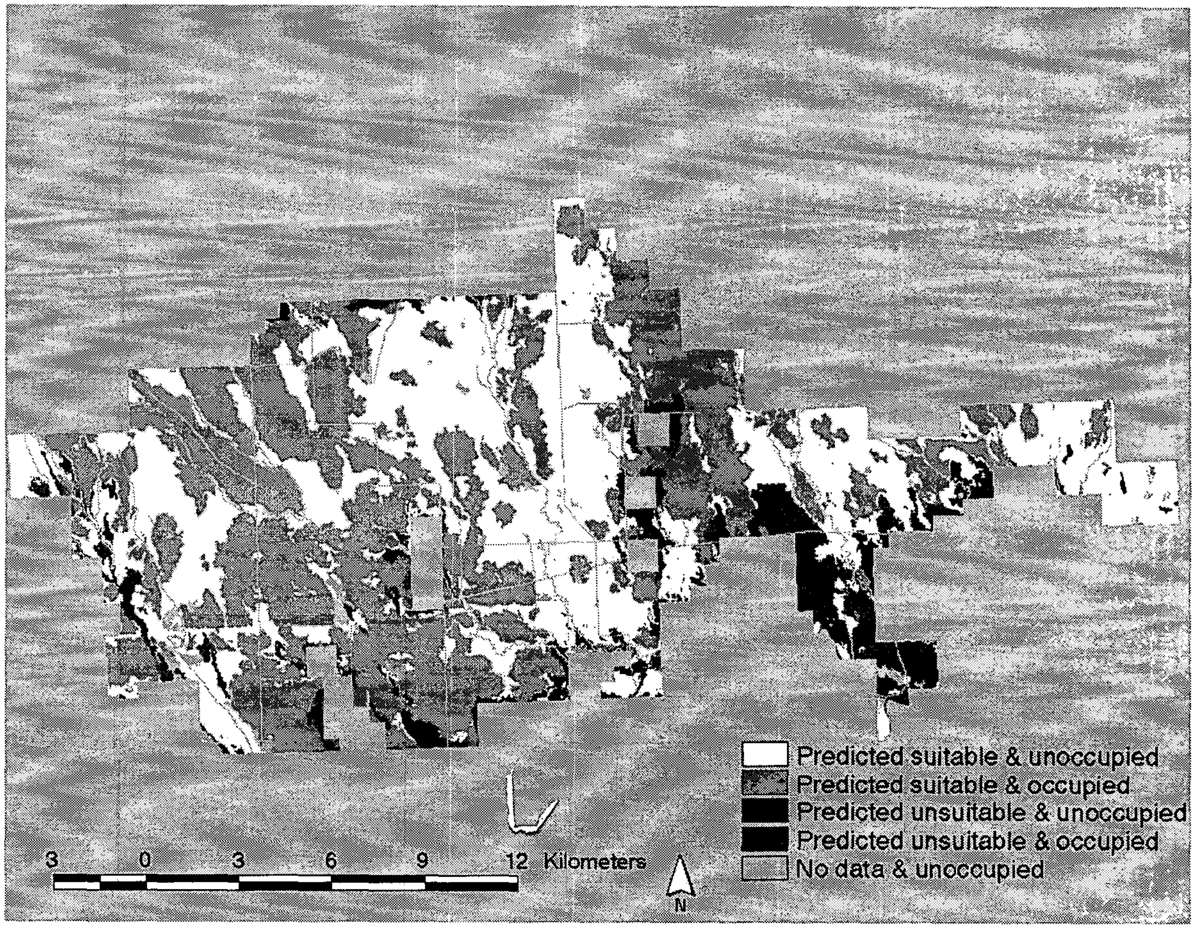


Figure 28. Resubstitution predictions for spatial model 5

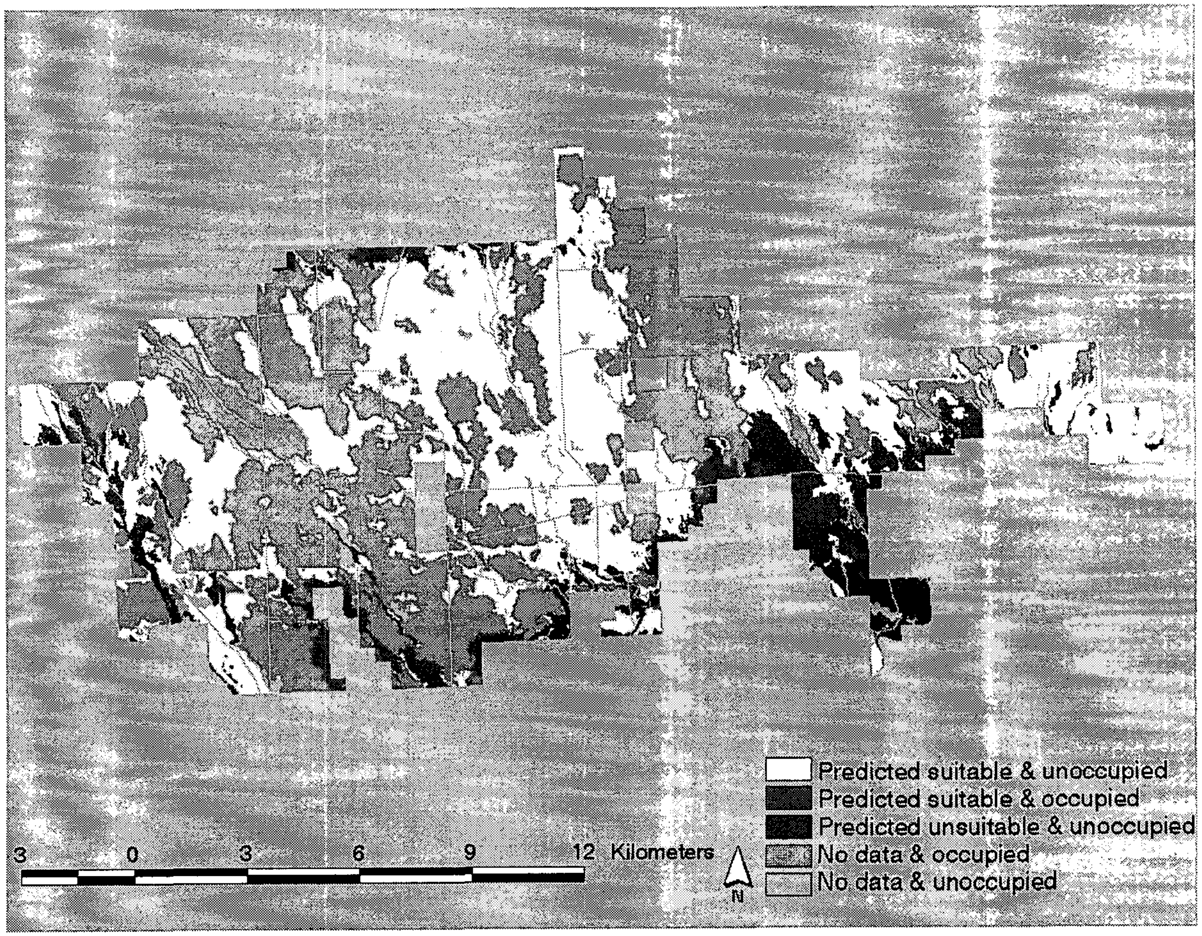


Figure 29. Resubstitution predictions for non-spatial model 7

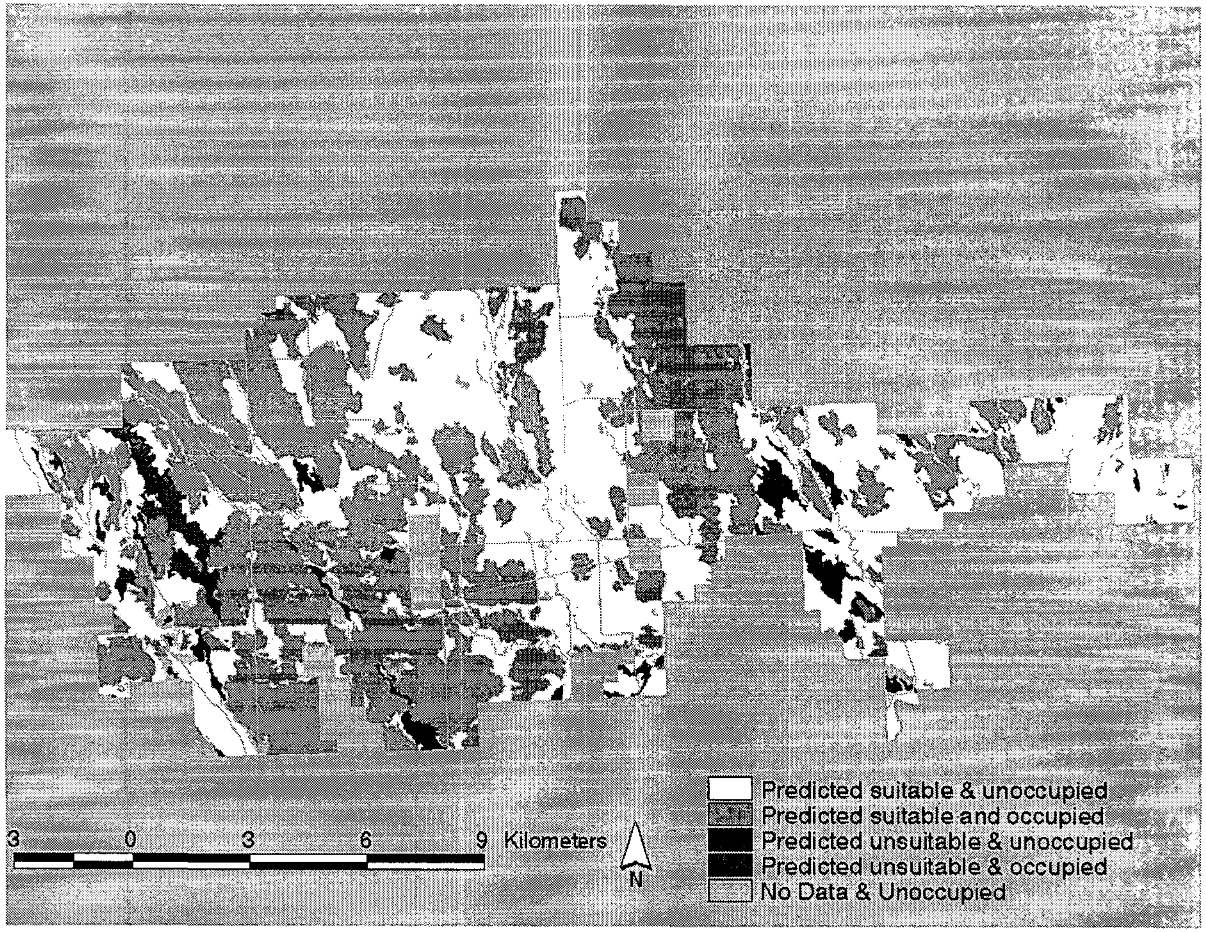


Figure 30. Resubstitution predictions for spatial model 7

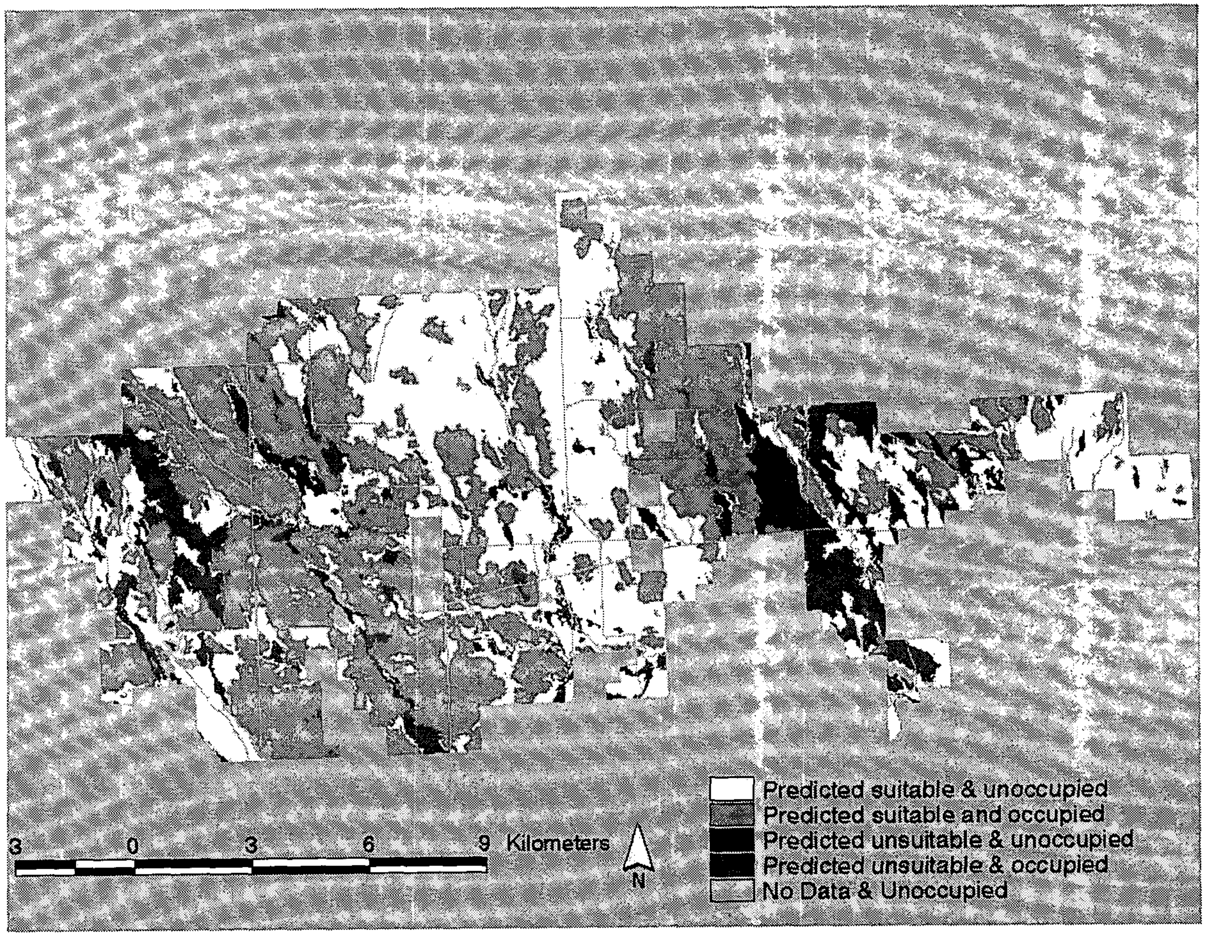


Figure 31. Predictions of spatial model 7 compared to maximum recorded habitat occupancy in the study area

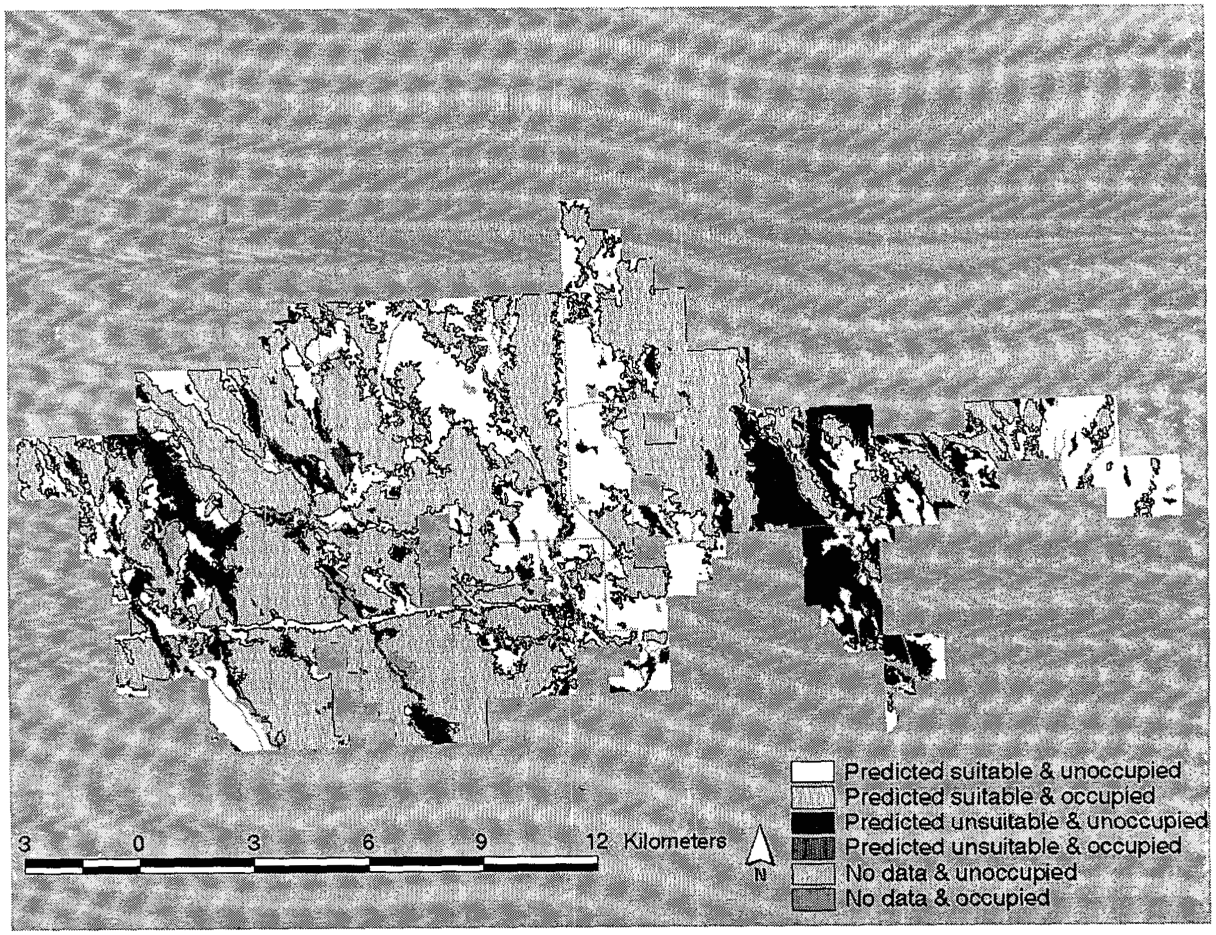


Figure 32. Predictions of spatial model 7 compared to maximum recorded habitat occupancy in an adjacent unmodeled area

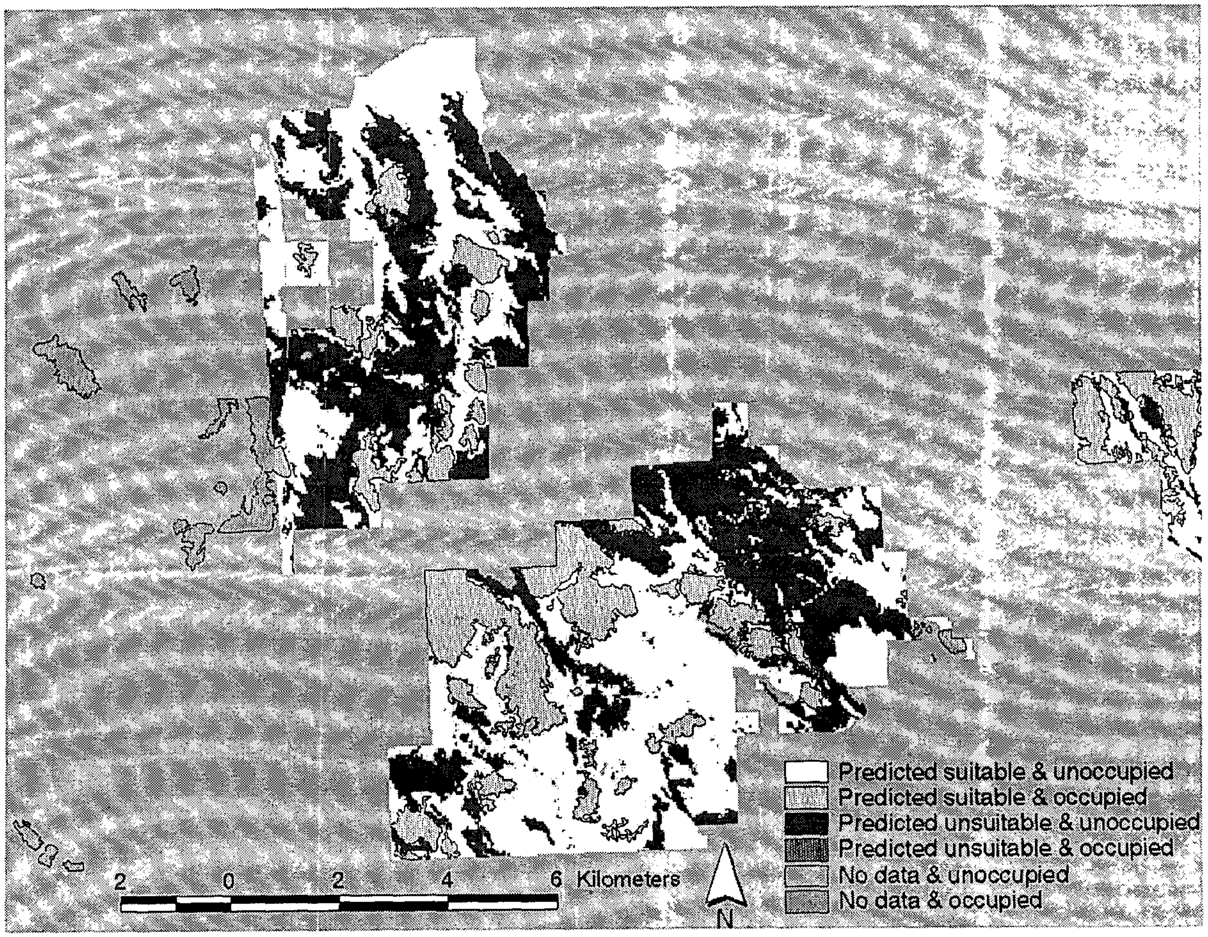
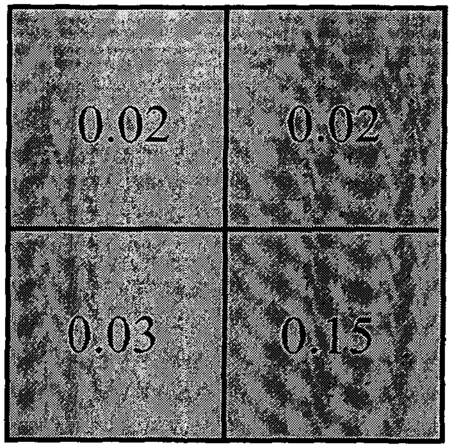
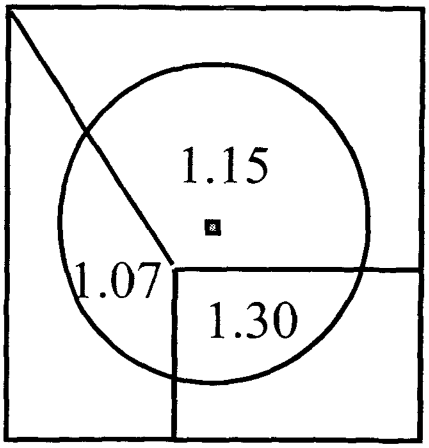


Figure 33. Habitat predictors may be ineffective if the unit of observation (green square) has significant variation within it, as with slope, or if the predictor is quantified at such a coarse scale that it varies little between observations across space, as with grazing intensity



Slope



Grazing Intensity

Appendix I. Macro used to calculate distance to water

```
/*******************************************************************
/* Program:  zonleucd.aml
/* Purpose:  Find zonal euclidean distance from points and polygons
/* Date:  November 2000
/* Author:  Jon Belak
/*******************************************************************
/*CREATE VARIABLES USED LATER TO CREATE FILE NAMES
&s count = 1
&s pasture = past
/*ASSIGNS PASTNAME.TXT TO THE VARIABLE FILE_A
/*PASTNAME.TXT LISTS POLYGONS TO BE USED AS ZONES
/*LISTS FIELDS AS THEY APPEAR IN .PAT
/*1 LINE PER ENTRY DOUBLESPACED
/*SORT ON NORTHING OR EASTING TO SIMPLIFY MOSAIC PROCESS
&s file_a = [open pastname.txt status -read]
&do &until %readstatus% = 102
  /*READ THE CURRENT LINE FROM FILE_A AND ASSIGN IT TO CUR_PAST
  &s cur_past = [read %file_a% readstatus]
  /*TYPE THE NAME OF THE CURRENT PASTURE
  &type %cur_past%
  /*RESELECT THAT POLYGON AND USE IT TO CREATE A COVERAGE
  reselect Pastures currpast  poly
  reselect Pastures#  = %cur_past%
  ~
  N
  N
  /*GIVE THE COVERAGE POLYGON TOPOLOGY
  build currpast poly
  grid
  setwindow ftpr_slope
  setcell ftpr_slope
  /*THE WATRSRC GRID IS THE UNION OF ALL POINT AND POLY SOURCES
  /*SOURCES HAVE NON-ZERO VALUES, ALL OTHER AREAS ARE NODATA
  /*CELL SIZE AND EXTENT ARE THE SAME AS DESIRED IN FINAL GRID
  /*CLIP TO RESTRICT DISTANCE ANALYSIS TO WITHIN-PASTURE SOURCES
  gridclip watrsrc currwatr cover currpast
  /*PERFORM DISTANCE ANALYSIS
  curreucd  = eucdistance(currwatr, #, #, 1000000, #)
  /*CLIP TO EXTENT OF CURRENT PASTURE
  /*OUTPUT FILES NAMED USING %PASTURE% COMBINED WITH %COUNT%
  gridclip curreucd  %pasture%%count% cover currpast
  quit
  /* KILL INTERMEDIATE COVERAGES
  kill currpast all
  kill currwatr all
  kill curreucd  all
  &s count = %count% + 1
  /*READ A BLANK LINE FROM THE FILE TO CHECK FOR EOF
  &s cover = [read %file_a% readstatus]
  &end
/* USING THE MOSAIC COMMAND AT THE GRID PROMPT, CUT AND PASTE
/* PASTURE NAMES FROM PASTNAME.TXT INTO THE APPROPRIATE EXPRESSION
```