2014

MODELING NORTHERN GOSHAWK (ACCIPITER GENTILIS) NESTING HABITAT ON THE LEWIS AND CLARK NATIONAL FOREST USING EIGENVECTOR FILTERS TO ACCOUNT FOR SPATIAL AUTOCORRELATION

Morganne Marie Lehr
The University of Montana

MODELING NORTHERN GOSHAWK (ACCIPITER GENTILIS) NESTING HABITAT ON THE

LEWIS AND CLARK NATIONAL FOREST USING EIGENVECTOR FILTERS TO ACCOUNT

FOR SPATIAL AUTOCORRELATION

By

Morganne Marie Lehr

B.S. Wildlife Biology, The University of Montana, Missoula, MT, 2008
Certificate in Geographic Information Science, The University of Montana, Missoula, MT, 2011

Thesis

presented in partial fulfillment of the requirements
for the degree of

Master of Science
in Geography

The University of Montana
Missoula, MT

December 2014

Approved by:

Sandy Ross, Dean of The Graduate School
Graduate School

Dr. David Shively, Chair
Department of Geography

Dr. Anna Klene, Member
Department of Geography

Dr. Beth Hahn, Outside Member
Aldo Leopold Wilderness Research Institute

Lehr, Morganne, M.S., December 2014   Geography

Modeling Northern Goshawk (*Accipiter gentilis*) Nesting Habitat on The Lewis and Clark National Forest using Eigenvector Filters to Account for Spatial Autocorrelation

Chairperson: Dr. David Shively

        Habitat suitability modeling has become a valuable tool for wildlife managers to identify areas of suitable habitats for management and conservation needs. The Northern goshawk (Accipiter gentilis) has been the focus of many modeling efforts, however, the current models guiding goshawk management on the Lewis and Clark National Forest may not fully capture the unique habitat characteristics that the goshawk is actually selecting for nesting habitat. Therefore, the first objective of this study was to explore the use of Maxent for modeling suitable goshawk nesting habitat on the Lewis and Clark National Forest in central Montana. However, goshawk territoriality and their use of alternate nest locations creates, spatial autocorrelation between the nest locations (nest locations that occur close to one another are not independent) and can complicate the development of a habitat suitability model. Spatial autocorrelation can have drastic effects on model prediction and can lead to false conclusions about ecological relationships, but when accounted for can lead to insights that may have been otherwise overlooked. As a result, this study also explored the use of eigenvector filters as additional explanatory variables to assist in "filtering" out the effects of spatial autocorrelation from the modeling effort. Furthermore, this study evaluated the difference in model outputs using different resampling methods (bootstrap and cross-validation) and number of variables to determine the differences between models. The results of the study showed that the use of eigenvector filters not only improved model performance and reduced commission error, but created more precise predictions of suitable habitat. Furthermore, this study also found that using bootstrap methods and all biologically relevant environmental variables (with the additional of eigenvector filters) provided the best overall model. However, wildlife managers should closely review the methods and results provided in this study and choose the model that best suits their available data and management needs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The Northern goshawk (*Accipiter gentilis*) is currently identified as a Management Indicator Species for old growth forest characteristics on the Lewis and Clark National Forest in central Montana, under the existing management plan that was established in 1986. The USDA Forest Service, along with state and local entities, has a legal obligation to protect this species, along with its habitat. Habitat suitability modeling can be a valuable tool for guiding goshawk management and conservation efforts across federal, state, and local government lands. In some instances, habitat suitability models (HSM) can help to identify key habitat components and provide insights into a species' habitat preferences. This information can be utilized to protect, enhance, or simply identify areas that provide suitable habitat for the Northern goshawk on the Lewis and Clark National Forest.

Ideally, HSM's are developed from presence/absence data collected using a standardized sampling method that randomly covers the full extent of the study area. However, these data (especially absence data) are not always attainable, due to many reasons including staffing, time, and budgetary constraints. Phillips et al. (2006) developed a machine learning program called Maxent that utilizes presence-only data, along with background pixels (pseudo-absences). Maxent has been the focus of many studies (e.g., Phillips et al., 2004; Elith et al., 2006; Phillips et al., 2006; Phillips and Dudık, 2008), and has consistently out-performed other modeling methods in its predictive performance. Maxent also provides options that allow for a more robust statistical model to be built, and for more nuanced assessment of results, than most other software packages.

Due to the goshawk's territorial tendencies and the use of multiple nest locations by the same individual within each territory, modeling goshawk nesting habitat can become complicated by spatial autocorrelation between nest locations (i.e., dependent variable). Spatial autocorrelation occurs when sample locations are spatially related to one another, which violates the statistical assumption of

independence of observations and can lead to confounding predictions (Dormann et al., 2007). In the past, accounting for spatial autocorrelation required complex statistical methods. However, recent studies (Diniz-Filho and Bini, 2005; Griffith & Peres-Neto, 2006; Dormann et al., 2007; Václavík et al., 2012) have explored the use of eigenvector filters as additional explanatory variables in species distribution and habitat suitability modeling. The philosophy behind this strategy is that the spatial autocorrelation can be accounted for by transforming the spatial arrangement of the dependent variable into a set of explanatory variables that represent the spatial structure at multiple scales.

The impetus of this study is to gain further knowledge of the nesting preferences of the Northern goshawk on the Lewis and Clark National Forest. This forest offers the opportunity to model diverse landscapes in two different sectors of this forest and to assess if goshawk nesting preferences differ between them. Furthermore, this study investigates three methodological strategies to improve model predictions. The objectives of this study are threefold: 1) to assess the use of eigenvector filters to reduce the effects of spatial autocorrelation among nest locations, 2) to investigate the differences between bootstrap and cross-validation replication methods and how they affect model results, and 3) to evaluate how reducing the number of environmental variables improves model performance. It is hoped that this study will not only help to describe and identify Northern goshawk habitat on the Lewis and Clark National Forest, but to also assess the level of expertise and computational requirements needed to successful carry out the methods described in the study and provide that information to wildlife managers to help guide them through the decision-making process of habitat suitability modeling.

LITERATURE REVIEW

Northern Goshawk Ecology

The Northern goshawk (hereafter referred to as goshawk) is the largest Accipitrine rapter in North America and inhabits the temperate parts of the northern hemisphere. This diurnal forest raptor is known for its ability to quickly and agilely maneuver under the dense forest canopy, due to its short, rounded wings and long tail (Mahon et al., 2008). Its nesting habitat has been the focus of several studies and habitat classifications vary widely based on the area of interest. For example, Clough (2000) studied nesting habitat in west central Montana, and classified nesting habitat as 40 acres (16 hectares) directly surrounding the nest tree. McGrath et al. (2003) defined nesting habitat in Northeastern Oregon as 74 acres (30 hectares); while Patla (1997) classified nesting habitat as 80 acres (32 hectares) on the Targhee National Forest. Reynolds et al. (1992) provides possibly the most widely used habitat classification from his comprehensive studies in the Southern Rockies; they describe the goshawks home range as approximately 6000 acres (2428 hectares), and further breaks down the home range into three classes: 1) nesting habitat (30-40 acres or 12-16 hectares), 2) post-fledgling family area (PFA; approximately 420 acres or 170 hectares), and 3) foraging habitat (approximately 5400 acres or 2185 hectares). Graham et al. (1994) suggests that nest areas may fluctuate in size and shape across geographic regions as a result of the topography, the availability of dense tree patches, and forest productivity.

The breeding season generally begins late-March to early-April and extends through late-July (Squires & Reynolds, 1997), with goshawks returning to the same nest area year after year and selecting one of several alternate nest locations established within the home range (Boyce et al., 2006; Squires & Reynolds, 1997).

Northern Goshawk Management

      The National Forest Management Act requires that all national forests create a Land Management Plan (LMP) that dictates management activities within that forest (Congress, 1976). To assist with the implementation of these plans, Forest Planning Rules are set in place to outline specific regulations and guidelines that will assure that the LMPs meet NFMA requirements. The 1982 Forest Planning Rule (USFS, 1982) requires that each national forest must designate a set of Management Indicator Species (MIS) because their population changes are believed to indicate the effects of management activities. There are five categories of MIS, including "plant or animal species selected because their population changes are believed to indicate the effects of management activities on other species of selected major biological communities or on water quality." The goshawk is currently designated as an MIS species on the Lewis and Clark National Forest (LCNF) due to its linkage to mature forest and old growth habitat.

      Samson (2006a) outlines three approaches to goshawk conservation in the USFS Northern Region (The Northern Region includes Northern Idaho, Montana, and portions of North and South Dakota): 1) to summarize the best available information and make predictions of available habitat amounts and distributions (Samson, 2006a,b); 2) follow a grid-based sampling framework (Woodbridge & Hargis, 2006) to provide occurrence data for use in probability estimates; and 3) provide information about goshawk activity and habitat guidelines for use at the project level (Samson, 2006a).

      Currently, many of the national forests are in the process of updating (or are planning to update) their LMPs to reflect a new 2012 Planning Rule in which MIS species are being replaced with a Focal Species concept. In particular,

> [f]ocal species are not surrogates for the status of other species. Focal species monitoring provides information regarding the effectiveness of the plan in providing the ecological

conditions necessary to maintain the diversity of plant and animal communities and the persistence of native species in the plan area (USFS, 2012, 21175).

Under the new planning rule it is likely that the goshawk will not be designated as a Focal Species for all the forests' LMPs. However, it is still likely that the goshawk will continue to be the focus of many legal challenges for the USFS. The goshawk has a long history of legal claims by frequent appellants, and although the goshawk may not be listed as a MIS or focal species on the LCNF, it is still protected under the Migratory Bird Treaty Act of 1918, as well as a variety of federal and state designations (e.g., Bureau of Land Management sensitive species, Montana status) (MTNHP, 2014). Therefore, the pattern of legal claims may continue and the need to track goshawk population, along with their habitat may not subside after the 2012 Planning Rule is implemented.

Northern Goshawk Studies

Although the goshawk has been the subject of much research, there is still a lot to learn about its habitat use within the LCNF, as the findings from previous studies may not be applicable to the preferences shown by nesting goshawk within this area of interest. Reynolds et al. (1992) performed extensive mark-recapture research in the southern Rockies to identify habitat components within a goshawk's home range. Squires and Kennedy (2006), have further studied the habitat needs of the goshawk in the western United States and found that goshawks show nesting preference for certain tree dominance types, tree size, tree canopy cover, vertical structure classes and total basal area. Additional research indicates that goshawks also prefer areas with relatively high canopy cover and open understories (Reynolds et al., 1982; Speiser and Bosakowski, 1987; Hayward and Escano, 1989; Reynolds and Joy, 1998; Cooper and Stevens, 2000; Penteriani 2002). While these studies show that goshawk have stronger nesting preferences towards larger vegetation classes, they may choose the less desirable vegetation classes if the prior is unavailable.

Samson (2006a) reviewed existing goshawk studies and used non-parametric bootstrapping statistical methods and Forest Inventory Analysis (FIA) data to model suitable goshawk habitat and calculate the confidence intervals and standard errors for each forest estimate in the Northern Region of the Forest Service. (Samson 2006). Bush and Lundberg (2008) updated Samson's (2006a) model to correct for several errors found in the original model. The updated model also used the most current version of FIA data, which included updates to the forest canopy cover estimates. After applying these updates to Samson's (2006a) model, Bush and Lundberg (2008) replicated Samson's methods to provide new estimates of suitable goshawk habitat. In 2009, Brewer et al. revised the classifications outlined by Samson (2006a) and Bush and Lundberg (2008) to incorporate local knowledge of the area, as well as current literature (specifically to reflect the knowledge and data collected from the 2005 regional goshawk surveys). Brewer used the new set of classifications to extract vegetation characteristics from the USFS Northern Region Existing Vegetation Mapping Program (VMap), as well as FIA data, to create a geospatial model that displayed areas of the region that are most likely to have goshawk habitat. Together, Samson (2006a), Bush and Lundberg (2008), and Brewer et al. (2009) are the primary studies guiding the management of goshawk habitat in the Northern Region.

In 2012, an extensive analysis of four of Region One's eastside forests (Custer, Helena, Lewis and Clark, and Gallatin National Forests) was conducted to assess whether the previous goshawk habitat classification accurately represented the goshawk's nesting preferences in the eastside forests. Using a combination of VMap and FIA data, Bush et al. (2012) extracted vegetation and stand characteristics from a 40-acre buffer around known nest locations. After evaluating the mean and standard deviation of vegetation classes within the known nest buffers, they found that the current nesting classifications used by Samson (2006a), Bush and Lundberg (2008), and Brewer et al. (2009)

showed narrower parameters, fewer canopy cover classes and fewer tree size classes than was observed for actual nesting goshawks in the eastside forest (Bush et al., 2012). Furthermore, this analysis also concluded that "goshawks are not exclusively using old growth forests for their breeding sites, despite their status as a MIS for old growth forests (Bush et al., 2012; 17)." This indicates that the Samson (2006a) and Brewer et al. (2009) models may have inconsistent predictive performance across the Northern Region. Bush et al. (2012) provided a new classification for eastside goshawk nesting habitat, which groups vegetation characteristics by dominance (Table 1).

Table 1: Vegetation classification for goshawk nesting habitat (modified from Bush et al., 2012)

Author	Data Source	Dominance Type	Tree Size	Canopy Cover	Total Basal Area	Vertical Structure Class	Predicted Acreage / Hectares on LCNF
Samson, (2006)	FIA	- Lodgepole pine - Ponderosa pine - Douglas-fir - Aspen - Hardwood mix - Shade-intolerant conifer mix	13-18"	34%-71%	121-253	1 or 2	19,462 / 7,876
Bush and Lundberg, (2008)	FIA	- Lodgepole pine - Ponderosa pine - Douglas-fir - Aspen - Hardwood mix - Shade-intolerant conifer mix	13-18"	34%-71%	121-253	1 or 2	23,953 / 9,693
Brewer, (2009)	R1 Vmap, FIA	- Lodgepole pine - Ponderosa Pine - Douglas-fir - Shade-intolerant conifer mix	10-15" 15" +	40-60% 60% +	N/A	N/A	Not calculated on LCNF
Bush et al., (2012)	R1Vmap, FIA	- Lodgepole pine - Douglas-fir - Ponderosa pine	5"+ 10"+ 10"+	40%+ 25%+ 25%+	N/A	N/A	526,816 / 213,195

Habitat Suitability Modeling

Habitat Suitability Modeling has become a useful method for managers to identify areas that could provide suitable habitat for a given species. Many modeling approaches have been developed to analyze presence-only data to create species distribution models, including envelope models such as BIOCLIM, ecological niche factor analyses such as BIOMAPPER, rule sets derived with genetic algorithms such as GARP, and multivariate distances such as DOMAIN and LIVES (Maxell, 2009). The maximum entropy algorithm used in the statistical software Maxent (Phillips et al., 2011) has been evaluated by several researchers (e.g., Phillips et al., 2004; Elith et al., 2006; Phillips et al., 2006; Phillips and Dudık, 2008) and found to be as good as or superior to most other presence-only and presence/absence methods due to its high performance of predicting species distribution and habitat suitability, and its wide array of statistical outputs used for determining variable importance.

Habitat suitability modeling strives to find the relationship between a species and its environment, to create a geographic prediction of the potential species distribution (Elith et al., 2006;); therefore, it strives to define the species' ecological niche. By extracting environmental conditions at the known nest localities, a sample of the species realized niche is provided (a set of conditions that the species is known to actually inhabit; Phillips et al., 2006). More specifically,

> [a] niche based model thus represents an approximation of the species' realized niche in the study area and environmental dimensions being considered… Areas that satisfy the conditions of a species' fundamental niche represent its potential distribution, whereas the geographic areas it actually inhabits constitutes its realized distribution (Phillips et al., 2006, 232).

Maxent Algorithm

The maximum entropy approach to habitat suitability modeling, as applied by the machine learning statistical program Maxent (version 3.3.3k), strives to find the maximum entropy

(dispersedness) between probability distributions in geographic space (Phillips et al., 2009). In other words,

> [o]ne maximizes the uncertainty (entropy) of the predicted distribution in order to obtain the most conservative estimate possible, given the data. Therefore, the predicted distribution contains as little information as possible about which cell is most likely to contain an individual, which corresponds to a prediction that is as uniform as possible in geographic space (Merow et al., 2013, 42).

Elith et al. (2011) provides a third equivalent explanation in which Maxent strives to minimize the relative entropy of probability distributions in environmental space (Elith et al., 2011). This explanation will be used to further understand the implementation of Maxent described below. The Maxent distribution has been described as being similar to the Gibbs distribution (Phillips et al., 2006; Elith et al., 2011), taking the form of:

$$f_1(z) = f(z)e^{\,n(z)}$$

where:

$$n(z) = \text{æ} + \text{ß} \bullet h(z)$$

æ is a normalizing constant that ensures that $f_1(z)$ integrates (sums) to 1.

the vector of feature is $h(z)$,

and the vector of coefficients, ß.

A background sample is created by selecting 10,000 random points which represent a range of environmental conditions available within the study area (Phillips et al., 2009). Using the environmental data from these 10,000 random points, $f(z)$, and the environmental data from our nest locations, $f_1(z)$, Maxent determines the ratio of $f_1(z)/f(z)$ which is known as Maxent's raw output. An estimate of $f_1(z)$ that is consistent with the occurrence data were used to determine the ratio, and since many distributions are possible, the distribution that is closest to $f(z)$ is chosen to minimize the

distance (relative entropy) between the two (Elith et al., 2011). Thus, "[t]he species' distribution is thus estimated by minimizing the distance between $f_1(z)$ and $f(z)$ subject to constraining the mean by f_1 (and the means of other covariates) to be close to the mean across presence locations" (Elith et al., 2011). The constraints are then extended from being constraints on the means of covariates to being constraints on the means of the "features." Because a species' response to the independent variables can be complex (or non-linear) a "feature" is derived from the original covariates as a form of transformation (Elith et al., 2011). Maxent currently has six features: linear, product, quadratic, hinge, threshold, and categorical (along with an auto feature which will automatically choose the best features given the data). However, Maxent will restrict the model to use only simple features[1] if the occurrence data are limited, as fewer samples limit the information for determining the relationships between the species and its environment (Barry and Elith, 2006). To avoid over-fitting, Maxent applies L1-Regularization so that the model focuses on the most important covariates and relaxes the output so that the values are close to the mean and not exactly equal to the mean (Dudik et al., 2006).

Spatial Eigenvector Mapping

Analyzing spatial autocorrelation has become an important consideration for species habitat modeling, as addressing the issue can lead to insights that may have been otherwise overlooked, while ignoring this issue can lead to false conclusions about ecological relationships (Lichstein et al., 2002). Spatial autocorrelation occurs when observations at nearby locations are not independent from one another (Dormann et al., 2007). In the past, accounting for spatial autocorrelation in habitat suitability modeling was not commonly applied and known methods were often complex. Recently, however, several methods are being utilized to account for spatial autocorrelation, including Spatial

---

[1] Linear features are always used, quadratic features are additionally used if the sample size is greater than 10, hinge features are used in conjunction with linear and quadratic features if the sample size is greater than 15, and threshold and product features are included with at least 80 samples (Elith et al., 2011).

Eigenvector Mapping (SEVM). The SEVM focuses on deconstructing the spatial structure of the dependent variable (in this instance, nest location) to extract different characteristics that describe the unique structure. This can be performed through a Principal Component Analysis (PCA) to identify the underlying structure of the data in multiple dimensions. The PCA defines the directions (or vectors) in which the most variability appears. In other words, if you were to take the nest structure and convert it into a multi-dimensional space, the first principal component would be the vector that describes the greatest variability within data (or the vector that defines the widest spread within the data). The vector, or line itself, is called the eigenvector and describes the direction of the principle component. Associated with the eigenvector is the eigenvalue, which quantifies the amount of variability in the data in the given direction (eigenvector); SEVM uses these methods to create a spatial filter that describes the spatial structure of each eigenvector. These spatial filters are often called eigenvector filters and can be included in habitat suitability modeling as additional explanatory variables to help "filter" out spatial autocorrelation, so the model can focus on deeper ecological relationships. This study utilized a stand-alone freeware application called Spatial Analysis in Macroecology (SAM) to perform the SEVM analysis (Rangel et al., 2010).

METHODS

Since the principles relating to Maxent and SEVM modeling have been discussed in the previous section, the study area, data, and methods related to the implementation of the software are described below. A flowchart summarizing these methods can be found in Figure 1 (located in the Appendix with all other figures).

Study Area

The LCNF is located in northwest and central Montana. The LCNF consists of nearly 1.8 million acres with elevation ranging from 4500 feet to 9362 feet. Nearly half of the forested land is designated as Wilderness areas including the Bob Marshall, Great Bear and Scapegoat Wildernesses (all in the Rocky Mountain Division). The LCNF also consists of seven mountain ranges, including island ranges such as the Big and Little Snowy Mountains, and the Castle Mountain Range. Glacier National Park borders the LCNF to the North (USFS, 2014; Montana Wilderness Association, 2014).

For the purpose of this study, the LCNF was split into two separate modeling areas: 1) The Rocky Mountain Division, and 2) the Jefferson Division. However, small portions of the Jefferson Division have been removed from the modeling effort due to their isolation from the main forest (Figure 2). The Rocky Mountain Division to the west and the Jefferson Division to the east. The Rocky Mountain division offers moist landscapes with high mountain peaks and coniferous forests, while the Jefferson Division offers wide-ranging grassy meadows edged with lodge pole and limber pine, along with isolated pockets of wooded forests and valley bottoms. Dry limestone hills and glaciated headlands can also be observed (USFS, 2014; Montana Wilderness Association, 2014).

The Forest Service oversees many management activities within the study area including timber, fire, recreation, grazing, and mineral programs. Fire, along with insects and disease, frequent the area causing widespread changes in the landscape.

Nest Occurrence Data

Nest data were collected on the Lewis and Clark National Forest during the 2006-2013 breeding seasons following the protocols in the Northern Goshawk Inventory and Monitoring Technical Guide (Woodbridge & Hargis, 2006). Stratified random samples were taken from within 1482 acre (600 hectare) primary sampling units (PSUs). This grid based framework was overlain on potential goshawk nesting and foraging habitat before sampling took place. The PSUs are approximately the same size as one goshawk territory; therefore positive nest detection within a PSU is likely the result of one breeding pair. When possible, each PSU was sampled twice within a given breeding season, once during the nestling period (late-May to early-July) and once during the juvenile stage (late-June to mid-August). Following standardized protocols, the PSUs were sampled using the acoustical broadcast call of the Northern goshawk. The sampling grid in each PSU

> "[c]omprises 120 call stations located on 10 transects that are 250 meters apart, with 12 call stations per transect. Call stations along each transect are 200 meters apart, and adjacent transect stations are offset 100 meters to maximize coverage. The objective is to provide complete survey coverage of the PSU so that all suitable goshawk habitats are within auditory detection distance (roughly 150 meters) of a call point. The procedure is to survey all potential goshawk habitats in the PSU until a detection is made or until all potentially suitable habitat within the PSU is completely surveyed. If a detection occurs, the PSU is recorded as having goshawk presence and the survey is ended. If a detection does not occur, the surveyors continue to survey at call points with increasingly less likelihood of goshawk presence" (Woodbridge & Hargis, 2006).

These surveys resulted in the identification of 156 nest locations on the Jefferson division, and 72 nest locations on the Rocky Mountain division. However, some of these known nest locations were not suitable for modeling efforts. Nests were removed from the modeling effort if 1) the nest

had more than 10% vegetation change within a forty acre buffer around the nest since its last recorded use, due to any combination of disturbances (fire, insects and disease, timber management, prescribed fire, etc. as indicated by the Forest Service Activity Tracking System (FACTS; Bush et al., 2012)); or 2) the nest was not able to be clearly identified as an active nest during the survey period. The survey identified several alternate nests that could have been active in previous years, but since the presence of nesting goshawks could not be confirmed, the nest was removed from the modeling effort to ensure accuracy in the sample set. Cooper's hawk, Great horned owls, and several other species use similar nest structures as the Northern goshawk. Therefore, this confirmation of goshawk occupancy becomes important. This analysis of nest data significantly reduced the number of nests of the Jefferson Division to 49 active nests, and 35 active nests on the Rocky Mountain division. All active nests within each dataset were modeled equally as each represents a sample of goshawk nesting preferences, even though the individual nests were acquired or occupied across multiple years and clusters of nests may have been used by the same or differing goshawks in alternating years.

Environmental Variables

As previously noted, Reynolds et al. (1992), Boyce et al. (2006) and Squires and Kennedy (2006) have all studied the habitat needs of the goshawk and have found that goshawks show nesting preference for specific tree dominance types, tree size, tree canopy cover, vertical structure classes and total basal area. Furthermore, topographic and climatic information at goshawk nest sites may also provide insights into the nest selection process. Mcgrath et al. (2003) found that goshawks showed preferences towards low topographic position and Dawson and Bortolotti (2000) found that climatic conditions may play a role in nest success and nest mortality for American kestrels, through direct and indirect effects of prey availability, predation, and fledgling growth. Additionally, both topographic and climatic variables may also drive vegetation characteristics. A preliminary list of

biologically relevant variables was developed from literature and expert opinions (Table 2). Due to data limitations, some vegetation variables had to be removed from this modeling effort.

Table 2: List of environmental variables

Environmental Variable	Source	Description
	Vegetation	
Tree Dominance Type	USFS R1 Existing Vegetation Mapping Program – Base Level: Dominance Group 6040 "Tree dominance group 6040 is based on two thresholds of tree abundance: 60% and 40%. If the single most abundant tree species comprises at least 60% of the total abundance of the classification attribute, the class assigned is the species" PLANTS code. If the most abundant species comprises less than 60% and at least 40% of the classification attribute, the class assigned is the species PLANTS code with a suffix of the tree lifeform subclass (USFS, 2009)."	NonForest represents all pixels which were classified as herbs, grasses, shrub, water, spveg or urban in the Dom_GRP_6040 category
		Aspen
		Aspen – shade tolerant mix*
		Douglas-fir
		Douglas-fir-shade intolerant mix
		Douglas-fir-shade tolerant mix
		Engleman spruce
		Engleman spruce-shade intolerant mix*
		Engleman spruce-shade tolerant mix
		Limber pine
		Limber pine – shade intolerant mix*
		Lodgepole pine
		Lodgepole pine-shade intolerant mix
		Lodgepole pine-shade tolerant mix
		Ponderosa pine**
		Ponderosa pine-shade intolerant mix**
		Shade Intolerant Mix*
		Shade Tolerant Mix*
		Subalpine fir
		Subalpine fir-shade intolerant mix
		Subalpine fir-shade tolerant mix
		Whitebark pine
		Whitebark pine – shade intolerant mix*
		Whitebark pine – shaded tolerant mix*
Tree Canopy Cover	USFS R1 Existing Vegetation Mapping Program – Base Level: "The proportion of the forest floor covered by the vertical projection of the tree crowns (USFS, 2009)."	Canopy cover of 10-24%
		Canopy cover of 25-39%
		Canopy cover of 40-59%
		Canopy cover of 60% or greater
Tree Size Class	USFS R1 Existing Vegetation Mapping Program – Base Level:	Diameter at Breast Height is 0-4"

	Diameter at Breast Height	Diameter at Breast Height is 5-9"
	"The proportion of the predominant diameter class of live trees within a setting (USFS, 2009)."	Diameter at Breast Height is 10-14"
		Diameter at Breast Height is 15" or greater
Tree Basal Area	*Removed from study due to lack of data*	
Tree Vertical Structure	*Removed from study due to lack of data*	
	Topography	
Slope		Slope in degrees
Aspect - Transformed	US Forest Service, R1 Geospatial Group	Transformed Aspect: values range from −1 to 1
Elevation		Digital Elevation Model (DEM) in feet
Roughness		Terrain variability
	Climate	
Mean Maximum Temperature	Prism Climate Group, Oregon State University	DEM-aided interpolation of Maximum Temperature (degrees Celsius)
Mean Temperature	30 year normal: 1981-2010	DEM-aided interpolation of Mean Temperature (degrees Celsius)
Mean Minimum Temperature	Temperature and Precipitation values are averages during the breeding season (March – July)	DEM-aided interpolation of Minimum Temperature (degrees Celsius)
Mean Precipitation		Precipitation (millimeters)

* indicates environmental variables that were used for modeling only on the Rocky Mountain division

** indicates variables only used on the Jefferson division

Topographical Data

All topographical data were derived from a 30-meter DEM (USGS) by the USFS Region One

Geospatial Group. Variables were clipped, resampled and snapped to the study area boundaries.

Climate Data

The climatic data were acquired from the PRISM climate group (Oregon State University,

2014), including the annual mean precipitation and annual mean maximum, mean, and minimum

temperature as 30-year normals between 1981 and 2010 at a 800 meter resolution. The temperature

data were resampled using the DEM-aided interpolation methods described by Willmott and

Matsuura (1995) to help incorporate temperature changes due to elevation. During this process, the temperature data were also resampled to a 30 meter resolution. The annual precipitation layer was clipped and resampled to a 30 meter resolution using bilinear interpolation. All climatic variables were snapped to the 30 meter DEM during the resampling methods to ensure consistency in cell size and extent among variables.

Vegetation Data

An initial compilation of potential vegetation variables were defined through literature review and expert opinion, and then included based on data availability. Tree dominance, canopy cover and tree size were included as data were readily available through VMap (USDA 2009). The LCNF VMap was created in 2005 and is derived from satellite and airborne imagery. The Base-level VMap (polygon feature class) was chosen for this study as it is a continuous dataset with a resolution of 10 meters and is designed for studies at the forest-level; FIA data were not considered for use in this study as they only represent plot information for specific tree stands and are not continuous across the study area.

To prepare the data for analysis, the VMap Base-level feature class was converted to a raster and then resampled and snapped to match the cell-size and extent of the 30 meter DEM. To reduce modeling inaccuracies, the categorical vegetation data were converted to a continuous dataset using methods described by Olson et al. (2014). Each individual category within all three vegetation layers was converted to its own binary environmental variable where presence is assigned a value of 1 and non-presence is assigned a value of 0. This created four canopy cover variables (10-24%, 25-39%, 40-59%, and 60% or greater), four tree size variables (0-4", 5-9", 10-14" and 15" or greater), and 24 tree dominance variables (Douglas-fir, ponderosa pine, lodgepole pine, etc). The pixel values within each of these newly created variables were then processed using the Focal Statistics tool in ArcGIS

10.1 and a circular moving window approach to represent the approximate proportion of each vegetation type present in a 3×3 neighborhood. For example, if 5 of the 9 pixels within the neighborhood had the value of 1, then the pixel was assigned a value of 0.55. This value means that approximately 55% of the pixels in the neighborhood were classified as a "Douglas-fir" dominance type.

Maxent Implementation

A point pattern analysis using the Ripley's K function was performed using the statistical analysis package R (R Core Group, 2014) to estimate the presence of spatial dependence among nest locations and the need to address this during the modeling process. Maxent was then used for the analysis to predict suitable goshawk nesting habitat. Four model scenarios were carried out for both the Rocky Mountain and the Jefferson Division: Scenario 1 is a full model using all environmental variables; Scenario 2 is a full model using all environmental variables plus 10 eigenvector filters; Scenario 3 is a model using the top 12 most significant variables as determined by scenario 1; and Scenario 4 is a model using the top 12 most significant variables as determined by scenario 1 plus the first 15 eigenvector filters.

For each of the scenarios described above, two models were created to examine the difference between bootstrap and cross-validation methods and determine if one method yields better results than the other. Bootstrapping methods create sample sets for replicated runs using sampling with replacement, while cross-validation methods divides the samples into folds (subsets), and each fold in turn is reserved for testing. Sub-sampling methods were not explored in this study as little literature was found on this method. Each model was set to replicate 10 times, and random seed was selected; when selected, a different random set of test/training partitions will be made and a different random subset of the background will be used (Phillips, 2008). However, random seed is automatically

turned off when cross-validation methods are used. Twenty-five percent of the occurrence location were set aside testing the model, leaving the remaining 75% for training. To ensure that the models had time to converge to the threshold of 0.00001, the maximum number of iterations was set to 5000. The aforementioned are settings within Maxent and were determined by following selection made in other literature (Václavík et al., 2012; Olson et al., 2014; Young et al., 2011; Phillips, 2008). A comparison of the 16 models developed in this study can be found in Table 3.

Table 3: Comparison of model setting for the four scenarios.

Division	Method	Scenario	Number of Variables	Number of Eigenvectors
Jefferson	Bootstrap	1	34	0
Jefferson	Bootstrap	2	34	10
Jefferson	Bootstrap	3	12	0
Jefferson	Bootstrap	4	12	15
Jefferson	Cross-validation	1	34	0
Jefferson	Cross-validation	2	34	10
Jefferson	Cross-validation	3	12	0
Jefferson	Cross-validation	4	12	15
Rocky Mountain	Bootstrap	1	37	0
Rocky Mountain	Bootstrap	2	37	10
Rocky Mountain	Bootstrap	3	12	0
Rocky Mountain	Bootstrap	4	12	15
Rocky Mountain	Cross-validation	1	37	0
Rocky Mountain	Cross-validation	2	37	10
Rocky Mountain	Cross-validation	3	12	0
Rocky Mountain	Cross-validation	4	12	15

The residuals from the models created in scenario 1 were submitted separately to a Moran's I analysis, using SAM, to determine the distances at which spatial autocorrelation remained significant. The distance at which spatial autocorrelation is no longer considered significant can be determined by the intercept of the Moran's I correlogram and can be used as a truncation threshold for the SEVM analysis (Václavík et al., 2012). By applying a truncation threshold, the SEVM focuses on the variation within the spatial structure that is related to spatial autocorrelation.

Ideally, all positive eigenvalues with a Moran's I coefficient greater than 0.5 would have been incorporated into Scenario 2 as additional explanatory variables to describe the spatial structure at multiple scales (eigenvectors with high eigenvalues describe more global variation, while low eigenvalues represent more local variation of the spatial structure (Griffith and Peres-Neto, 2006; Terribile and Diniz-Filho, 2009; Václavík et al., 2012). However, due to computational limitations, only the first 10 eigenvectors were included as additional environmental variables in a second run of the full model.

Scenarios 3 and 4 were created to incorporate more eigenvectors into the model, as these would theoretically further describe the spatial structure of the nest locations. Several environmental variables had little contribution to the models created in Scenario 1. Therefore, to reserve computational power for the addition of eigenvector filters, only 12 environmental variables were incorporated into Scenario 3. These 12 variables were carefully selected based on the response curves, percent contribution, permutation importance, and explanatory power from the results of Scenario 1. Scenario 3 is therefore a baseline for Scenario 4 which adds the first 15 eigenvectors to Scenario 3.

Eigenvector Analysis

SAM 4.0 provides a user friendly SEVM module which utilizes a grid that describes nest richness across the study area and applies a truncation threshold to calculate the eigenvectors from a centered-distance matrix. A truncation value of 6 km was used for both modeling methods on the Rocky Mountain Division, along with the cross-validation models on the Jefferson Division. Bootstrapping methods on the Jefferson Division, however, had a truncation threshold of 8 km likely due to the difference in nest locations used for testing and training. Over 1700 spatial filters were created for each of the full models created in Scenario 1. Generally accepted guidelines for

determining which eigenvector filters to use in habitat modeling have not yet been defined in literature, though several methods have been explored. Václavík et al. (2012) used the first 50 positive eigenvectors and entered them successively as additional environmental variables, retaining only those that improved the models fit or had a significant relationship with the dependent variable; Terribile and Diniz-Filho (2009) selected the positive eigenvector filters in the first distance class which had a Moran's I coefficient greater than 0.1; Blach-Overgaard et al. (2010) used a fixed number of eigenvector filters (either the first 7 or first 14) for further analysis and De Marco et al. (2008) utilized the first 5 positive eigenvector filters.

Eigenvectors which produced a Moran's I coefficient greater than 0.5 were selected for further modeling, resulting in 25-27 eigenvector filters. An attempt to incorporate all these eigenvectors into Scenario 2 was made, but due to the number and high resolution of the initial environmental variables and computational limitations, the model could only handle 10 additional variables, leading to the need development of Scenarios 3 and 4 to further explore the influence of the eigenvector filter. The eigenvector filters were resampled and snapped to match the 30 meter DEM in ArcMap 10.1 prior to further Maxent modeling.

Assessing Model Performance

A comparison of the models performance to accurately predict suitable goshawk nesting habitat was made by evaluating the Area Under the [Receiver Operator Characteristic] Curve (AUC; provided as a Maxent output). The AUC is a measure of model performance with values ranging between 0 and 1, where values lower than 0.5 suggests the model performs no better than random and a value of 1 indicates a perfect prediction. A value of 0.7 or above is considered a well performing model (Hanley & McNeil, 1982).

While the AUC has been a widely used and acceptable method for measuring model performance, several studies suggest that using the AUC as a sole measure of performance can be insufficient, especially when using presence/pseudo-absences such as Maxent does (Allouche et al., 2006; Lobo et al., 2008). Furthermore, the AUC is a threshold-independent measure so that when threshold is applied to determine a binary suitable vs. unsuitable habitat, a secondary threshold-dependent measure is needed, such as the Cohen's Kappa Statistic or True Skills Statistic (TSS; also known as the Hanssen-Kuipper Skill Score; Allouche et al., 2006). For this study, TSS was chosen as a secondary measure of model performance, as TSS is independent of prevalence (unlike Kappa).

Furthermore, TSS accounts for both omission (false negatives, which are cells of known presence being classified as unsuitable habitat, leading to under-prediction) and commission errors (false positives, or areas of known absence being classified as suitable habitat, leading to over-prediction), unlike the AUC which weights omission and commission errors equally (Allouche et al., 2006; Lobo et al., 2008). The TSS utilizes a confusion matrix and takes the form of:

TSS = [(ad) − (bc)] / [(a+c)(b+d)]

where:
 a is the number of true positives,
 b is the number of false positives,
 c is the number of true negatives, and
 d is the number of false negatives.

TSS values range from −1 to +1, where +1 indicated perfect performance and values less than 0 represent performance that is no better than random. Values above a 0.5 are considered to indicate good models. TSS values, along with omission and commission rates were calculated in Microsoft Excel.

Model Residuals

       Model residuals were evaluated to assess whether the inclusion of eigenvector filters helped in reducing spatial autocorrelation. Sample prediction values across all replicated runs were averaged for each model and residuals were calculated as (1 – prediction value; Václavík et al., 2012). Model residuals were submitted to SAM and the Moran's I analysis and values were calculated using geographic distances among pairs of samples across 15 distance classes. The test significance was evaluated based using 200 permutations (Václavík et al., 2012).

RESULTS AND DISCUSSION

Model Performance

The AUC values calculated for the Maxent model runs are represented by box and whisker plots in Figure 3. While the initial bootstrap and cross-validation models (Scenario 1) for the Rocky Mountain performed well with an AUC of 0.9423 and 0.9248, adding the first 10 eigenvector filters (Scenario 2) improved the average AUC to 0.9711 and 0.9475, respectively. Furthermore, the models run using just the 12 most influential biological variables from the initial bootstrap and cross-validation models (Scenario 3), also performed well with AUC values of 0.9247 and 0.8979, and both increased to 0.9381 and 0.9499 by including the first 15 eigenvectors (Scenario 4). The Jefferson Division saw similar results with the initial bootstrap and cross-validation models (Scenario 1) performing well with an AUC of 0.9475 and 0.9092. Including the first 10 eigenvectors (Scenario 2) in these models resulted in an increase of AUC to 0.9548 and 0.9525, respectively. Furthermore, the bootstrap and cross-validation models ran in Scenario 3 also performed well with AUC values of 0.9161 and 0.9201. Including the first 15 eigenvectors in these models (Scenario 4), improved the AUC to 0.9363 and 0.9485, respectively.

The TSS values were calculated across all replicated runs for each scenario (Figure 4) based upon the threshold that maximized model specificity (true negatives) and sensitivity (true positives). The models that included eigenvector filters usually outperformed the associated initial model, with the exception of scenario 4 on the Rocky Mountain division which used cross-validation. This model had a slightly lower TSS value of 0.6177, while its associated initial model (Scenario 3) had a value of 0.6432. This inconsistency could be caused by the low number of nest locations set aside for testing. Due to the low sample size, setting aside 25% of the nest locations resulted in just 3 or 4 nest location available for testing. In several instances, the difference between a nest location's prediction

value being classified as a true positive (as indicated by the given threshold) was only off by 0.0001 during the TSS calculation, leading to a higher omission rate. Therefore, what could have been considered as 3 true positives and an omission error of 0.00 was assessed as 2 true positive with an omission rate of 0.33. With small test samples, one false negative can have a large effect on model performance evaluation as higher omission and commission rates lead to a lower TSS value and therefore lower model performance. This could also explain the higher levels of variability in TSS scores using cross-validation methods versus bootstrap methods. Furthermore, comparing TSS values from Scenario 1 and 3 for all methods indicates that variable reduction did not significantly improve model performance for this study.

Maxent provides several thresholds in which "suitable vs. unsuitable" habitat can be defined. Figures 5-8 show the average training omission rate and predicted area at various cumulative thresholds across the replicated runs as provided by Maxent. While bootstrapping methods produced models with less variability across the replicated runs, the test omission rate falls further from the predicted omission rate. This typically occurs when the test and training data are not independent of one another (Phillips, 2008). Even though 25% of the data was withheld for testing, bootstrapping methods which employ sampling with replacement may not produce independent subsets of test and training data among replicates when spatial autocorrelation exists amongst the nest location. Cross-validation on the other hand, shows a large amount of variation in the omission rate across replicated runs, but the average omission rate lies closer to the predicted rate.

Omission and commission rates were also determined at the threshold that maximized model specificity and sensitivity (Figures 9 and 10). Comparing omission and commission rates among the model scenarios, the inclusion of spatial structure through eigenvector filters appeared to lower the average commission rates when compared to the initial models. Omission rates do not show a clear

improvement across models, but as mentioned previously, omission rates are highly sensitive to small sample sizes.

Model Residuals

Models residuals were evaluated for each study area and method (Figure 11 and 12). The Moran's I correlograms show that while significant autocorrelation still exists, the inclusion of eigenvector filters may have helped reduce this effect. For example, Figure 10b shows a slight reduction in Moran's I values around 20 km (for Scenario 1, the Moran's I value dropped from 0.226 to 0.088). As the eigenvector filters represent spatial autocorrelation at multiple scales, this could suggest that one of the filters used in this model, captured the spatial structure of autocorrelation at 20 km. More research is needed to determine if the inclusion of more eigenvector filters in Maxent would further reduce spatial autocorrelation across all distances.

Furthermore, the spatial relationship displayed in the model residuals could suggest that the model is missing an explanatory variable that is important to describing goshawk nesting habitat. Due to data unavailability, basal area and vertical structures of the nest stands were not able to be included in the modeling process, even though these variables were described in literature as being important to goshawk nesting habitat (Boyce et al., 2006; Squires and Kennedy, 2006; Brewer et al., 2009). Inclusion of these variables may help reduce spatial autocorrelation among model residuals.

Variable Importance

Several outputs were provided by Maxent to assess variable importance, including response curves; the percent contribution each variables has in creating the model; and three jackknife analyses that evaluated the level of contribution the test and training data have on the overall model gain and AUC. However, reviewing these outputs in isolation may not yield the best determination of variable

importance as each provides valuable information about the variable's role in model predictions. Therefore, all outputs were reviewed in turn, and variable importance was determined if 1) the response curve was biologically reasonable, 2) the variable had a high contribution to the models test and training gain, and 3) the variable showed a reduction in the model's performance if removed. Variables from all scenarios were evaluated for each method on an individual basis, and a list of the most important variables from each model is shown in Table 3. These variables were further broken down into three sub-categories representing their level of contribution to each model; variables symbolized in bold represent the highest contribution; italic variables represent very high contribution; and variables without symbolization represents high contribution to the overall model. Bootstrap and cross-validation methods revealed similar variable importance.

Of the potential environmental variables evaluated in this study, twelve were found to be important to goshawk nesting habitat in Scenario 1 on the Jefferson division. Precipitation during the breeding season, tree size in the 10-14" range, canopy cover between 10-24%, and elevation were the most important and provided the highest contribution to the model, followed closely by Douglas-fir, and slope.

Scenario 2 revealed that the addition of eigenvector filters can affect variable importance. While the variables with the highest contribution from Scenario 1 remained relatively the same (with the exclusion of elevation), Scenario 2 revealed an additional five variables that were also important to modeling goshawk nesting habitat (canopy cover ≥ 60%, canopy cover 25-39%, maximum temperature, roughness, and tree size ≥ 15") and removed two variables from the original list of important variables (minimum temperature and tree size between 5-9").

Table 4. Important variables as determined by evaluating the response curves, percent contribution, and jackknife tests.

Jefferson Division				Rocky Mountain Division			
Scenario 1	*Scenario 2*	*Scenario 3*	*Scenario 4*	*Scenario1*	*Scenario2*	*Scenario 3*	*Scenario 4*
Precipitation	Precipitation	Precipitation	Tree Size 10-14"	Elevation	Elevation	Douglas-fir	Douglas-fir
Tree Size 10-14"	Tree Size 10-14"	Tree Size 10-14"	Precipitation	Douglas-fir	Douglas-fir	Elevation	Elevation
Canopy Cover 10-24%	Canopy Cover 10-24%	Canopy Cover 10-24%	Elevation	Canopy Cover 25-39%	Precipitation	Precipitation	Precipitation
Elevation	*Douglas-fir*	Douglas-fir	Douglas-fir	Canopy Cover ≥ 60%	Canopy Cover 25-39%	Canopy Cover ≥ 60%	Canopy Cover ≥ 60%
Douglas-fir	*Elevation*	*Elevation*	Canopy Cover 10-24%	*Minimum Temperature*	Canopy Cover ≥ 60%	Canopy Cover 25-39%	Canopy Cover 25-39%
Slope	*Canopy Cover 25-39%*	*Aspect*	Tree Size 5-9"	*Precipitation*	Canopy Cover 10-24%	Minimum Temperature	Canopy Cover 10-24%
Canopy Cover 25-39%	*Limber Pine*	Slope	Aspect	Canopy Cover 10-24%	Minimum Temperature	Canopy Cover 10-24%	Tree Size 5-9"
Lodge-pole pine: Shade Intolerant Mix	Aspect	Minimum Temperature	Limber Pine	Tree Size 5-9"	Roughness	Tree Size 5-9"	Minimum Temperature
Limber Pine	Maximum Temperature	Canopy Cover 25-39%	Lodge-pole pine: Shade Intolerant Mix	Aspect	Slope	Roughness	Roughness
Minimum Temperature	Slope	Lodge-pole pine: Shade Intolerant Mix	Canopy Cover 25-39%	Roughness	Lodge-pole pine: Shade Intolerant Mix	Aspect	Slope
Tree Size 5-9"	Lodge-pole pine: Shade Intolerant Mix	Tree Size 5-9"	Slope	Tree Size ≥ 15"	Aspect	Slope	Aspect
Aspect	Canopy Cover 40-59%	Limber Pine	Minimum Temperature	Slope	Tree Size 5-9"	Tree Size ≥ 15"	Tree Size ≥ 15"
	Roughness				Tree Size 10-14"		
	Canopy Cover ≥ 60%				Canopy Cover 40-59%		
	Tree Size ≥15"						

When spatial autocorrelation is present, observations have characteristics that are similar to one another (and therefore does not bring forth one full degree-of-freedom; Legendre, 1993). Therefore, important characteristics about each individual observation can be missed. By quantifying the autocorrelation among nest locations, the addition of eigenvector filters allows the model to more clearly examine the relationships between individual nest locations and environmental variables. This could explain why the addition of eigenvector filters had a shift on variable importance, and identified additional important variables.

The Rocky Mountain Division saw slightly different results, with the most important variables remaining fairly stable across all scenarios. For Scenario 1, Douglas-fir dominance type, elevation, and canopy cover between 25-39% and greater than 60% had the highest contribution, followed closely by minimum temperature and precipitation during the breeding season. Scenario 2 had the same important variables but placed more importance on precipitation and less on minimum temperature during the breading season. Scenario 2 also identified three new important variables including canopy cover between 40-59%, lodge-pole pine dominance type with a mix of shade intolerant species, and tree size between 10-14".

Scenarios 3 and 4 for both divisions may provide a more definitive selection of the most important variables as the model focuses purely on creating a prediction using only the top variables. Therefore, the variable contribution from the less important variables in Scenario 1 and 2 are redistributed, revealing a more definitive selection of important variables.

In comparing variable importance, it becomes apparent that goshawks show different preferences across the two divisions. Model results for the Rocky Mountain division showed much stronger preferences towards high canopy cover classes than those for the Jefferson division, which showed a stronger preference towards a combination of tree size and canopy cover classes. As mentioned previously, the two divisions are comprised of quite different landscapes. The Jefferson division is not as heavily forested as the Rocky Mountain division and generally produces smaller tree size classes. This can lead to lower canopy cover classes in which the goshawk can utilize a larger tree size class to obtain a more desirable canopy cover around the nest. While this selection may not be highly preferable to goshawks, it provides the best habitat available within this division, suggesting that goshawks may be opportunistic in their habitat selection. Furthermore, elevation,

precipitation and Douglas-fir dominance type remained consistently important across both divisions. As mentioned previously, high levels of precipitation can reduce nestling survival rates.

Furthermore, this study has determined a different range of important vegetation variables as compared to other study efforts (Table 5.) This study identified a substantial reduction in important tree dominance types than has been identified in previous studies, with Douglas-fir and lodgepole pine (shade-intolerant mix) having the highest importance. While tree size values remained fairly similar to those identified in previous studies, canopy cover revealed a wider range of classes in this study. This may further support the conclusion that goshawks show preferences towards high canopy cover classes but may select lower canopy covers if the preferred is not available. Furthermore, previous studies modeled the Lewis and Clark National Forest as a whole and did not account for spatial autocorrelation.

In addition to vegetation variables, this study also revealed that goshawks may choose nesting habitat based on other factors, such as climate and topography. In all the models, both elevation and the amount of precipitation received during the breeding season played an important role in determining suitable nesting habitat.

Model Predictions

Figures 13-16 display the habitat predictions created by the Maxent models. The models that incorporated spatial structure through eigenvector filters (Scenarios 2 and 4) appear to have produced less conservative predictions in terms of suitable versus unsuitable habitat. This is likely due to the reduction of commission error.

In order to calculate the acreage of suitable habitat predicted by each model, the average

Table 5: A comparison of vegetation parameters describing nesting goshawk habitat modified from Bush et al., (2012)*

Author	Dominance Type	Tree Size	Canopy Cover	Predicted Acreage / Hectares on LCNF	
Samson (2006)	- Lodgepole pine - Ponderosa pine - Douglas-fir - Aspen - Hardwood mix - Shade-intolerant conifer mix	13-18"	34%-71%	19,462 / 7,876	
Bush and Lundberg (2008)	- Lodgepole pine - Ponderosa pine - Douglas-fir - Aspen - Hardwood mix - Shade-intolerant conifer mix	13-18"	34%-71%	23,953 / 9,693	
Brewer (2009)	- Lodgepole pine - Ponderosa Pine - Douglas-fir - Shade-intolerant conifer mix	10-15" 15" +	40-60% 60% +	Not calculated on the LCNF	
Bush et al. (2012)	- Lodgepole pine	5"+	40%+	526,816 / 213,195	
	- Douglas-fir	10"+	25%+		
	- Ponderosa pine	10"+	25%+		
This study	Jefferson	Jefferson	Jefferson	Jefferson	Entire Study Area
	- Douglas-fir - Limber pine - Lodgepole pine: shade intolerant mix	10-14" 15" +	10-24% 25-39% 40-59% 60% +	*40,999 / 16,591	*51733 / 20,935
	Rocky Mountain	Rocky Mountain	Rocky Mountain	Rocky Mountain	
	- Douglas-fir - Lodgepole pine: shade intolerant mix	5-9" 10-14"	10-24% 25-39% 40-59% 60% +	*10,734 / 4,344	

*Vegetation types and acreage determined from Scenario 2 using bootstrap methods. It is important to note that the entire Jefferson division was not included in this modeling effort due to areas of isolation. Therefore comparison between acreage estimates in this study and previous efforts should not be made directly.

logistic prediction was converted to a binary (suitable vs. unsuitable) variable using the threshold (generated by Maxent) that maximized sensitivity and specificity. Total acreage for each model can be compared in Figure 17 and Table 6. The inclusion of eigenvectors resulted in a substantial reduction in habitat estimates, ranging from 38-87% fewer acres. Maxent also provides additional thresholds for which suitable vs. unsuitable habitat can be define, however these thresholds were not explored in this study.

Model Selection

While all of the 16 models produced in this study were considered to be good, those that analyzed all of the biologically relevant variables using bootstrap replication methods and eigenvector filters provided the best overall results (Scenario 2). While this model may not be the most parsimonious, it provides insight to goshawk nesting preferences and provides managers with information about the most important environmental variables for identifying potential goshawk nesting habitat. While incorporating eigenvector filters nearly always improved the initial model results, the amount of time, geoprocessing and computational demands were quite high. If computing resources and the time needed to produce and utilize these filters are not available, one might consider excluding eigenvector filters from the modeling effort as this study has shown that these models still produce good results (Table 6). However, not utilizing eigenvector filters may lead to an over-prediction of suitable goshawk nesting habitat.

Table 6: Summary of model results

Division	Method	Scenario	Number of Variables	Number of Eigenvectors	AUC	TSS	Omission Rate	Commission Rate	Acreage/ Hectares
Jefferson	Bootstrap	1	34	0	0.9475	0.7454	0.1583	0.09629	122,190 / 49,449
Jefferson	Bootstrap	2	34	10	0.9548	0.8105	0.1167	0.07288	40,999 / 16,592
Jefferson	Bootstrap	3	12	0	0.9161	0.7182	0.1333	0.14845	107,990 / 43,702
Jefferson	Bootstrap	4	12	15	0.9363	0.8084	0.1000	0.09159	59,428 / 24,050
Jefferson	Cross-validation	1	34	0	0.9092	0.7233	0.0850	0.1917	156,030 / 63,143
Jefferson	Cross-validation	2	34	10	0.9525	0.8185	0.1000	0.08154	48,385 / 19,581
Jefferson	Cross-validation	3	12	0	0.9201	0.7436	0.1050	0.15137	97,812 / 39,583
Jefferson	Cross-validation	4	12	15	0.9485	0.7681	0.1450	0.08686	61,047 / 24,705
Rocky Mountain	Bootstrap	1	37	0	0.9423	0.743	0.1500	0.10700	81,872 / 33,132
Rocky Mountain	Bootstrap	2	37	10	0.9711	0.8335	0.1000	0.06651	10,734 / 4,344
Rocky Mountain	Bootstrap	3	12	0	0.9247	0.7343	0.1250	0.14067	107,491 / 43,500
Rocky Mountain	Bootstrap	4	12	15	0.9381	0.7665	0.1750	0.05850	22,484 / 9099
Rocky Mountain	Cross-validation	1	37	0	0.9248	0.6585	0.1833	0.15818	81,288 / 32,896
Rocky Mountain	Cross-validation	2	37	10	0.9475	0.804	0.0833	0.11269	39,350 / 15,924
Rocky Mountain	Cross-validation	3	12	0	0.8979	0.6432	0.1250	0.23181	139,214 / 56,338
Rocky Mountain	Cross-validation	4	12	15	0.9499	0.6177	0.24177	0.14061	53,078 / 21,480

CONCLUSION

While habitat suitability modeling is an effective way for managers to determine areas that are suitable for management and conservation efforts, the methods used to produce the predictions can yield drastically different results. The findings from this study showed that by including spatial structure, (through the use of eigenvector filters as additional environmental variables) model performance can be improved, commission rates can be lowered, and less conservative habitat predictions can be produced. Furthermore, using eigenvector filters may help to reduce spatial autocorrelation among a model's residuals, though substantial spatial autocorrelation might still be present. This could result from the omission of important explanatory variables needed to fully model the type of habitat in question (e.g., basal area, vertical structure, etc.).

In this study, bootstrapping methods revealed slightly higher AUC and TSS values, along with lower omission and commission rates, than cross-validation methods. Additionally, there was considerably less variability among the results of the replicated runs. This suggests that bootstrapping methods may provide better model performance for modeling suitable goshawk nesting habitat. Furthermore, reducing the number of environmental variables did not significantly improve model performance. This is inconsistent with other studies (Burbach, 2011), though it is possible that the reduced models still included too many environmental variables.

In terms of goshawk nesting habitat, the most interesting finding of this study was the difference in variable importance compared to previous studies. The wider range of vegetation parameter values used on the LCNF may suggest that goshawks are more opportunistic than previously thought, selecting areas that provide the best habitat available within the area, even though they may not be highly preferable. While regional goshawk models are good in theory, this study

shows the importance of creating individual models for each forest to identify key environmental variables that are important for goshawk nesting habitat and create more localized predictions.

In light of this study, there are several recommendations that may further improve these methods for future modeling efforts to predict suitable goshawk nesting habitat:

1) Explore the addition of prey availability into the modeling effort;

2) Further reduce the number of environmental variables to 5 or 6, and evaluate model performance;

3) If using cross-validation methods for replication, increase the number of samples reserved for testing to ensure that the rate of omission error will not be heavily inflated;

4) If using bootstrap methods for replication, ensure that test samples are independent from training samples;

Given the results described above, this study suggests using all biologically relevant variables to model suitable goshawk nesting habitat in Maxent, along with bootstrap replication methods. Furthermore, if significant spatial autocorrelation exists, it is recommended that the modeler incorporate spatial structure through eigenvector filters to help describe a deeper relationship among nesting goshawks and their environment, as well as provide less conservative habitat suitability predictions. However, if time, expertise, and computational power are limited, then it may be acceptable to exclude eigenvector filters, as such models developed in this study still had high performance. While many wildlife managers are becoming competent in ArcGIS, it is recommended that they work closely with a GIS specialist if they decided to move forward with methods similar to those used in this study. The geoprocessing involved in this study can be quite extensive and it is important for the individual to understand the concepts involved with each step. Furthermore, they

should work together to review the methods and results provided in this study and choose the model

that best suits their management needs and expertise.

LITERATURE CITED

Allouche O., Tsoar A., & Kadmon R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.

Barry S. & Elith J. (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413–423.

Blach-Overgaard A., Svenning J.-C., Dransfield J., Greve M., & Balslev H. (2010) Determinants of palm species distributions across Africa: the relative roles of climate, non-climatic environmental factors, and spatial constraints. *Ecography*, no–no.

Boyce D.A., Reynolds R.T., & Graham R.T. (2006) Goshawk Status and Management: What Do We Know, What Have We Done, Where Are We Going? *Studies in Avian Biology*, **31**, 312–325.

Brewer L.T., Canfield J.E., Dohmen A.R., Anderson S., Barber J., Hahn B., Johnson W., Kowalski S., Lockman D., Robert D.J., Swisher K.S., & Tomasik E. (2009) Northern Goshawk, Northern Region Overview, Key Findings and Project Considerations. *USDA Forest Service, Northern Region*, 1–54.

Burbach T. (2011) *The Influence of Environmental Variables on Predicting Rare-Plant Habitat in the Nez Perce National Forest.*

Bush R., Hahn B., Barber J., Lehr M., Weldon H., Weston J., & Wolf A. (2012) Eastside Goshawk Nest Analysis. 1–21.

Bush R. & Lundberg R. (2008) Wildlife Habitat Estimate Updates fo rthe Region 1 Conservation Assessment. *USDA Forest Service, Northern Region*, 1–22.

Clough L. (2000) Nesting habitat selection and productivity of northern goshawks in west-central Montana. 95.

Congress (1976) National Forest Management Act Of 1976. **16 U.S.C.**, 1–13.

Cooper J. & Stevens V. (2000) A review of the ecology, management and conservation of the northern goshawk in British Columbia. 31.

Dawson R. & Bortolotti G. (2000) Reproductive success of American kestrels: the role of prey abundance and weather. *The Condor*, 814–822.

Diniz-Filho J.A.F. & Bini L.M. (2005) Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecology and Biogeography*, **14**, 177–185.

Dormann C.F., M. McPherson J., B. Araújo M., Bivand R., Bolliger J., Carl G., G. Davies R., Hirzel A., Jetz W., Daniel Kissling W., Kühn I., Ohlemüller R., R. Peres-Neto P., Reineking B.,

Schröder B., M. Schurr F., & Wilson R. (2007a) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.

Dudik M., Schapire R.E., & Phillips S.J. (2006) Correcting sample selection bias in maximum entropy density estimation. *Advances in neural information processing systems*, **18**, 323–330.

Elith J., Graham C.H., & Al. E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith J., Phillips S.J., Hastie T., Dudík M., Chee Y.E., & Yates C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

Graham R.T., Reynolds R.T., Reiser M.H., Basset R.L., & Boyce D.A. (1994) SUSTAINING FOREST HABITAT FOR THE NORTHERN GOSHAWK : A QUESTION OF SCALE. *Studies in Avian Biology*, 12–17.

Griffith D.A. & Peres-Neto P.R. (2006) Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*, **87**, 2603–2613.

Hanley J.A. & McNeil B.J. (1982) The Meaning and Use of the Area under a Receiver Operating (ROC) Curvel Characteristic. *The Radiological Society of North America, Inc.*, **143**, 29–36.

Hayward G.D. & Escano R.E. (1989) Goshawk Nest-Site Characteristics in Western Montana and Idaho. *The Condor*, **91**, 476–479.

Legendre P. (1993) Spatial Autocorrelation: Trouble or New Paradigm. *Ecology*, **74**, 1659–1673.

Lichstein J.W., Simons T.R., Shriner S.A., & Franzreb K.E. (2002) Spatial Autocorrelation and Autoregressive Models in Ecology. *Ecological Monographs*, **72**, 445–463.

Lobo J.M., Jiménez-Valverde A., & Real R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.

Mahon T., Mcclaren E.L., & Doyle F.I. (2008) Parameterization of the Northern Goshawk (Accipiter gentilis laingi) Habitat Model for Coastal British Columbia . Nesting and Foraging Habitat Suitability Models and Territory Analysis Model. 1–52.

De Marco P., Diniz-Filho J.A.F., & Bini L.M. (2008) Spatial analysis improves species distribution modelling during range expansion. *Biology letters*, **4**, 577–80.

Maxell B.A. (2009) *State-wide assessment of status, predicted distribution, and landscape-level habitat suitability of amphibians and reptiles in Montana.* The University of Montana,

Mcgrath M.T., DeStefano S., Riggs R.A., Irwin L.L., & Roloff J. (2003) Spatially Explicit Influences on Northern Goshawk Nesting Habitat in the Interior Pacific Northwest. *Wildlife Society Monographs*, **154**, 1–63.

Merow C., Smith M.J., & Silander J. a. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, no–no.

Montana Natural Heritage Program (2014) Animal Species of Concern. Accessed July 2014. Available at: http://mtnhp.org/SpeciesOfConcern/?AorP=a&OpenFolders=S&Species=Birds.

Montana Wilderness Association (2014) National Forests: Lewis and Clark National Forest. Accessed July 2014. Available at: http://wildmontana.org/discover-the-wild/montanas-public-lands/national-forests/.

Olson L.E., Sauder J.D., Albrecht N.M., Vinkey R.S., Cushman S. a., & Schwartz M.K. (2014) Modeling the effects of dispersal and patch size on predicted fisher (Pekania [Martes] pennanti) distribution in the U.S. Rocky Mountains. *Biological Conservation*, **169**, 89–98.

Oregon State University (2014) PRISM Climate Group. Accessed on March 2014. Available at: http://prism.oregonstate.edu/.

Patla S.M. (1997) *Nesting ecology and habitat of the northern goshawk in undisturbed and timber harvest areas on the Targhee National Forest, Greater Yellowstone Ecosystem.* Idaho State University, Pocatello, Idaho, USA,

Penteriani V. (2002) Goshawk nesting habitat in Europe and North America: a review. *Ornis Fennica*, **79**, 149–163.

Phillips S., Dudik M., & Schapire R. (2011) Maxent: A program for maximum entropy modelling of species geographic distributions.

Phillips S., Dudík M., & Schapire R. (2004) A maximum entropy approach to species distribution modeling. 655–662.

Phillips S.J. (2008) A Brief Tutorial on Maxent.

Phillips S.J., Anderson R.P., & Schapire R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Phillips S.J. & Dudık M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

Phillips S.J., Dudík M., Elith J., Graham C.H., Lehmann A., Leathwick J., & Ferrier S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological applications : a publication of the Ecological Society of America*, **19**, 181–97.

R Core Group (2014) A Language and Enironment for Statistical Computing. Available at: http://www.r-project.org.

Rangel T.F., Diniz-Filho J.A.F., & Bini L.M. (2010) SAM: a comprehensive application for Spatial Analysis in Macroecology. *Ecography*, **33**, 46–50.

Reynolds R. & Joy S. (1998) *Distribution, territory occupancy, dispersal, and demography of northern goshawks on the Kaibab plateau, Arizona.*

Reynolds R.T., Graham R.T., & Reiser M.H. (1992) Management Recommendations for the Northern Goshawk in the Southwestern United States. General Technical Report RM-217. *USDA Forest Service, Rocky Mountain Forest and Range Experiment Station*, 90.

Reynolds R.T., Meslow E.C., & Wight H.M. (1982) Nesting Habitat of Coexisting Accipter in Oregon. *The Journal of Wildlife Management*, **46**, 124–138.

Samson F.B. (2006a) A Conservation Assessment of the Northern Goshawk , Black-backed Woodpecker , Flammulated Owl , and Pileated Woodpecker in the Northern Region. *USDA Forest Service, Northern Region*, 161.

Samson F. B. (2006b) Habitat Estimates For Maintaining Viable Populations of the Northern Goshawk, Black-backed Woodpecker, Flammulated Owl, Pileated Woodpecker, American Marten, and Fisher Habitat Estimates for Maintaining Viable Populations. *USDA Forest Service, Northern Region*, 1–24.

Speiser R. & Bosakowski T. (1987) Nest Site Selection by Northern Goshawks in Northern New Jersey and Southeastern New York. *The Condor*, **89**, 387–394.

Squires J.R. & Kennedy P.L. (2006) Northern Goshawk ecology: an assessment of current knowledge and information needs for conservation and management. *Studies in Avian Biology*, **31**, 8–62.

Squires J.R. & Reynolds R.T. (1997) Northern Goshawk. *The Birds of North America*, 34.

Terribile L.C. & Diniz-Filho J.A.F. (2009) Spatial patterns of species richness in New World coral snakes and the metabolic theory of ecology. *Acta Oecologica*, **35**, 163–173.

USDA Forest Service (1982) 1982 Rule: National Forest System Land and Resource Management Planning. 171.

USDA Forest Service (2009) Region 1 Vegetation Mapping Program (R1 Vmap). Accessed March 2014. Available at: http://www.fs.usda.gov/detailfull/r1/landmanagement/gis/?cid=stelprdb5331054&width=full.

USDA Forest Service (2012) 2012 Rule: National Forest System Land and Management Planning. 36 CFR Part 219. *Federal Register*, **77**, 21162–21276.

USDA Forest Service (2014) Lewis and Clark National Forest: Welcome to our Forest. Accessed July 2014. Available at: http://www.fs.usda.gov/main/lcnf/home.

Václavík T., Kupfer J. a., & Meentemeyer R.K. (2012) Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM). *Journal of Biogeography*, **39**, 42–55.

Willmott C.J. & Matsuura K. (1995) Smart Interpolation of Annually Averaged Air Temperature in the United States. *Journal of Applied Meteorology*, **34**, 2577–2586.

Woodbridge B. & Hargis C. (2006) Northern Goshawk Inventory and Monitoring Technical Guide. Gen. Tech. Rep. WO-71. 80.

Young N., Evangelista P., & Carter L. (2011) A Maxent Model v3.3.1 Tutorial (ArcGIS v10). 1–20.

Appendix
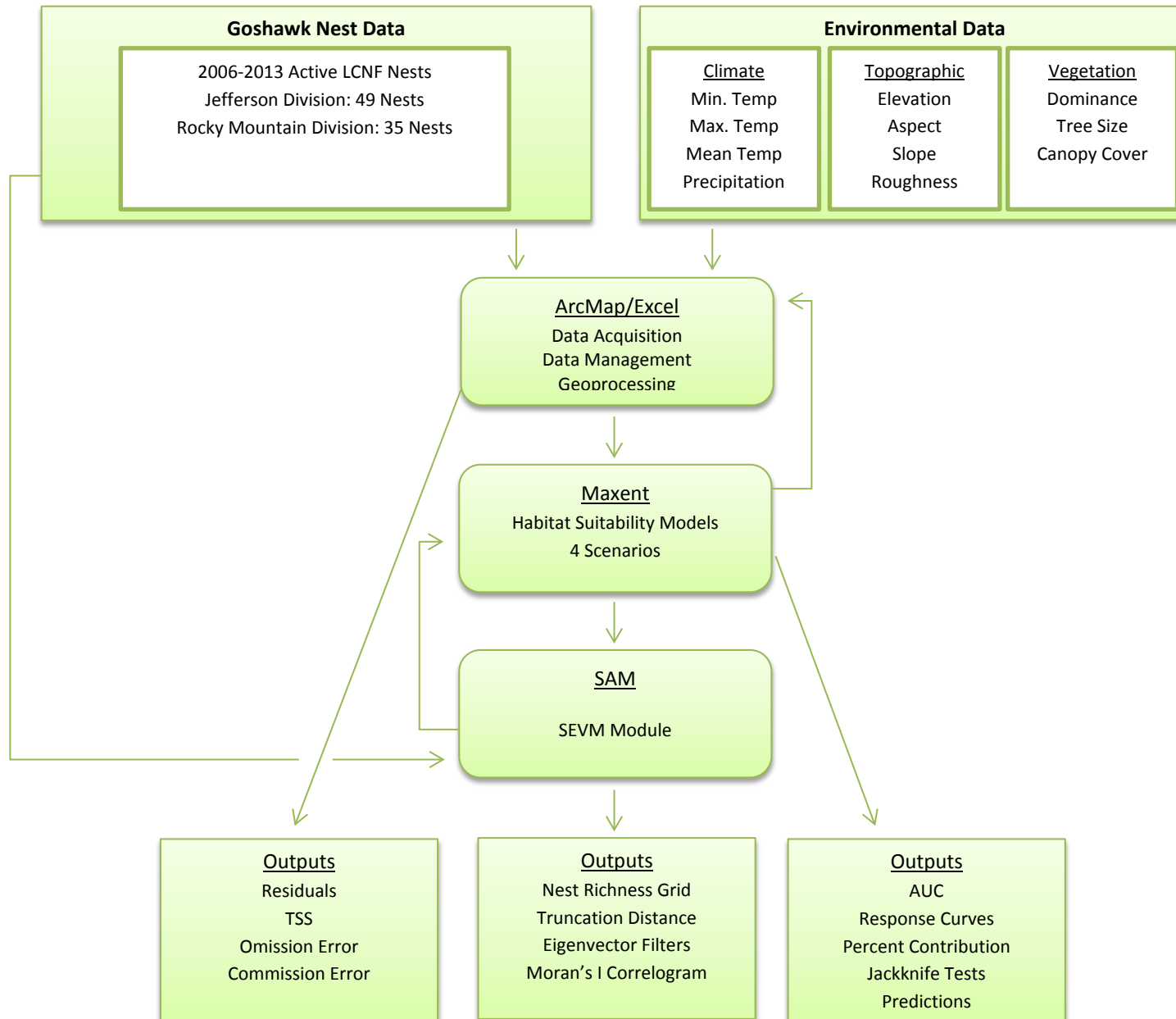
Figure 1: Flow diagram showing study methodology
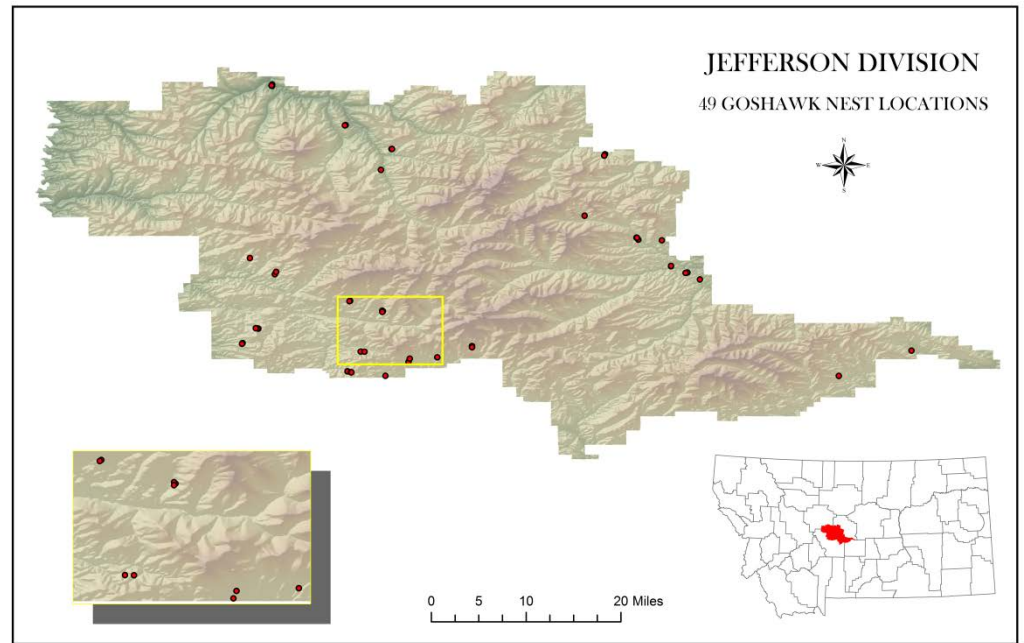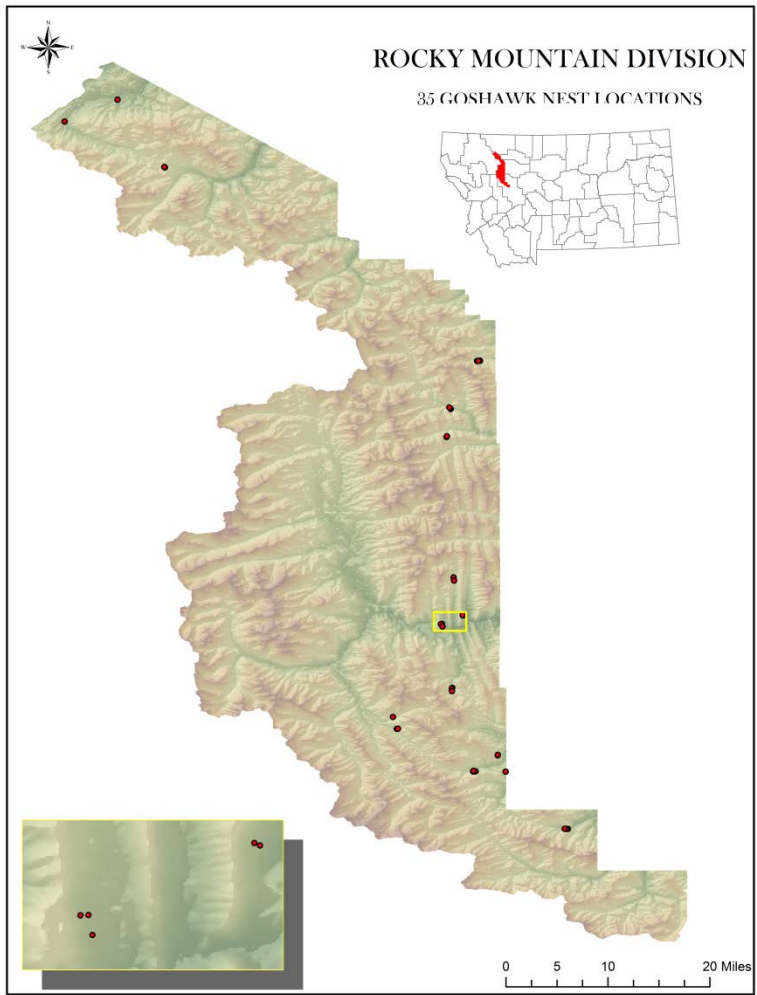
Figure 2: Study area
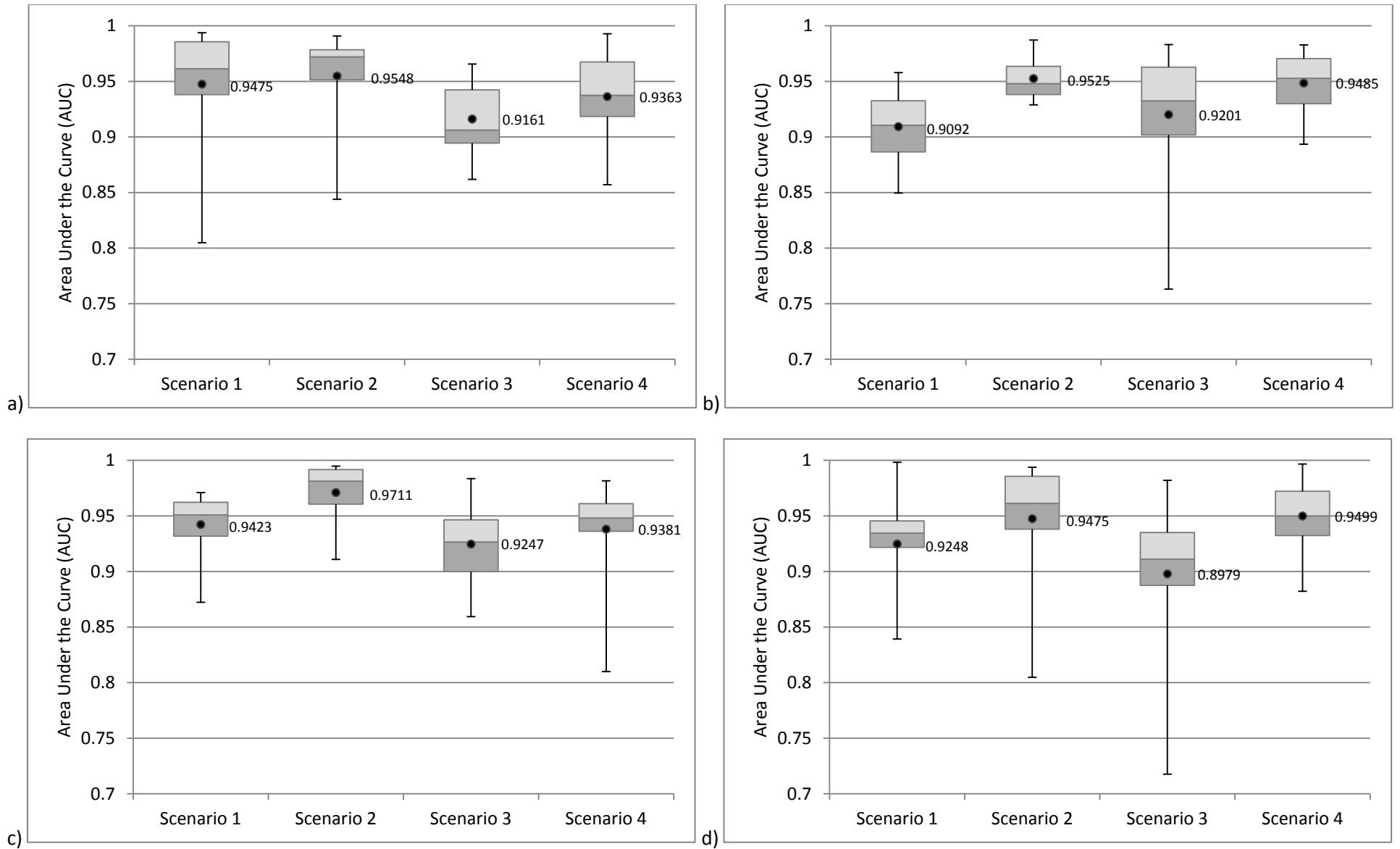
Figure 3: Area under the curve (AUC): a) Jefferson Division with bootstrap methods, b) Jefferson Division with cross-validation methods, c) Rocky Mountain with bootstrap methods, and d) Rocky Mountain with cross-validation methods.

44
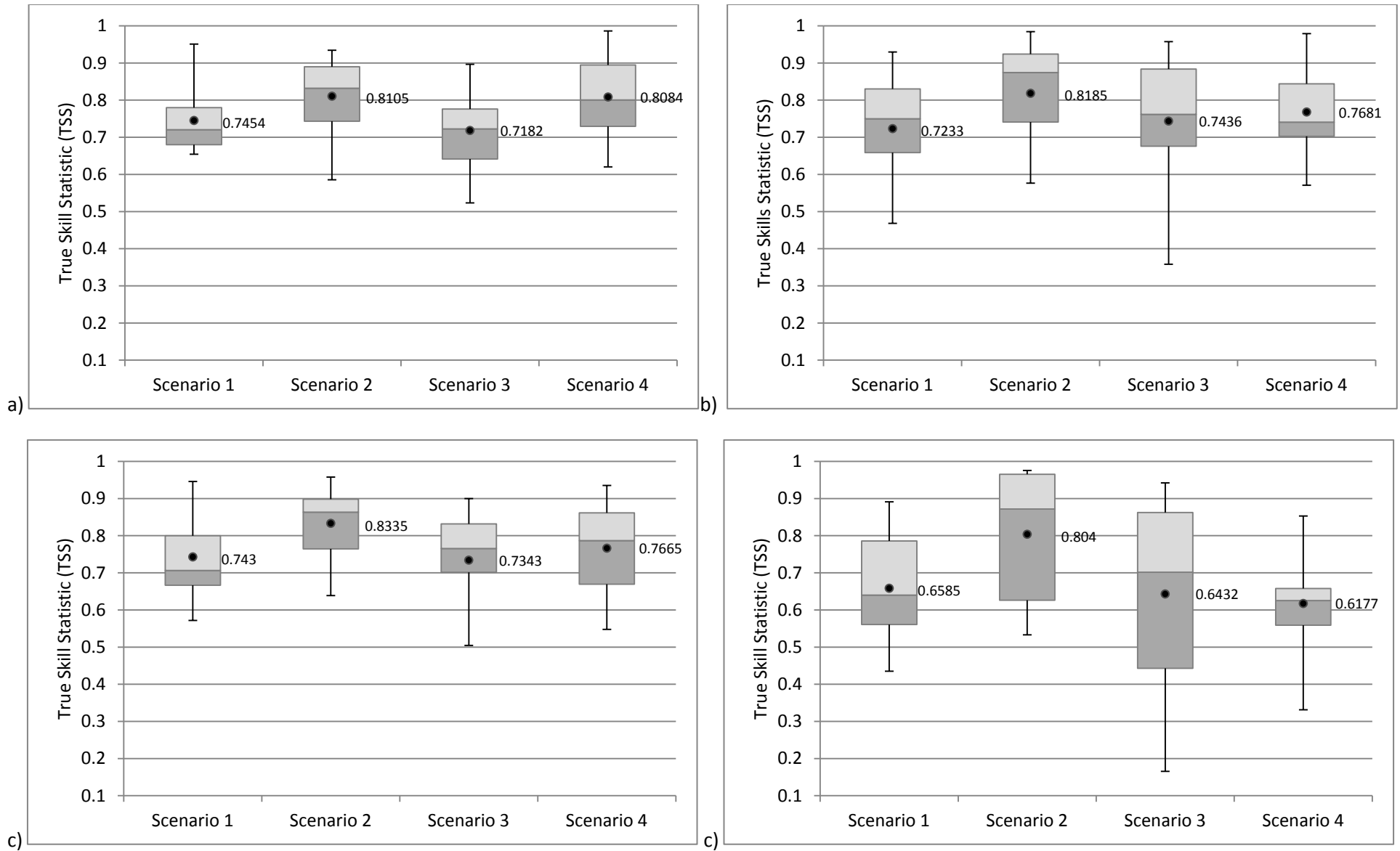
Figure 4: True skill statistic (TSS): a) Jefferson Division with bootstrap methods, b) Jefferson Division with cross-validation methods, c) Rocky Mountain with bootstrap methods, and d) Rocky Mountain with cross-validation methods.
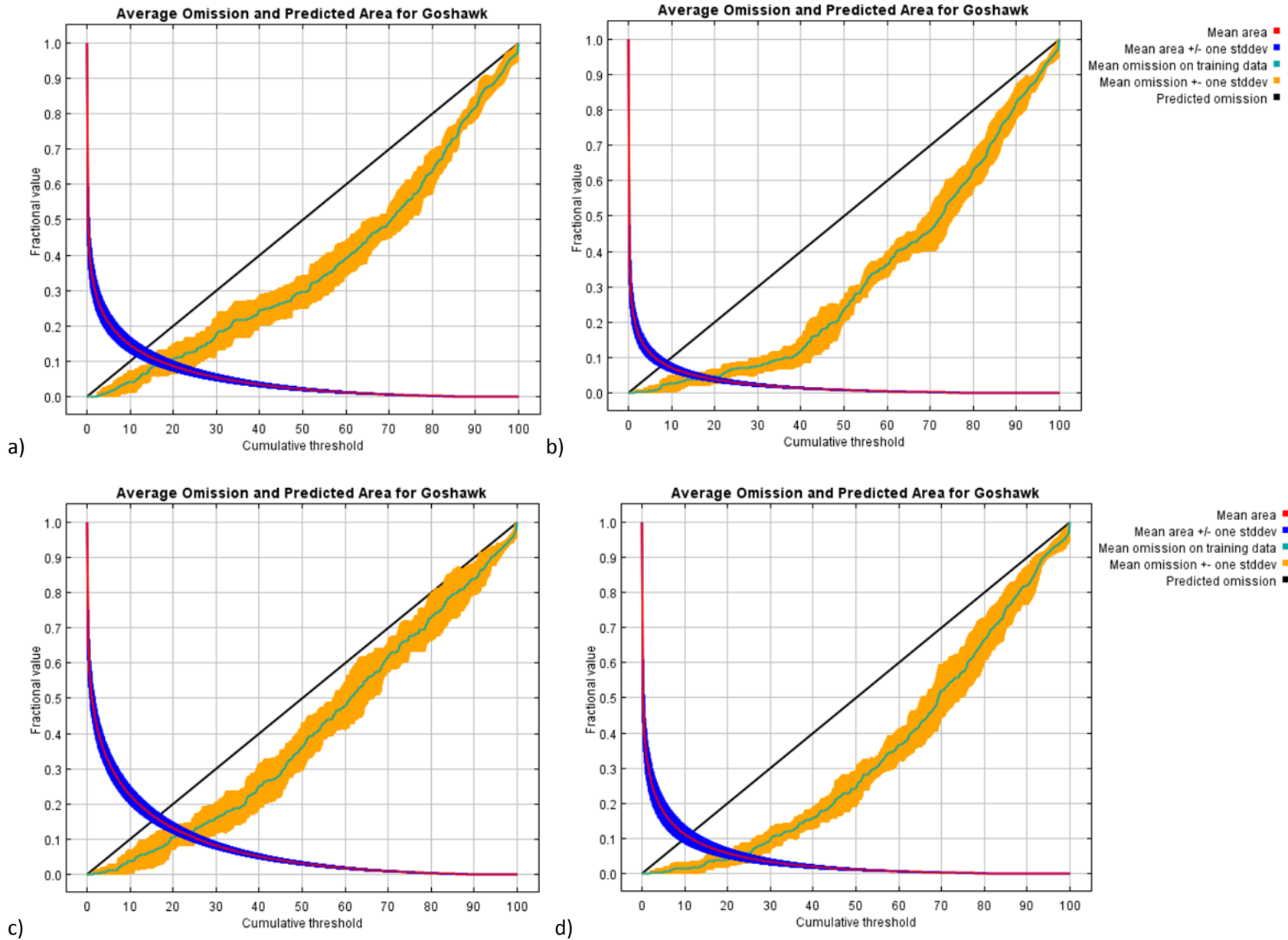
Figure 5. Omission/commission analysis provided by Maxent for the Jefferson Division using bootstrap methods: a) Scenario 1, b) Scenario 2, c) Scenario 3, and d) Scenario 4.
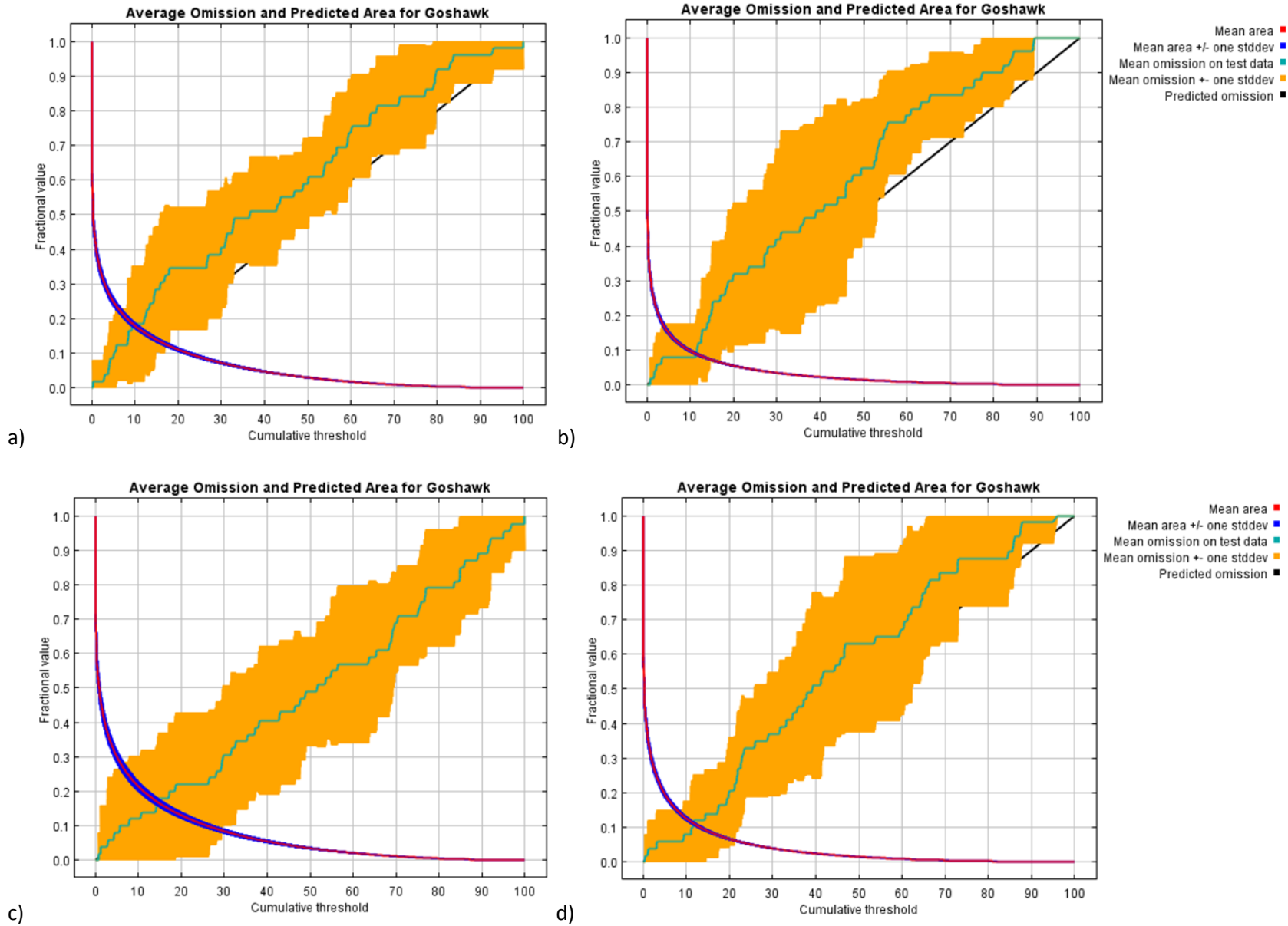
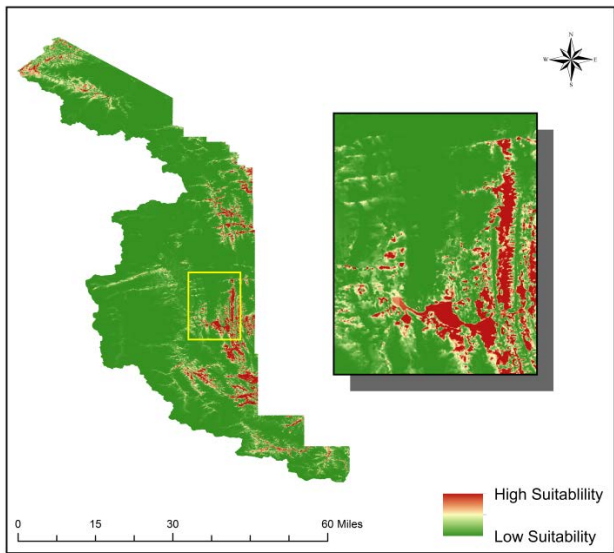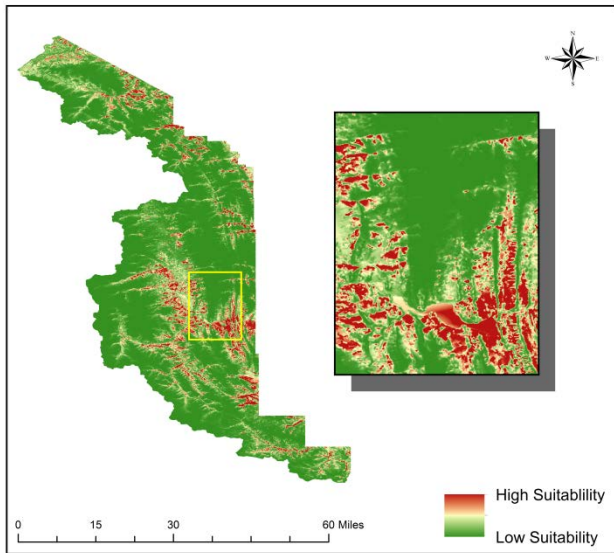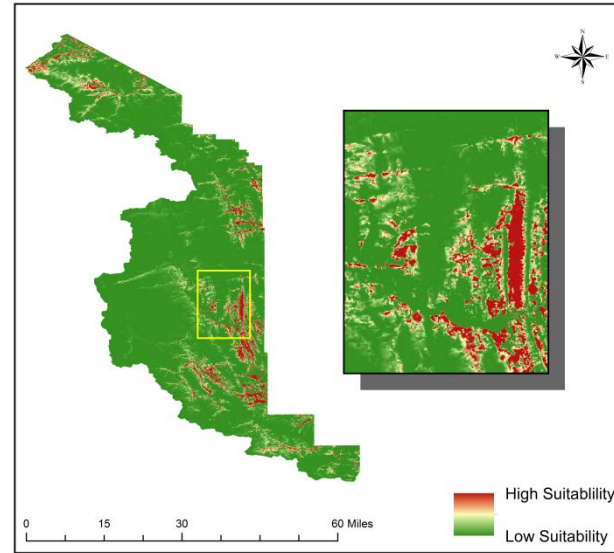**Figure 6. Omission/commission analysis provided by Maxent for the Jefferson Division using cross-validation methods: a) Scenario 1, b) Scenario 2, c) Scenario 3, and d) Scenario 4.**
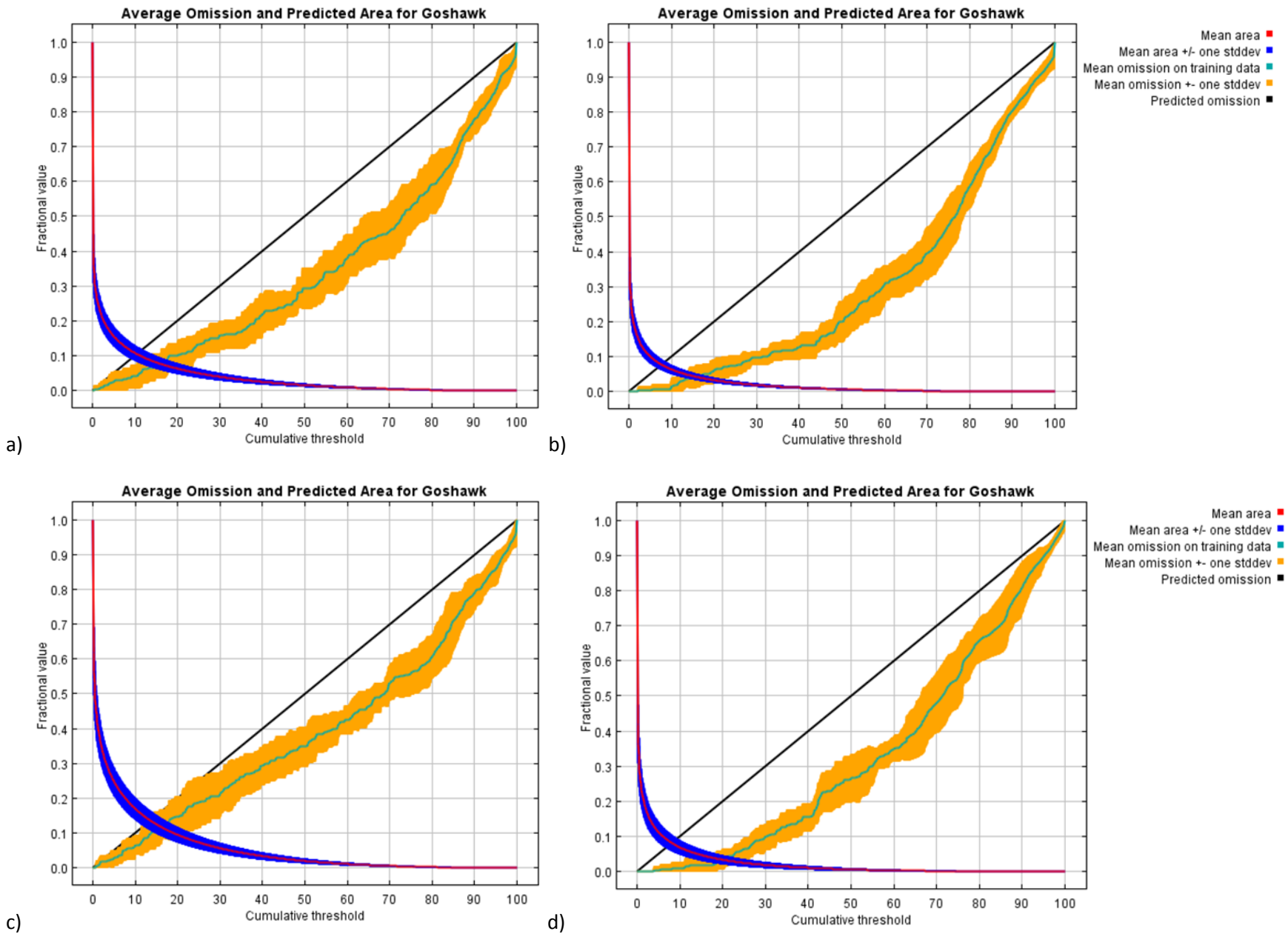
Figure 7. Omission/commission analysis provided by Maxent for the Rocky Mountain division using bootstrap methods: a) Scenario 1, b) Scenario 2, c) Scenario 3, and d) Scenario 4.
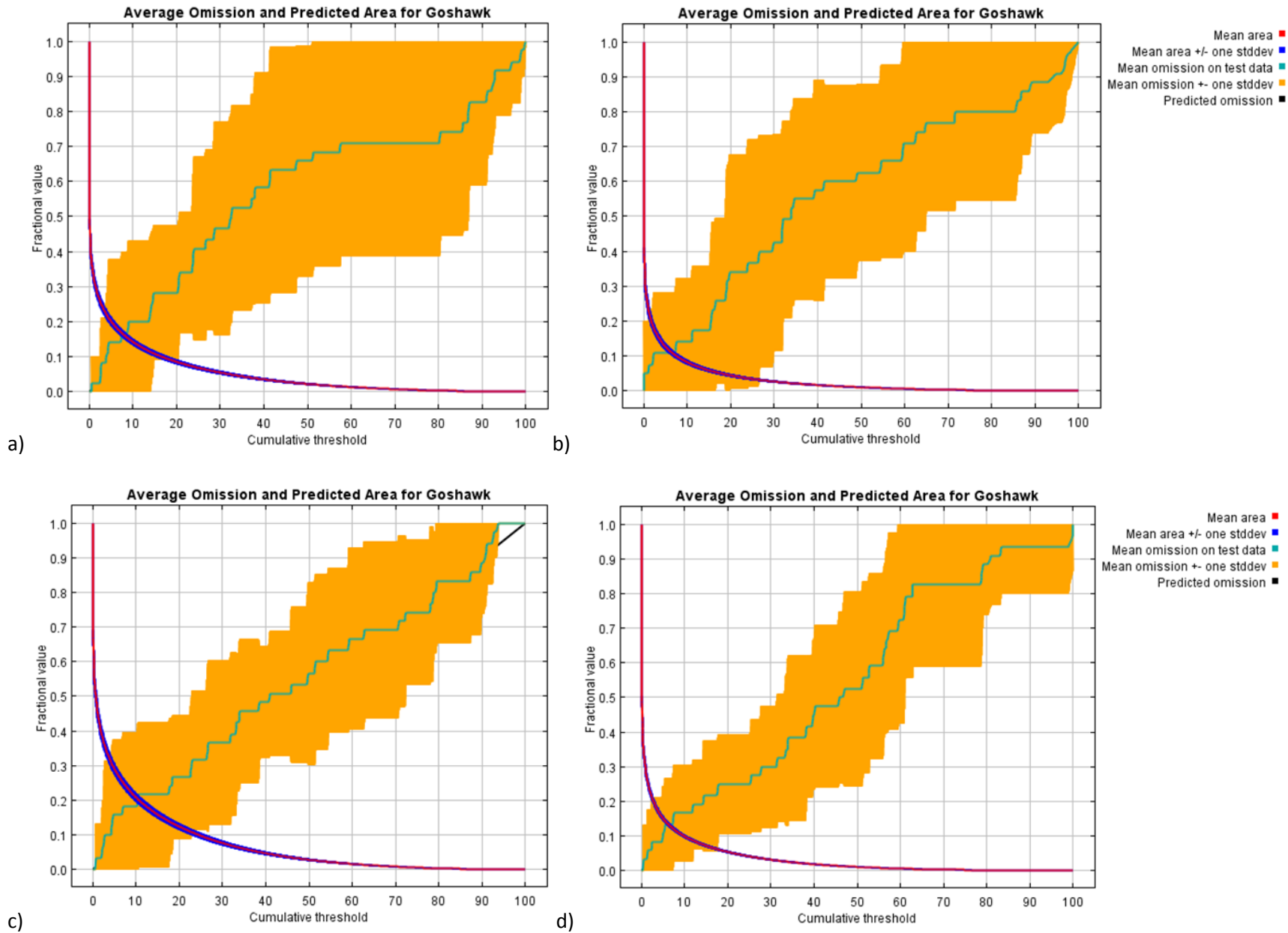
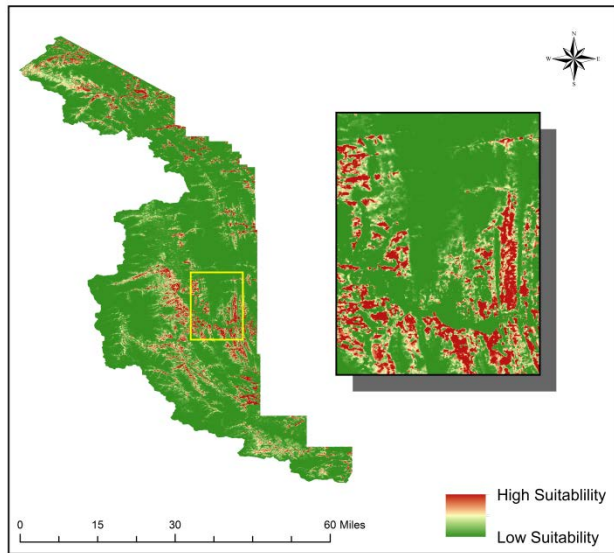**Figure 8. Omission/commission analysis provided by Maxent for the Rocky Mountain Division using cross-validation methods: a) Scenario 1, b) Scenario 2, c) Scenario 3, and d) Scenario 4.**
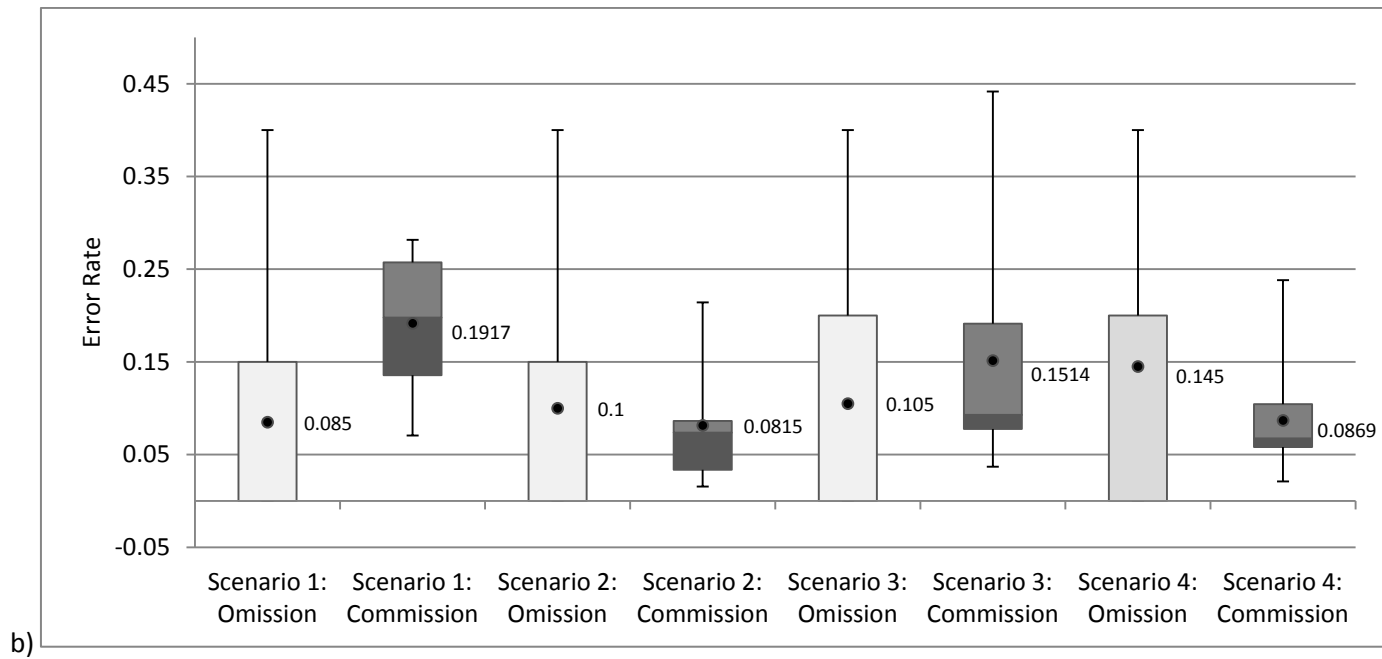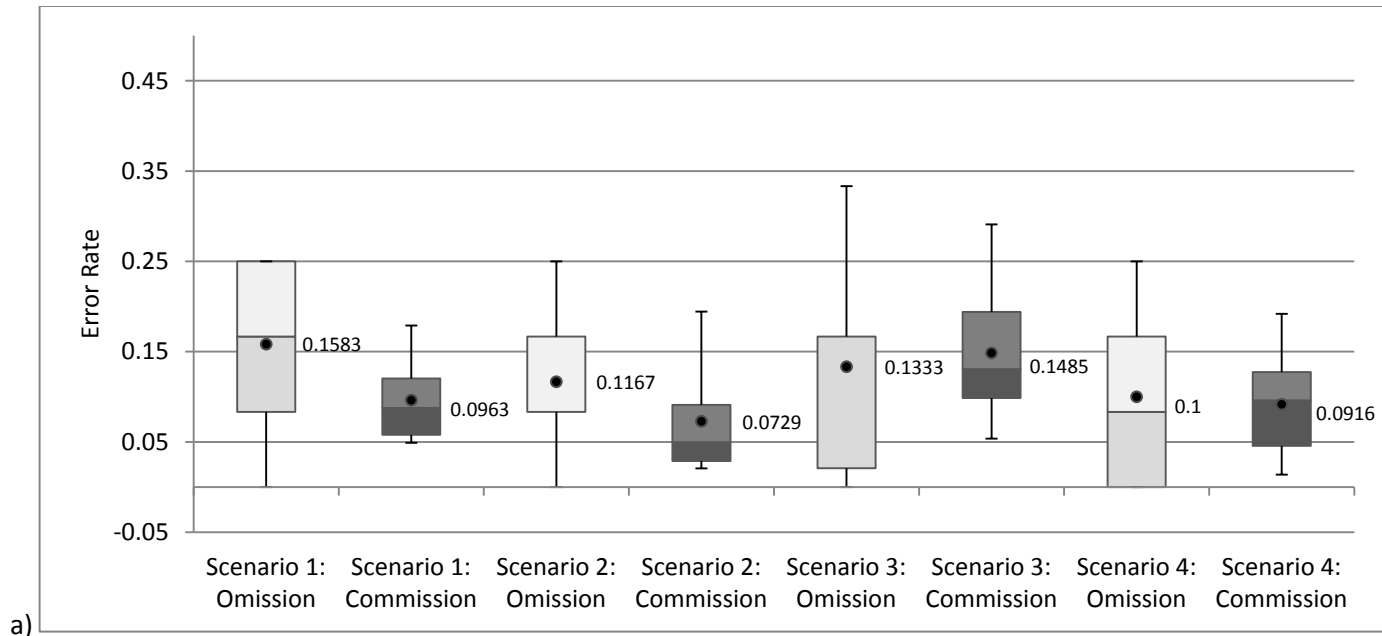
Figure 9: Omission/commission error rate for the Jefferson Division, using: a) bootstrap methods, and b) cross-validation methods.
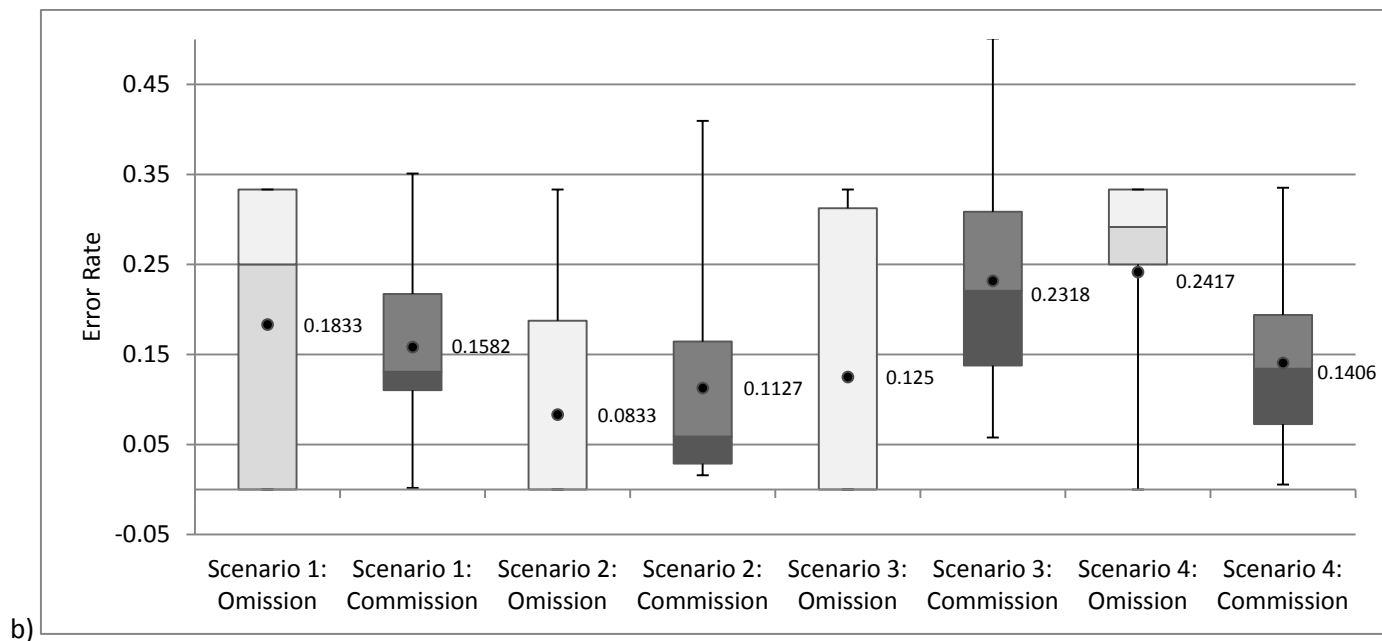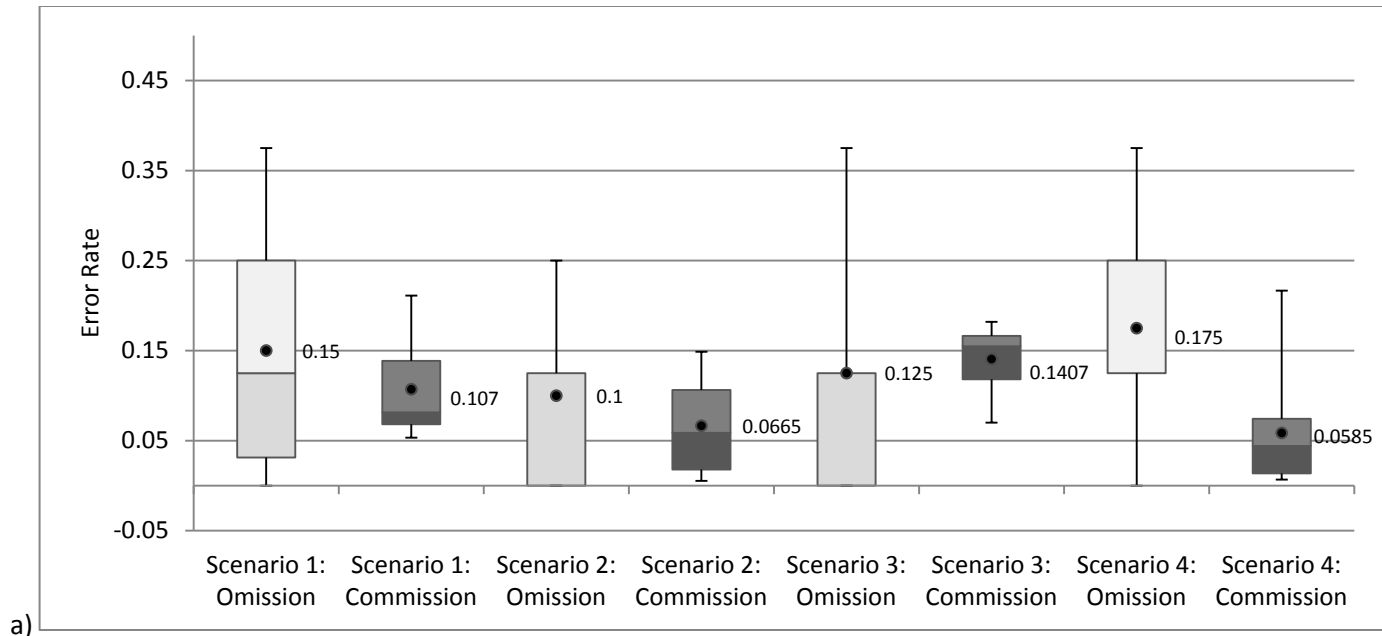
Figure 10: Omission/commission error rate for the Rocky Mountain Division, using: a) bootstrap methods, and b) cross-validation methods.

a)

b)
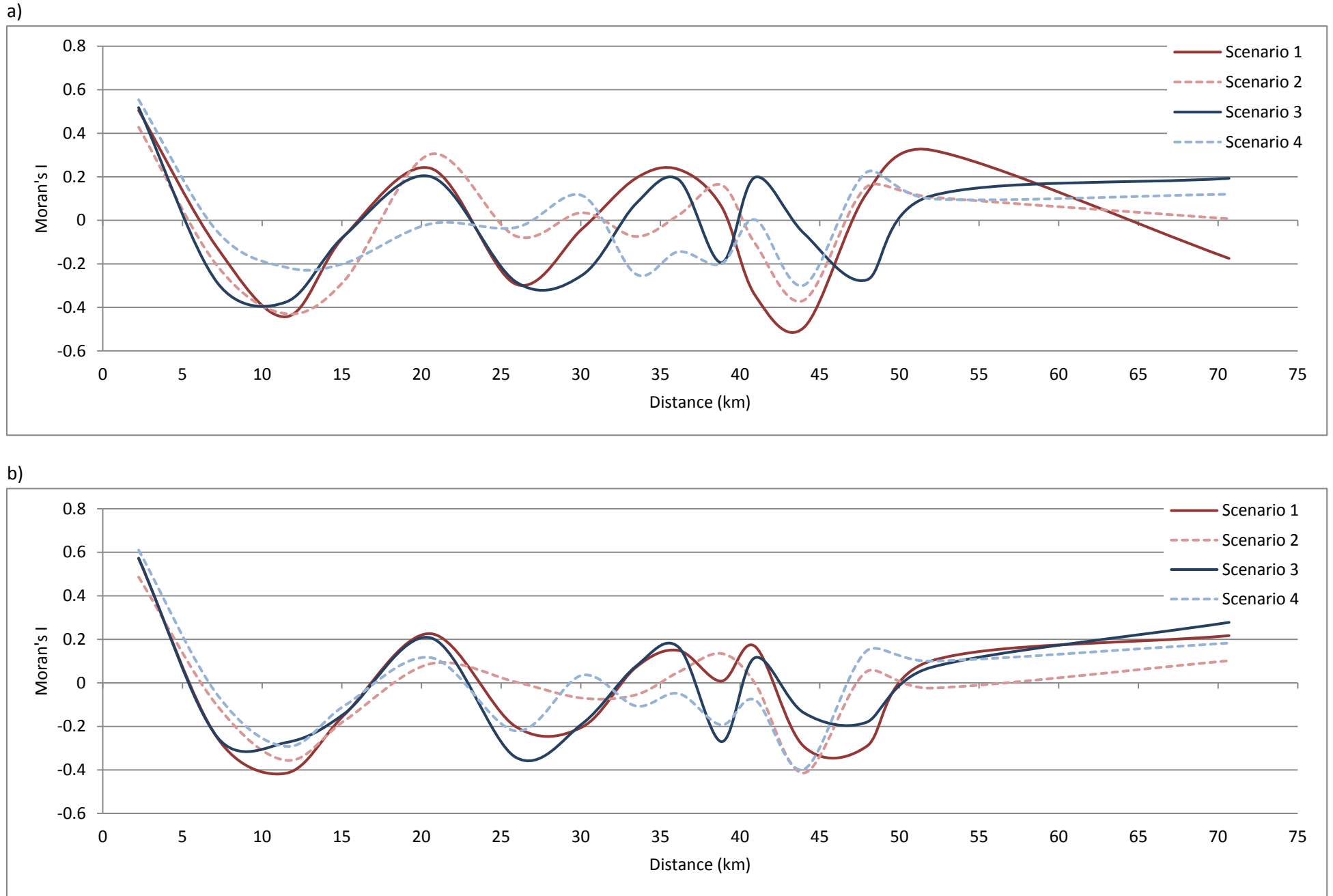
Figure 11: Evaluation of model residuals for the Jefferson Division, using: a) bootstrap methods, and b) cross-validation methods.

a)

b)
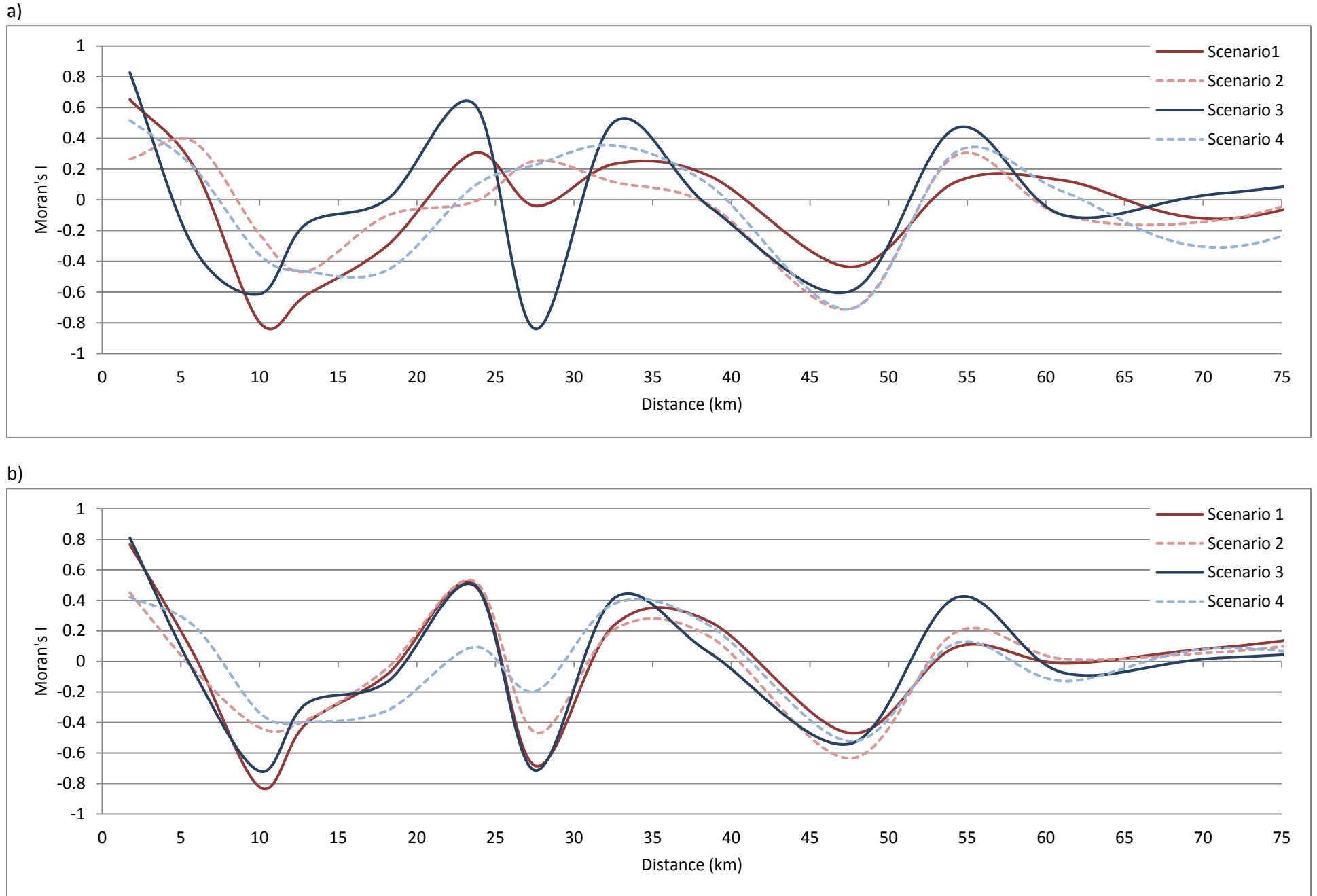
**Figure 12: Evaluation of model residuals for the Rocky Mountain Division, using: a) bootstrap methods, and b) cross-validation methods.**
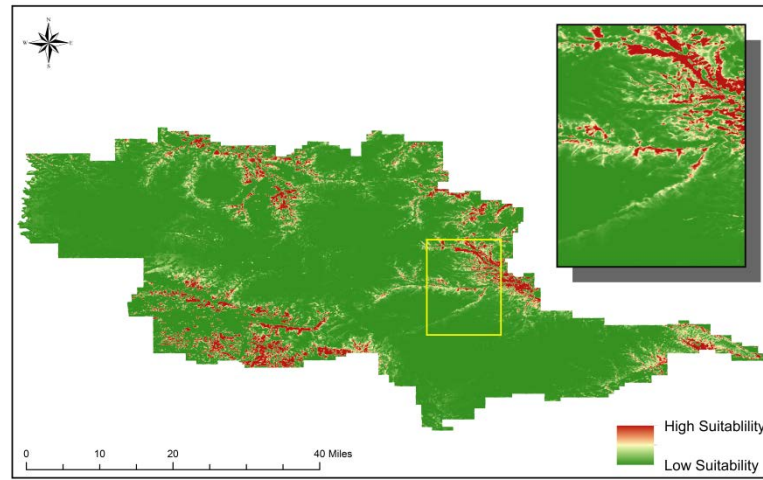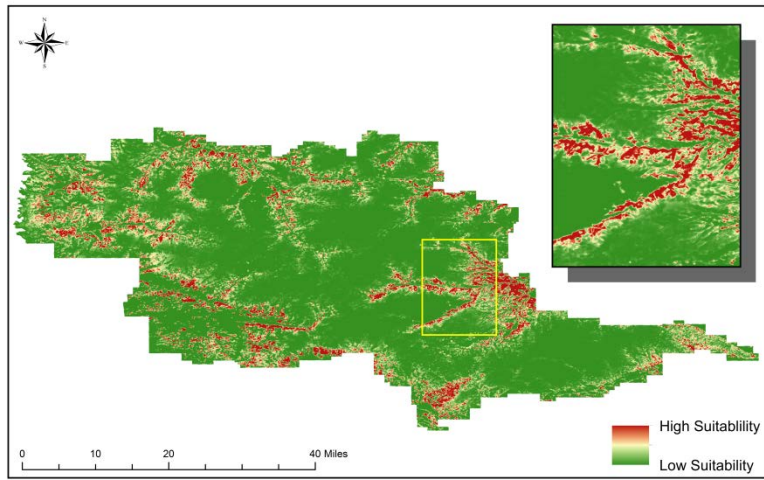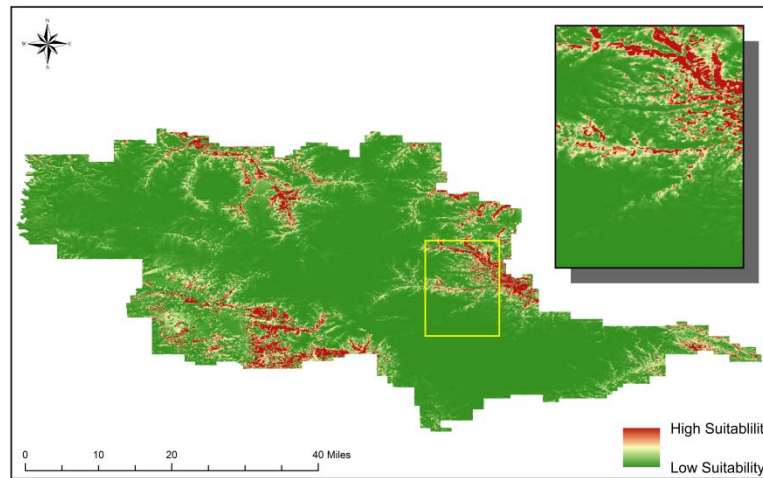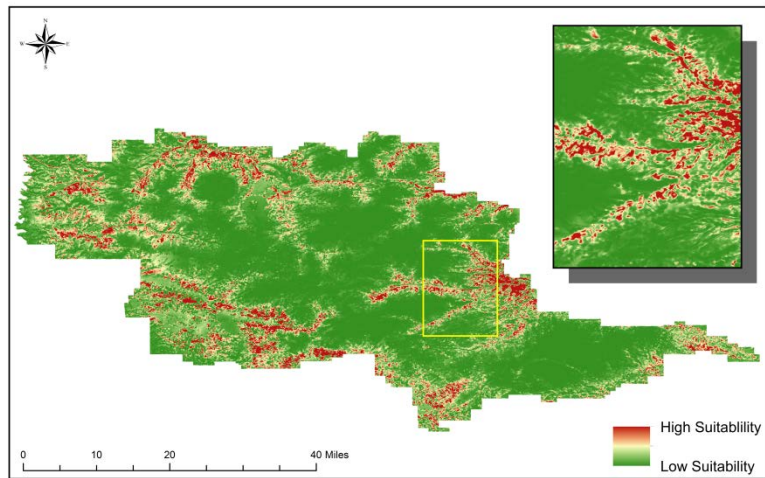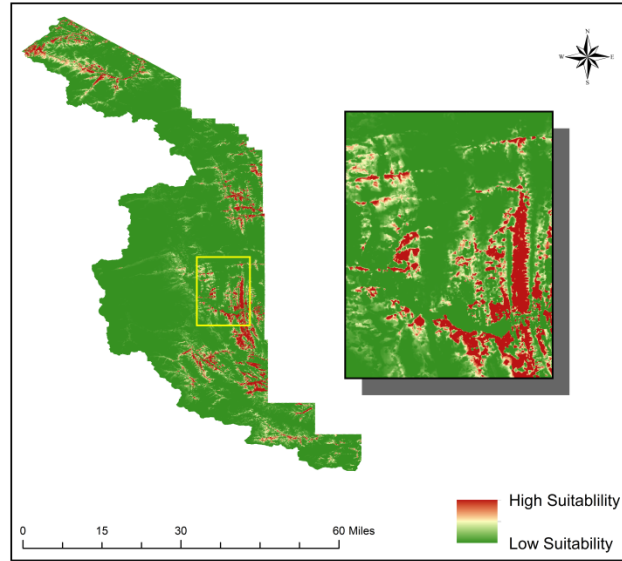
Figure 13: Model predictions averaged across replicated runs for Jefferson Division using bootstrap methods: a) Scenario 1, b) Scenario 2, c) Scenario 3, and d) Scenario 4.

54
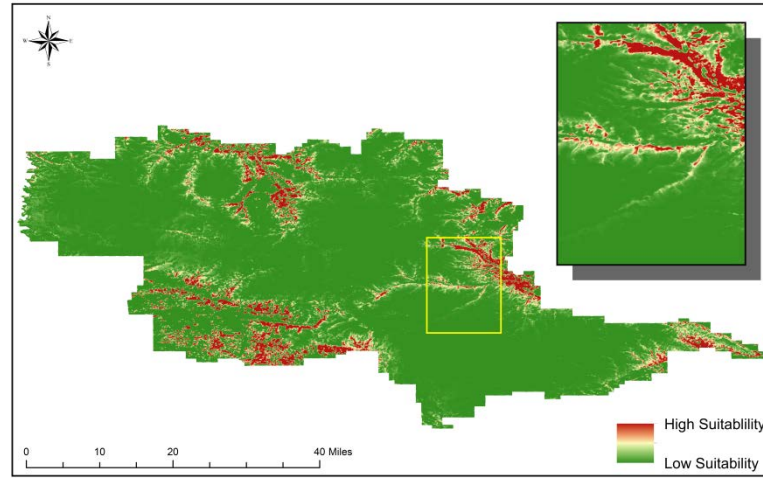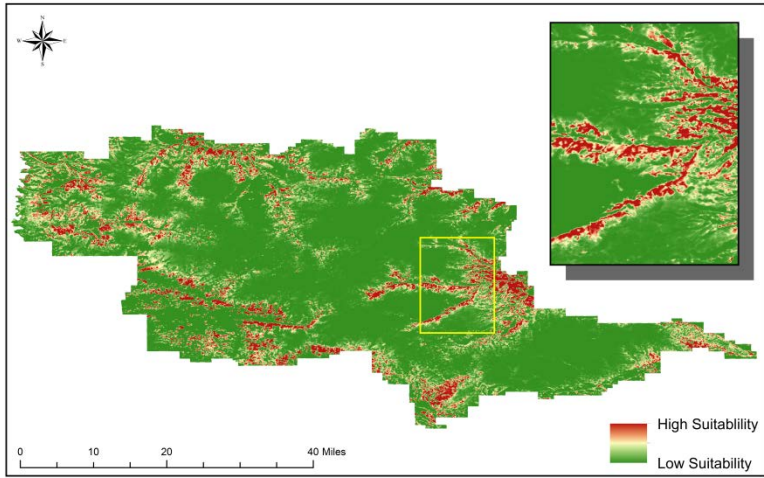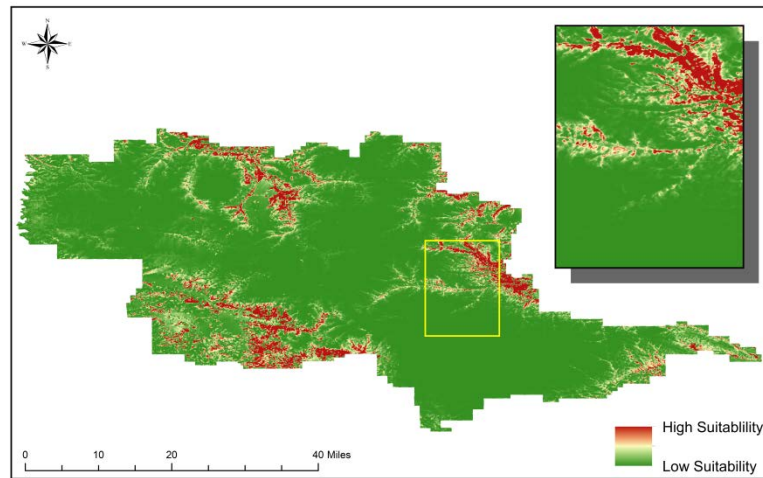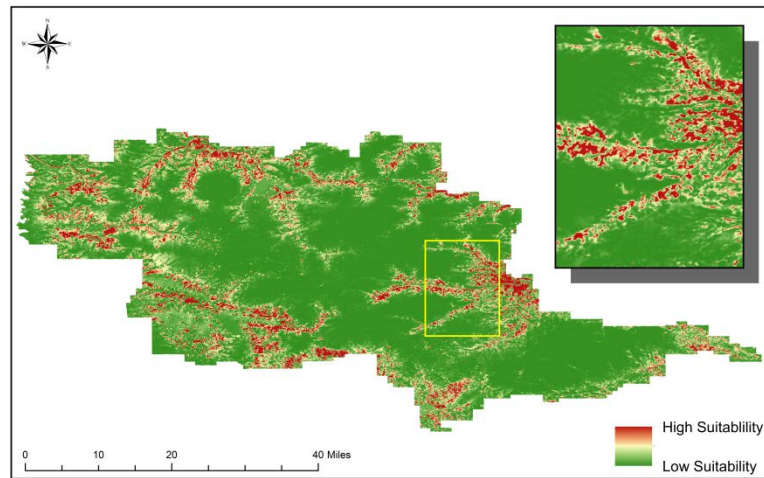
Figure 14: Model predictions averaged across replicated runs for Jefferson Division using cross-validation methods: a) Scenario 1, b) Scenario 2, c) Scenario 3, and d) Scenario 4.

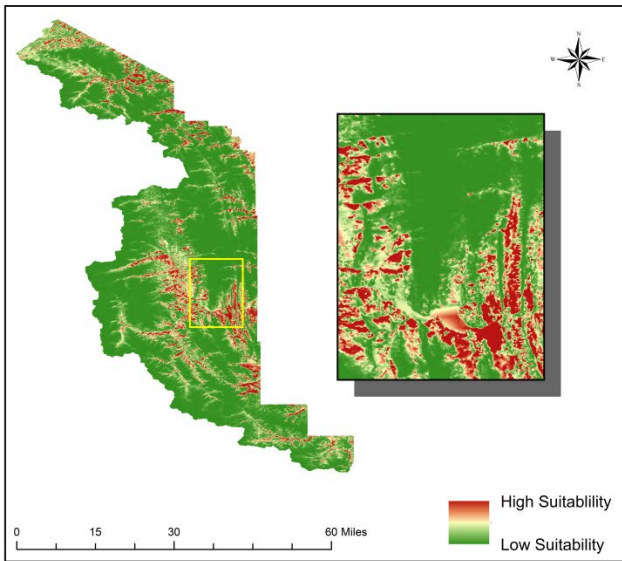**Figure 15: Model predictions averaged across replicated runs for Rocky Mountain Division using bootstrap methods: a) Scenario 1, b) Scenario 2, c) Scenario 3, and d) Scenario 4.**
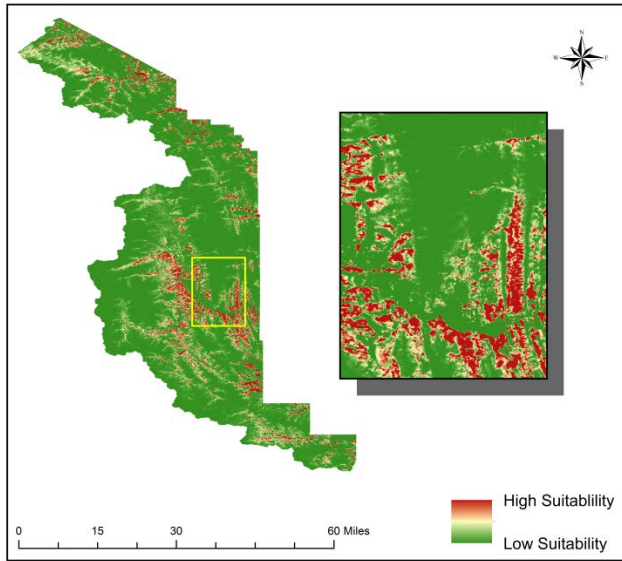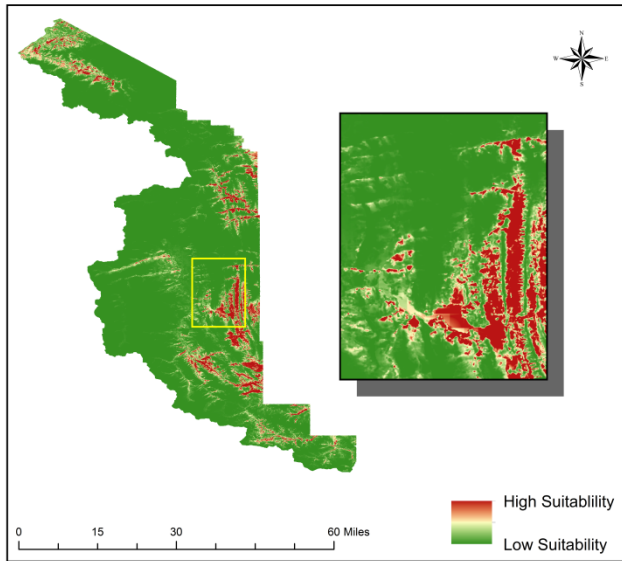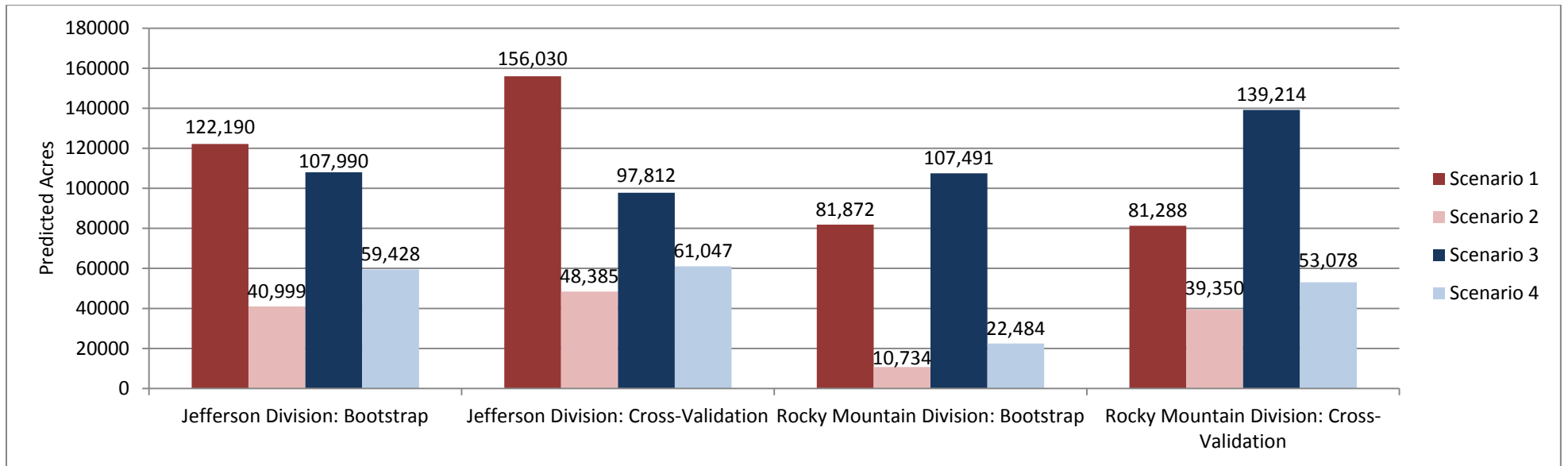
Figure 16: Model predictions averaged across replicated runs for Rocky Mountain Division using cross-validation methods: a) Scenario 1, b) Scenario 2, c) Scenario 3, and d) Scenario 4.

Figure 17. Predicted acreage for each model calculated by using the average maximum test sensitivity plus specificity logistic threshold.