Modeling Forest Bird Species' Likelihood of Occurrence in Utah With Forest Inventory and Analysis and Landfire Map Products and Ecologically Based Pseudo-Absence Points

Phoebe L. Zarnetske[1], **Thomas C. Edwards, Jr.**[2], **and Gretchen G. Moisen**[3]

Abstract.—Estimating species likelihood of occurrence across extensive landscapes is a powerful management tool. Unfortunately, available occurrence data for landscape-scale modeling is often lacking and usually only in the form of observed presences. Ecologically based pseudo-absence points were generated from within habitat envelopes to accompany presence-only data in habitat classification models (HCM) for the northern goshawk (*Accipiter gentilis atricapillus*). We built models at two resolutions, using predictor variables derived from 250-m Forest Inventory and Analysis map products, 30-m U.S. Department of Agriculture Landfire map products, and digital elevation models. Cross-validation provided an assessment of models' predictive capabilities. Use of ecologically based pseudo-absence points to accompany extant presence points in HCM can be a powerful asset for species conservation.

Introduction

Species habitat classification models (HCMs) and their associated habitat suitability maps are valuable assets to species monitoring, conservation, and land-use planning, particularly across broad landscapes where intensive surveying and monitoring is difficult. Species metapopulations are often spread across broad landscapes, justifying management across spatial extents which approximate their population ranges. Recent advances in statistical modeling, Geographic Information System (GIS), and remote sensing have enabled researchers to more accurately model species-habitat relationships in a spatial context (Levin *et al.* 1997, McNoleg 1996, Scott *et al.* 2002). In addition, the creation of broad-scale continuous map products that portray land-cover variables have become valuable resources for species habitat modeling, particularly across large spatial extents.

The U.S. Department of Agriculture (USDA) Forest Service Forest Inventory and Analysis (FIA) program recently created map products of forest attributes across the central Utah highlands at 250-m resolution (table 1) (Blackard *et al.* 2004). Similar map products of forest attributes have been created by the USDA Landfire program as well, although at a finer resolution of 30-m (table 2) (Keane *et al.* 2002, Landfire 2005). When spatially intersected with species occurrence data (i.e., spatially explicit presence/absence points), spatial map products produce predictor variables useful in species HCMs. The statistical model underlying HCMs can then distinguish (i.e., classify) suitable habitat from nonsuitable habitat. Generalized linear models (GLMs), generalized additive models, and classification trees (Breiman *et al.* 1984) are among the most frequently used techniques. GLMs are a popular choice because they have proven to be robust and stable (Brotons *et al.* 2004; Engler *et al.* 2004; Guisan *et al.* 1999, 2002; Manel *et al.* 1999; Pearce and Ferrier 2000; Thuiller *et al.* 2003).

Landscape-scale studies often lack absence data due to the amount of resources needed to collect absence data points

[1] Graduate student, Oregon State University, Department of Zoology, 3029 Cordley Hall, Corvallis, OR 97331–3029. E-mail: zarnetsp@science.oregonstate.edu.

[2] Assistant Unit Leader and Professor, Utah Cooperative Fish and Wildlife Research Unit, College of Natural Resources, Utah State University, Logan, UT 84322–5230. E-mail: t.edwards@gis.usu.edu.

[3] Research Forester, U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, Ogden, UT 84401. E-mail: gmoisen@fs.fed.us.

across an extensive study area. In addition, occurrence databases containing recent and historical data (e.g., from censuses, field studies, museum and herbaria records) usually contain only presences. This lack of absence data precludes the use of many statistical techniques in HCMs. Two basic methods exist for handling the problem of presence-biased datasets: (1) using profile-type models, which incorporate the presence-only bias into the model and control its effects on the resulting predictions (e.g., Hirzel et al. 2002, Stockwell and Peterson 2002b, Zaniewski *et al.* 2002), and (2) generating pseudo-absence points to be used in place of unknown absence data (e.g., Engler *et al.* 2004, Guisan and Thuiller 2005, Stockwell and Peterson 2002b, Zaniewski *et al.* 2002).

Profile-type models contrast attributes of presence-only data with background levels of environmental variables across the study area. Factor analysis (Hirzel *et al.* 2002), fuzzy set theory (Busby 1991, Chicoine *et al.* 1985, Robertson *et al.* 2004), artificial intelligence methods (Stockwell and Noble 1991), and statistical mechanics (Phillips *et al.* 2006) are modeling techniques used in profile-type models. Unfortunately, several drawbacks exist for profile-type models, including overprediction (e.g., Ecological Niche Factor Analysis [ENFA]—Brotons *et al.* 2004, Engler *et al.* 2004), and assumptions that all predictor variables are equally important in determining species' distribution (e.g., Fuzzy Envelope Model—Robertson *et al.* 2004), that the presence data is unbiased (e.g., ENFA—Hirzel *et al.* 2002), that the presence-only data originates only from source habitat (e.g., Maxent—Phillips *et al.* 2006), and difficulty in interpretation and statistical assessment. An alternative to profile-type models is to generate so-called pseudo-absences to pair with known presences (Ferrier and Watson 1997, Stockwell and Peters 1999, Stockwell and Peterson 2002b, Zaniewski *et al.* 2002). Traditional techniques of generating pseudo-absence points involve randomly selecting pseudo-absences from broadly defined species ranges, such as an entire study, and excluding locations where presence points exist. To constrain pseudo-absence point selection, Stockwell and Peters (1999) suggested creating a pseudo-survey region that confines the extent from which the pseudo-absence points can be generated.

This study improves on existing pseudo-absence point generation techniques by incorporating biological knowledge concerning the species-habitat relationship to constrain the region from which pseudo-absence points are selected. The northern goshawk (*Accipiter gentilis atricapillus*), considered a management indicator species in many national forests of the Intermountain West, was chosen as a study species because of its well-documented habitat associations (Brotons *et al.* 2004), and sufficient recent and extant presence points in Utah, but lack of absence points. The objectives of this study were (1) to develop methods to incorporate species' ecology into the generation of pseudo-absence points, (2) to apply these methods to produce HCMs and output likelihood of occurrence maps for northern goshawk nest sites and nest areas across Utah's central highlands, and (3) to test the utility of FIA and Landfire vegetation map products for wildlife habitat applications.

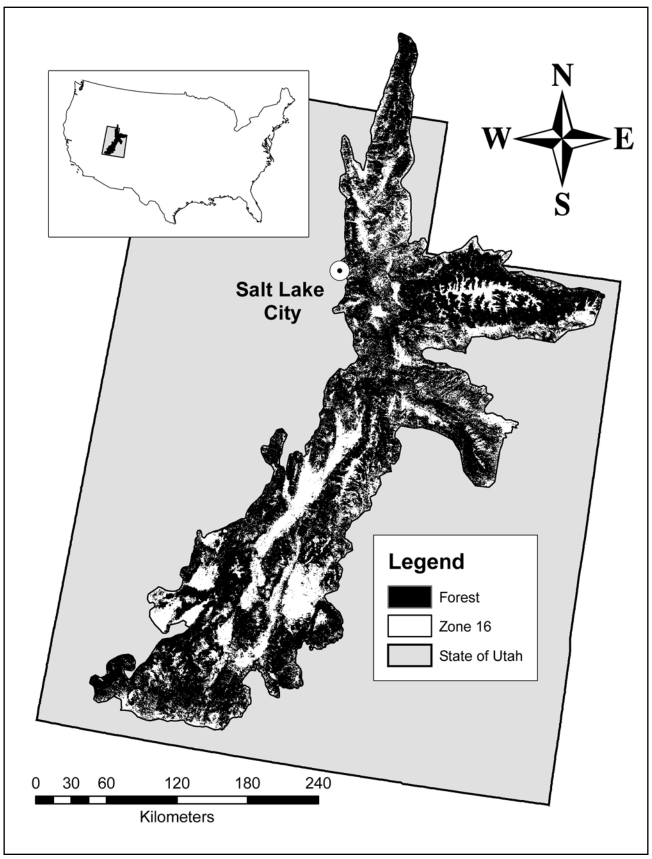
To incorporate biological knowledge, and constrain the region from which pseudo-absence points are randomly selected, spatial habitat envelopes based on known goshawk habitat associations were created. We define a habitat envelope as an ecological representation of a species, or species feature's (e.g., nest), observed distribution (i.e., realized niche) based on a single attribute, or the spatial intersection of multiple attributes. No existing computer programs (e.g., GARP) were employed to assure complete control over the generation of habitat envelopes and the selection of ecologically based pseudo-absence points. Known ecological associations of northern goshawk nest location to habitat variables were translated into increasingly complex habitat envelopes. Pseudo-absence points were then paired with extant presence points and used in logistic regression to model the likelihood of occurrence of northern goshawk nest site (at 30-m resolution) and nest area (at 250-m resolution) as a function of habitat predictor variables. Models were evaluated with accuracy metrics via 10-fold cross-validation. Top models were translated into likelihood of occurrence maps across the study area, creating habitat suitability maps for each resolution. These envelope-based models were then compared to traditional models based on current practices for generating pseudo-absences, which involve selection of absences from broadly defined species ranges.

Methods

Study Region

The study occurred in forested regions of Zone 16, comprising the Wasatch and Uinta mountain ranges in Utah, southeast Idaho, and southern Wyoming (fig. 1). The zone is approximately 55 percent forested. Elevation ranges from 386 m to 3,978 m (1,266 to 13,051 ft). Zone 16 was selected because FIA and Landfire map products are complete for this region, thereby providing a large set of digital predictor layers for modeling purposes.

Figure 1.—*Zone 16 study region.*



Study Species

The northern goshawk is the largest accipiter in North America, is an apex predator, and is holarctic in distribution (Squires and Reynolds 1997). Homerange is approximately 2,400 ha, and consists of three components: nest area (~12 ha), post-fledging area (~170 ha), and the foraging area (~2,200 ha) (Reynolds *et al.* 1992). Multiple nest areas exist within a home range (Reynolds *et al.* 1994), and multiple (satellite) nests occur within each nest area (Reynolds *et al.* 2005, Squires and Reynolds 1997). In general, large trees (> 40.6 cm diameter at breast height) (Beier and Drennan 1997) arranged in a clump (Graham *et al.* 1999) with dense canopy cover (Beier and Drennan 1997, Bright-Smith and Mannan 1994, Graham *et al.* 1999, Reynolds *et al.* 1992, Stephens 2001) are preferred for nesting.

For modeling purposes, nest site is defined as the habitat immediately surrounding the nest (nest tree to 0.10 ha area surrounding the nest tree) (Reynolds *et al.* 1982, Squires and Reynolds 1997). Nest area is defined as habitat around the nest, 10 to 12 ha area in size that includes the nest tree, adult roosts, and prey plucking sites (Newton 1979, Reynolds *et al.* 1992). Nest site habitat characteristics were derived from the spatial intersection of 30-m resolution (0.09-ha) predictor variables with nest presence points (for suitable nest sites) and nest pseudo-absence points (for unsuitable nest sites) in a GIS (Zarnetske 2006). Nest area habitat characteristics were derived from the spatial intersection of 250-m resolution (6.25-ha) predictor variables with nest presence points (for suitable nest areas) and nest pseudo-absence points (for unsuitable nest areas) in a GIS (Zarnetske 2006). Hereafter, 30-m resolution (representing nest sites) and 250-m resolution (representing nest areas) models using habitat envelopes are referred to as "30-m habitat envelope models" and "250-m habitat envelope models," respectively. Models using traditional techniques of generating pseudo-absence points are referred to as "30-m traditional models" and "250-m traditional models."

Predictor Variables

The FIA plot-based inventories of forested land across Zone 16 were combined with 250-m MODIS imagery using regression tree modeling techniques to produce a suite of vegetation map products (Blackard *et al.* 2004) (tables 1 and 2). Landfire 30-m resolution map products for Zone 16 were created with plot inventories (including FIA plot data), biophysical gradients, and 30-m Landsat Thematic Mapper imagery using regression tree modeling as well (Keane *et al.* 2002, Landfire 2005)

Table 1.—*Descriptions of continuous spatial map products from Landfire and FIA.*

Variable	Abbreviation	Description
Landfire map product		
Canopy bulk density	CBD	kg/m
Canopy base height	CBH	m to live canopy
Forest canopy cover	FCC	% canopy cover
Forest height	FHT	m
Herbaceous canopy cover	HCC	% canopy cover
Herbaceous height	HHT	m
Shrub canopy cover	SCC	% canopy cover
Shrub height	SHT	m
Elevation	ELEV	m
Slope	SLP	% rise
Transformed aspect	TASP	0-1
FIA		
Stand age	AGE	age (yrs)
Basal area	BA	m²/ha
Forest biomass	BIO	tons/ha
Crown cover	CC	% crown cover
Forest growth	GRW	m³/ha
Quadratic mean diameter	QMD	cm
Stand density index	SDI	ha
Trees per hectare	TPH	trees > 2.54 cm DBH per ha
Forest volume	VOL	m³/ha
Weighted height	WHT	m (weighted by larger trees)
Elevation	ELEV	m
Slope	SLP	% rise
Transformed aspect	TASP	0-1

(tables 1, 2). Zone 16 Landfire map products included forested and nonforested areas; for modeling and comparison with FIA map products, nonforested regions in Landfire map products were excluded. Elevation (m), slope (%), and aspect (degrees) derived from 30-m and 250-m digital elevation models (DEMs) originating from the U. S. National Elevation Dataset. Aspect was rescaled to a scale from 0 to 1, where high values were assigned to north-northeast facing slopes (Roberts and Cooper 1989):

$$TASP = [1-\cos(aspect-30)] / 2.$$

FIA map products and the 250-m DEM were restricted to use in the 250-m models while Landfire map products and the 30-m DEM were restricted to use in the 30-m models.

Table 2.—*Descriptions of categorical spatial map products from Landfire and FIA. All units were converted to the metric system for consistency.*

Variable	Abbreviation
Landfire map product	
Cover type (dominant cover type)	COV
ponderosa pine	PP
lodgepole pine	LPP
high elevation pine[a]	HEP
Douglas fir	DF
white fir	WF
spruce/fir	SF
pinyon-juniper	PJ
juniper	J
riparian and other hardwoods	RH
aspen/birch	AB
Structure (height, forest canopy cover)	STR
forest, height <= 10m, canopy <= 40%	STR1
forest, height <= 10m, canopy > 40%	STR2
forest, height > 10m, canopy <= 40%	STR3
forest, height > 10m, canopy > 40%	STR4
FIA map product	
Forest type (dominant forest type)	FT
rocky mountain juniper[b]	PJ
juniper woodland[b]	PJ
pinyon-juniper woodland	PJ
Douglas fir	DF
ponderosa pine	PP
white fir	WF
Engelmann spruce	ES
Engelmann spruce/subalpine fir	ES/SAF
subalpine fir[c]	ES/SAF
blue spruce[d]	DF
lodgepole pine	LPP
foxtail pine/bristlecone pine[c]	ES/SAF
limber pine[c]	ES/SAF
aspen	QA
deciduous oak woodland	DO
cerocarpus woodland	CW
intermountain maple woodland[e]	DO
misc. western hardwood woodlands[e]	DO

[a] *Reassigned to lodgepole pine.*
[b] *Reassigned to pinyon-juniper woodland.*
[c] *Reassigned to Engelmann spruce/subalpine fir.*
[d] *Reassigned to Douglas fir.*
[e] *Reassigned to deciduous oak woodland.*

Northern Goshawk Nest Presence Points

Recent (1993–2005) northern goshawk nest point locations were obtained from the Ashley, Dixie, Fishlake, Manti-La Sal, Uinta, and Wasatch-Cache National Forests, and the Utah Division of Wildlife Natural Heritage database (n = 564). Only the most recently active nest per northern goshawk territory was selected to reduce spatial autocorrelation (n = 285, 1994–2005) (Reynolds 1983, Speiser and Bosakowski 1991, Woodbridge and Detrich 1994). Four nest locations were associated with Landfire nonforested pixels and six nest locations were associated with FIA nonforested pixels. These locations were eliminated, reducing the total number of nests to 281 for the 30-m models and 279 for the 250-m models.

Ecologically Based Pseudo-Absence Points

Pseudo-absence points for both the 30-m and 250-m habitat envelope models were randomly selected from within ecologically based habitat envelopes. Habitat envelopes represent the spatial extent of a species' gross habitat needs, contain a large percentage of extant presence points, and are derived from single or the intersection of multiple spatial variables. Viable northern goshawk habitat envelopes contained at least 90 percent of extant presence points (≥ 508 of 564). Habitat envelopes were created in a GIS based on existing habitat associations reported from 30 northern goshawk studies in the Western United States (Zarnetske 2006).

These northern goshawk studies provided 30-m resolution nest site and 250-m resolution nest area habitat characteristics during the breeding season. Geometric means (geomeans) of published minima and maxima values of habitat characteristics (i.e., percent canopy closure, tree height) across studies were used to set the lower and upper limits of the habitat envelopes. If minima and maxima habitat characteristics were not reported, the 95 percent confidence interval habitat characteristic values, or values of ± 1 standard deviation (SD) were used instead. If unavailable, standard deviation was calculated as $SD = SE \times \sqrt{(N)}$.

Single variable habitat envelopes were created by extracting the grid cells for a particular habitat variable that fell within geomeans of published minima and maxima. The spatial intersection of two or more envelopes produced multivariable habitat envelopes. Habitat envelopes at each model resolution containing less than 90 percent of all presence points (n < 507) were discarded. Up to three habitat envelopes from each type (1, 2, and 3 variable) which contained the highest percentages of presence points, were chosen for pseudo-absence point generation (tables 3 and 4).

Before selection of pseudo-absence points from habitat envelopes, northern goshawk nest areas and post-fledging areas (a total of 182 ha centered on each nest) for all 564 nests were removed from the habitat envelopes so that pseudo-absence points would not be selected from areas where known nests and defended territories occur (Reynolds *et al.* 1992).

Unbalanced ratios of presence to absence points can affect the accuracy of classification models (Manel *et al.* 2001, Stockwell

Table 3.—*Northern goshawk 30-m nest site habitat envelopes from Landfire map products. "∩" refers to the spatial intersection of multiple single-variable habitat envelopes. All units converted to the metric system for consistency.*

Envelopes	Values	% of all nest points contained (n = 564)	% cover of zone 16 forested area
CONASP	All conifers and aspen[a]	97.7	95.0
ELEV	1,828–3,048 m	96.0	88.0
FHT	8.97–25.6 m	96.9	79.3
CONASP ∩ FHT	All conifers and aspen[a] FHT: 8.97–25.6 m	94.9	60.1
CONASP ∩ ELEV	All conifers and aspen[a] ELEV: 1,828–3,048 m	93.7	84.6
ELEV ∩ FHT	ELEV: 1,828–3,048 m FHT: 8.97–25.6 m	92.8	65.2
CONASP ∩ ELEV ∩ FHT	All conifers and aspen[a] ELEV: 1,828–3,048 m FHT: 8.97–25.6 m	90.9	53.1

[a] *All conifers and aspen, including ponderosa pine, lodgepole pine, high elevation pine, Douglas fir, white fir, spruce/fir, pinyon/juniper, juniper, aspen/birch.*

Table 4.—*Northern goshawk 250-m nest area habitat envelopes from FIA map products. "∩" refers to the spatial intersection of multiple single-variable habitat envelopes.*

Envelopes	Values	% of all nest points contained (n = 564)	% cover of zone 16 forested area
QMD	11.5–77 cm	100.0	99.6
WHT	5–21 m	99.8	81.3
SDI	337–2,021 (in ha)	97.9	90.3
QMD ∩ WHT	QMD: 11.5–77 cm WHT: 5–21 m	99.8	80.9
QMD ∩ SDI	QMD: 11.5–77 cm SDI: 337–2,021 (in ha)	97.9	89.8
SDI ∩ WHT	SDI: 337–2,021 (in ha) WHT: 5–21 m	97.7	74.9
QMD ∩ SDI ∩ WHT	QMD: 11.5–77 cm SDI: 337–2,021 (in ha) WHT: 5–21 m	97.7	74.6
ELEV ∩ QMD ∩ WHT	ELEV: 1,828–3,048 m QMD: 11.5–77 cm	95.9	72.4
QMD ∩ CONASP ∩ WHT	QMD: 11.5–77 cm All conifers and aspen[a] WHT: 5–21 m	95.9	68.5

[a] *All conifers and aspen including pinyon-juniper, Douglas fir, ponderosa pine, white fir, Engelmann spruce, sub-alpine fir, lodgepole pine, and aspen.*

and Peterson 2002a); consequently, the proportion of presence to absence points was balanced to ensure that predicted distributions are as accurate as possible. One hundred sets of pseudo-absence points were randomly selected from each habitat envelope (n = 281 for 30-m habitat envelopes; n = 279 for 250-m habitat envelopes). A one-sample *t*-test compared the distribution of 100 samples against the habitat envelope population to test for biased samples ($P > 0.05$, one-sample *t*-test).

Traditional Pseudo-Absence Points

A set of traditional pseudo-absence points was also generated for each of the 30-m and 250-m traditional models. Instead of generating habitat envelopes, pseudo-absence points were randomly selected from the entire study region (in this case, all forested area within Zone 16). Existing northern goshawk nest areas and post-fledging areas were removed from the study area so that pseudo-absence points would not be assigned within existing territories.

Habitat Envelope Model Creation

One set of pseudo-absence points was randomly selected for each habitat envelope (Zarnetske 2006), paired with the presence points, and modeled as the response using logistic regression in R (R Project 2006). Variables used to create a habitat envelope were not included in the logistic regression models (i.e., QMD was not included as a predictor variable if the pseudo-absence points were generated from the QMD habitat envelope). Some cover types in models at 30-m resolution and some forest types in models at 250-m resolution were so low in occurrences of presences or pseudo-absences that logistic regression could not converge. Consequently, "high elevation pine" was reassigned as "lodgepole pine" within the Landfire cover type. Within FIA forest type, "rocky mountain juniper" and "juniper woodland" were reassigned "pinyon-juniper woodland," "sub-alpine fir," "limber pine," and "bristlecone pine" were reassigned "Engelmann spruce/sub-alpine fir," "blue spruce" was reassigned "Douglas fir," and "intermountain maple woodland" and "misc. western hardwoods" were reassigned "deciduous oak woodland." Decisions to reassign types were based on associated forest cover following Burns and Honkala (1990). In this manner, the presence/pseudo-absence data sets remained balanced and logistic regression was able to converge.

Ten-fold cross-validation was performed to assess model predictive capability, and associated model metrics were calculated (i.e., sensitivity, specificity, kappa, percent correctly classified [PCC], and receiver operating characteristic plot's area under the curve [AUC]). Models produced from the same set of absence points and habitat envelope were ranked according to Akaike's information criterion (AIC) (Akaike 1973), cross-validation error rate, and adjusted deviance (D^2_{adj}) to form a list of top candidate models (4 for each traditional model resolution, 18 for 30-m models, and 33 for 250-m models) (Zarnetske 2006). Because these pooled top models were not nested models containing the same variables, AIC was not an appropriate measure of performance for comparison among models. To determine the top model of each resolution, models were selected first by low cross-validation error rate (all models < 1 percent from the lowest cross-validation error rate), and then by the highest D^2_{adj}. If D^2_{adj} was equal for two or more top models, the more parsimonious model was chosen as the top model (tables 5 and 6).

Traditional Model Creation

Traditional models were created following the same methodology as described above for the habitat envelope with exceptions noted here. One set of pseudo-absence points was randomly selected for each resolution, paired with the presence points, and modeled as the response with predictor variables using logistic regression in R (Zarnetske 2006). All traditional models per resolution were ranked according to AIC. Top candidate models were each assessed for fit (D^2_{adj}) and predictive capability (sensitivity, specificity, kappa, PCC, AUC) on the training data, and internally validated by 10-fold cross validation. Models were ranked and the top model per resolution was chosen following the methodology above (tables 5 and 6).

Table 5.—*Top habitat envelope and traditional models from the list of competing top models. Only significant cover types and forest types are shown. Direction of variable influence is indicated by "+" or "–" preceding the variable. AIC and ΔAIC are reported but these models cannot be compared with AIC because they contain different sets of pseudo-absence points.*

Model name	Model	AIC	ΔAIC
30-m resolution			
TRAD30-4	+CBH*** –HCC* +FHT*** –SLP***	583.40	4.68
CONASP ∩ ELEV3	+CBH*** +FHT*** –SLP***	546.95	2.96
250-m resolution			
TRAD250-3	+ FT(–PJ*** +DF* +PP*** +WF*** +LPP*) +GRW –SLP***	584.10	3.28
QMD3	+FT(–PJ*** +DF* +PP*** +WF***) +GRW*** –SLP***	542.53	3.80

AIC = Akaike information criterion.
** P < 0.05; ** P < 0.01; *** P < 0.001*
Forest type codes: PJ = pinyon/juniper, DF = Douglas fir, PP = ponderosa pine, WF = white fir, LPP = lodgepole pine.

Table 6.—*Top models' fit and predictive capability statistics. ( ) = standard error. Error rate represents 10-fold cross-validation error rate.*

Model name	D^2_{adj}	Sensitivity	Specificity	Kappa	PCC	Error rate	AUC
30-m resolution							
TRAD30-4	0.74	0.81	0.74	0.56	0.78	23.39	0.82 (0.02)
CONASP ∩ ELEV3	0.69	0.84	0.77	0.62	0.81	19.46	0.85 (0.02)
250-m resolution							
TRAD250-3	0.72	0.78	0.70	0.48	0.74	27.03	0.82 (0.02)
QMD3	0.67	0.83	0.76	0.59	0.79	22.68	0.84 (0.02)

AUC = area under the curve; PCC = percent correctly classified.

Habitat Suitability Maps

Top habitat envelope 30-m and 250-m models' coefficients were translated into likelihood of occurrence GIS maps across forested regions of Zone 16 using StatMod extension (Garrard 2002) in ArcView 3.2 (ESRI 2005) (figs. 2 and 3). Maps depict likelihood of occurrence of nest site (30-m) or nest area (250-m) ranging from 0 to 1, where 0 to 0.25 is very unsuitable habitat (light grey), 0.25 to 0.50 is unsuitable habitat (grey), 0.50 to 0.75 is suitable habitat (dark grey), and 0.75 to 1.0 is highly suitable habitat (black).

Results

Top habitat envelope models always outperformed top traditional models in terms of model fit and predictive capability (table 6). Slope was always significantly negative in all top candidate models (table 5). Top traditional and habitat envelope 30-m models all contained forest height variables (i.e., CBH and FHT) that were significantly positive. Top traditional and habitat envelope 250-m models all contained FT either within the habitat envelope or as a predictor variable. At both resolutions, four-variable habitat envelope models always had the lowest AIC scores, but similar three-variable models were usually within $\Delta AIC < 2$, suggesting that addition of a fourth predictor variable did not improve models significantly (Zarnetske 2006).

The top 30-m traditional model (TRAD30-4) had four parameters and included a negative association with HCC, a significantly negative association with SLP, and a significantly positive association with both CBH and FHT (table 5). The top 250-m traditional model (TRAD250-3) was the most parsimonious of top competing models with low cross-validation error rates. This model had 12 parameters, including a positive association

Figure 2.—*Likelihood of goshawk nest site occurrence across Zone 16 based on top 30-m habitat envelope model.*

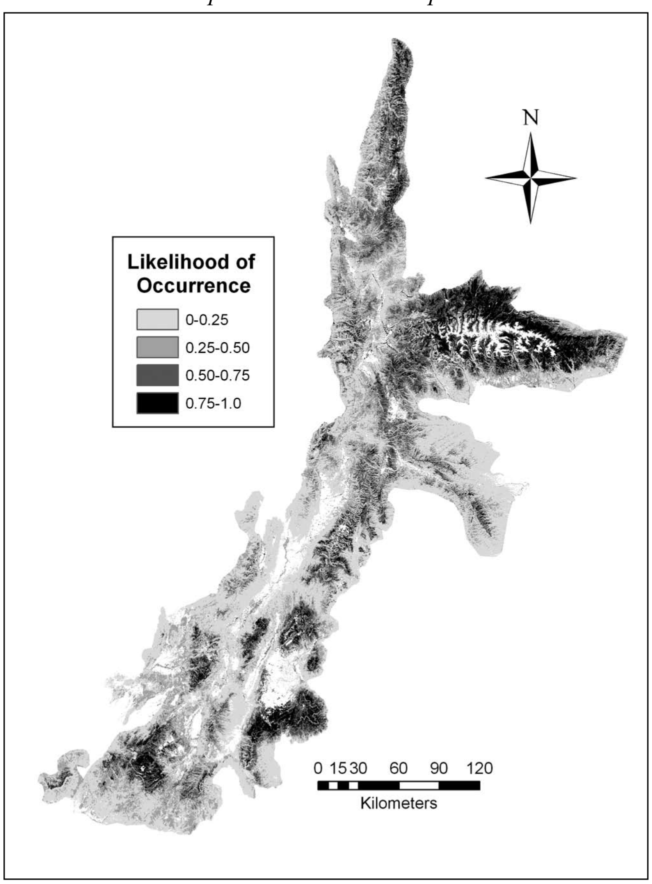
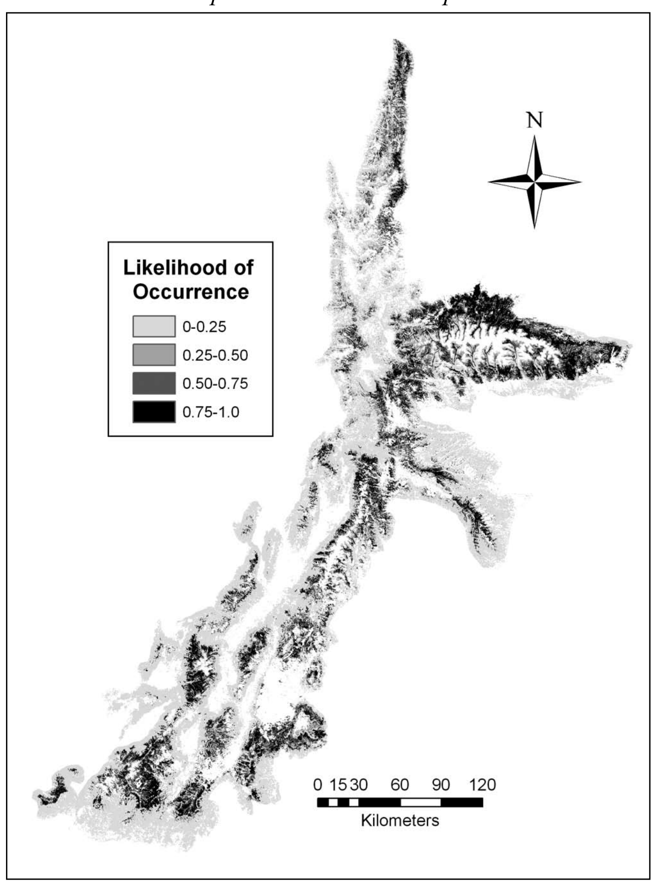


Figure 3.—*Likelihood of goshawk nest area occurrence across Zone 16 based on top 250-m habitat envelope model.*



with forest types appropriate for northern goshawk nest areas (i.e., ponderosa pine, white fir, lodgepole pine), a negative association with low suitability forest types (pinyon/juniper), a significantly positive association with GRW, and a significantly negative association with SLP (Zarnetske 2006).

The CONASP ∩ ELEV habitat envelope produced the pseudo-absence points used in the top 30-m habitat envelope model. These pseudo-absences were less variable than those used in the top 30-m traditional model for most variables. CONASP ∩ ELEV3 was the most parsimonious of competing models with only three parameters (CBH, FHT, SLP). QMD3 was the top 250-m habitat envelope, and pseudo-absence points in this model were also less variable than those used in TRAD250-3 for most variables (Zarnetske 2006).

Discussion

The ecology of the species-habitat relationship was successfully incorporated into the generation of ecologically based pseudo-absence points by creating habitat envelopes. This method can be a powerful asset for extracting information from extant species occurrence databases, which are often underutilized because traditional modeling techniques require both presence and absence points. Rare, threatened, and sensitive species such as the northern goshawk can benefit from the production of landscape-scale habitat suitability maps using ecologically based pseudo-absence points. These habitat suitability maps provide information for species range shift studies, censuses, reserve designs, species reintroduction, habitat restoration, and biodiversity conservation. Ecologically based pseudo-absence points can be applied to any species, ecosystem, data resolution, and spatial extent, given some *a priori* knowledge concerning the species-habitat relationship. In addition, the use of readily available software (e.g., Program R and ArcGIS) allows easy application and the flexibility to work with user-defined data structures and map products.

Representation of underlying ecological relationships will always improve when observed and measured biological relationships are incorporated into modeling and extrapolation (Belovsky *et al.* 2004). The consistent improvement in model fit and predictive capability of habitat envelope models is a testament to using ecologically-based pseudo-absence points over traditional pseudo-absences. In addition, reliance on sound statistical modeling with pseudo-absences may be more appropriate than obscure profile-type models that are difficult to interpret and assess for model fit and predictive capability. In cases where field-collected absences are suspected to be false (Graham *et al.* 2004; Hirzel *et al.* 2001, 2002), ecologically based pseudo-absences could provide more robust absences.

Incorporating ecology into pseudo-absence point generation should also decrease the chance of a biologically inappropriate top model because habitat envelopes already constrain the regions of pseudo-absence point generation to preferred habitat. Pseudo-absences from habitat envelopes will be less variable on average than those generated from an entire study area due to the reduced region from which they are selected. In this study, this allowed the statistical models to improve classification of northern goshawk highly suitable and moderately suitable habitat, and to create more statistically and ecologically robust models. The distinction of highly suitable from moderately suitable is essential to species conservation and habitat management, particularly for rare, threatened, and sensitive species susceptible to habitat fragmentation and degradation. Researchers interested in well-studied species such as the northern goshawk will be able to focus on highly suitable habitat within the range of suitable habitat. Lesser-known species will still benefit from ecologically based pseudo-absence points because presumably some aspect of their habitat association is known (i.e., they prefer forested over nonforested areas).

A successful habitat envelope will maximize the percent of presence points contained while at the same time reducing the area from which pseudo-absence points are selected. Based on the high number of presence points contained, the reduced area from which pseudo-absence points are selected, and the high model fit and predictive capability statistics, the CONASP ∩ ELEV habitat envelope appears to be the best habitat envelope at 30-m resolution for the northern goshawk. The top 250-m habitat envelope model included the same parameters (FT, GRW, SLP) as the top traditional model, and almost the same area as the entire study region (99 percent of the study region). The similarity in model parameters and spatial extent caused more heterogeneous ecologically based pseudo-absence points than expected, but the top 250-m habitat envelope model was still an improvement in model fit and predictive capability over the top 250-m traditional model. High AUC values from these top habitat envelope models indicate that the models have useful application across the study area because they are insensitive to threshold cut-off values (Fielding and Bell 1997, Swets 1988), and have few false positives. Due to the immense resources needed to census goshawk metapopulations across an area the size of Zone 16, false positives would be detrimental to census efforts. The high sensitivity, specificity, kappa, and PCC, and low cross-validation error rate for both top habitat envelope models suggests that sampling high likelihood of occurrence sites may lead to the discovery of new nest sites and nest areas.

Northern goshawk habitat associations were pooled from the entire Western United States, consisting of a variety of forested ecosystems. If more northern goshawk habitat studies were available for regions similar to Utah's central highlands (i.e., Intermountain West field studies), and only this habitat information was considered in creating habitat envelopes, more two- and three-variable habitat envelopes could have provided robust sets of pseudo-absence points. In addition, the entire study area already represents a habitat envelope of sorts because it is constrained to forests in Zone 16. If traditional models were produced over all of Zone 16 (i.e., 55 percent forest and 45 percent nonforest), they would better reflect the current methods of generating pseudo-absence points.

The models contain error from a variety of sources. As models themselves, each FIA and Landfire map product contributes spatial and classification error. Because FIA and Landfire map products of Zone 16 incorporate plot-based inventory data and satellite imagery spanning from 1998–2003, certain cells within the map products may inaccurately reflect current conditions (Blackard *et al.* 2004, FIA 2005, Landfire 2005). DEMs and spatial error associated with nest points are additional sources of error. No known methods exist to incorporate the inherent error of spatial layers into statistical classification models.

The assumptions that habitat is saturated (Capen *et al.* 1986), and that the species modeled is in equilibrium with its environment (Austin 2002), are often ignored in broad-scale habitat modeling because knowing the locations of all individuals or the individuals' attribute (i.e., a nest) is nearly impossible across a broad landscape, particularly at one time step. It is probable that not all northern goshawk territories in Zone 16 have been identified and that some of the nests used in the models are in sink habitats. Some nests used in modeling may be in sink habitat due to incomplete nest activity information. Most large-scale species occurrence datasets include data collected by a variety of survey methods and do not have complete data point attribute information.

The likelihood of occurrence of nest sites and nest areas across Zone 16 reflects environmental habitat variables only. Prey abundance, nest productivity, and interannual climatic variability are important variables driving nest success in addition to appropriate habitat type (Doyle and Smith 1994, Keane 1999, Reynolds *et al.* 2006, Salafsky *et al.* 2005, Wiens *et al.* 2006). These variables were not available for each nest location across Zone 16; consequently, northern goshawk nest placement does not necessarily correlate directly with nest productivity. It is for this reason that the nest site and nest area habitat suitability maps should be used as guides to locate new highly suitable habitat, and new nests and territories. Intensifying sampling in highly suitable areas while decreasing sampling effort in low suitability areas should increase sampling efficiency.

As new nests are located, the model can be adjusted to incorporate more recently active nests. Ideally, a northern goshawk HCM would use 1 year's nest activity information across Zone 16 so that unoccupied known nests could be treated as absences and occupied nests could be treated as presences. On a landscape scale such as Zone 16, this could be achieved on a per-season basis by identifying successful nests (i.e., those that produced fledglings) during existing routine national forest nest monitoring, determining competitor presence and prey abundance through distance sampling (following Salafsky *et al.* 2005), using spatial habitat data such as FIA and Landfire map products, incorporating disturbance extents and severity ratings (such as timber sale, fire, and beetle kill), and recording weather extremes and climatic events in proximity to each nest. In this manner, data collected on a per-year basis would allow for per-breeding season analysis of suitable and successful nesting habitat. The likelihood of occurrence of active nests could be modeled following Reich *et al.* (2004), but using pseudo-absence points to accompany known presence points. If these assessments continue for several years in a row, the variability between years will be captured and a more complete assessment of across-landscape habitat suitability and nest success will be clear.

Researchers and conservationists will gain insight into the level of habitat saturation throughout Zone 16 through yearly monitoring of new nests found with the help of this model, and monitoring existing nests for activity. Combining this knowledge of nest activity with age of adult breeders would help determine the stability of the Zone 16 population (Kenward *et al.* 1999, Reynolds *et al.* 2006, Reynolds and Joy 2006). Sympatric species such as the sharp-shinned hawk (*Accipiter striatus*), spotted owl (*Strix occidentalis*), and barred owl (*Strix varia*), have similar nesting habitat requirements to the northern goshawk (Bildstein and Meyer 2000, Gutiérrez *et al.* 1995, Mazur and James 2000) and will likely benefit from northern goshawk habitat conservation. If possible, northern goshawks should be assessed on a bioregional scale, incorporating population demographics across the Western United States (Woodbridge and Hargis 2006). HCMs using ecologically based pseudo-absence points with FIA and Landfire map products to locate new territories and nests across the Western United States will assist this bioregional assessment.

Conclusions

FIA and Landfire map products will be useful in habitat assessments to a range of species, particularly rare, threatened, and sensitive species. As alteration to U.S. landscapes continue, it is becoming increasingly important to insure connectivity among ecosystems and available habitat for species' metapopulations across entire ecoregions. The production of these map products across the ecoregions of the United States will greatly assist species habitat assessments, land-use planning, and ecosystem conservation over broad spatial extents. Ecologically-based pseudo-absence points in combination with extensive land cover map products such as FIA and Landfire have the potential to assist a wide variety of species' habitat assessments and increase the utility of database presence points.

Acknowledgments

The authors thank the Utah National Forest Biologists, Utah Division of Wildlife Resources, the FIA Ogden office, and the USDA Landfire Program for their cooperation in this project.

Literature Cited

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N.; Csaki, F., eds. Proceedings, second international symposium on information theory. Budapest, Hungary: Academia Kiado: 267-281.

Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling. 157(2-3): 101-118.

Beier, P.; Drennan, J.E. 1997. Forest structure and prey abundance in foraging areas of northern goshawks. Ecological Applications. 7(2): 564-571.

Belovsky, G.E.; Botkin, D.B.; Crowl, T.A.; Cummins, K.W.; Franklin, J.F.; Hunter, M.L., Jr.; Joern, A.; Lindenmayer, D.B.; MacMahon, J.A.; Margules, C.R.; Scott, J.M. 2004. Ten suggestions to strengthen the science of ecology. Bioscience. 54(4): 345-351.

Bildstein, K.L.; Meyer, K. 2000. Sharp-shinned hawk (*Accipiter striatus*). In: Poole, A.; Gill, F., eds. The birds of North America, No. 482. Philadelphia, PA: Academy of Natural Sciences; Washington, DC: American Ornithologists' Union. http://bna.birds.cornell.edu/BNA/account/Sharp-shinned_Hawk/. (1 March 2005).

Blackard, J.A.; Moisen, G.G.; Tymcio, R.P. 2004. Modeling investigations into mapping of forest attributes with FIA data for the State of Wyoming. Ogden, UT: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, Rocky Mountain Research Station.

Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. 1984. Classification and regression trees. Belmont, CA: Wadsworth International Group. 358 p.

Bright-Smith, D.J.; Mannan, R.W. 1994. Habitat use of breeding male northern goshawks in northern Arizona. Studies in Avian Biology. 16: 58-65.

Brotons, L.; Thuiller, W.; Araujo, M.B.; Hirzel, A.H. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography. 27(4): 37-448.

Burns, R.M.; Honkala, B.H., tech. coords. 1990. Silvics of North America: 1. conifers; 2. hardwoods. Agriculture Handbook 654. Washington, DC: U.S. Department of Agriculture, Forest Service.

Busby, J.R. 1991. BIOCLIM—a bioclimatic analysis and prediction system. In: Margules, C.R.; Austin, M.P., eds. Nature conservation: cost effective biological surveys and data analysis. Canberra, Australia: CSIRO: 64-68.

Capen, D.E.; Fenqick, J.W.; Inkley, D.B.; Boynton, A.C. 1986. Multivariate models of songbird habitat in New England forests. In: Verner, J.A.; Morrison, M.L.; Ralph, C.J., eds. Wildlife 2000: modelling habitat relationships of terrestrial vertebrates. Madison, WI: University of Wisconsin Press: 171-175.

Chicoine, T.K.; Fay, P.K.; Nielsen, G.A. 1985. Predicting weed migration from soil and climate maps. Weed Science. 3(1): 57-61.

Doyle, F.I.; Smith, J.M.N. 1994. Population response of northern goshawks to the 10-year cycle in numbers of snowshoe hares. Studies in Avian Biology. 16: 122-129.

Engler, R.; Guisan, A.; Rechsteiner, L. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology. 41(2): 263-274.

Environmental Research Systems Institute, Inc. (ESRI). 2005. Redlands, CA: Environmental Research Systems Institute, Inc.

Ferrier, S.; Watson, G. 1997. An evaluation of the effectiveness of environmental surrogates and modeling techniques in predicting the distribution of biological diversity. Canberra, Australia: Environment Australia.

Fielding, A.H.; Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation. 24(1): 38-49.

Forest Inventory and Analysis (FIA). 2005. Arlington, VA: Forest Inventory and Analysis National Program. http://www.fia.fs.fed.us. (1 September 2005).

Garrard, C. 2002. Statmod ArcView Extension. http://www.gis.usu.edu/~chrisg/avext. (30 March 2005).

Graham, C.H.; Ferrier, S.; Huettman, F.; Moritz, C.; Peterson, A.T. 2004. New developments in museum-based informatics and application in biodiversity analysis. Trends in Ecology and Evolution. 19: 497-503.

Graham, R.T.; Rodriguez, R.L.; Paulin, K.M.; Player, R.L.; Heap, A.P.; Williams, R. 1999. The northern goshawk in Utah: habitat assessment and management recommendations. Gen. Tech. Rep. RMRS-GTR-22. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 48 p.

Guisan, A.; Edwards, T.C., Jr.; Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling. 157(2-3): 89-100.

Guisan, A.; Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters. 8(9): 993-1009.

Guisan, A.; Weiss, S.B.; Weiss, A.D. 1999. GLM versus CCA spatial modeling of plant distribution. Plant Ecology. 143(1): 107-122.

Gutiérrez, R.J.; Franklin, A.B.; Lahaye, W.S. 1995. Spotted owl (*Strix occidentalis*). In: Poole, A.; Gill, F., eds. The birds of North America, No. 179. Philadelphia, PA: Academy of Natural Sciences; Washington, DC: American Ornithologists' Union. http://bna.birds.cornell.edu/BNA/account/Spotted_Owl/. (1 March 2005).

Hirzel, A.H.; Hausser, J.; Chessel, D.; Perrin, N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data. Ecology. 83(7): 2027-2036.

Hirzel, A.H.; Helfer, V.; Métral, F. 2001. Assessing habitat-suitability models with a virtual species. Ecological Modelling. 145(2-3): 111-121.

Keane, J.J. 1999. Ecology of the northern goshawk in the Sierra Nevada, California. Davis, CA: University of California. 123 p. Ph.D. dissertation.

Keane, R.E.; Rollins, M.G.; McNicoll, C.H.; Parsons, R.A. 2002. Integrating ecosystem sampling, gradient modeling, remote sensing, and ecosystem simulation to create spatially explicit landscape inventories. RMRS-GTR-92. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 61 p.

Kenward, R.E.; Marcstrom, V.; Karlbom, M. 1999. Demographic estimates from radiotagging: models of age-specific survival and breeding in the goshawk. Journal of Animal Ecology. 68(5): 1020-1033.

Landfire. 2005. U.S. Department of Agriculture, Landfire Program. http://www.landfire.gov. (2 October 2004).

Levin, S.A.; Grenfell, B.; Hastings, A.; Perelson, A.L. 1997. Mathematical and computational challenges in population biology and ecosystems science (in bioinformatics). Science, New Series. 275(5298): 334-343.

Manel, S.; Dias, J.M.; Buckton, S.T.; Ormerod, S.J. 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. Journal of Applied Ecology. 36(5): 734-747.

Manel, S.; Williams, H.C.; Ormerod, S.J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology. 38(5): 921-931.

Mazur, K.M.; James, P.C. 2000. Barred owl (*Strix varia*). In: Poole, A.; Gill, F., eds. The birds of North America, No. 508. Philadelphia, PA: Academy of Natural Sciences; Washington, DC: American Ornithologists' Union. http://bna.birds.cornell.edu/BNA/account/Barred_Owl/. (1 March 2005).

McNoleg, O. 1996. The integration of GIS, remote sensing, expert systems and adaptive co-kriging for environmental habitat modeling of the highland haggis using object-oriented, fuzzy-logic and neural-network techniques. Computers and Geosciences. 22(5): 585-588.

Newton, I. 1979. Population ecology of raptors. Vermillion, SD: Buteo Books. 399 p.

Pearce, J.L.; Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling. 133(3): 225-245.

Phillips, S.J.; Anderson, R.P.; Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling. 190(3-4): 231-259.

R Project. 2006. R Project for statistical computing. http://www.r-project.org. (8 January 2004).

Reich, R.M.; Joy, S.M.; Reynolds, R.T. 2004. Predicting the location of goshawk nests: modeling the spatial dependency between nest locations and forest structure. Ecological Modelling. 176(1-2): 109-133.

Reynolds, R.T. 1983. Management of western coniferous forest habitat for nesting accipiter hawks. Gen. Tech. Rep. RM-102. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 7 p.

Reynolds, R.T.; Graham, R.T.; Reiser, M.H.; Bassett, R.L.; Kennedy, P.L.; Boyce, D.A., Jr.; Goodwin, G.; Smith, R.; Fisher, E.L. 1992. Management recommendations for the northern goshawk in the southwestern United States. Gen. Tech. Rep. RM-217. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 184 p.

Reynolds, R.T.; Joy, S.M. 2006. Demography of northern goshawks in northern Arizona, 1991–1996. Studies in Avian Biology. 31: 63-74.

Reynolds, R.T.; Joy, S.M.; Leslie, D.G. 1994. Nest productivity, fidelity, and spacing of northern goshawks in Arizona. Studies in Avian Biology. 16: 106-113.

Reynolds, R.T.; Meslow, E.C.; Wight, H.M. 1982. Nesting habitat of coexisting accipiters in Oregon. Journal of Wildlife Management. 46(1): 124-138.

Reynolds, R.T.; Wiens, J.D.; Joy, S.M.; Salafsky, S.R. 2005. Sampling considerations for demographic and habitat studies of northern goshawks. Journal of Raptor Research. 39(3): 274-285.

Reynolds, R.T.; Wiens, J.D.; Salafasky, S.R. 2006. A review and evaluation of factors limiting northern goshawk populations. Studies in Avian Biology. 31: 260-273.

Roberts, D.W.; Cooper, S.V. 1989. Concepts and techniques of vegetation mapping. In: Ferguson, D.; Morgan, P.; Johnson, F.D., eds. Land classifications based on vegetation: applications for resources management. Gen. Tech. Rep. INT-257. Ogden, UT: U.S. Department of Agriculture, Forest Service: 90-96.

Robertson, M.P.; Villet, M.H.; Palmer, A.R. 2004. A fuzzy classification technique for predicting species' distributions: applications using invasive alien plants and indigenous plants. Diversity and Distributions. 10(5-6): 461-474.

Salafsky, S.R.; Reynolds, R.T.; Noon, B.R. 2005. Patterns of temporal variation in goshawk reproduction and prey resources. Journal of Raptor Research. 39(3): 237-246.

Scott, J.M.; Heglund, P.J.; Morrison, M.L.; Haufler, J.B.; Raphael, M.G.; Wall, W.A.; Samson, F.B., eds. 2002. Predicting species occurrences, issues of accuracy and scale. Washington, DC: Island Press. 868 p.

Speiser, R.; Bosakowski, T. 1991. Nesting phenology, site fidelity, and defense behavior of northern goshawks in New York and New Jersey. Journal of Raptor Research. 25(4): 132-135.

Squires, J.R.; Reynolds, R.T. 1997. Northern goshawk (*Accipiter gentilis*). In: Poole, A.; Gill, F., eds. The birds of North America, No. 298. Philadelphia, PA: Academy of Natural Sciences; Washington, DC: American Ornithologists' Union. http://bna.birds.cornell.edu/BNA/account/Northern_Goshawk/. (10 February 2005).

Stephens, R.M. 2001. Migration, habitat use, and diet of northern goshawks that winter in the Uinta Mountains, Utah. Laramie, WY: University of Wyoming. M.S. thesis.

Stockwell, D.R.B.; Noble, I.R. 1991. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. Mathematics and Computers in Simulation. 32: 249-254.

Stockwell, D.R.B.; Peters, D. 1999. The GARP modeling system: problems and solutions to automated spatial prediction. International Journal of Geographic Information Science. 2: 143-158.

Stockwell, D.R.B.; Peterson, A.T. 2002a. Effects of sample size on accuracy of species distribution models. Ecological Modelling. 148(1): 1-13.

Stockwell, D.R.B.; Peterson, A.T. 2002b. Controlling bias in biodiversity data. In: Scott, J.M.; Heglund, P.J.; Morrison, M.L.; Haufler, J.B.; Raphael, M.G.; Wall, W.A.; Samson, F.B., eds. Predicting species occurrences: issues of accuracy and scale. Covelo, CA: Island Press: 537-546.

Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. Science. 240(4857): 1285-1293.

Thuiller, W.; Araújo, M.B.; Lavorel, S. 2003. Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. Journal of Vegetation Science. 14(5): 669-680.

Wiens, J.D.; Noon, B.R.; Reynolds, R.T. 2006. Effects of fluctuating food and weather conditions on post-fledging survival of juvenile northern goshawks. Ecological Applications. 16: 406-418.

Woodbridge, B.; Hargis, C.D. 2006. A design for monitoring northern goshawks (*Accipiter gentilis*) at the bioregional scale. Studies in Avian Biology. 31: 274-287.

Woodbridge, B.P.; Detrich, J. 1994. Territory occupancy and habitat patch size of northern goshawks in the southern Cascades of California. Studies in Avian Biology. 16: 83-87.

Zaniewski, A.E.; Lehmann, A.; Overton, J.M. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling. 157(2-3): 261-280.

Zarnetske, P.L. 2006. Landscape-scale habitat classification modeling with incomplete data: the generation of ecologically-based pseudo-absence points. Logan, UT: Utah State University. 168 p. M.S. thesis.