*Ecological Applications*, 17(6), 2007, pp. 1714–1726


# HABITAT CLASSIFICATION MODELING WITH INCOMPLETE DATA: PUSHING THE HABITAT ENVELOPE

PHOEBE L. ZARNETSKE,[1,4] THOMAS C. EDWARDS, JR.,[2] AND GRETCHEN G. MOISEN[3]

[1]*Ecology Center and Department of Forest, Range, and Wildlife Sciences, Utah State University, Logan, Utah 84322-5230 USA*
[2]*USGS Utah Cooperative Fish and Wildlife Research Unit, College of Natural Resources, Utah State University, Logan, Utah 84322-5290 USA*
[3]*USDA Forest Service, Rocky Mountain Research Station, 507 25th Street, Ogden, Utah 84401 USA*

*Abstract*. Habitat classification models (HCMs) are invaluable tools for species conservation, land-use planning, reserve design, and metapopulation assessments, particularly at broad spatial scales. However, species occurrence data are often lacking and typically limited to presence points at broad scales. This lack of absence data precludes the use of many statistical techniques for HCMs. One option is to generate pseudo-absence points so that the many available statistical modeling tools can be used. Traditional techniques generate pseudo-absence points at random across broadly defined species ranges, often failing to include biological knowledge concerning the species–habitat relationship. We incorporated biological knowledge of the species–habitat relationship into pseudo-absence points by creating habitat envelopes that constrain the region from which points were randomly selected. 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. We created HCMs for Northern Goshawk (*Accipiter gentilis atricapillus*) nest habitat during the breeding season across Utah forests with extant nest presence points and ecologically based pseudo-absence points using logistic regression. Predictor variables were derived from 30-m USDA Landfire and 250-m Forest Inventory and Analysis (FIA) map products. These habitat-envelope-based models were then compared to null envelope models which use traditional practices for generating pseudo-absences. Models were assessed for fit and predictive capability using metrics such as kappa, threshold-independent receiver operating characteristic (ROC) plots, adjusted deviance ($D^2_{adj}$), and cross-validation, and were also assessed for ecological relevance. For all cases, habitat envelope-based models outperformed null envelope models and were more ecologically relevant, suggesting that incorporating biological knowledge into pseudo-absence point generation is a powerful tool for species habitat assessments. Furthermore, given some a priori knowledge of the species–habitat relationship, ecologically based pseudo-absence points can be applied to any species, ecosystem, data resolution, and spatial extent.

*Key words:* Accipiter gentilis atricapillus; *Forest Inventory and Analysis (FIA); habitat classification model; habitat envelope; Landfire; management indicator species; Northern Goshawk; pseudo-absence.*

## INTRODUCTION

One of the more challenging issues in ecology is evaluating the utility of extant data for use in resolving crucial conservation problems, especially for rare, sensitive, threatened, and endangered species management. Many state and federal agencies, non-government organizations, museum collections, and herbaria have a wealth of existing species occurrence information from which ecological insights could be extracted and applied towards a broad set of conservation problems, such as the identification of crucial habitats or landscape-scale species distribution models. Unfortunately, many of these data were collected in fashions not well suited for many of today's sophisticated forms of statistical analysis (e.g., generalized linear models [GLMs], McCullagh and Nelder 1989; generalized additive models [GAMs], Hastie and Tibshirani 1990; classification trees, Breiman et al. 1984), especially when used for classification purposes such as habitat modeling (Guisan et al. 2002). The more common concerns directed towards extant data typically center on design issues, with non-probabilistic forms of sampling (e.g., purposive sampling; Schreuder et al. 2001) and the associated biases (Edwards et al. 2006) being one of the more common. Nonetheless, these data contain important biological information that can be applied towards numerous conservation issues.

Here we explore a means for pairing extant, presence-only data with so-called pseudo-absences (Ferrier and Watson 1997) generated from ecological backgrounds, thereby facilitating the use of classification tools for

Manuscript received 2 August 2006; revised 1 March 2007; accepted 5 March 2007. Corresponding Editor: R. L. Knight.

[4] Present address: Department of Zoology, 3029 Cordley Hall, Oregon State University, Corvallis, Oregon 97331 USA. E-mail: zarnetsp@science.oregonstate.edu

modeling species habitat and likelihoods of species occurrence (known as habitat classification models (HCMs) or species distribution models). Presence-only data often arise when biologists target a particular ecological attribute considered important for species conservation and management, such as the occurrence of nesting sites, and ignore locations where the targeted element is absent. This lack of absence data precludes the use of many classification techniques mentioned above for modeling crucial habitats and the likelihoods of species occurrence in those habitats. Two basic approaches exist for handling presence-biased datasets for HCM: (1) incorporating the presence-only bias into the model and controlling its effects on the resulting predictions (e.g., Hirzel et al. 2002, Stockwell and Peterson 2002, Phillips et al. 2006); or (2) generating pseudo-absence points to be used in place of unknown absence data (e.g., Ferrier and Watson 1997, Zaniewski et al. 2002, Engler et al. 2004, Graf et al. 2005, and Lütolf et al. 2006).

Although bias is associated with presence-only data, they can still be used for modeling purposes when true absences are lacking. Often referred to as profile-type models, example approaches include environmental niche factor analysis (ENFA; Hirzel et al. 2002), ecological niche modeling (with Mahalanobis $D^2$; Rotenberry et al. 2006), the use of set theory in simple and fuzzy envelope techniques (SEMs and FEMs in BIOCLIM; Chicoine et al. 1985, Busby 1991), artificial intelligence methods for use in genetic algorithm for rule-set production (GARP; Stockwell and Peters 1999), and statistical mechanics in maximum entropy modeling (Maxent; Phillips et al. 2006). The resulting output is typically a habitat suitability index or a species potential distribution.

Several authors have suggested generating so-called pseudo-absences to pair with extant presences as an alternative to profile-type models (Ferrier and Watson 1997, Stockwell and Peters 1999, Zaniewski et al. 2002). The most common technique for generating pseudo-absences involves randomly selecting absence points from the entire study region, excluding where the species is found (Parra-Olea et al. 2005, Stockman et al. 2006). These techniques do not typically incorporate ecological knowledge of the species-habitat relationship; for the most part, pseudo-absences are selected from broadly defined areas, such as the study region or simple range maps.

Recently, researchers have applied different methods to constrain the region of pseudo-absence point selection, and tested these variations for habitat modeling in Switzerland. Engler et al. (2004) found that the best GLMs of the endangered perennial hemi-cryptophyte (*Eryngium alpinum*) used randomly generated pseudo-absence points from ENFA-defined areas where the species was unlikely to occur. Similarly, Graf et al. (2005) created habitat models for the capercaillie (*Tetrao urogallus*) using a variety of predictor variable spatial scales, and found that the best models used pseudo-absence points generated from non-presence areas. Finally, Lütolf et al. (2006) found that the best models for three target butterfly species used pseudo-absence points generated within regions where neither target species nor sympatric species occurred for a century. These techniques generate pseudo-absence points from outside the presence areas or within low-suitability habitat, rather than across the entire study region.

Here we incorporate ecology into the selection of pseudo-absences, improve classification of suitable habitat (i.e., identify suitable vs. highly suitable), and reduce the inherent variability in pseudo-absences drawn from an entire study region. Our technique relies on external published knowledge of the species–habitat relationship to create habitat envelopes, which we define here as ecological representations 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. Habitat envelopes constrain the region from which species pseudo-absence points are generated to within the species distribution. These habitat envelopes can range from simple to complex, depending on the amount of ecological knowledge they incorporate. We refer to pseudo-absence points generated from within habitat envelopes as "ecologically based" pseudo-absence points. A key distinction of our method (as opposed to methods of Engler et al. 2004, Graf et al. 2005, and Lütolf et al. 2006, which randomly generate pseudo-absence points outside the species known distribution) is that it generates pseudo-absence points from *within* the species distribution (excluding its known presence areas), enabling improved differentiation between suitable and highly suitable habitat, which is extremely important for conservation of rare, sensitive, threatened, and endangered species.

Our goal was to determine whether the incorporation of known ecological relationships into habitat envelopes (from which ecologically based pseudo-absence points were selected) could increase the classification ability in a HCM. We selected a habitat specialist species, the Northern Goshawk (*Accipiter gentilis atricapillus*), as our test species. The Northern Goshawk was an ideal species to model with ecologically based pseudo-absence points because of its high degree of nest habitat specificity, its status as a management indicator species and as a sensitive species by various management agencies in the Intermountain West of the United States, and the availability of a relatively large extant data set of nest presence points.

Known ecological associations of Northern Goshawk nest locations to habitat variables were translated into increasingly complex habitat envelopes from which ecologically based pseudo-absence points were randomly selected. Ecologically based pseudo-absence points were then paired with extant presence points in logistic regression to model the likelihood of occurrence of

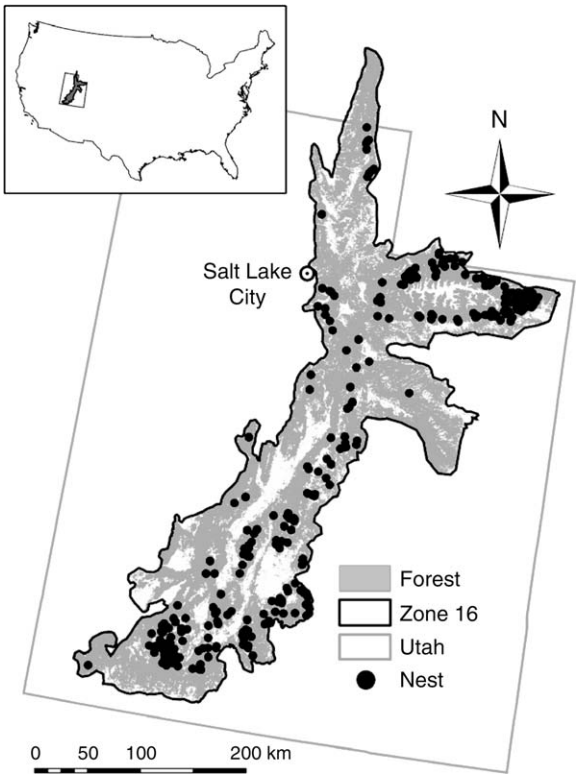


FIG. 1. Zone 16 study region map. Forested areas are in black; white areas within Zone 16 represent non-forest. Black circles are Northern Goshawk nest presence points ($n = 281$ nests) used in the nest site models. The same Northern Goshawk nest presence points were used in the nest area models, less two nest presence points in the Uinta Mountains (northeast region of Zone 16).

Northern Goshawk nest site and nest area as a function of habitat predictor variables. These likelihood of occurrence models were next translated into spatially explicit habitat suitability maps across the study region for evaluation. Null models, based on pseudo-absence points selected from the entire study region, were developed in a similar fashion. We compared the habitat envelope-based models with the null models using a variety of accuracy metrics via ten-fold cross-validation, and tested these accuracy metrics between models with a rank test.

We predicted that models using ecologically based pseudo-absence points would improve classification and produce more ecologically relevant models because any time observed and measured ecological relationships are incorporated into modeling and extrapolation, the resulting outcomes are more representative of true underlying ecological relationships (Belovsky et al. 2004).

## METHODS

### *Study region and species*

Our study region covered the forested land of the Wasatch and Uinta mountain ranges in the central Utah highlands (hereafter Zone 16, Homer and Gallant 2001; Fig. 1). Fifty-five percent of Zone 16 is forest ($\sim 4 \times 10^6$ ha) and includes aspen (*Populus tremuloides*), Douglas-fir (*Pseudotsuga menziesii*), Engelmann spruce (*Picea engelmannii*), blue spruce (*P. pungens*), subalpine fir (*Abies lasiocarpa*), white fir (*A. concolor*), lodgepole pine (*Pinus contorta*), high elevation pines (limber pine; *P. flexilis,* and bristlecone pine; *P. longaeva*), ponderosa pine (*P. ponderosa*), pinyon pine (*P. edulis*), juniper (*Juniperus osteosperma* and *J. scopulorum*), Gambel oak (*Quercus gambelii*), bigtooth maple (*Acer grandidentatum*), and cercocarpus woodlands. Elevation in Zone 16 ranges from 386 to 3978 m. The study region was confined to only forested land to represent generally appropriate nesting habitat for the study species (i.e., non-forest area was excluded).

The Northern Goshawk is the largest accipiter in North America, and is holarctic in distribution. Home range is ~2370 ha, and consists of three components: nest area (~12 ha), post-fledging area (~170 ha), and the foraging area (~2188 ha; Reynolds et al. 1992). Multiple nest areas typically exist within a home range, and multiple satellite nests occur within each nest area (Reynolds et al. 1994, Squires and Reynolds 1997). In general, large trees (>40.6 cm diameter at breast height [dbh]) arranged in a clump with dense canopy cover (Hayward and Escaño 1989, Bright-Smith and Mannan 1994, Beier and Drennan 1997, Graham et al. 1999) are preferred for nesting. High territory occupancy and high nesting rates are associated with increased tree density and canopy closure in the nest area (Finn et al. 2002).

We modeled two ecological components of Northern Goshawk nesting habitat based on in-field measurement resolutions. The first, the nest site, was defined as the nest tree and habitat immediately surrounding the nest tree, an area 0.10 ha in size (Reynolds et al. 1982, Squires and Reynolds 1997). The second was nest area, which was defined as habitat 0.10–12 ha beyond the nest site, including adult roosts and prey plucking sites (Reynolds et al. 1992). Nest site and nest area were both used to test the ability of ecologically based pseudo-absence points to outperform traditional pseudo-absence points at fine and coarse resolutions, respectively.

### *Habitat predictor variables*

Two resolutions of predictor variables were used in our analyses to match the two resolutions of Northern Goshawk nesting habitat, and to test the utility of two map product sets. Coarse resolution predictor variables originated from the U.S. Department of Agriculture Forest Inventory and Analysis (FIA) Program (data *available online*).[5] In addition to reporting numerous silvicultural metrics for forest management, FIA recently generated 250-m resolution vegetation map products using regression-tree modeling techniques of forest plot-

[5] ⟨http://www.fia.fs.fed.us⟩

based inventories and 250-m MODIS imagery across Zone 16 (Blackard et al. 2004; and see footnote 5). FIA digital representations of forest attributes were used to model the nest area. Finer resolution predictor variables came from a multi-partner ecosystem and wildland fire mapping project called Landfire (involving the USDA Forest Service Fire Lab, United States Geological Survey [USGS], and The Nature Conservancy; project description *available online*).[6] Regression tree modeling with plot inventories (including FIA plot data), biophysical gradients, and 30-m Landsat TM imagery created 30-m resolution Landfire prototype map products for Zone 16 (Rollins et al. 2006). Zone 16 Landfire map products included forested and non-forested areas; for modeling and comparison with FIA map products, non-forested regions in Landfire map products were excluded. Landfire map products were used to model the nest site.

Eleven FIA and 10 Landfire map products were used as Northern Goshawk habitat predictor variables (Table 1). Topographic variables (i.e., elevation [m], slope [%]. aspect [degrees]) were derived from 30-m and 250-m digital elevation models (DEMs) originating from the U.S. National Elevation Data set (Table 1). Aspect was transformed to a scale from 0 to 1, where the highest values were assigned to north-northeast-facing slopes (Roberts and Cooper 1989), $\text{TASP} = [1 - \cos(\text{aspect} - 30)]/2$. To reflect the habitat distinctions of the nest site and nest area, map products at 250-m resolution (FIA and DEM) and 30-m resolution (Landfire and DEM) were restricted to the 250-m nest area and 30-m nest site models, respectively.

TABLE 1. Predictor variables derived from FIA (250-m resolution) and Landfire (30-m resolution) spatial map products, and digital elevation models (DEMs).

| Variable | Abbreviation | Description |
|---|---|---|
| FIA | | |
| Stand age | AGE | yr |
| Basal area | BA | $m^2$/ha |
| Forest biomass | BIO | Mg/ha |
| Crown cover | CC | % |
| Forest type† | FT | 18 dominant tree types |
| Forest growth | GRW | $m^3$/ha |
| Quadratic mean diameter | QMD | cm |
| Stand density index | SDI | ha |
| Trees per hectare | TPH | no. trees >2.54 cm dbh per ha |
| Forest volume | VOL | $m^3$/ha |
| Weighted height | WHT | m (weighted by larger trees) |
| Landfire | | |
| Canopy bulk density | CBD | kg/m |
| Canopy base height | CBH | height (m) to live canopy |
| Cover type‡ | COV | 10 dominant cover types |
| Forest canopy cover | FCC | % |
| Forest height | FHT | m |
| Herbaceous canopy cover | HCC | % |
| Herbaceous height | HHT | m |
| Shrub canopy cover | SCC | % |
| Shrub height | SHT | m |
| Structure§ | STR | four classes |
| Topographic¶ | | |
| Elevation | ELEV | m |
| Slope | SLP | rise, % |
| Transformed aspect | TASP | 0–1 |

*Notes:* All units were converted to the metric system for consistency. Vegetation predictor variables are from forested area only.

† Source: Bechtold and Patterson (2005), and see footnote 5.

‡ Source: Rollins et al. (2006), and see footnote 6.

§ Class 1, forest height ≤10 m, canopy ≤40%; class 2, forest height ≤10 m, canopy >40%; class 3, forest height >10 m, canopy ≤40%; class 4, forest height >10 m, canopy >40%.

¶ From DEMs of the U.S. National Elevation Data set.

### *Response data*

*Extant presence points.*—Locations of nests used by breeding pairs of Northern Goshawks from 1994 to 2005 were obtained from National Forests in Zone 16 and the Utah Division of Wildlife Natural Heritage database ($n$ = 572 nests). These nests were distributed across the entire study region, with 38% concentrated in the region of the Uinta Mountains in the northeastern portion of the study region. Eight nests which intersected non-forest pixels (as defined by 30-m Landfire or 250-m FIA) were eliminated, and territories with multiple nests were condensed to the most recently active nest, leaving 281 nests for 30-m nest site models, and 279 nests for 250-m nest area models. These reduced numbers of nests were more evenly distributed across Zone 16 (Fig. 1).

*Habitat envelopes and their ecologically based pseudo-absence points.*—Ecologically based pseudo-absence points were randomly selected within the defined nest site and nest area habitat envelopes (Tables 2 and 3). To create habitat envelopes, we first surveyed the literature for published field-based studies of Northern Goshawk nest habitat associations during nest occupancy (April to August; Reynolds et al. 2005). We used habitat associations reported in 25 published studies set in the western United States (Appendix A). We took geometric means (geomeans) of the published minima and maxima habitat characteristic values for different variables (e.g., forest canopy cover, tree height) and then applied these as the lower and upper bounds of single-variable habitat envelopes.

Single-variable habitat envelopes were created in ArcGIS 9.0 (ESRI, Redlands, California, USA) by retaining digital values from the FIA, Landfire, and DEM map products that fell within the geomean of the published ranges for a given predictor habitat variable. Values outside the ranges were excluded from the habitat envelope. Spatial intersection of two or more habitat envelopes produced multi-variable habitat envelopes (e.g., a two-variable habitat envelope, QMD ∩ ELEV, where ∩ is the spatial intersection of single-

[6] ⟨http://www.landfire.gov⟩

TABLE 2. Northern Goshawk 30-m nest site habitat envelopes from Landfire map products.

| Envelopes | Values | Portion of all nest points contained (%) ($n = 564$ nests) | Cover of Zone 16 forested area (%) |
|---|---|---|---|
| CONASP | All conifers and aspen† | 97.7 | 95.0 |
| ELEV | 1828–3048 m | 96.0 | 88.0 |
| FHT | 8.97–25.6 m | 96.9 | 79.3 |
| CONASP ∩ FHT | All conifers and aspen†<br>FHT: 8.97–25.6 m | 94.9 | 60.1 |
| CONASP ∩ ELEV | All conifers and aspen†<br>ELEV: 1828–3048 m | 93.7 | 84.6 |
| ELEV ∩ FHT | ELEV: 1828–3048 m<br>FHT: 8.97–25.6 m | 92.8 | 65.2 |
| CONASP ∩ ELEV ∩ FHT | All conifers and aspen†<br>ELEV: 1828–3048 m<br>FHT: 8.97–25.6 m | 90.9 | 53.1 |

*Notes:* Abbreviations are: CONASP, all conifers and aspen; ELEV, elevation; FHT, forest height. The spatial intersection of two or more habitat envelopes is indicated by ∩.

† All conifers and aspen from Landfire map product COV documented in the nest site according to Graham et al. (1999). These include aspen–birch, Douglas-fir, spruce–fir, white fir, lodgepole pine, high-elevation pine, ponderosa pine, pinyon–juniper, and juniper.

variable habitat envelopes quadratic mean diameter [QMD] and elevation [ELEV]). Single-variable habitat envelopes were larger in area (contained more pixels) than multi-variable habitat envelopes. Because habitat envelopes must encompass appropriate nesting habitat, only those envelopes (single or multi-variable) containing at least 90% of all extant presence points ($\geq$508 of 564 nests for both nest site and nest area models) were used for subsequent generation of ecologically based pseudo-absence points. Up to three habitat envelopes from each combination type (one, two, and three variable) containing the highest percentages of presence points were chosen to generate the ecologically based pseudo-absence points (Tables 2 and 3). Northern Goshawk nest areas and post-fledging areas (a total of 182 ha centered on each nest) of all extant 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).

Habitat envelope values and unique cell identifiers were exported to SAS 9.1 (SAS Institute 2003) for the following analyses. Pseudo-absence points were randomly generated from all cells of a given habitat envelope. The number of pseudo-absence points from each habitat envelope was set equal to the number of presence points given that unbalanced ratios of presence to absence points can affect the accuracy of classification models (Manel et al. 2001). To determine if randomized pseudo-absence point selection was biased and unrepre-

TABLE 3. Northern Goshawk 250-m nest area habitat envelopes from FIA map products.

| Envelopes | Values | Portion of all nest points contained (%) ($n = 564$ nests) | Cover of Zone 16 forested area (%) |
|---|---|---|---|
| QMD | 11.5–77 cm | 100 | 99.6 |
| WHT | 5–21 m | 99.8 | 81.3 |
| SDI | 337–2021 ha | 97.9 | 90.3 |
| QMD ∩ WHT | QMD: 11.5–77 cm<br>WHT: 5–21 m | 99.8 | 80.9 |
| QMD ∩ SDI | QMD: 11.5–77 cm<br>SDI: 337–2021 ha | 97.9 | 89.8 |
| SDI ∩ WHT | SDI: 337–2021 ha<br>WHT: 5–21 m | 97.7 | 74.9 |
| QMD ∩ SDI ∩ WHT | QMD: 11.5–77 cm<br>SDI: 337–2021 ha<br>WHT: 5–21 m | 97.7 | 74.6 |
| ELEV ∩ QMD ∩ WHT | ELEV: 1828–3048 m<br>QMD: 11.5–77 cm<br>WHT: 5–21 m | 95.9 | 72.4 |
| QMD ∩ WHT ∩ CONASP | All conifers and aspen†<br>QMD: 11.5–77 cm<br>WHT: 5–21 m | 95.9 | 68.5 |

*Notes:* Abbreviations are: QMD, quadratic mean diameter; WHT, weighted height; SDI, stand density index; ELEV, elevation; CONASP, all conifers and aspen. The spatial intersection of two or more habitat envelopes is indicated by ∩.

† All conifers and aspen from FIA map product FT documented in the nest area according to Graham et al. (1999). These include aspen, Douglas-fir, Engelmann spruce, subalpine fir, white fir, lodgepole pine, ponderosa pine, and pinyon–juniper.

sentative of the source population (the habitat envelope), we randomly generated 100 sets of pseudo-absence points from each habitat envelope, and used a two-sample Kolmogorov-Smirnov statistic to determine the likelihood that any set of pseudo-absence points differed from the source population. All randomly generated pseudo-absence points were within the 95% CI of the envelope population means ($P > 0.05$, all tests), indicating no bias in pseudo-absence point generation.

*Null envelopes and their traditional pseudo-absence points.*—Null envelope models were created to compare our habitat envelope method with traditional methods of pseudo-absence point generation. Here, pseudo-absence points were randomly generated from within null envelopes representing the entire study region extent (i.e., forested regions of Zone 16), minus the nest areas and post-fledging areas, at 30-m and 250-m resolutions. The number of pseudo-absence points generated from null envelopes was set equal to the number of presence points of the model resolution. Two-sample Kolmogorov-Smirnov tests were run on the 100 sets of traditional pseudo-absences; all were within the 95% CI of the null envelope population means ($P > 0.05$, all tests). These pseudo-absence points were used in the nest site and nest area null models, respectively.

### *Model building and evaluation*

One set of pseudo-absence points was randomly selected from each of the 100 sets generated from the habitat and null envelopes, paired with the presence points, and modeled as the response using logistic regression in R 2.1.1 (program *available online*).[7] Variables used to create habitat envelopes were not included as predictor variables in any of the corresponding logistic models (i.e., if quadratic mean diameter [QMD] was used to generate a QMD envelope from which pseudo-absence points were selected, QMD was not included as a predictor variable in the subsequent logistic model). Some cover and forest type categories were so low in occurrences of presences or pseudo-absences that the logistic models could not converge. Consequently, these categories were reassigned to more common types based on associated forest cover following Burns and Honkala (1990). Highly correlated predictor variables (as defined by Pearson correlation coefficient $>0.70$ at $P < 0.05$) were not included together in one model; the more ecologically relevant variable was included.

Models were run with up to four predictor variables included, using Northern Goshawk ecology to guide variable selection. Fifty-nine top candidate habitat and null envelope models (Tables 11–14 in Zarnetske 2006) were all assessed for fit (adjusted deviance, $D^2_{adj}$) and predictive capability (sensitivity, specificity, kappa) on the training data, and internally validated by 10-fold cross validation. Receiver operating characteristic plots (ROC) and associated area-under-curve (AUC) were performed on the training data and predicted values. Models produced from the same set of pseudo-absence points were ranked according to Akaike's information criterion (AIC; Akaike 1973), cross-validation error rate, and $D^2_{adj}$ to form a list of top competing models (Tables 4 and 5). Because many of the top competing models used different pseudo-absence points, they were not nested models and therefore AIC was not an appropriate method for model performance assessment. Therefore, the best-fit model for each category (nest site and nest area null and habitat envelope models) was selected first on lowest cross-validation error rate (all models $<1\%$ from the lowest cross-validation error rate were considered), and then by the highest $D^2_{adj}$. If $D^2_{adj}$ was equal for two or more top candidate models, the more parsimonious model was chosen.

Comparisons among best-fit null and habitat envelope models were made using a one-tailed Kruskal-Wallis rank test, testing our prediction that habitat envelope models would perform better than null models across the range of selected model fit and predictive metrics. To test if traditional pseudo-absence points were more variable than ecologically based pseudo-absence points, a two-sample test of variance was run on pseudo-absence points from each type (100 sets pooled) with the alternative hypothesis that the variance of ecologically based pseudo-absence points divided by the variance of traditional pseudo-absence points was greater than 1.

Coefficients of best-fit model parameters were translated into likelihood of occurrence maps across forested regions of Zone 16 using StatMod extension for ArcView 3.2 (ESRI, Redlands, California, USA; *available online*).[8] Map output likelihood of occurrence ranged from 0 to 1 and was categorized into four levels of habitat suitability: very low (0–0.25), low (0.26–0.5), moderate (0.51–0.75), and high (0.76–1.0).

## Results

### *Characteristics of competing top models*

Top competing habitat envelope models outperformed null envelope models both in model fit and predictive capabilities for each model resolution (Fig. 2). Results of the rank test indicated that all habitat envelope accuracy metrics except sensitivity were significantly better than those of null models (one-tailed Kruskal-Wallis rank test, $P = 0.05$, Fig. 2). The higher sensitivity and specificity of habitat envelope models indicated that presences and absences, respectively, were classified better than those in null envelope models. Additionally, the top competing null envelope models did not reflect known Northern Goshawk nest habitat relationships as well as the top competing habitat

[7] ⟨http://www.r-project.org⟩

[8] ⟨http://arcscripts.esri.com/details.asp?dbid=12502⟩

TABLE 4. AIC and ΔAIC for the top four nest site null and habitat envelope models.

| Model name | Model | AIC | ΔAIC |
|---|---|---|---|
| Null envelope | | | |
| NS-Null1 | + CBH + COV(+ P − LPP − DF − SF − PJ − J − RH − AB) + FHT − SLP | 578.72 | 0 |
| NS-Null2 | − CBD + CBH + COV(+ P − LPP − HEP − SF − PJ − J − RH − AB) − SLP | 581.02 | 2.30 |
| NS-Null3 | − CBD + CBH + FHT − SLP | 581.31 | 2.59 |
| NS-Null4 | + CBH − HCC + FHT − SLP | 583.40 | 4.68 |
| Habitat envelope | | | |
| NS-ELEV1 | + CBH − COV(+ PP − SF − PJ − J − AB) + FHT − SLP | 512.53 | 0 |
| NS-CONASP3 | + CBH + FHT − SLP | 552.82 | 1.90 |
| NS-CONASP ∩ ELEV2 | + CBH + FHT − HHT† − SLP | 545.73 | 1.74 |
| NS-CONASP ∩ ELEV3 | + CBH + FHT − SLP | 546.95 | 2.96 |

*Notes:* Cover type (COV) codes: PP, ponderosa pine; LPP, lodgepole pine; HEP, high-elevation pine; DF, Douglas-fir; SF, spruce–fir; PJ, pinyon–juniper; J, juniper; RH, riparian and other hardwoods; AB, aspen–birch. Other predictor variables are: CBH, canopy base height; FHT, forest height; SLP, slope; CBD, canopy bulk density; HCC, herbaceous canopy cover. Nest site models were selected from four candidate top null-envelope models and 18 candidate top habitat-envelope models. Model names derive from the null or habitat envelope that generated their pseudo-absence points; NS refers to nest site. Multiple models per habitat envelope are distinguished with 1, 2, 3, and so on following the habitat envelope name (i.e., NS-ELEV2). Only significant Landfire cover types are shown. Direction of variable influence is indicated by "+" or "−" preceding the variable. All nest site top null models can be compared with AIC because they contain the same set of pseudo-absence points. Nest site habitat envelope models listed here (with the exception of NS-CONASP ∩ ELEV2 and NS-CONASP ∩ ELEV3) cannot be compared with AIC because they each use a different set of ecologically based pseudo-absence points from their particular habitat envelope.

† Not significant.

envelope models; null envelope models selected some predictor variables tied less strongly to Northern Goshawk ecology (Tables 4 and 5). Ecologically based pseudo-absence points used in habitat envelope models were less variable than traditional pseudo-absence points (although not always significantly less variable via two-sample test of variance at $P = 0.05$), allowing better classification of suitable and non-suitable nest habitat for both nest site and nest area. Null envelope models never had more than one model with $\Delta AIC < 2$, indicating that null models were statistically different from one another (Burnham and Anderson 2002). In contrast, models of the same habitat envelope often had more than one model with $\Delta AIC < 2$, indicating several statistically similar models (Burnham and Anderson 2002) and that using ecologically based pseudo-absence points may lead to more competing top models.

Slope was a significantly negative parameter in both the null and habitat envelope top competing models (Tables 4 and 5), reflecting findings of Northern Goshawk preference for low to moderate slopes in the western United States (e.g., Squires and Ruggiero 1996). Top competing nest site null and habitat envelope models all contained significantly positive forest height parameters (CBH and FHT [see parameter abbreviations in Table 1]). Top competing nest area null and

TABLE 5. AIC and ΔAIC for the top four nest area null and habitat envelope models.

| Model name | Model | AIC | ΔAIC |
|---|---|---|---|
| Null envelope | | | |
| NA-Null1 | + FT(− PJ + PP + WF) + GRW − SLP + WHT | 580.82 | 0 |
| NA-Null2 | − AGE† + FT(− PJ + DF + PP + WF + LPP) + GRW − SLP | 583.03 | 2.21 |
| NA-Null3 | + FT(− PJ + DF + PP + WF + LPP) + GRW − SLP | 584.10 | 3.28 |
| NA-Null4 | + BIO + FT(− PJ + PP + WF + LPP) − SLP + WHT | 584.44 | 3.62 |
| Habitat envelope | | | |
| NA-QMD1 | + FT(− PJ + PP + WF) + GRW − SLP + WHT | 538.73 | 0 |
| NA-QMD2 | − AGE + FT(− PJ + DF + PP + WF) + GRW − SLP | 539.78 | 1.05 |
| NA-QMD3 | + FT(− PJ + DF + PP + WF) + GRW − SLP | 542.53 | 3.80 |
| NA-QMD ∩ SDI4 | − AGE† + FT(− PJ + PP + WF) + GRW − SLP | 546.06 | 1.89 |

*Notes:* Forest type (FT) codes: PJ, pinyon–juniper; DF, Douglas-fir; PP, ponderosa pine; WF, white fir; LPP, lodgepole pine. Other predictor variables are: GRW, forest growth; SLP, slope: WHT, weighted height: AGE, stand age; BIO, forest biomass. Nest area models were selected from four candidate top null-envelope models and 33 candidate top habitat-envelope models. Model names derive from the null or habitat envelope which generated their pseudo-absence points; NA refers to nest area. Multiple models per habitat envelope are distinguished with 1, 2, 3, and so on following the habitat envelope name (i.e., NA-QMD2). Only significant FIA forest types are shown. Direction of variable influence is indicated by "+" or "−" preceding the variable. All nest area top null models can be compared with AIC because they contain the same set of pseudo-absence points. NA-QMD nest area habitat envelope models listed here can be compared with AIC because they use the same set of ecologically based pseudo-absence points; NA-QMD ∩ SDI4 uses a different set of ecologically based pseudo-absence points and cannot be compared with the NA-QMD models using AIC.

† Not significant.

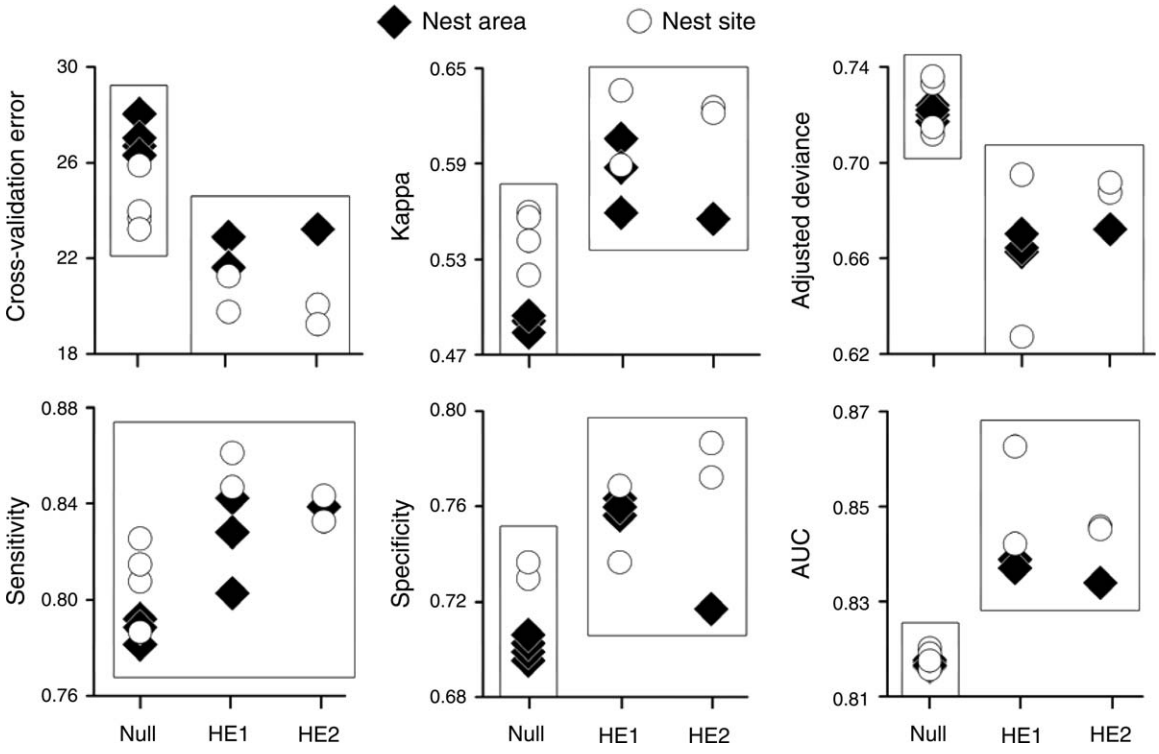


FIG. 2. Comparison of model accuracy metrics associated with the top null envelope and habitat envelope models for nest site and nest area listed in Tables 4 and 5: "Null," null envelope models using traditional pseudo-absence points; "HE1," habitat envelope models using ecologically based pseudo-absence points from a one-variable habitat envelope (e.g., QMD); "HE2," habitat envelope models using ecologically based pseudo-absence points from a two-variable habitat envelope (e.g., CONASP ∩ ELEV). Open circles represent nest site models; black diamonds represent nest area models. Means of metric values within boxes are not significantly different at $P < 0.05$; means between boxes are significantly different at $P < 0.05$, one-tailed Kruskal-Wallis test. "AUC" is area under the curve.

habitat envelope models contained the same parameters (FT, GRW, SLP), but FT category parameters in the habitat envelope models better reflected Northern Goshawk nest area habitat in Utah (i.e., lodgepole pine (LPP) is not as preferable; Graham et al. 1999). The top competing nest site and nest area habitat envelope models were ecologically sound and contained parameters which agreed strongly with published Northern Goshawk habitat associations. When nonsensical parameters were included in top competing habitat envelope models, they were insignificant, or were an artifact of categorical loading (e.g., −HHT in NS-CONASP ∩ ELEV2, −AGE in NA-QMD2 and NA-QMD ∩ SDI4, and −COV categories in NS-ELEV1).

### *Best-fit nest site and nest area models*

The best-fit nest site habitat envelope model was CONASP ∩ ELEV3. This complex habitat envelope met our expectations that incorporation of more biological knowledge (both forest type and elevation), and a constrained sampling region, would yield more ecologically and statistically robust models. The two-variable habitat envelope (CONASP ∩ ELEV) which generated the ecologically based pseudo-absence points in the CONASP ∩ ELEV3 model contained 85% of the study region and 94% of the extant presence points (Table 2). The single variable nest area habitat envelope, QMD, produced the ecologically based pseudo-absence points used in the best-fit nest area habitat envelope model (NA-QMD3). While QMD did not constrain the habitat envelope by much, retaining 99% of the study region, it did generate points used in models which outperformed null models in terms of fit and predictive capability (Fig. 2, Table 3).

Nest site and nest area habitat envelope models with four predictor variables always had the lowest AIC scores, but similar models with three predictor variables were usually within $\Delta AIC < 2$, suggesting that addition of a fourth predictor variable did not improve models significantly. High AUC values from the best-fit habitat envelope models show that these models have useful application across the study region because they are insensitive to threshold cut-off values, and have few false positives (Fig. 2).

FIA forest type and Landfire cover, both representations of dominant tree species, were always included in the best-fit habitat envelope models (as parameters in NA-QMD3 and as the habitat envelope in NS-CONASP ∩ ELEV3). The parameters of the NS-CONASP ∩ ELEV3 model (conifer and aspen forests between 1830

and 3050 m elevation, with moderate canopy base height, tall trees, and moderate slopes) reflect known Northern Goshawk habitat associations at the nest site (Squires and Ruggiero 1996, McGrath et al. 2003). Suitable nest area parameters of the NA-QMD3 model, dominated by Douglas-fir, ponderosa pine, or white fir, with moderate slopes, and moderate to high quadratic mean diameter (11.5–77 cm), reflect published ecological relationships of the nest area (Reynolds et al. 1982, McGrath et al. 2003).

The best-fit null envelope models (NS-Null4 and NA-Null3) did not reflect known Northern Goshawk-habitat relationships as well as the best-fit habitat envelope models. The negative association with HCC in NS-Null4 has never been documented as an important variable for Northern Goshawk nest site selection; –HCC likely reflects the denser over-story typical of nest sites which allow less light to penetrate to the forest floor. The best-fit nest area null and habitat envelope models used pseudo-absence points drawn from similar spatial extents (i.e., 100% of the study region for NA-Null3 and 99% of the study region for NA-QMD3; Table 3). These best-fit models also had similar model parameters. However, inappropriate nest area forest types, deciduous oak woodland and cercocarpus woodland, had twice the positive influence on suitable nest area habitat in the NA-Null4 model as in the NA-QMD3 model (Appendix B). Additionally, other top competing null models contained spurious parameters (e.g., nonsensical COV categories in NS-Null1, NS-Null2, and –AGE in NA-Null2).

### *Spatial depictions*

The best-fit habitat envelope nest site and nest area model outputs both predicted high likelihood of occurrence in appropriate Northern Goshawk habitat throughout Zone 16 (Fig. 3). The portion of the study region predicted to be of very low and low habitat suitability (0–0.50; blue and green) is approximately the same for the best-fit nest site and nest area null and habitat envelope models (nest site null envelope model: 73.7%; nest site habitat envelope model: 76.0%; nest area null envelope model: 71.3%; and nest area habitat envelope model: 73.7%). However, within the range of "suitable" (0.51–1.0; yellow and red) the best-fit habitat envelope models contain a slightly larger proportion of highly suitable (0.76–1.0; red) area than their best-fit null envelope model counterparts (8.1% for nest site null envelope vs. 10.0% for nest site habitat envelope; 8.6% for nest area null envelope vs. 12.0% for nest area habitat envelope; Zarnetske 2006).

## Discussion

As the amount of biological knowledge incorporated into habitat envelope generation increased, model performance improved due to both the reduced variability of ecologically based pseudo-absence points and the enhanced ecological relevance (Fig. 2). This, in turn, allowed the classification model to better distinguish suitable from unsuitable habitat. We found that increasing the complexity of habitat envelopes (i.e., one- and two-variable habitat envelopes) increased the precision and decreased the error rate of top competing models when compared with the null envelope technique (Fig. 2). However, increasing the complexity of habitat envelopes to 3-variable habitat envelopes so reduced the area from which ecologically based pseudo-absence points were generated (Tables 2 and 3) that these pseudo-absence points became too similar to presence points. In this case, the statistical model had difficulty creating HCMs which were both precise and low in error. This suggests that for a given species, there exists a threshold of habitat envelope complexity, beyond which resulting HCMs do not improve and may worsen. For the present study, this threshold was achieved at three-variable habitat envelopes (Zarnetske 2006).

The decision to restrict the study region to only forested regions was based on the ecological knowledge that Northern Goshawks do not nest in non-forested areas within our study region. By restricting this area to within the Northern Goshawk nesting habitat range, we ask the models to tease out the habitat variable values which are most important for determining highly suitable nesting habitat. Other techniques of pseudo-absence point generation select these points from outside of the species known distribution, thus generating pseudo-absence points in unsuitable habitat (e.g., Engler et al. 2004, Graf et al. 2005, Lütolf et al. 2006). These methods are excellent for determining suitable from non-suitable, but in order to recognize what habitat the species prefers within its known range, it is necessary to incorporate our ecological knowledge a priori into habitat envelopes, restricting the region from which pseudo-absence points can be selected.

Here best-fit habitat-envelope models not only have improved classification and are more ecologically robust, they identify highly suitable habitat within the range of suitable (0.51–1.0) habitat more accurately when compared to best-fit null envelope models using heterogeneous traditional pseudo-absence points (as indicated by the habitat envelope models' higher values of sensitivity and AUC; Fig. 2). Distinguishing highly suitable from moderately suitable habitat may be essential to species conservation and habitat management, particularly for rare, sensitive, threatened, and endangered species susceptible to habitat fragmentation and degradation.

Our assessment of this technique at two spatial resolutions provided additional testing of the method itself but also provided an analysis of the utility of Landfire and FIA map products for HCMs. We found that both Landfire and FIA map products as model predictor variables provided accurate representations of Northern Goshawk nesting habitat. From these results, we are confident that national Landfire and FIA map products are excellent sources of predictor variables for

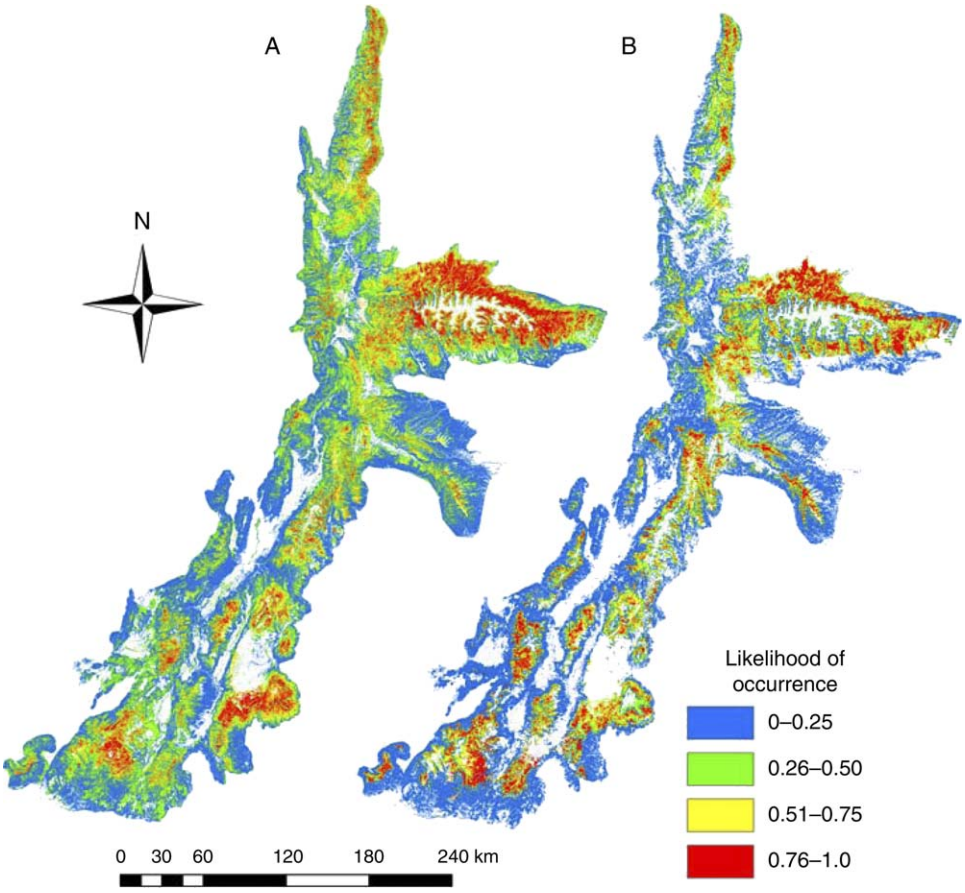


FIG. 3. Likelihood of nest site occurrence across Zone 16 based on best-fit habitat envelope models: (A) best-fit nest site habitat envelope model (NS-CONASP ∩ ELEV3); (B) best-fit nest area habitat envelope model (NA-QMD3). Likelihood of occurrence ranges from 0 to 1, where habitat suitability levels are: very low (0–0.25; blue), low (0.26–0.50; green), moderate (0.51–0.75; yellow), and high (0.76–1.0; red).

species HCMs, when used at the appropriate resolution for the study species. This technique, like other HCM methods, will provide robust ecologically relevant model results only if there is careful matching of resolution and extent among the species ecology, predictor variables, and study region.

Even with abundant and robust predictor variables, deficient knowledge of the species-habitat relationship will hinder the creation of accurate HCMs. However, the habitat envelope technique can still be applied to species whose habitat relationships are less familiar. Presumably some knowledge of species-habitat relationships exists for most species, even if it is simply that they prefer forested over non-forested habitat. For example, a lesser-known species with similar management indicator status, the American Three-toed Woodpecker (*Picoides dorsalis*), feeds on beetle-infested trees, often post-burn, and seems to prefer spruce forest (Bock and Bock 1974, Murphy and Lenhausen 1998, Hobson and Schieck 1999). Habitat envelopes created with this limited habitat information resulted in improved woodpecker HCMs (statistically and ecologically) when compared to the null technique (P. Zarnetske, *unpublished data*).

Additionally, if a strong association of occurrence exists between a less-studied species and a well-studied surrogate species, the surrogate's habitat associations could also be used to create habitat envelopes for the less-studied species, as long as discretion is taken in interpreting the resulting habitat suitability map (similar to Lütolf et al. 2006). The reduction of even a few error rate percentage points by habitat envelope models is an important improvement for rare, sensitive, threatened, and endangered species whose conservation assessments often rely on incomplete data. If the modeler is concerned about selecting one top model produced with deficient ecological knowledge, model averaging could allow the inclusion of several top models (Burnham and Anderson 2002). Regardless of the ecological knowledge base, including some biological knowledge for a given species in the generation of pseudo-absence points is both important for maintaining ecological relevance in HCMs, and improving model performance. This inclusion of ecological knowledge should also decrease the

chance of a biologically inappropriate top model because habitat envelopes already constrain the sampling region to that of generally preferred habitat.

Although a variety of techniques exist for generating pseudo-absence points, recent studies suggest that superior models use pseudo-absence points generated from constrained areas where the variability of points is reduced (e.g., Engler et al. 2004, Graf et al. 2005, Lütolf et al. 2006, and the present study). The consistent improvement in model fit and predictive capability of habitat envelope models demonstrated here, further supports the use of constrained ecologically based pseudo-absence points over traditional pseudo-absences generated over entire study regions. While profile-type models were not explicitly tested against our habitat envelope models, use of robust statistical modeling tools (e.g., GLMs, GAMs, classification trees) with ecologically based pseudo-absence points and extant presence points may be more appropriate than profile techniques, which (1) assume that all predictor variables are equally important in determining species distribution (e.g., FEM; Robertson et al. 2004), (2) assume presence data is unbiased (e.g., ENFA; Hirzel et al. 2002), (3) assume the presence-only data originates only from source habitat (e.g., Maxent; Phillips et al. 2006), (4) overpredict (e.g., ENFA; Brotons et al. 2004, Engler et al. 2004), and (5) are difficult to interpret and assess statistically. Furthermore, if field-collected absence points in a dataset are expected to contain false or spurious absence points (Graham et al. 2004, Hirzel et al. 2001, 2002), including additional ecologically based pseudo-absence points may provide more robust absences for modeling.

Common problems with this type of modeling include assumptions of habitat saturation and lack of complete species distribution data. The present technique is no exception. The inherent HCM assumptions that habitat is saturated (Capen et al. 1986), and that the species modeled is in equilibrium with its environment (Austin 2002, Guisan and Thuiller 2005) are often ignored in broad-scale habitat modeling because knowing the locations of all individuals or the attribute of interest (i.e., a nest site) is nearly impossible across a large spatial extent, particularly at one time step. Because landscape-scale census and monitoring is not complete for Northern Goshawks within Zone 16 (and is one reason why these models were created), we cannot assume that habitat is saturated, especially because it is probable that not all territories have been identified. Additionally, due to incomplete activity information for some nests throughout 1994–2005, some nest locations may exist in so-called sink habitat (Pulliam 1988) with poor habitat quality or lack of abundant prey items nearby.

Predictor variable error is another recurring issuing with HCMs, particularly at broad spatial scales. No matter how well any habitat model fits the data from which it is developed, it contains error from a variety of sources (Edwards et al. 1996, Pearce and Ferrier 2000). FIA and Landfire map products are models of vegetation attributes which contain error too, error which was then transferred to the Northern Goshawk HCMs. Additionally, because FIA and Landfire map products of Zone 16 incorporate plot-based inventory data and satellite imagery spanning from 1998 to 2003, certain cells within the map products may not reflect current conditions, although validation of these map products indicates low error (Blackard 2004, Rollins et al. 2006, and see footnotes 5 and 6). Unfortunately, no method exists to incorporate the inherent error of spatial predictor layers into classification models such as GLMs, even if the error is known. Conservation recommendations based on HCMs should take into account the uncertainty inherent in the models and their map outputs so that variability within the system is recognized. However, even with inherent error, the good fit, predictive capability, and ecological relevance of the Northern Goshawk best-fit habitat envelope models suggest that sampling high habitat suitability areas may lead to the discovery of new nest sites and nest areas in Zone 16.

Here we use habitat as an example for constraining the region of selection to refine the classification; other ecologically relevant variables could be used to constrain envelopes. For a given species, these might include spatially explicit representations of: predator, competitor, prey, or resource abundance; disturbances (e.g., fire, pollutants, and drought); resource proximity; territory boundaries; degree of isolation; and demographic parameters. Inclusion of other ecologically relevant Northern Goshawk variables (in habitat envelopes and as predictor variables) may have improved classification further, predicting likelihood of occurrence for suitable and successful nests (i.e., suitable conditions for successful fledglings). Prey abundance, competitor presence, nest-specific fledgling survival, forest characteristics of post-fledging and foraging areas, territoriality, forest disturbances (i.e., bark beetles, fire, timber harvest), and inter-annual climatic variation have all been shown to be important variables for nest success (Squires and Reynolds 1997, Salafsky et al. 2005, Reynolds et al, 2006, Wiens et al. 2006). However, as is the case for many other species of concern, these data were either not available or were incomplete. Yearly surveys of existing nests and regions of high likelihood of occurrence will help provide these valuable data for refined models of suitable and successful Northern Goshawk nests.

## Conclusions

This study contributes to research aimed at advancing habitat and species distribution modeling techniques. Here we improved upon traditional techniques of generating pseudo-absence points by incorporating the ecology of the species–habitat relationship. Habitat envelopes provide an ecologically based method for

generating less variable pseudo-absence points which can produce more statistically and ecologically robust HCMs. Moreover, given some a priori knowledge of the species–habitat relationship, ecologically based pseudo-absence points can be created for any species, ecosystem, data resolution, and spatial extent.

Acknowledgments

The USDA Forest Service Rocky Mountain Research Station Forest, Inventory, and Analysis Program, Ogden, Utah, funded this research. USDA Forest Service biologists and Utah Division of Wildlife Resources personnel provided Northern Goshawk nest presence points. We thank Richard Cutler and James MacMahon for their contributions to this project. Additionally, we thank two anonymous reviewers and T. B. Murphy for their helpful comments.

Literature Cited

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 *in* B. N. Petrov and F. Csaki, editors. Second International Symposium on Information Theory. Academia Kiado, Budapest, Hungary.

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

Bechtold, W. A., and P. L. Patterson, editors. 2005. The enhanced forest inventory and analysis program: national sampling design and estimation procedures. General Technical Report SRS-80. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, North Carolina, USA.

Beier, P., and J. E. Drennan. 1997. Forest structure and prey abundance in foraging areas of Northern Goshawks. Ecological Applications 7:564–571.

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

Blackard, J. A., G. G. Moisen, and R. P. Tymcio. 2004. Modeling investigations into mapping of forest attributes with FIA data for the state of Wyoming. Report, Forest Inventory and Analysis. Rocky Mountain Research Station, USDA Forest Service, Ogden, Utah, USA.

Bock, C. E., and J. H. Bock. 1974. On the geographical ecology and evolution of the Three-toed Woodpeckers, *Picoides tridactylus* and *P. arcticus*. American Midland Naturalist 92: 397–405.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. Classification and regression trees. Wadsworth International Group, Belmont, California, USA.

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

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

Burnham, K. P., and D. Anderson. 2002. Model selection and multi-model inference. Springer Science and Business Media, New York, New York, USA.

Burns, R. M., and B. H. Honkala. 1990. Silvics of North America. 1. Conifers; 2. Hardwoods. Agriculture handbook 654. Volume 2. U.S. Department of Agriculture, Forest Service, Washington, D.C., USA.

Busby, J. R. 1991. BIOCLIM: a bioclimatic analysis and prediction system. Pages 64–68 *in* C. R. Margules and M. P. Austin, editors. Nature conservation: cost effective biological surveys and data analysis. CSIRO, Canberra, Australia.

Capen, D. E., J. W. Fenqick, D. B. Inkley, and A. C. Boynton. 1986. Multivariate models of songbird habitat in New England forests. Pages 171–175 *in* J. A. Verner, M. L. Morrison, and C. J. Ralph, editors. Wildlife 2000: modelling habitat relationships of terrestrial vertebrates. University of Wisconsin Press, Madison, Wisconsin, USA.

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

Edwards, T. C., Jr., D. R Cutler, N. E. Zimmermann, L. Geiser, and G. G. Moisen. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. Ecological Modelling 199:132–141.

Edwards, T. C., Jr., E. Deshler, D. Foster, and G. G. Moisen. 1996. Adequacy of wildlife habitat relation models for estimating spatial distributions of terrestrial vertebrates. Conservation Biology 10:263–270.

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

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

Finn, S. R., D. E. Varland, and J. M. Marzluff. 2002. Does Northern Goshawk breeding occupancy vary with nest-stand characteristics on the Olympic Peninsula, Washington? Journal of Raptor Research 36:265–279.

Graf, R. F., K. Bollmann, W. Suter, and H. Bugmann. 2005. The importance of spatial scale in habitat models: capercaillie in the Swiss Alps. Landscape Ecology 20:703–717.

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

Graham, R. T., R. L. Rodriguez, K. M. Paulin, R. L. Player, A. P. Heap, and R. Williams. 1999. The Northern Goshawk in Utah: habitat assessment and management recommendations, General Technical Report RMRS-GTR-22. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, Utah, USA.

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

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8:993–1009.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized additive models. Chapman and Hall/CRC, New York, New York, USA.

Hayward, G. D., and R. E. Escaño. 1989. Goshawk nest-site characteristics in western Montana and northern Idaho. Condor 91:476–479.

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

Hirzel, A. H., V. Helfer, and F. Métral. 2001. Assessing habitat-suitability models with a virtual species. Ecological Modelling 145:111–121.

Hobson, K. A., and J. Schieck. 1999. Changes in bird communities in boreal mixed wood forest: harvest and wildfire effects over 30 years. Ecological Applications 9: 849–63.

Homer, C. G., and A. Gallant. 2001. Partitioning the conterminous United States into mapping zones for Landsat TM land cover mapping. USGS Draft White Paper, U.S. Department of the Interior, U.S. Geological Survey, EROS Data Center, Sioux Falls, South Dakota, USA.

Lütolf, M., F. Kienast, and A. Guisan. 2006. The ghost of past species occurrence: improving species distribution models for presence-only data. Journal of Applied Ecology 43:802–815.
Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38:921–931.
McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Second edition. Chapman and Hall/CRC, New York, New York, USA.
McGrath, M. T., S. DeStephano, R. A. Riggs, L. L. Irwin, and G. L. Roloff. 2003. Spatially explicit influences on Northern Goshawk nesting habitat in the interior Pacific Northwest. Wildlife Monographs 154:1–63.
Murphy, E. C., and W. A. Lenhausen. 1998. Density and foraging ecology of woodpeckers following a stand-replacement fire. Journal of Wildlife Management 62:1359–1372.
Parra-Olea, G., E. Martínez-Meyer, and G. Pérez-Ponce de León. 2005. Forecasting climate change effects on salamander distribution in the highlands of central Mexico. Biotropica 37:202–208.
Pearce, J. L., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling 133:225–245.
Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231–259.
Pulliam, H. R. 1988. Sources, sinks, and population regulations. American Naturalist 132:652–661.
Reynolds, R. T., R. T. Graham, M. H. Reiser, R. L. Bassett, P. L. Kennedy, D. A. Boyce, Jr., G. Goodwin, R. Smith, and E. L. Fisher. 1992. Management recommendations for the Northern Goshawk in the southwestern United States. General Technical Report RM-217. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, USA.
Reynolds, R. T., S. M. Joy, and D. G. Leslie. 1994. Nest productivity, fidelity, and spacing of Northern Goshawks in Arizona. Studies in Avian Biology 16:106–113.
Reynolds, R. T., E. C. Meslow, and H. M. Wight. 1982. Nesting habitat of coexisting accipiters in Oregon. Journal of Wildlife Management 46:124–138.
Reynolds, R. T., J. D. Wiens, S. M. Joy, and S. R. Salafsky. 2005. Sampling considerations for demographic and habitat studies of Northern Goshawks. Journal of Raptor Research 39:274–285.
Reynolds, R. T., J. D. Wiens, and S. R. Salafasky. 2006. A review and evaluation of factors limiting Northern Goshawk populations. Studies in Avian Biology 31:260–273.
Roberts, D. W., and S. V. Cooper. 1989. Concepts and techniques of vegetation mapping. Pages 90–96 *in* D. Ferguson, P. Morgan, and F. D. Johnson, editors. Land classifications based on vegetation: applications for resources management. General Technical Report INT-257. U.S. Department of Agriculture, Forest Service, Ogden, Utah, USA.
Robertson, M. P., M. H. Villet, and A. R. Palmer. 2004. A fuzzy classification technique for predicting species' distributions: applications using invasive alien plants and indigenous plants. Diversity and Distributions 10:461–474.
Rollins, M. G., R. E. Keane, Z. Zhu, and J. P. Menakis. 2006. Executive summary. Pages 1–4 *in* M. G. Rollins and C. K. Frame, technical editors. The LANDFIRE Prototype Project: nationally consistent and locally relevant geospatial data for wildland fire management. General Technical report RMRS-GTR-175. USDA Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, USA.
Rotenberry, J. T., K. L. Preston, and S. T. Knicks. 2006. GIS-based niche modeling for mapping species' habitat. Ecology 87:1458–1464.
Salafsky, S. R., R. T. Reynolds, and B. R. Noon. 2005. Patterns of temporal variation in goshawk reproduction and prey resources. Journal of Raptor Research 39:237–246.
SAS Institute. 2003. SAS 9.1. SAS Institute, Cary, North Carolina, USA.
Schreuder, H. T., T. G. Gregoire, and J. P. Weyer. 2001. For what applications can probability and non-probability sampling be used? Environmental Monitoring and Assessment 66:281–291.
Squires, J. R., and R. T. Reynolds. 1997. Northern Goshawk (*Accipiter gentilis*). *In* A. Poole and F. Gill, editors. The birds of North America, No. 298. Academy of Natural Sciences, Philadelphia, Pennsylvania, USA, and American Ornithologists' Union, Washington, D.C., USA.
Squires, J. R., and L. F. Ruggiero. 1996. Nest-site preference of Northern Goshawks in southcentral Wyoming. Journal of Wildlife Management 60:170–177.
Stockman, A. K., D. A. Beamer, and J. E. Bond. 2006. An evaluation of a GARP model as an approach to predicting the spatial distribution of non-vagile invertebrate species. Diversity and Distributions 12:81–89.
Stockwell, D. R. B., and D. Peters. 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., and A. T. Peterson. 2002. Controlling bias in biodiversity data. Pages 537–546 *in* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: issues of accuracy and scale. Island Press, Covelo, California, USA.
Wiens, J. D., B. R. Noon, and R. T. Reynolds. 2006. Post-fledging survival of Northern Goshawks: the importance of prey abundance, weather, and dispersal. Ecological Applications 16:406–418.
Zaniewski, A. E., A. Lehmann, and J. M. Overton. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling 157:261–280.
Zarnetske, P. L. 2006. Landscape-scale habitat classification modeling with incomplete data: the generation of ecologically based pseudo-absence points. Thesis. Utah State University, Logan, Utah, USA.

## APPENDIX A

Literature studies and values used in creating 30-m nest site and 250-m nest area habitat envelopes for Northern Goshawk (*Ecological Archives* A017-071-A1).

## APPENDIX B

Best-fit Northern Goshawk nest site (30-m) and nest area (250-m) habitat and null envelope model coefficients and their significance (*Ecological Archives* A017-071-A2).