ORIGINAL PAPER

Gyrfalcon nest distribution in Alaska based on a predictive GIS model

Travis L. Booms · Falk Huettmann · Philip F. Schempf

Abstract The gyrfalcon (*Falco rusticolus*) is an uncommon, little studied circumpolar Arctic bird that faces conservation concerns. We used 455 historical nest locations, 12 environmental abiotic predictor layers, Geographic Information System (ArcGIS), and TreeNet modeling software to create a spatially explicit model predicting gyrfalcon breeding distribution and population size across Alaska. The model predicted that 75% of the state had a relative gyrfalcon nest occurrence index value of <20% (where essentially no nests are expected to occur) and 7% of the state had a value of >60%. Areas of high predicted occurrence were in northern and western Alaska. The most important predictor variable was soil type, followed by subsurface geology and vegetation type. Nine environmental factors were useful in predicting nest occurrence, indicating complex multivariate habitat relationships exist. We estimated the breeding gyrfalcon population in Alaska is 546 ± 180 pairs. The model was 67% accurate at predicting nest occurrence with an area under the curve score of 0.76 when assessed with independent data; this is a good result when considering its application to the entire state of Alaska. Prediction accuracy estimates were as high as 97% using 10-fold cross validation of the training data. The model helps guide science-based management efforts in times of increasing and global pressures for this species and Arctic landscapes.

Keywords Alaska · Arctic · Breeding distribution · Conservation biology · *Falco rusticolus* · Gyrfalcon · Predictive modeling

T. L. Booms · F. Huettmann
Department of Biology and Wildlife, Institute of Arctic Biology,
University of Alaska Fairbanks Biology,
211 Irving I, Fairbanks, AK 99775, USA

F. Huettmann
e-mail: fffh@uaf.edu

P. F. Schempf
U.S. Fish and Wildlife Service,
Migratory Bird Management-Raptors,
3000 Vintage Blvd. Suite 240, Juneau, AK 99801, USA
e-mail: phil_schempf@fws.gov

T. L. Booms (✉)
Alaska Department of Fish and Game, Nongame Program,
1300 College Road, Fairbanks, AK 99701, USA
e-mail: travis.booms@alaska.gov; travisbooms@hotmail.com

Introduction

Species distribution is essential knowledge for conservation biology (Araujo and Guisan 2006). The distribution of a species is informed by its ecological niche, defined by Hutchinson (1957) as the set of biotic and abiotic conditions in which a species is able to persist and maintain stable population sizes. The ecological niche is both the fundamental niche (defined by abiotic factors) and the realized niche (defined by abiotic and biotic factors). Learning what variables contribute to defining the boundaries of either of these types of niches informs our understanding of the species' ecology, can be used to predict the distribution of the species, and is often relevant for specific management actions (Peterson 2001).

Understanding where species occur temporally and spatially across large geographic areas is important to conserving, monitoring, and managing species effectively (Wu and Smeins 2000). However, detailed species distribution data spanning large areas is rarely available, especially for remote Arctic areas. Extrapolating beyond areas of known presence, using predictive modeling, helps to estimate

distribution, particularly for rare or endangered species in remote areas (Peterson 2001; Pearce and Boyce 2006). It is a convenient and cost-efficient approach making use of data collected during previous decades. Such predictive models are valuable for guiding conservation actions and planning (Heglund 2002). For example, a model that evaluated habitat suitability for the endangered timber wolf (*Canis lupis*) was useful in the recovery of the species because it gave managers a realistic idea of future population size and distribution (Mladenoff et al. 1995, 1999).

The gyrfalcon occurs at low densities across the circumpolar Arctic (Cade 1982) where it breeds above 55°N. It is an important apex predator to the Arctic ecosystem that feeds on ptarmigan (*Lagopus* spp.) (Booms et al. 2008). It relies on cliff and cliff-like structures for laying its eggs on rock ledges or in stick nests built by other bird species (Palmer 1988). Our understanding of the variables that influence gyrfalcon distribution is limited and based on findings from small geographic areas. The global breeding population estimate is roughly 8,000–11,000 pairs (Potapov and Sale 2005) and though the gyrfalcon is not listed as endangered or threatened in North America, it is a C.I.T.E.S. Appendix 1 species. Current best estimates of its breeding distribution in Alaska are based on expert opinion and largely extrapolated from a few areas that have been well surveyed (Swem et al. 1994). However, a large portion of the state has not been surveyed, and the state's full potential for gyrfalcon nesting habitat is largely unknown (Swem et al. 1994). Data from the few areas that are regularly surveyed in Alaska have not been combined to address statewide conservation issues for a synthesis or to investigate ecological questions beyond local scales.

Though gyrfalcons inhabit remote areas, the species will likely face serious conservation threats from resource development issues and global warming through changes in vegetation and prey species (Booms et al. 2008). Arctic portions of Alaska are believed to contain the second largest deposit of oil and oil-equivalent natural gas in the world (U.S. Geological Survey 2008), and development activities could potentially affect Alaska's gyrfalcon population. The potential for wind turbine developments along coastal Alaska is another threat to the species. Coastal areas that are classified as having "outstanding" or "superb" potential for wind development (U.S. Department of Energy 2008) are also important to gyrfalcons (Britten et al. 1995). Because wind turbines are known to kill large numbers of birds (Johnson et al. 2001) including falcons (Smallwood and Thelander 2004), and have the potential to reduce populations of resident raptors (Hunt et al. 1999), wind turbines could impact Alaska's gyrfalcon population. Therefore, identifying potential hotspots for breeding (and potential conflict with development) is important to conserving the species. Gyrfalcons will be impacted by global

warming because Arctic habitats are predicted to be significantly affected (Booms et al. 2008). In Alaska, the mean annual temperature has warmed by as much as 2.2°C in the past 50 years (Stafford et al. 2000) and such warming has been associated with deleterious changes in bird nesting phenology (Crick 2004). Shrub growth is increasing in Arctic Alaska (Tape et al. 2006) and drying is expected, changing the structure of the open tundra used by gyrfalcons for hunting. Increased shrubs may provide additional cover for ptarmigan, reduce gyrfalcon foraging efficiency, and could impact gyrfalcon population growth and distribution.

We compiled gyrfalcon nest locations in Alaska and created a model to predict current breeding locations for gyrfalcons (its fundamental niche) based on factors measured at historical nest sites. Guided by the resulting model, we then collected independent, spatially explicit evaluation data to assess model accuracy. To our knowledge, this is the first predictive, spatial model of breeding gyrfalcons anywhere and a step towards developing a global conservation effort to assess uncertainties. We also present an innovative technique to model sensitive data (nest locations) without having to obtain the actual locations. Our specific research goals were: (1) estimate the breeding distribution and population size of gyrfalcons in Alaska, (2) determine the relative importance of a suite of environmental variables that explain the breeding distribution, and (3) assess the accuracy and utility of the model. The results help inform us about the factors influencing gyrfalcon nesting and guide future sampling, surveying, and conservation efforts across the state.

Materials and methods

Environmental layers

We chose 12 environmental variables to develop the model based on availability and our knowledge of gyrfalcon ecology and published literature (Booms et al. 2008). All data layers were publicly available and had statewide coverage (Table 1). We re-projected layers into Clark 1866 Albers (in meters) and merged them for a consistent statewide coverage. All geographic information system (GIS) operations were conducted in ArcMap 9.2. and 9.3 (Environmental Systems Research Institute 2008). Slope and aspect layers were derived in ArcMap from an official state-wide digital elevation model. We calculated the distance to fresh water, coastline, and human structures using the Euclidean distance tool in ArcMap 9.2. Aspect data were categorized into four directional groups ($N = 316$–$45°$, $E = 46$–$135°$, $S = 136$–$225°$, $W = 226$–$315°$, and flat) and used as a categorical variable. We used average April temperature and precipitation because we expected breeding gyrfalcons

Table 1 Environmental GIS layers used to predict gyrfalcon nest occurrence across Alaska and their relative importance

Environmental layer	Relative importance[a]	Pixel size	Variable type	Number of levels	Source	Citation layer based on	Website
Soil type	100	Polygon	Categorical	83	Alaska Geospatial Data Clearinghouse	Rieger et al. (1979)	http://agdc.usgs.gov/data/usgs/erosafo/soil/soil.html
Sub-surface geology	58	Polygon	Categorical	54	Alaska Geospatial Data Clearinghouse	Beikman (1980)	http://agdc.usgs.gov/data/usgs/geology/index.html
Vegetation type	24	Polygon	Categorical	21	Alaska Geospatial Data Clearinghouse	Fleming (1997)	http://agdc.usgs.gov/data/usgs/erosafo/veg/vegetation.html
Surface geology	16	Polygon	Categorical	25	Alaska Geospatial Data Clearinghouse	Karlstrom et al. (1964)	http://agdc.usgs.gov/data/usgs/erosafo/surfgeol/surfgeol.html
Slope	15	300 m	Continuous	–	Alaska Geospatial Data Clearinghouse	Derived from digital elevation model	http://agdc.usgs.gov/data/akdb/dem/dem.html
Distance to ocean coast	10	60 km	Continuous	–	World Coastline Extractor—US Geological Survey	World Coastline Extractor	http://rimmer.ngdc.noaa.gov
Mean April temperature	8	1 km	Continuous	–	Worldclim	Higmans et al. (2005)	www.worldclim.com
Digital elevation model	7	300 m	Continuous	–	Alaska Geospatial Data Clearinghouse	U.S. Geological Survey (1997)	http://agdc.usgs.gov/data/akdb/dem/dem.html
Aspect	3	300 m	Categorical	5	Alaska Geospatial Data Clearinghouse	Derived from digital elevation model	http://agdc.usgs.gov/data/akdb/dem/dem.html
Distance to fresh water	0	300 m	Continuous	–	Global Lakes and Wetlands Database	Lehner and Doll (2004)	www.worldwildlife.org/science/data/item1877.html
Distance to human development	0	1 km	Continuous	–	Center for International Earth Science Information Network	Sanderson et al. (2003)	www.sedac.ciesin.columbia.edu/wildareas/downloads.jsp
Mean April precipitation amount	0	1 km	Continuous	–	Worldclim	Higmans et al. (2005)	www.worldclim.com

[a] Scores taken from TreeNet

would be most influenced by these parameters in April during territory establishment and early incubation.

Training data

We obtained 414 gyrfalcon nest locations spanning all types of Gyrfalcon nesting habitats in Alaska between 1972 and 2007 from collaborators to use as training data (Fig. 1). We did not differentiate between successful nests (potentially higher quality sites) and unsuccessful nests (lower quality sites). Nests were found during a 36-year period including some from the 1960s or earlier (Cade 1960), making the dataset one of the largest and longest term collections of raptor nest locations used for predictive modeling in Alaska, and likely elsewhere. Nest locations were converted to WGS-84 datum, re-projected into Clarke 1866 Albers, and imported as a shapefile layer into GIS. Gyrfalcon nest locations in Denali National Park were provided to us as the extracted environmental data (see "Methods" below). A total of 455 nests were used in our model.

We created 10,000 random points across Alaska using the freely available Hawth's Tools in ArcGis (Beyer 2008). We used these points as a measure of available habitats against which we compared the 455 nest locations (Manly et al. (2002); Engler et al. 2004). The ratio of 455 presence versus 10,000 pseudo-absence points is commonly used in

Fig. 1 The approximate historical gyrfalcon nest locations (*circles*) used to create the predictive model and areas used to assess model accuracy (*diamonds*). Study areas used to assess model accuracy included parts of the Togiak National Wildlife Refuge, Lisburne Peninsula, and Arctic National Wildlife Refuge. The number of nests (n) used to create the predictive model from each area is stated within each circle. The N *arrow* indicates north. Parts of southeast Alaska and the Aleutian chain are not shown because no nest data were used from these locations

the modeling literature (Craig and Huettmann 2008) and the uneven ratio is corrected for by using balanced weight settings in TreeNet (see below).

Modeling approach

We used presence-available modeling to predict nest occurrence following design II in Manly et al. (2002) (Pearce and Boyce 2006). We extracted information from environmental layers at historical nest sites and random points in ArcMap 9.3 using Hawth's Tools. We subjected these data to stochastic gradient boosting algorithms using program TreeNet 2.0 (Salford 2002). Stochastic gradient boosting is part of regression tree analysis (Friedman 2002) that creates binary trees by recursively partitioning data into two data sets based on predictor variables while trying to minimize variation within each dataset. Subsequent trees are constructed for the prediction of the residuals from the previous trees and results are computed from the entire group of trees (Friedman 2002).

We constructed our model in TreeNet using binary logistic regression and the balanced class weights option to account for unequal sample sizes of presence and available points. Otherwise, we used default setting in TreeNet and allowed it to optimize the number of trees in the model. Because the optimal number of trees was less than 150, there was no need to build additional trees to further optimize the model (Salford 2002).

For prediction-to data, we created a point lattice grid of 18,000 regularly spaced points across Alaska (approximately 7×7 km spacing), and extracted information from the 12 environmental layers (Table 1) described above for each point. We then used the optimized model to predict nest presence at each of the 18,000 points based on the extracted environmental data at each point. Predicted presence was scaled from 0 to 100% and interpreted as the relative index of occurrence (Keating and Cherry 2004; Araujo and Williams 2000). We imported the dataset of spatially referenced predictions into GIS as a raster file and interpolated between the regular points using inverse distance weighting (IDW) to obtain a smoothed predictive map of gyrfalcon nest distribution.

To estimate the state's breeding population, we assigned density estimates to each predicted category (Nielsen et al. 2008; Onyeahialam et al. 2005) from the range of published nesting densities found in Alaska (1 nest per 200–1000 km^2) (Swem et al. 1994) (Table 2). Based on our experience with the species, we assumed no gyrfalcons bred in areas with predicted occurrence levels ≤40%. We multiplied the total area of Alaska covered by each prediction category by the corresponding density estimate and summed the totals to estimate the state population, similar to Boyce and McDonald (1999) (Table 2). We multiplied the final estimate by the
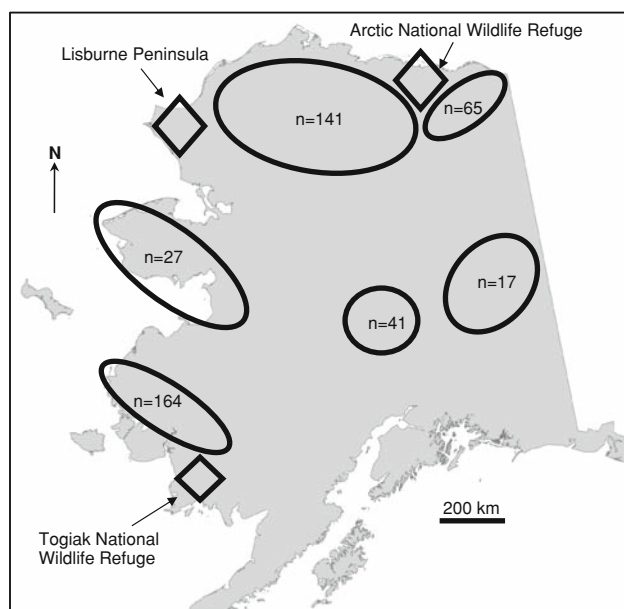
Table 2 Estimated number of breeding gyrfalcon pairs extrapolated from model-predicted nest occurrence categories and the amount of area each category covers in Alaska

Predicted gyrfalcon nest occurrence category (%)	Area in Alaska (km^2)	Estimated gyrfalcon nesting density (nest/km^2)[a]	Total # of estimated nests
0–20	1,113,000 (75%)	0	0
21–40	141,000 (10%)	0	0
41–60	114,000 (8%)	1/1,000	114
61–80	80,000 (5%)	1/300	267
81–100	33,000 (2%)	1/200	165
	1,481,000		546

[a] Nest densities from Swem et al. (1994)

model's accuracy, as determined by independent survey data, to provide a measure of error around the estimate. The total area of Alaska used in our analysis is 1,481,000 km^2 (U.S. Census Bureau 2004).

Model evaluation

We evaluated the model thoroughly using internal training data as well as two types of independent data: (1) the number of potential nest cliffs in plots and (2) the presence or absence of an occupied gyrfalcon nest in plots. We defined a potential nest cliff as any rock structure with a vertical rock face ≥4 m, based on published descriptions of gyrfalcon nest cliffs (Booms et al. 2008) and 8 years of field experience working with the species. We assumed a priori that the second measure of accuracy in the independent data (occupied nests) would be biased low because it was influenced by biotic variables that we did not attempt to model. This sets-up a mismatch between modeling the fundamental niche and measuring model accuracy with the realized niche. However, we included this measure of accuracy for completeness because it would be biologically informative to learn how well the model-predicted both occupied and potential nest sites.

Evaluation with internal training data

We used the aspatial 10-fold cross validation procedure in TreeNet which divided the original training data into ten groups and used nine of the groups as training data. The remaining group was used as testing data. This was done 10 times and a different group of data was withheld for testing each time. Testing results were then averaged across the ten iterations and the area under the curve (AUC) estimate in the receiver operating curve (ROC) plot was taken directly from TreeNet to assess prediction accuracy. We

considered AUC scores <0.7 indicated low model accuracy, 0.7–0.9 moderate accuracy, and >0.9 high accuracy (Swets 1988).

Independent spatial data

We evaluated the model with spatially explicit independent data collected after the model was built and believe this approach provides the most reliable assessment of accuracy. Many studies fail to test model accuracy spatially and in the field using additional, independent data (Heglund 2002) and instead only evaluate model accuracy by re-sampling or partitioning training data (Manel et al. 2001). Data partitioning methods such as k-fold partitioning reduce the sample size of training data and though they are an acceptable method of evaluation, they are less effective and meaningful than using independent and truly spatial data (Verbyla and Litaitis 1989; Fielding and Bell 1997; Fielding 2002).

We ground-truthed the model by conducting landscape-scale aerial surveys in model-predicted areas to learn if predictions correctly classified gyrfalcon nest occurrence in survey plots in May and June 2008. We selected three study areas for which the model predicted high gyrfalcon nest occurrence (>80%) but for which no training data were present: the Lisburne Peninsula in northwest Alaska, parts of the Arctic National Wildlife Refuge in northeast Alaska, and parts of the Togiak National Wildlife Refuge in southwest Alaska (Fig. 1).

We placed circular, 50-km^2 plots (4 km radius) in each study area within each of three predicted occurrence categories: "high" (predicted >80% occurrence, n = 5), "moderate" (60–40%, n = 6), and "low" (<20%, n = 7). Survey plots were paired and located within 10 km of each other within each prediction category to reduce travel time between plots (Fig. 2). We subjectively located paired plots within 160 km of a runway to provide plane access and refueling options. No other information was used to determine plot location, and we had no prior knowledge of gyrfalcon occurrence in the study areas. We attempted to survey ten plots in each predicted occurrence category across the state (30 total), but poor weather allowed only 18 plots to be surveyed. Additionally, we had planned to conduct repeat surveys on the 30 plots to estimate detectability and correct for imperfect detection, but poor weather prevented this.

We chose the 4-km radius plots because that was the approximate pixel size of the predictive layer and because the size allowed us to survey two plots per flight based on plane fuel capacity and consumption. Our ArcGIS plot and survey maps were transferred to Google Earth to be publicly and easily available to our survey pilots. Each plot was surveyed by TB and a pilot using a two-seat Piper Super
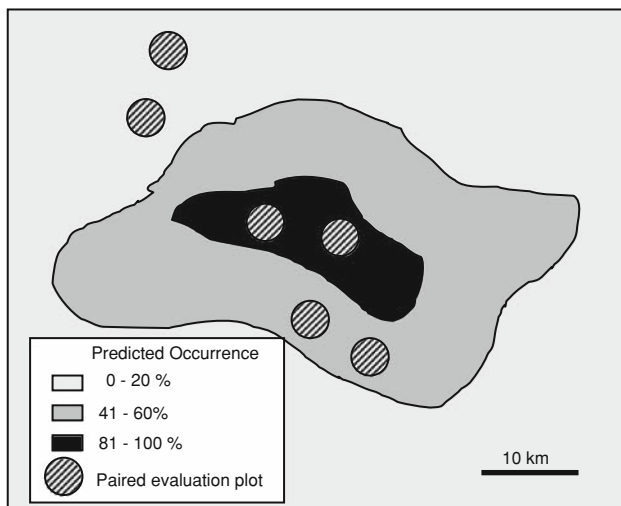
Fig. 2 Schematic diagram of paired evaluation plots surveyed to collect independent testing data for the predictive model. Each pair of plots was placed within one of three predicted occurrence categories (0–20%, 41–60%, and 81–100%) and surveyed for gyrfalcon nest cliffs and occupied nests

Cub or Aviat Husky fixed-wing plane for 40–120 min, depending on the geographic complexity of the plot. The plane was flown as low and slow as was safely possible given conditions, typically 60 m above the ground (range 30–200 m) at 110–130 km/h. We recorded the number of occupied gyrfalcon nests (a nest with eggs, young, or territorial adults) and the number of potential nest cliffs found in each plot.

Gyrfalcons may not breed every year even in good quality habitat because of natural fluctuations in prey, weather conditions, and other stochastic variables (Nielsen and Cade 1990a). Therefore, it was possible that a plot could have been occupied by breeding gyrfalcons in some years but not in the year we surveyed it. Additionally, detecting raptors during aerial surveys can be difficult (Andersen 2007, TLB unpublished data), and it is possible that we failed to detect a few occupied sites during our surveys. Hence, collecting information on the presence of both occupied and potential nest sites on plots provided us with a more comprehensive understanding of model accuracy in terms of fundamental and realized niches, imperfect survey detectability, and the underlying biological mechanisms.

For evaluation purposes, we considered a plot as occupied (true positive finding) if it contained an occupied gyrfalcon nest or had >5 potential nest cliffs on the plot. We chose five cliffs as a cut-off value after completing the surveys because that was the minimum number of cliffs found on a plot in which an occupied gyrfalcon nest was also detected. We assessed model accuracy by comparing the predicted gyrfalcon nest occurrence value of each plot (high 0.9, moderate 0.5, low 0.1) to the occupancy status of

the plot determined by aerial surveys. We then created a confusion matrix for each dataset using counts of true positive, false positive, true negative, and false negative results when comparing predicted versus observed data. We used ROC graphs and AUC scores to interpret model accuracy (Fielding and Bell 1997). Calculations were performed in publicly available online program ROC plot (Schroeder 2004) for the independent survey data.

General methods

Several logical biases deserve attention. First, presence data were obtained opportunistically and may not capture the full spectrum of variation in gyrfalcon nesting preferences or its ecological niche. This could introduce bias if presence data were not representative of most of the natural variation and gradients in nesting areas. However, presence data came from all regions of Alaska where gyrfalcons are documented to breed (Fig. 1). Second, it is possible that gyrfalcon nesting distribution may have changed over the course of the 36+ year dataset such that the resulting model may not apply to the current breeding population. A large-scale shift in nesting occurrence over this period of time is unlikely because gyrfalcons are relatively long-lived birds (Nielsen and Cade 1990a) and likely have high nest site fidelity (Nielsen 1991) with many nesting areas used repeatedly by generations of gyrfalcons. For example, some historical gyrfalcon nest sites in Greenland have been occupied by gyrfalcons and other raptors for the last 2,500 years (K. Burnham, unpublished data). Third, was the scale (extent and pixel size; Wu and Hobbs 2002; Huettmann and Diamond 2006; Guisan et al. 2007) appropriate for the work? We selected Alaska as our extent because it was the appropriate political management unit for implementing conservation actions and it is large enough to include much of the inherent variation in nesting occurrence for the species, making results informative about the species as a whole and on a global level. We assume the entire state was available to gyrfalcons because they move across the state (Britten et al. 1995) and into Canada and the lower 48 United States (Sanchez 1993). Pixel size was determined by environmental layers publicly available with statewide coverage; most were 1 km or less and therefore adequate for modeling occupied nest locations (which were typically separated by 10–100 km).

Results

Nine of 12 environmental variables predicted gyrfalcon nest occurrence (Table 1). The most important predictor variable was soil type, followed by sub-surface geology and vegetation type. Gyrfalcon nests were most commonly

Table 3 Partial dependency scores of the three most influential variables predicting gyrfalcon nest occurrence in Alaska taken from response curves provided in TreeNet

Predictor variable	Soil, geology, or vegetation type	Partial dependency score
Soil	Pergelic Cryaquepts	2.60
$n = 267$[a]	Pergelic Cryumbrepts	2.60
	Histic Pergelic Cryaquepts	2.50
	Pergelic Cryoborolls	2.40
Subsurface geology	Quaternary Mafic Volcanic Rocks—Basalt	1.00
$n = 182$[a]	Ordovician Rocks—limestone and shale	1.00
	Precambrian Z undifferentiated volcanic rocks	0.90
	Upper Cretaceous Continental Deposits	0.85
Vegetation	Dwarf shrub tundra	0.35
$n = 24$[a]	Alpine Tundra and Barrens	0.30
	Ocean Coast	0.30
	Tussock sedge/dwarf shrub tundra	0.28
	Tall shrub	0.28

[a] Denotes the total number of types within each predictor variable used for predictive modeling

Actual values are presented instead of the response curves for easier interpretation. The range of scores within each predictor variable varied from 2.6 to −0.8 for soil, 1.0 to −1.1 for subsurface geology, and 0.35 to −4.1 for vegetation. Positive partial dependency scores denote a positive association with gyrfalcon nests; negative scores indicate a negative association. Only the four highest partial dependency scores from the numerous levels within each for the three predictor variables are listed here for brevity

associated with pergelic cryaquepts, soils that were typically wet, frozen, had high organic content, and had gravelly, steep slopes (Rieger et al. 1979) (Table 3). The most common subsurface geology associated with nest sites were Quaternary mafic volcanic rock such as basalt and Ordovician limestone and shale. Nests were most associated with dwarf shrub tundra, alpine tundra and barrens, and ocean coast vegetation types (Table 3).

The optimized TreeNet model contained 48 statistical trees and predicted gyrfalcon nesting occurrence (Fig. 3). Approximately 75% of the state was predicted to have an index of relative occurrence <20%; 7% of the state was predicted to have an index of >60% (Table 2). Areas of high predicted occurrence (>80%) were patchy and widely dispersed, located in southwest, west, northwest, and northern Alaska. They included well-known breeding areas on the Seward Peninsula and in parts of the Brooks Mountain range and northern foothills. Areas of high predicted
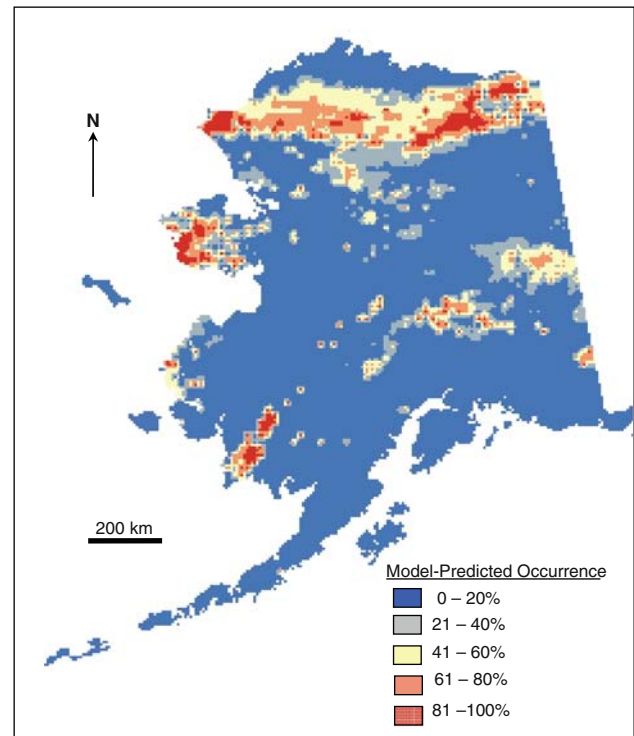
Fig. 3 Model-predicted map of gyrfalcon nest occurrence in Alaska (0–100% relative occurrence). Parts of southeast Alaska and the Aleutian chain are not shown, though these areas were in the lowest prediction category

occurrence not well-known as gyrfalcon breeding areas included parts of the Togiak National Wildlife Refuge in southwest Alaska, the Lisburne Peninsula in northwest Alaska (though see White and Boyce 1977), and parts of the Brooks Mountain range within the Arctic National Wildlife Refuge in northeast Alaska. These areas varied from 35 to 80 km in width. Using the range of published nesting densities (Swem et al. 1994), the area of each predicted occurrence category, and model accuracy as a measure of error, we estimated approximately 546 ± 180 breeding pairs (using our model accuracy estimate (67%) from evaluation plots as a measure of error) occur in the state in any given year (Table 2).

Using 10-fold cross validation on the training data, the optimized model was 97% and 93% accurate in aspatial terms and when assigning presence and absence, respectively. The AUC was 0.96, indicating very high prediction accuracy of the data used. The model's accuracy in predicting potential nest cliffs in the independent and spatial evaluation data was 67% and had an AUC score of 0.76. This indicated the model was moderately accurate and useful when predicting potential nest cliffs in the real world. The model was 36% accurate for predicting occupied gyrfalcon nests in the evaluation data with an AUC score of 0.38,

indicating the model performed worse than if presence/absence were assigned randomly when tested with data from the realized niche.

Discussion

It is useful to think of the variable with the lowest predictive value (aspect) as defining the outer boundary of the species' fundamental niche in multidimensional space. Each subsequent predictor variable (in this case, digital elevation, April temperature, distance to coast, etc.) sequentially shrinks the niche in multidimensional space as each variable increases in its predictive ability, until all abiotic factors have been considered. The resulting space is the fundamental niche and is most restricted by the variables with the most predictive influence: vegetation type, subsurface geology, and soil type for gyrfalcons. The fundamental niche could be further sequentially refined using biotic variables that influence the species' ability to persist such as prey, competing species, and individual fitness. The core of this multidimensional space is the species' realized niche in space and time and its description is the ultimate conservation biology goal. Here we contribute to this effort by describing the fundamental niche because of its relative simplicity compared to the complex realized niche.

The importance of soil type, sub-surface geology, and vegetation type in the model likely reflects their relative importance in the species' ecology. It is important to note, however, that the variables we found to be relatively unimportant in this study may play an important role in falcon ecology at other scales, locations, or systems. For example, Urios and Martinez-Abrain (2006) found that elevation, aspect, slope, and distance to human developments were important in describing nest site preferences of Eleonora's Falcons (*Falco eleonorae*) on a Mediterranean island but they were of relatively low importance in this study.

Gyrfalcon nests were commonly found on pergelic cryaquepts soils that support tundra and dwarf shrub vegetation on which the gyrfalcon's prey (ptarmigan) depend. Sub-surface geology ranked high as a predictor variable because the underlying geology greatly determined cliff occurrence on the landscape. Gyrfalcon nests are associated with volcanic rocks (basalt) and sedimentary rocks (limestone and shale), which probably produce more cliffs than other types (alluvial deposits). The predominant vegetation types associated with nests match our current understanding of gyrfalcons as an obligate tundra breeder. They provide gyrfalcons ptarmigan and the open environments needed to capture them. Nests were also associated with ocean coastline vegetation, which may seem surprising. However, we suspect gyrfalcons are selecting such areas in some regions because they provide direct access to

seabirds, a common alternative food source (Nielsen and Cade 1990b). This finding is particularly interesting because coastal habitats are also valuable to non-breeding gyrfalcons (Britten et al. 1995) and have high potential for wind turbine developments that may cause conservation conflicts (U.S. Department of Energy 2008).

Our Alaska gyrfalcon population estimate was higher than Swem et al. (1994) estimate of 375–635 pairs. Issues that influenced our population estimate include: (1) we subjectively assigned nesting densities to predicted occurrence categories based on published estimates and our own knowledge base with the species. Future population estimates may change if the density estimates or biological knowledge changes. (2) Our attempts to assess model accuracy with independent data are only first steps in validating the model. Clearly, increasing the number and distribution of evaluation plot surveys would improve confidence in our accuracy estimate. (3) The actual population varies by year because of stochastic events and because gyrfalcons respond numerically to fluctuating ptarmigan populations (Nielsen 1999). (4) Estimating the area that should be used for density estimate extrapolations is complex, e.g. whether lakes are to be excluded, and can affect population estimates.

Results from our accuracy assessments suggested the model was highly accurate (93–97%) when using re-sampling methods and moderately accurate (67%) when using independent data. Studies using data-mining techniques with internal accuracy assessments reported very high accuracy assessments, especially when compared to more traditional modeling techniques such general linear models (Elith et al. 2006). However, models of complex biological systems with very high prediction accuracy are unusual and should be viewed with scrutiny because the complex nature of biological systems makes them difficult to capture and predict. Typical ecological models such as general linear models have much lower prediction accuracies because of this (Fielding 2002). Evaluating a model spatially with independent field data, as we did, provides a more realistic and accurate assessment of model accuracy and is preferred (Fielding and Bell 1997; Manel et al. 1999). Tests with independent data often reduce initial accuracy; our accuracy assessments support this finding and that predictive models need to be evaluated with independent data to assess true accuracy (Heglund 2002) and to gain credibility among managers and other decision makers.

The best measure of our model accuracy was the count of potential nest cliffs obtained from plot surveys. This measure was not influenced by complex biotic variables, was in line with our attempts to model the fundamental niche (as determined by abiotic factors), and was a true measure of model performance in the real world. The model's AUC score was 0.76 when assessed with counts of

potential nest cliffs, which is considered to be moderately accurate (Swets 1988) and 'useful' (Elith and Burgman 2002). Our model should not be viewed as optimal, but rather an important first step towards refining our understanding of abiotic and biotic factors influencing gyrfalcons.

Model accuracy was low using presence/absence of occupied nests on plots, but this is unsurprising for a number of reasons. First, survey plot size was relatively small (50 km^2) compared to gyrfalcon breeding density estimates in Alaska (1 pair per $170–1,000 \text{ km}^2$) (Swem et al. 1994), and therefore under-sampling may have occurred. Low breeding densities make reliable accuracy evaluation difficult regardless of model accuracy (Henebry and Merchant 2002). Increasing plot size lowers sample size beyond desired levels because of fuel and weather limitations. Hence, plot size was a compromise between restrictive logistics and the likelihood of a plot including an occupied nest if one was present on the landscape. Second, some nests may have already failed and were unavailable to be detected during surveys. This is particularly true for the 2008 breeding season, when we observed some of the lowest occupancy and productivity rates in a long term study area in recent times in southwest Alaska (TB unpublished data). Third, not all occupied raptor nests are detected during aerial surveys (TB unpublished data), and some occupied nests could have been missed (Boyce et al. 2005). Fourth, evaluating a model that attempts to predict the fundamental niche of a species with data on the realized niche is probably overly conservative. We did not include biological predictor variables that influenced breeding distribution because they were too complex to measure and were unavailable for statewide coverage. For example, stochastic ptarmigan densities influence gyrfalcon nest occupancy (Nielsen 1999) but are unavailable on a statewide or temporal basis. Therefore, the accuracy of the model using occupied nest data alone is probably not highly informative.

Wiens (2002) suggested that model accuracy is as good as the performance of the environmental layers with which it was produced. Though we used the best available layers with statewide coverage, some layers have not been rigorously ground-truthed, metadata were lacking, and their accuracy was not always known quantitatively. It is unlikely that any model using these layers can truly achieve 93–97% accuracy in the real world. Therefore, a model that captures gyrfalcon nesting ecology, distribution, and population size in one quantitative formula with an accuracy of 67% across the entire state is a significant step forward in our knowledge.

Our spatially explicit, non-linear model offers a number of advantages over non-modeling methods (Table 4) and linear models. First, it helps us understand complex systems in simple, transparent terms. It also provides discrete

Table 4 Advantages of modeling over non-modeling approaches for predicting species occurrence and population size

(A)	Quantitative
(B)	Repeatable
(C)	Objective
(D)	Fast
(E)	Convenient
(F)	Nest distribution summarized by one algorithm
(G)	Provides habitat response curves
(H)	Includes multivariate interactions and responses
(I)	Compilation of all relevant data into one dataset
(J)	Brings experts together
(K)	Stimulates discussion
(L)	Improves hypotheses
(M)	Broadly applicable across remote, inaccessible areas
(N)	Represents best available science
(O)	Represents complex interactions with simple numerics

measures of relative variable importance, breeding distribution, and population size in readily interpretable formats based on objective, best available science. Second, non-linear modeling captures complex multivariate relationships not possible with linear methods (Elith et al. 2006). Criticisms of non-linear analyses such as classification and regression trees (CART) include concerns that precision is difficult to estimate, optimal trees may not be found, and results may be sensitive to small changes in data (Anderson et al. 2000; Hastie et al. 2001; Elith et al. 2006). However, stochastic gradient boosting is a refinement of the traditional regression tree analysis that addresses these concerns and improves model performance (Friedman 2002). It excels at modeling non-linear data common in ecological studies, can handle large numbers of categorical and continuous predictor variables, performs at a faster rate than traditional techniques, and is robust to datasets that contain up to 30% faulty data in some instances (Craig and Huettmann 2008). Machine learning models such as stochastic gradient boosting produce highly accurate predictions that perform faster, are more informative, and are similar to or better than, the accuracy of traditional linear modeling approaches (Elith et al. 2006; Breiman 2001). Finding that 9 of 12 environmental variables influenced gyrfalcon distribution indicates that complex multivariate habitat and environmental relationships exist for this species and that using non-linear modeling is prudent. Third, model-based estimates enable dynamic, near real-time population estimation (instead of static ones fixed in time) and facilitates further ecological research. For example, we could include real-time data on ptarmigan distribution and population levels (if/when such become available) in modeling efforts to produce dynamic gyrfalcon population estimates. The implications for

guiding fieldwork and research design are considerable including investigating spatial distribution patterns of species (fragmentation and source/sink dynamics). None of our model components are fully explored, yet deserve more attention to understand spatially explicit population dynamics.

Our model should be useful to managers addressing conservation issues in Alaska. For example, the model could be combined with existing regionalized IPCC climate models to forecast future gyrfalcon population size, distribution, and changes under varying climate scenarios (Seavy et al. 2008). Or, distribution maps could overlay maps of current and predicted locations of oil, gas, minerals, and wind resources to identify areas of potential future conflict, estimate the potential size or severity of impacts caused by a specific activity, and prioritize conservation strategies geographically.

Our modeling efforts represent a significant collating of sensitive nest location data from collaborators concerned about potential negative nest disturbance. Gyrfalcons are highly sought after by birders, and a substantial illegal harvest occurs in parts of the globe for falconry (Lobkov 2000). If made available, nest location data could negatively impact the resource. For example, we encountered resistance in Europe about sharing nest locations; so much so that it prevented us from conducting research. Our work in Alaska is an example of the need to build partnerships, establish trust, and creatively solve problems to maintain data security while not preventing scientific learning through meta-analysis of data. Our innovative method of having a data-holder extract the relevant information from the GIS layers and provide that information to researchers instead of actual nest locations (as suggested by C. McIntyre), should be useful for others to further conservation and science when dealing with sensitive data.

References

Andersen DE (2007) Survey techniques. In: Bird D, Bildstein K (eds) Raptor research and management techniques. Hancock House Publishers, Blaine, WA, pp 89–100

Anderson DR, Burnham KP, Thompson WL (2000) Null hypothesis testing: problems, prevalence, and an alternative. J Wildl Manage 64:912–923

Araujo MB, Guisan A (2006) Five (or so) challenges for species distribution modeling. J Biogeo 33:1677–1688

Araujo MB, Williams PH (2000) Selecting areas for species persistence using occurrence data. Biol Conserv 96:331–345

Beikman HM (1980) Geologic map of Alaska. U.S. Geological Survey special publication # SG0002-1T and SG0002-2T, Washington, DC

Beyer HL (2008) Hawth's analysis tools for ArcGIS. http://www.spatialecology.com/htools. Accessed 12 Dec 2008

Booms TL, Cade TJ, Clum NJ (2008) Gyrfalcon (*Falco rusticolus*). In: Poole A (ed) The birds of North America online. Cornell Lab of Ornithology. doi:10.2173/bna.114. http://bna.birds.cornell.edu/bna/species/114. Accessed 10 Jan 2009

Boyce MS, McDonald LL (1999) Relating populations to habitats using resource selection functions. Trends Ecol Evol 14:268–272

Boyce DA, Kennedy PL, Beier P, Ingraldi MF, MacVean SR, Siders MS, Squires JR, Woodbridge B (2005) When are goshawks not there? Is a single visit enough to infer absence at occupied nest areas? J Raptor Res 39:296–302

Breiman L (2001) Random forests. Mach Learn 45:5–32

Britten MW, McIntyre CL, Kralovec M (1995) Satellite radio telemetry and bird studies in national parks and preserves. Park Sci 15:20–24

Cade TJ (1960) Ecology of the peregrine and gyrfalcon populations in Alaska. Univ Calif Pub Zool 63:151–290

Cade TJ (1982) The falcons of the world. Cornell University Press, Ithaca, NY

Craig E, Huettmann F (2008) Using blackbox algorithms such as Tree-Net and random forests for data-mining and for finding meaningful patterns, relationships, and outliers in complex ecological data: an overview, an example using golden eagle satellite data and an outlook for a promising future. In: Hsiao-fan W (ed) Intelligent data analysis: developing new methodologies through pattern discovery and recovery. IGI Global, Hershey, PA, pp 65–84

Crick HQ (2004) The impact of climate change on birds. Ibis 146:48–56

Elith J, Burgman M (2002) Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. In: Scott JM, Heglund PJ, Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB (eds) Predicting species occurrences. Island Press, Washington, pp 303–314

Elith J, Graham C, Anderson RP et al (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151

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. J Appl Ecol 41:263–274

Environmental Systems Research Institute (2008) ArcMap 9.3 resource center. http://www.resources.esri.com. Accessed 12 Dec 2008

Fielding AH (2002) What are the appropriate characteristics of an accuracy measure? In: Scott JM, Heglund PJ, Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB (eds) Predicting species occurrences. Island Press, Washington, pp 303–314

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24:38–49

Fleming M (1997) A statewide vegetation map of Alaska using pheno-logical classification of AVHRR data. In: Walker DA, Lillie AC (eds) The second circumpolar Arctic vegetation mapping workshop and the CAVM-North American workshop, pp 25–26

Friedman JH (2002) Stochastic gradient boosting. Comput Stat Data Anal 38:367–378

Guisan A, Graham CH, Elith J, Huettmann F (2007) Sensitivity of predictive species distribution models to change in grain size: insights from a multi-models experiment across five continents. Divers Distrib 13:332–340

Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning: data mining, inference, and prediction. Springer, New York

Heglund PJ (2002) Foundations of species-environment relations. In: Scott JM, Heglund PJ, Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB (eds) Predicting species occurrences. Island Press, Washington, pp 35–42

Henebry GM, Merchant JW (2002) Geospatial data in time: limits and prospects for predicting species occurrences. In: Scott JM, Heglund PJ, Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB (eds) Predicting species occurrences. Island Press, Washington, pp 291–302

Higmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climat 25:1965–1978

Huettmann F, Diamond AW (2006) Large-scale effects on the spatial distribution of seabirds in the Northwest Atlantic. Landsc Ecol 21:1089–1108

Hunt WG, Jackman RE, Hunt TL, Driscoll DE, Culp L (1999) A population study of golden eagles in the Altamont Pass wind resource area 1994–1997. Report to National Renewable Energy Laboratory, Subcontract XAT-6-16459-01. University of California, Santa Cruz

Hutchinson GE (1957) A treatise on limnology. Wiley, New York

Johnson GD, Erickson WP, Strickland MD, Shepherd MF, Shepherd DA (2001) Avian monitoring studies at the Buffalo Ridge Wind Resource area, Minnesota: results of a 4-year study. Technical report prepared for Northern States Power Co, Minneapolis, MN

Karlstrom TNV, Coulter HW, Fernald AT et al (1964) Surface geology of Alaska. Miscellaneous geologic investigations map I-357. Washington, DC

Keating KA, Cherry S (2004) Use and interpretation of logistic regression in habitat selection studies. J Wildl Manage 68:774–789

Lehner B, Doll P (2004) Development and validation of a global database of lakes, reservoirs and wetlands. J Hydrol 296:1–22

Lobkov EG (2000) Illegal trapping and export of gyrfalcons from Kamchatka is a threat to the very existence of the Kamchatka population. Abstracts of the first conference on conservation of biodiversity in Kamchatka and its coastal waters. Kamchatsk NIRO, Petropavlosk-Kamchatskiy

Manel S, Dias JM, Ormerod SJ (1999) Comparing discriminate analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecol Model 120:337–347

Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence–absence models in ecology: the need to account for prevalence. J Appl Ecol 38:921–931

Manly BFJ, McDonald LL, Thomas DL, McDonald TL, Erickson WP (2002) Resource selection by animals, statistical design and analysis for field studies, 2nd edn. Kluwer Academic Publishers, London

Mladenoff DJ, Sickley TA, Haight RG, Wydeven AP (1995) A regional landscape analysis and prediction of favorable gray wolf habitat in the northern great lakes region. Conserv Biol 9:279–294

Mladenoff DJ, Sickley TA, Wydeven AP (1999) Predicting gray wolf landscape recolonization: logistic regression models vs. new field data. Ecol Appl 9:37–44

Nielsen ÓK (1991) Age of first breeding and site fidelity of gyrfalcons. Náttúrufæðingurinn 60:135–143

Nielsen ÓK (1999) Gyrfalcon predation on ptarmigan: numerical and functional responses. J Anim Ecol 68:1034–1050

Nielsen ÓK, Cade TJ (1990a) Annual cycle of the gyrfalcon in Iceland. Nat Geo Res 6:41–62

Nielsen ÓK, Cade TJ (1990b) Seasonal changes in food habits of gyrfalcons in northeast Iceland. Ornis Scand 21:202–211

Nielsen SE, Stenhouse GB, Beyer HL, Huettmann F, Boyce MS (2008) Can natural disturbance-based forestry rescue a declining population of grizzly bears? Biol Conserv 141:2193–2207

Onyeahialam A, Huettmann F, Bertazzon S (2005) Modeling sage grouse: progressive computational methods for linking a complex set of local biodiversity and habitat data towards global conservation statements and decision support systems. Lecture Notes in Computer Science 3482, International Conference on Computational Science and its Applications Proceedings Part III, pp 152–161

Palmer RS (1988) Handbook of North American birds, vol 5. Vail-Ballou Press, Binghamton, NY

Pearce JL, Boyce MS (2006) Modeling distribution and abundance with presence-only data. J Appl Ecol 43:405–412

Peterson AT (2001) Prediction species' geographic distributions based on ecological niche modeling. Condor 103:599–605

Potapov E, Sale R (2005) The gyrfalcon. Yale University Press, New Haven, CT

Rieger S, Schoephorster DB, Furbush CE (1979) Exploratory soil survey of Alaska. U.S. Department of Agriculture. Natural Resource Conservation Service, Fort Worth, Texas

Salford Systems (2002) TreeNet Version 2.0. http://www.salford-systems.com/treenet. Accessed 30 Oct 2008

Sanchez GH (1993) The ecology of wintering gyrfalcons (*Falco rusticolus*) in central South Dakota. Master's Thesis, Boise State University

Sanderson EW, Jaiteh M, Levy MA, Redford KH, Wannebo AV, Woolmer G (2003) The human footprint and the last of the wild. Biosci 52:891–904

Schroeder B (2004) ROC plot. http://brandenburg.geoecology.uni-potsdam.de/users/schroeder/download.html. Accessed 30 Oct 2008

Seavy NE, Dybala KE, Snyder MA (2008) Climate models and ornithology. Auk 125:1–10

Smallwood KS, Thelander CG (2004) Developing methods to reduce bird mortality in the Altamont Pass Wind Resource Area. PIER-EA contract no. 500-01-019, Sacramento, CA

Stafford J, Wendler G, Curtis J (2000) Temperature and precipitation of Alaska: 50 year trend analysis. Theor Appl Climat 67:33–44

Swem T, McIntyre C, Ritchie RJ, Bente PJ, Roseneau DG (1994) Distribution, abundance, and notes on the breeding biology of gyrfalcons (*Falco rusticolus)* in Alaska. In: Meyburg B-U, Chancellor RD (eds) Raptor conservation today: proceedings of the IV world conference on birds of prey and owls, Berlin, Germany, May 10–17, 1992. World Working Group on Birds of Prey and Owls, London, pp 437–444

Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240:1285–1293

Tape K, Sturm M, Racine C (2006) The evidence for shrub expansion in northern Alaska and the pan-Arctic. Glob Change Biol 12:686–702

Urios G, Martinez-Abrain A (2006) The study of nest-site preferences in Eleonora's falcon (*Falco eleonorae*) through digital terrain models on a western Mediterranean island. J Ornithol 147:13–23

US Census Bureau (2004) 2000 census of population and housing, United States summary. Report PHC-3-1, Washington DC

US Department of Energy (2008) U.S. wind resources map. http://www.windpoweringamerica.gov/wind_maps.asp. Accessed 22 Oct 2008

US Geological Survey (2008) Circum-Arctic resource appraisal: estimates of undiscovered oil and gas north of the Arctic Circle. USGS Fact Sheet 2008–3049

US Geological Survey (1997) Alaska 300 m digital elevation model. US Geological Survey Alaska Field Office, Anchorage, Alaska

Verbyla DL, Litaitis JA (1989) Resampling methods for evaluating classification accuracy of wildlife habitat models. Environ Manage 13:783–787

White CM, Boyce DA (1977) Distribution and ecology of raptor habitat studies for the Kilbuck Mountains, Anvik, Unalakleet, and northwestern Arctic regions of Alaska. US Bureau of Land Management Report, Anchorage, AK

Wiens JA (2002) Predicting species occurrences: progress, problems, and prospects. In: Scott JM, Heglund PJ, Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB (eds) Predicting species occurrences. Island Press, Washington, pp 739–749

Wu J, Hobbs R (2002) Key issues and research priorities in landscape ecology: an idiosyncratic synthesis. Landsc Ecol 17:355–365

Wu XB, Smeins FE (2000) Multiple-scale habitat modeling approach for rare plant conservation. Landsc Urban Plan 51:11–28