Addressing biased occurrence data in predicting potential Sierra Nevada red fox habitat for survey prioritization

Casey Cleve , John Perrine , Barbara Holzman , Ellen Hines

ABSTRACT: The Sierra Nevada red fox *Vulpes vulpes necator* is listed as a threatened species under the California Endangered Species Act. It originally occurred throughout California's Cascade and Sierra Nevada mountain regions. Its current distribution is unknown but should be determined in order to guide management actions. We used occurrence data from the only known population, in the Lassen Peak region of northern California, combined with climatic and remotely sensed variables, to predict the species' potential distribution throughout its historic range. These model predictions can guide future surveys to locate additional fox populations. Moreover, they allow us to compare the relative performances of presence-absence (logistic regression) and presence-only (maximum entropy, or Maxent) modeling approaches using occurrence data with potential false absences and geographical biases. We also evaluated the recently revised Maxent algorithm that reduces the effect of geographically biased occurrence data by subsetting background pixels to match biases in the occurrence data. Within the Lassen Peak region, all models had good fit to the test data, with high values for the true skill statistic (76–83%), percent correctly classified (86–92%), and area under the curve (0.94–0.96), with Maxent models yielding slightly higher values. Outside the Lassen Peak region, the logistic regression model yielded the highest predictive performance, providing the closest match to the fox's historic range and also predicting a site where red foxes were subsequently detected in autumn 2010. Subsetting background pixels in Maxent reduced but did not eliminate the effect that geographically biased occurrence data had on prediction results relative to the Maxent model using full background pixels.

INTRODUCTION

Determining the composition and distribution of suitable habitat is crucial for the successful management of rare or endangered species (Guisan & Zimmermann 2000). Species distribution models have become an important tool to identify optimal survey areas for these species and to increase the probability of locating previously unknown populations (Engler et al. 2004, Edwards et al. 2005, Guisan et al. 2006, Rachlow & Svancara 2006, Ferreira de Siqueira et al. 2009, Menon et al. 2010, Rebelo & Jones 2010). Because a rare species will likely be absent from most sample locations in a simple random sample, targeting survey locations to areas with high probability of species occurrence represents a more efficient use of limited conservation resources (Guisan et al. 2006).

The Sierra Nevada red fox *Vulpes vulpes necator* is a medium-sized canid that historically occurred at low

densities throughout the high elevations of the Sierra Nevada and southern Cascade mountain ranges of California and Oregon, USA (Grinnell et al. 1937, Hall 1981, Sacks et al. 2010). Within this broad range, Grinnell et al. (1937) reported 3 population centers: the Mount Shasta/Lassen Peak region in northern California, the central Sierra near Mono Lake and Yosemite National Park, and the southern Sierra near Mount Whitney (now mostly in the Sequoia and Kings Canyon National Parks). In 1980, due to a noticeable decline in numbers, the Sierra Nevada red fox was listed as a state 'threatened' species; the factors causing its decline are unknown (CDFG 1996, 2004). Despite its former extent, the verified detections since 1991 have all been in the Lassen Peak region of northern California (Perrine et al. 2007). A recent conservation assessment for this species (Perrine et al. 2010) recommended that targeted surveys for the Sierra Nevada red fox be conducted throughout its historic range to determine whether any additional populations exist. Prior attempts to detect this species as part of multi-species carnivore inventories have been unsuccessful, even in the Lassen Peak region where the species has persisted (Zielinski et al. 2005). Surveys targeting the Sierra Nevada red fox and focusing on areas with high probabilities of the species' occurrence may increase detection probability and survey efficiency.

Our goal was to predict the extent and distribution of suitable habitat for the Sierra Nevada red fox throughout its historic range, based on the characteristics of occupied habitat in its current known range. The available data were collected during a comprehensive ecological study of the Lassen Peak population (Perrine 2005). Unfortunately, these detection data contained 2 important biases. First, the survey data contained potential 'false absences;' surveys have failed to detect this species in areas where populations were known to be present. Absence data for rare species may result from the rarity of the species rather than its true absence (MacKenzie et al. 2006). False absences are problematic in species distribution modeling because they do not indicate unsuitable habitat or confirm that the species does not occur at a given site (Guisan & Zimmermann 2000, Engler et al. 2004, MacKenzie et al. 2006). Second, the only available occurrence data for the Sierra Nevada red fox were from the Lassen Peak region, which represents only a small portion of the species' historic range. Researchers have cautioned against using geographically biased occurrence data or transferring models to broad unsampled regions (Peterson et al. 2007, Barbosa et al. 2009); however, such projections have successfully increased detection rates (Guisan et al. 2006). General linear models (GLMs) may transfer quite well to unsampled areas (Randin et al. 2006,

Barbosa et al. 2009); presence-only models such as Maxent have made improvements in their transferability (Phillips 2008), with their performance being positively correlated with the similarity between the occurrence data region and the projection area (Bulluck et al. 2006).

Addressing these biases provided an opportunity to explicitly compare the performance of 2 different modeling approaches: logistic regression based on presence-absence data versus maximum entropy based on presence-only data. Logistic regression has been widely used in species distribution modeling (Mladenoff et al. 1999, Nielsen et al. 2002, Johnson et al. 2004, Posillico et al. 2004, Olivier & Wotherspoon 2006). Logistic regression uses presence-absence data to model the probability of species occurrence as a function of its predictor variables, which can be continuous or categorical (MacKenzie et al. 2006). Its output is confined to values between 0.0 and 1.0.

Maximum entropy methods, although relatively new and not as widely used as logistic regression, can outperform logistic regression (Elith et al. 2006) and have successfully identified locations of previously undiscovered populations (Rebelo & Jones 2010). The most widely used maximum entropy approach, in the program Maxent (Phillips et al. 2006), estimates the species' probability distribution that is most dispersed within the constraints of the target population information. Like logistic regression, Maxent can use both categorical and continuous predictor variables, and the output can provide information on the relative contribution of each predictor variable (Phillips et al. 2006). Maxent can utilize presence-only data, sidestepping the problem of false absences, but geographical and environmental biases in the occurrence data can introduce considerable error in presence-only models (Phillips 2008). Presence-only models such as Maxent draw background pixel values from the entire study region, while presence data values are drawn only from a small portion of the study area (Phillips et al. 2006). The resulting predictions may therefore underrepresent habitat suitability outside of the occurrence data area (Peterson et al. 2007, Phillips 2008). Substantial improvements can be made in Maxent models derived from biased occurrence data by selecting background data with similar biases as the occurrence data (Phillips 2008). This approach has recently been implemented in Maxent (Phillips 2008), but its effects have had little empirical validation.

Here we compared the performance and output of a presence-absence logistic regression model versus presence-only Maxent models with and without the transferability improvements. We then combined the output from the logistic regression and updated Maxent model to generate an ensemble prediction, lever-

aging the strengths of each model while minimizing their respective weaknesses (Araújo & New 2007, Stohlgren et al. 2010). In addition, we explored the use of unclassified spectral data as a predictor variable in place of predetermined classification schemes (e.g. vegetation or canopy cover categories). Although classified maps are commonly used predictor variables, wildlife may respond to continuous environmental gradients that are not captured in the classification schemes (Laurent et al. 2005). By using unclassified spectral data, species' occurrence can be predicted by spectrally detectable components of their habitat, rather than predetermined classification schemes that may inaccurately delineate boundaries between cover types and under-represent habitat heterogeneity (St-Louis et al. 2006). Using unclassified spectral reflectance in the distribution model may minimize errors in the resulting predictive maps (Laurent et al. 2005).

MATERIALS AND METHODS

Study area. Our model prediction area covers the area within and immediately surrounding the historic range of the Sierra Nevada red fox (Fig. 1). This includes the Sierra Nevada and the southernmost extent of the Cascade Range in California. The prediction area spans 2 Major Land Resource Area (MLRA) ecoregions: the Sierra Nevada ecoregion and the California portion of the Eastern Cascade Slopes and Foothills ecoregion (USDA-NRCS 2006). The Sierra Nevada ecoregion extends from just south of Lassen Peak to the Tehachapi Pass near Bakersfield. The majority of this ecoregion is comprised of elevations ranging from 450 to 2750 m, with the highest peak being Mount Whitney (4419 m). The California portion of the Eastern Cascade Slopes and Foothills ecoregion represents the southernmost extent of the Cascade Mountain Range, extending from the Central Cascade Mountains to the Sierra Nevada. The majority of this ecoregion is comprised of elevations ranging from 450 to 2500 m, with the highest peak being Mount Shasta (4318 m).

Sierra Nevada red fox survey data were collected from the area surrounding Lassen Peak (3187 m), the southernmost peak in the Cascade Range. The Lassen

Peak region (6455 km^2) includes Lassen Volcanic National Park (LVNP), the surrounding Lassen National Forest (LNF), and the immediately adjacent lands of various ownerships (Fig. 1). This montane area is dominated by conifers such as Jeffrey pine *Pinus jeffreyi* and Ponderosa pine *P. ponderosa*, red fir *Abies magnifica* and white fir *A. concolor*, and mountain hemlock *Tsuga mertensiana*, along with wet alpine meadows and talus slopes. This area has a Mediterranean climate with warm dry summers and cold wet winters. Most of the annual precipitation occurs as snow from November through April, with snowpacks at the higher elevations often exceeding 3 m in depth and persisting into the summer months.

Fox survey data. Sierra Nevada red fox locations were determined using 4 detection methods: radio telemetry, scats (feces), and camera surveys using opportunistic and stratified random sampling designs. Each detection method contained biases.
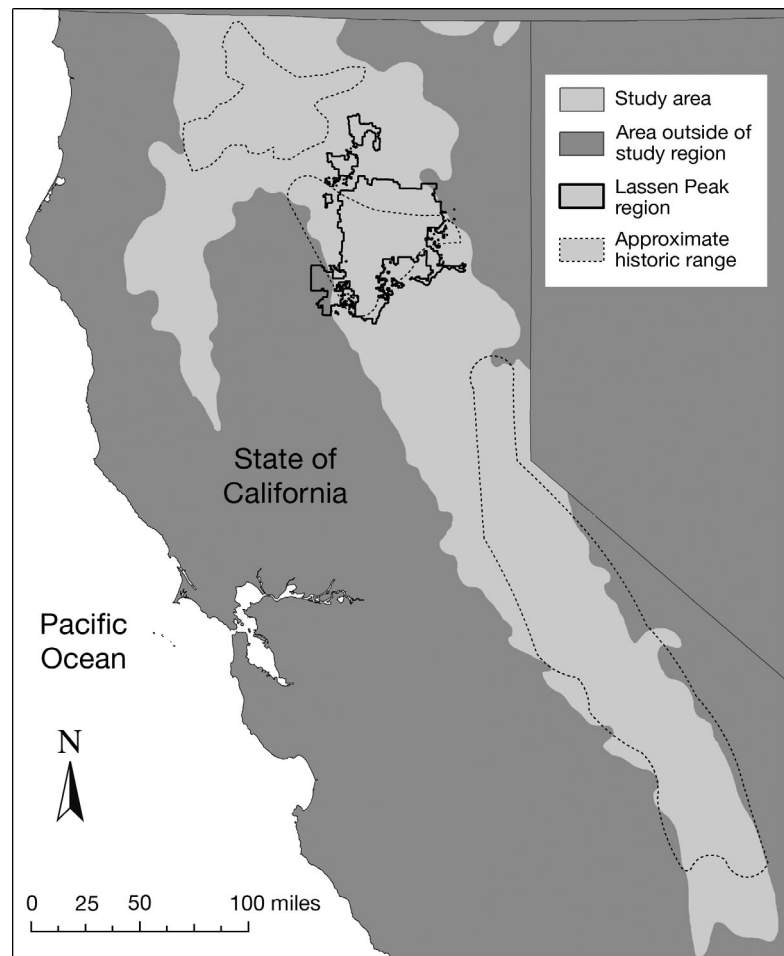

Fig. 1. *Vulpes vulpes necator.* Model prediction area for the Sierra Nevada red fox relative to its historic range (Grinnell et al. 1937) and the Lassen Peak region. Note that scale bar is in miles (1 mile = ca. 1.6 km)

Five Sierra Nevada red foxes were captured and fitted with VHF collars and tracked by aerial and ground-based telemetry from 1998 through 2002 (Perrine 2005). All capture and handling activities were in accordance with California Department of Fish and Game and University of California Berkeley protocols. The field team collected 586 independent ground telemetry locations using a Trimble GeoExplorer II GPS and LOCATE II telemetry software (Nams 2001), and aerial telemetry provided 123 additional locations. In addition, a total of 227 Sierra Nevada red fox scats were collected opportunistically from June 1998 through December 2002, primarily in association with ground telemetry. The telemetry and scat locations were clustered in the western half of LVNP and the adjacent LNF lands (Perrine 2005).

The opportunistic camera station survey was conducted between 1992 and 2002 and consisted of 968 baited TrailMaster (Goodson and Associates) camera stations throughout LVNP and the LNF (Perrine 2005). This survey was conducted primarily by Park and Forest Service biologists following the standard protocol for surveying forest carnivores (Zielinski & Kucera 1995). Although the cameras were widely distributed throughout the region, sampling biases arose due to the opportunistic nature of this survey. For example, the southwest portion of LVNP and the LNF east of the Caribou Wilderness were heavily sampled, whereas the northern portion of the region had the least sampling effort. Samples were also biased toward roads. This survey yielded 50 Sierra Nevada red fox detection locations, with multiple detections at some locations.

The stratified random camera survey was conducted in the summers of 2001 and 2002. This survey consisted of 24 sites stratified by elevation and randomly placed throughout the Lassen Peak region. Each site contained 2 baited TrailMaster cameras approximately 1.6 km apart, following the standard protocol for forest carnivores (Zielinski & Kucera 1995). This survey yielded 3 red fox detection locations (Perrine 2005).

We combined and subsampled these occurrence data to use as the species response variable. Combining these data reduced the effect of the sampling biases inherent in each method, but new errors arose as a result of combining 4 different collection methods that spanned multiple years. Some locations contained 2 presence points, because a fox was detected in the same area by 2 different methods. These pseudoreplicates violate the assumption that training data points are independent, which in turn can bias model results (Guisan & Zimmermann 2000). Additionally, false absences likely occurred at some camera locations due to the elusiveness of this species. Within the 10 yr sampling period, several camera locations detected foxes in one year but not another. Having detection and non-detection records at the same location can introduce conflicting information in the model, which may lower its predictive power. To remove pseudoreplicates and correct for conflicting information, we used ArcGIS (ESRI) to label a location as a presence if a fox was detected there at any time during the 10 yr period. Similarly, we deleted duplicate detections at the same location.

After pooling these data, the sampling intensity varied between habitat types. Since survey data were plentiful (~2000 records), we randomly subsampled the data from each environmental zone to balance the sampling intensity (Guisan & Zimmermann 2000, Araújo & Guisan 2006). The environmental zones were based on a combination of 10 elevation zones and 19 California Wildlife Habitat Relationship (CWHR; Mayer & Laudenslayer 1988) types. This subsampling reduced the dataset from 2000 to 1200 points (600 per presence and absence, respectively) and reduced but did not eliminate sample clustering. To reduce clustering bias, we used Thiessen polygons (Rhynsburger 1973) to downweight points that occurred close together.

Environmental predictor variables. We based our environmental predictor variables on Sierra Nevada red fox ecology and the availability of digital data. The Sierra Nevada red fox is associated with high-elevation conifer forests, subalpine woodlands, talus slopes, and barren areas above treeline (Grinnell et al. 1937, Schempf & White 1977, Perrine 2005). To represent these environmental conditions, we used a variety of GIS layers containing vegetation, climate, hydrology, and forest structure data. Specifically, to represent vegetation and forest stand structure, we used CalFire Fire Resource and Assessment Program's Multi-Source Land Cover Data (MSLC), which contained CWHR vegetation type, total tree canopy closure, tree size class, and tree density class attributes (www.frap.cdf.ca.gov). We also derived Tasseled-cap greenness and wetness from Landsat 5 imagery as an additional variable (software: ERDAS 2008 Leica geosystems geospatial imaging, Atlanta, GA). Tasseled-cap transformation variables represent a continuous environmental gradient that is highly correlated with stand age and structural complexity (Hansen et al. 2001). Pixels containing high greenness and wetness values are associated with dense vegetation having high leaf area index, while lower values indicate sparsely vegetated areas such as barren areas or regions of snow and ice (White et al. 1997, Waring & Running 1998). We used Spatial Analyst (ESRI) to calculate the Euclidean (straight-line) distance from the center of each raster cell to the nearest water feature from the National Hydrography dataset (http://nhd.usgs.gov) and to derive slope from a 30 m digital elevation model (USGS 2000). The Sierra Nevada red fox's elevational limits, relationship to snow pack, and phylogeography (Aubry et al. 2009) suggest

an affinity for specific climatic conditions. We used gridded climate data derived from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) to predict the species' response to different climatic conditions; specifically, we used mean monthly precipitation and monthly average daily minimum and maximum temperatures from 1971 through 2000 (Daly et al. 1994).

After selecting the initial environmental variables based on red fox ecology, we used the R statistical package (R Development Core Team 2005) to determine correlations between continuous variables and to identify interaction terms. If 2 variables were correlated (Pearson's correlation coefficient > 0.3), only the variable with the lower p value was retained. We then identified pairwise interaction terms using a classification and regression tree (Miller & Franklin 2002). Classification and regression trees determine a set of if–then statements that define class membership, and can express complex non-linear and non-additive relationships among the predictor variables (Miller & Franklin 2002). We included 9 environmental variables in the classification and regression tree model: CWHR type, total tree canopy closure, tree size class, tree density class, slope, February precipitation, minimum December temperature, Tassled-cap greenness, and distance to water.

After removing correlated predictor variables and determining interactions, we selected significant predictor variables using iterative manual stepwise logistic regression; at each run, the least significant variable was removed until only significant variables remained (Hastie & Pregibon 1992, Hosmer & Lemeshow 2000). In addition, at each step the Akaike information criterion (AIC) was used to select the best fitting model. AIC is a standardized score used to compare models for best fit relative to the number of parameters in the model; lower AIC values indicate better fit (Burnham & Anderson 2002). The 9 environmental variables listed in the previous paragraph, along with the 2 pairwise interaction terms identified in the classification and regression tree model (see 'Results'), were included in the stepwise logistic regression weighted by Thiessen polygon area. The resulting set of significant predictor variables was used in both the Maxent and logistic regression models. This allowed for direct comparison between the 2 modeling approaches. This 2-step method of using a GLM to select predictor variables followed by Maxent modeling has been shown to create predictions with very high area under the receiver operating characteristic (ROC) curve (AUC) values (Wollan et al. 2008). High AUC values indicate low error, while lower values indicate lower predictability (Pearce & Ferrier 2000).

Distribution models and model evaluation. We generated 3 distribution models: a presence-only maximum entropy method (Maxent) with full region background pixels (hereafter, MFB), a Maxent model using a subset of background pixels with similar biases as the occurrence data (MSB), and a spatially-weighted presence–absence logistic regression model (LRW). We used the default parameters for both Maxent models and generated outputs in the logistic regression format.

For each approach, we developed the model with a random subset of 70% of the data and withheld the remaining 30% for model evaluation. To determine the classification accuracy of each model, we used the evaluation data to identify the optimum cutoff value that corresponded with high red fox habitat suitability. Optimum cutoff values were determined by calculating the true skill statistic (TSS) across the entire range of potential cutoff values, and the cutoff value that corresponded with the highest TSS was selected as the optimum cutoff (Allouche et al. 2006, Jones et al. 2010). We then calculated the AUC to assess predictive performance (Buckland et al. 1997).

Predictive accuracy could only be tested in the Lassen Peak region because of the limited geographic extent of the available data. To compare model performance outside the Lassen Peak region, we compared the distribution and abundance of each model's suitable habitat to the other models and to the historic range map for the Sierra Nevada red fox (Grinnell et al. 1937). We used the optimum cutoff value to determine the appropriate suitable habitat threshold. To create the ensemble prediction, the values of each model that fell below the optimum cutoff value were given a value of 0.0. The mean probability value from both models was then assigned to each cell of the study area.

RESULTS

The classification and regression tree identified 2 significant pairwise interaction terms: February precipitation with minimum December temperature, and February precipitation with image greenness. The stepwise logistic regression weighted by Thiessen polygon area reduced our environmental variable set to the following 6 variables: February precipitation, minimum December temperature, Tasseled-cap greenness, distance to water, the interaction between February precipitation and minimum December temperature, and the interaction between February precipitation and greenness. We used these 6 variables to generate the species distribution models via the 3 approaches described in the above section 'Distribution models and model evaluation'.

Within the Lassen Peak region, all models had good fit to the test data, as indicated by high values for TSS, percent correctly classified, and AUC (Table 1). For all 3

Table 1. Summary of model results for all 3 models of suitable habitat for Sierra Nevada red fox *Vulpes vulpes necator*. MFB: Maxent with full background pixels; MSB: Maxent with subset background pixels; LRW: spatially-weighted logistic regression; AUC: area under the curve; TSS: true skill statistic; LPR SH: suitable habitat (ha) for Lassen Peak Region; SA SH: suitable habitat (ha) for entire study area; SA-LPR SH: suitable habitat (ha) outside the Lassen Peak Region. Suitable habitat is defined as the area that contains a probability of red fox occurrence equal to or greater than the optimum cutoff value. Optimum cutoff values were determined by calculating the TSS across the entire range of potential cutoff values, and the cutoff value that corresponded with the highest TSS was selected as the optimum cutoff

Model	Correctly classified (%)	AUC	TSS (%)	Optimum cutoff	LPR SH (ha)	SA SH (ha)	SA-LPR SH (ha)
MFB	91.5	0.9579	83.1	0.333	92 187	660 479	568 292
MSB	90.8	0.9537	81.9	0.157	79 754	935 362	855 608
LRW	86.0	0.9438	75.8	0.184	115 530	1 546 021	1 430 491

accuracy assessment statistics, the MFB model yielded the highest values, with the MSB model slightly lower, followed by the LRW model. All 3 models had low optimum cutoff values, ranging from 0.157 for the MSB model to 0.333 for the MFB model (Table 1).

Despite their similar accuracy, the 3 models varied in the location and extent of the suitable habitat area they predicted (Table 1, Fig. 2). Within the Lassen Peak region, the MFB and MSB models predicted approximately 20 and 30% less suitable habitat, respectively, than the LRW model (Table 1). This pattern was inten-

sified outside of the Lassen Peak region (Table 1), where the MSB and MFB models predicted approximately 40 to 60% less suitable habitat than the LRW model. All 3 models approximated the historic range boundary for the Sierra Nevada red fox, with the LRW model yielding the closest fit and the MFB model having the sparsest fit (Fig. 2).

The Maxent models (MFB and MSB) selected regions with lower minimum December temperatures and higher February precipitation than the entire study region (Table 2). The interaction between
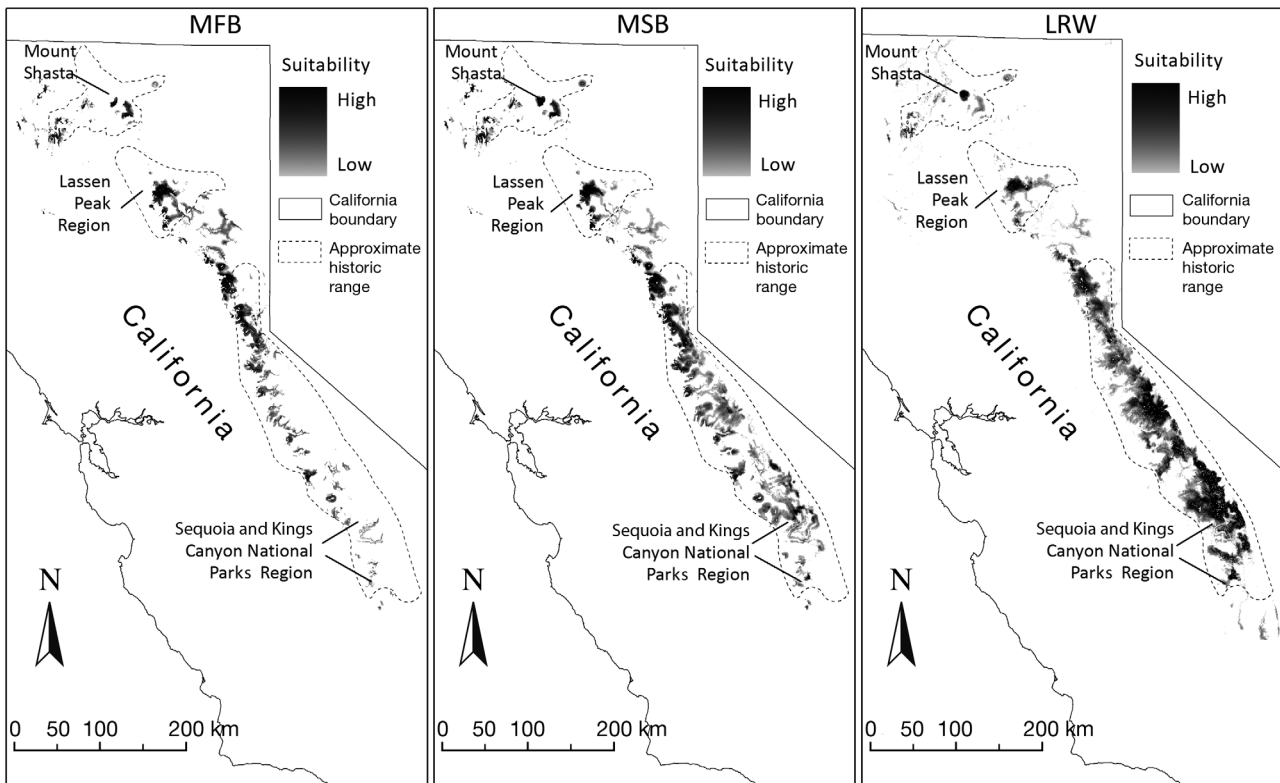

Fig. 2. *Vulpes vulpes necator*. Predicted suitable habitat for Sierra Nevada red fox based on 3 models: Maxent full background pixels (MFB), Maxent with subsetted background pixels (MSB), and spatially-weighted logistic regression (LRW). The dashed line represents the historic range (Grinnell et al. 1937). Figs. 3 to 5 display in detail the 3 geographic regions identified in this map

Table 2. Range, mean, and SD of environmental variables (minimum temperature in December, precipitation in February, greenness, and distance to nearest body of water) for the entire study area and for suitable habitat areas for Sierra Nevada red fox *Vulpes vulpes necator* as predicted by each of the 3 models. MFB: Maxent with full background pixels; MSB: Maxent with subset background pixels; LRW: spatially-weighted logistic regression

Environmental variables (units)	Study area		MFB		MSB		LRW	
	Min/Max	Mean ± SD	Min/Max	Mean ± SD	Min/Max	Mean ± SD	Min/Max	Mean ± SD
MIN.TEMP. DEC (°C)	−15.3/5.2	−3.9 ± 3.4	−10.4/−2.4	−5.3 ± 1.0	−15.3/−3.6	−6.8 ± 2.3	−15.3/5.3	−7.4 ± 2.6
FEB.PRECIP (mm)	26.8/481.6	183.3 ± 80.5	137.7/471.3	240.7 ± 64.0	147.7/471.3	239.3 ± 52.8	34.3/471.3	223.9 ± 57.6
GREENNESS (Derived)	−2266/1205	52 ± 247	−1906/1030	39 ± 217	−2134/881	−125 ± 375	−2134/896	−168.6 ± 356.3
DIST.TO WATER (m)	0/23 350	3912 ± 3402	0/10 331	2051 ± 1483	0/10 657	2038 ± 1441	0/13 829.3	2715.6 ± 2110.2

December minimum temperature and February precipitation accounted for 50% of the predictor variable contribution to the Maxent models (Table 3). The LRW model also included this interaction and predicted areas with a lower minimum temperature than the study region as a whole. But unlike the Maxent models, the LRW model had a February winter precipitation range similar to that of the entire study area, which included areas that received lower precipitation (Table 2). The remaining environmental variable summary statistics were comparable across the 3 models: suitable habitat was within 2 to 3 km of a water feature, and greenness values were lower than the study area mean (Table 2). These variables represented 12 and 2%, respectively, of the predictor variable contribution in the Maxent models (Table 3).

DISCUSSION

We used presence-absence (logistic regression) and presence-only (maximum entropy) methods to create the first spatially explicit habitat suitability model, based on climatic and remotely sensed variables, for the Sierra Nevada red fox, a threatened species under the California Endangered Species Act. The resulting model predictions can guide future surveys to locate additional populations of this rare subspecies, and also illustrate the relative performance of these modeling approaches. Our analysis is one of the first empirical tests of the revised algorithm in Maxent to reduce the effect of geographically biased occurrence data and improve its ability to transfer to new study regions (Phillips 2008).

Predictor variables

Our findings support the conclusion by Laurent et al. (2005) that wildlife may have a response to the continuous environmental gradient present in unclassified spectral data that are not captured in predetermined classification schemes. In our study, image greenness emerged as a predictor variable over 3 predetermined classification schemes: tree density, tree size class, and CWHR category. This suggests that the vegetation associations of the Sierra Nevada red fox are better characterized by the unclassified spectral data than by these predetermined classification schemes. This finding is significant because few studies leverage the benefits of remotely-sensed data in their species distribution modeling (Turner et al. 2003, Gillespie et al. 2008). In addition, predetermined classification schemes are often not available, are time consuming and expensive to produce, and are often inconsistent between regions. Not having to rely on vegetation maps for habitat modeling can greatly reduce needed resources because satellite imagery is often free and readily available. Additionally, using satellite imagery instead of predetermined classification schemes creates a more parsimonious model by decreasing the number of variables needed.

Both the logistic regression and maximum entropy modeling approaches indicated that climate was a major component of habitat suitability for the Sierra Nevada red fox. Recent phylogenetic analyses have indicated that the Sierra Nevada red fox and its conspecifics in the Cascade and Rocky Mountains (*Vulpes vulpes cascadensis* and *V. v. macroura*, respectively) comprise a distinct genetic lineage separate from much of the rest of North America (Aubry et al. 2009). This lineage was more widespread during the height of the Pleistocene glaciation but retracted to the high elevations of the western mountains when the glaciers retreated. These historic range expansions and contractions coincide with regional climate change, indicating that the Sierra Nevada red fox may be physiologically and ecologically constrained to subalpine climate zones (Aubry et al. 2009).

Both modeling approaches identified areas with low winter minimum temperatures, but the models varied

Table 3. Variable coefficients for logistic regression (LR) model and percent contributions for Maxent models. See Table 2 for definitions of variables and units. na: not applicable

Variable	LR: coefficient	Maxent: % contribution
Constant	0.3067	na
FEB.PRECIP	−0.0001483	30.4
MIN.TEMP.DEC	0.008398	5.3
GREENNESS	−0.00483	2.0
DIST.TO WATER	−0.0003856	11.9
Interaction: FEB.PRECIP × MIN.TEMP.DEC	−0.000000636	50.3
Interaction: FEB.PRECIP × GREENNESS	0.000000144	0.1

in how they represented the effect of precipitation. The Maxent models predicted areas with winter precipitation well above the regional minimum, whereas the LRW model predicted areas with lower winter precipitation (Table 2). This likely accounted for much of the spatial difference between the model predictions. Within the Lassen Peak region, for example, the LRW model predicted more suitable habitat than the Maxent models, and predicted more habitat in the eastern portion of the area (Fig. 3). Field surveys occasionally detected red fox east of Lassen Peak, but far less frequently than in the western portion of the area region, where the projections of the Maxent models were concentrated (Perrine 2005).

Outside of the Lassen Peak region, the differences in prediction area were exacerbated, both on the scale of the species' entire range and at a more localized scale of the original population centers of the Sierra Nevada red fox as described by Grinnell et al. (1937). Although all 3 models predicted high habitat suitability throughout the historic range of the Sierra Nevada red fox, the MSB model predicted slightly more habitat than the MFB model, and the LRW model predicted far more suitable habitat than either Maxent model. For example, north of Lassen Park, the LRW model predicted all of Mount Shasta to be suitable habitat, whereas the MSB model selected Shasta's eastern slope and a small portion of its peak and western slope, and the MFB model selected only its eastern slope (Fig. 4). The discrepancy became more pronounced with increased distance from Lassen Peak, the location of the occurrence data. In the southernmost historic population center, the Sequoia and Kings Canyon National Parks region of the southern Sierra Nevada, the LRW identified most of the region as suitable habitat (196 855 ha; Fig. 5). In contrast, the MFB predicted only a small amount of suitable habitat (23 234 ha), and the MSB model predicted an intermediate amount (84 496 ha). The pattern of Maxent and logistic regression models yielding similar AUC values but predicting slightly different suitable habitat areas is consistent with prior findings (Gibson et al. 2007).
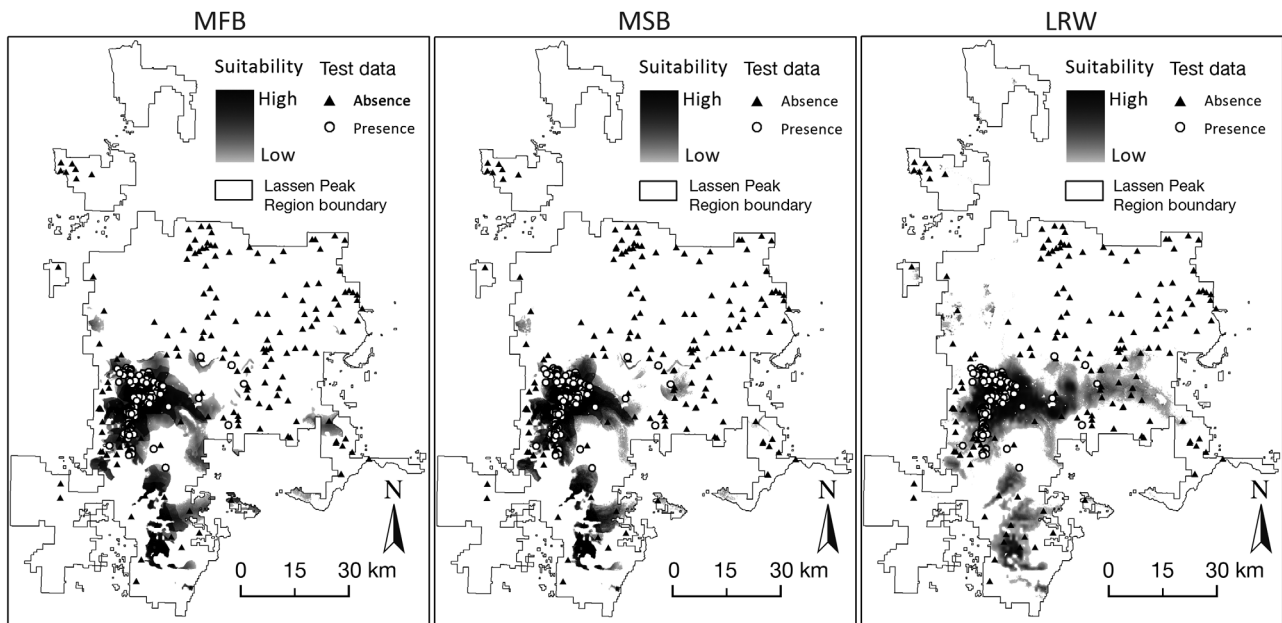

Fig. 3. *Vulpes vulpes necator.* Predicted suitable habitat for Sierra Nevada red fox in the Lassen Peak region, relative to presence-absence test data. MFB: Maxent with full background pixels; MSB: Maxent with subset background pixels; LRW: spatially-weighted logistic regression. See Fig. 2 for the location of the Lassen Peak region in California
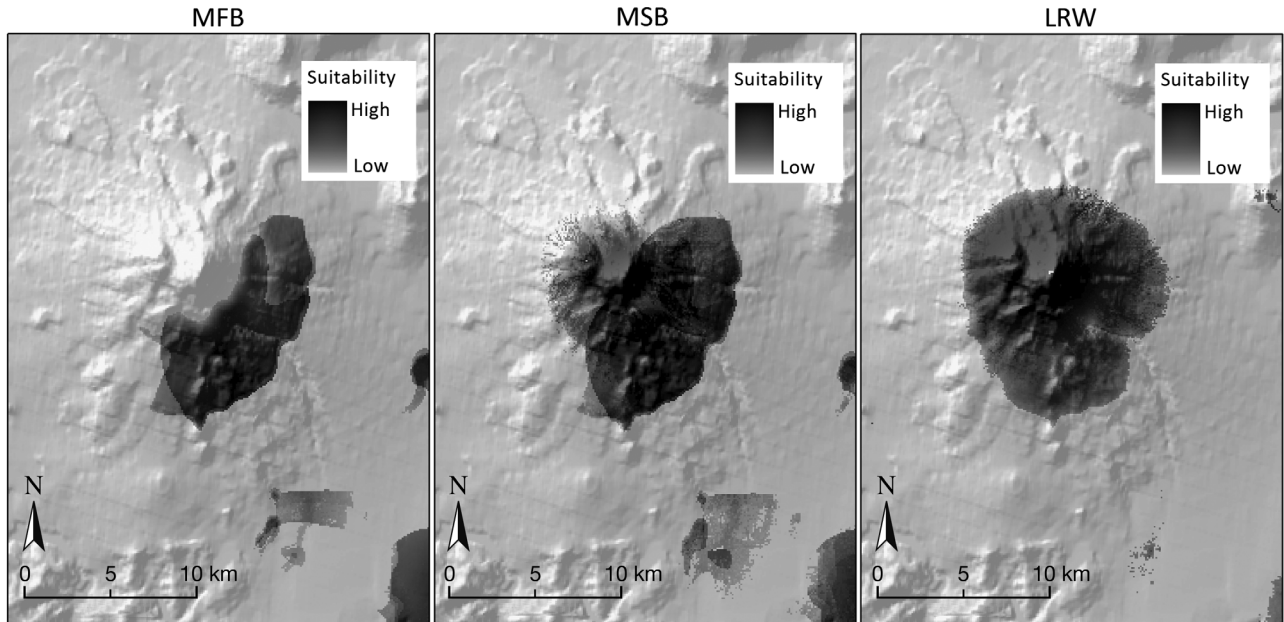
Fig. 4. *Vulpes vulpes necator.* Predicted suitable habitat for Sierra Nevada red fox in the Mount Shasta region of northern California. MFB: Maxent with full background pixels; MSB: Maxent with subset background pixels; LRW: spatially-weighted logistic regression. See Fig. 2 for the location of Mount Shasta in California
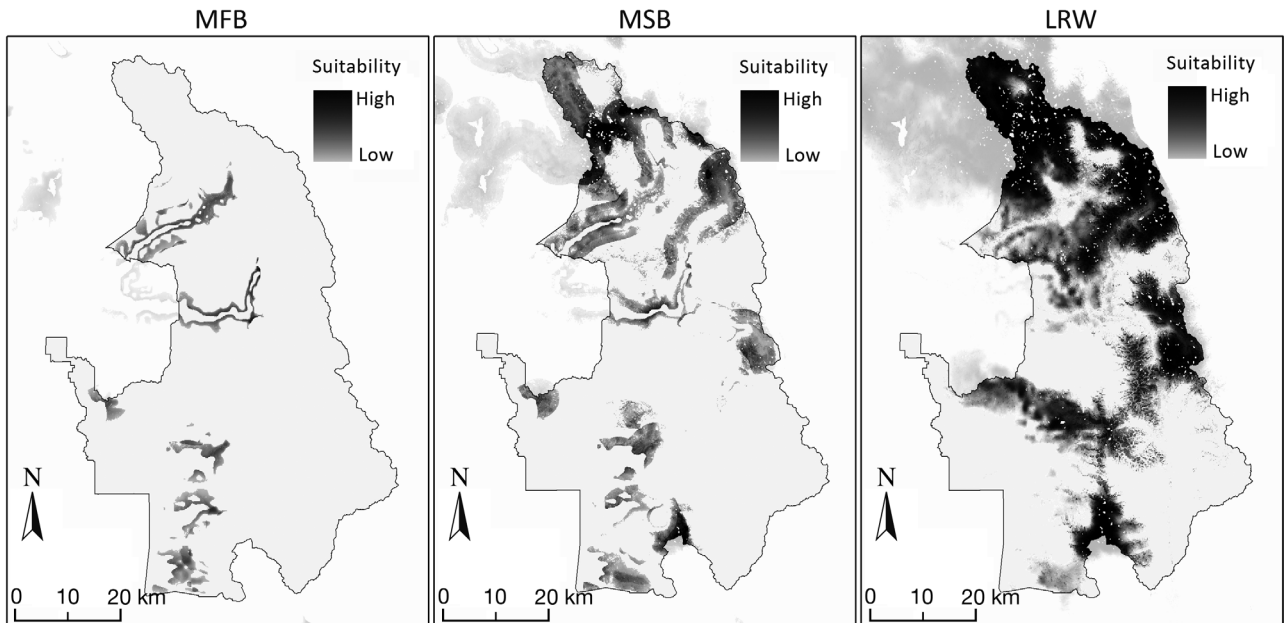

Fig. 5. *Vulpes vulpes necator.* Predicted suitable habitat for Sierra Nevada red fox in the Sequoia and Kings Canyon National Parks region in the southern Sierra. MFB: Maxent with full background pixels; MSB: Maxent with subset background pixels; LRW: spatially-weighted logistic regression. See Fig. 2 for the location of the Sequoia and Kings Canyon National Parks region in California

Because evaluation data were available only within the Lassen Peak region, our assessment of model performance beyond this region was limited to comparisons between models and with the species' historic range (Grinnell et al. 1937). The MSB model, which used subsetted background pixels to match the biases in the occurrence data, yielded a prediction area and extent that was intermediate between the MFB model and the LRW model (Fig. 2). Although the MFB model yielded high accuracies within the occurrence data region, it predicted less suitable habitat in areas farther away from the occurrence data than the MSB model.

These results are consistent with general expectations for geographically biased presence data (Phillips 2008). The MSB model, which used a subset of background pixels, greatly reduced the effect of geographic bias. In contrast to the Maxent models, the LRW model predicted abundant suitable habitat outside of the Lassen region, confirming prior findings that logistic regression models can transfer well to similar study areas (Randin et al. 2006, Barbosa et al. 2009).

Our range predictions for the Sierra Nevada red fox, like all species distribution models, are hypotheses that should be tested by the collection of new data, especially in predicted areas that were previously unsampled (Olivier & Wotherspoon 2006, Parra & Monahan 2008, Wollan et al. 2008, Costa et al. 2009). The increased use of automatic camera stations for carnivore inventory and monitoring on public lands throughout the Sierra Nevada holds great promise for additional detections of Sierra Nevada red fox. Fortuitously, in autumn 2010 Forest Service biologists obtained photographs of several putative Sierra Nevada red foxes in the Sonora Pass region of the northern Sierra Nevada, approximately 75 km south of Lake Tahoe (A. Rich and S. Lisius pers. comm.). Although final confirmation via molecular genetic methods is still underway, preliminary analyses and the fact that several foxes were detected in close proximity indicate that these animals represent the first verified population of Sierra Nevada red fox detected out of the Lassen area in nearly 2 decades (B. Sacks pers. comm.). These detections lend further support to the logistic regression model over the 2 Maxent models. The logistic regression identified these detection sites as high quality habitat for Sierra Nevada red fox; the subalpine woodland is virtually identical structurally to occupied Sierra Nevada red fox habitat in the Lassen region, despite being composed of whitebark pine *Pinus albicaulus* as opposed to mountain hemlock in the Lassen region. The Maxent model with subsetted background pixels identified the Sonora Pass site as medium-quality habitat for Sierra Nevada red fox, whereas the original Maxent algorithm did not identify it as potential habitat at all.

The ability to subset background pixels is relatively new to Maxent (Phillips 2008), and few studies to date have used or evaluated these methods (Anderson & Raza 2010). Our findings support the conclusion of Anderson & Raza (2010) that calibrating the study region or subsetting background pixels to the occurrence data area yields a Maxent model with a larger predicted area that is less concentrated around the occurrence data region. Our study emphasizes the importance of subsetting background pixels in Maxent when using geographically biased presence data, which are common with rare species. Moreover, our results indicate that while subsetting significantly reduces the effects of geographically biased presence data in Maxent, it does not completely resolve the problem. Despite the risk of false absences, the LRW model did quite well predicting both within and outside of the Lassen Peak region, leading us to conclude that in this case the potentially false absences were less problematic than the transferability issues inherent to the Maxent models.

Management implications

As a carnivore closely associated with montane tree-line habitats, the Sierra Nevada red fox may be acutely affected by climate changes. Since 1920, California's average annual temperature has warmed by 1.7°C, with greater warming occurring with daily minimum temperatures and at higher elevations (Kapnick & Hall 2009). In the central Sierra Nevada, December minimum temperatures have increased by 0.23°C decade^{-1} in Yosemite Valley (1220 m elevation) and 0.32°C decade^{-1} at Lake Tahoe (1900 m; Thorne et al. 2006). Recent surveys in the Sierra Nevada have documented range shifts by small mammals (Moritz et al. 2008), birds (Tingley et al. 2009), butterflies (Forister et al. 2010), and conifers (Thorne et al. 2006), consistent with responses to climate warming. Despite the recent detection of at least 1 Sierra Nevada red fox in the northern Sierra, the lack of recent documented detections in the southern extent of its historic range (Perrine et al. 2010) may indicate that its range has retracted northward in response to climate change. It is unclear whether climate has a direct or indirect impact, such as facilitating coyotes *Canis latrans* or other competitors or changing understory structure through altered fire regimes.

If such range retraction has already occurred, the Sierra Nevada red fox may have little future in California. Climate warming is expected to continue if not accelerate in the coming century, although forecasts of the amount and rate of change depend greatly on the specific climate model and emissions scenario used (Hayhoe et al. 2004). Nevertheless, montane regions are likely to experience the greatest warming (Snyder et al. 2002), with mean annual temperatures in the Sierra Nevada and southern Cascades projected to rise 3.0 to 3.5°C by 2070 to 2099 (Ackerly et al. 2010). These changes will likely have a profound effect on the ranges, elevations, and associations of California's biota (e.g. Loarie et al. 2008, Parra & Monahan 2008, Wiens et al. 2009, Ackerly et al. 2010, Forister et al. 2010) The Sierra Nevada red fox may follow the pattern of another alpine-associated carnivore, the wolverine *Gulo gulo*, which disappeared from Califor-

nia by the 1930s, with populations persisting only in higher latitude states such as Washington, Montana, and Idaho (Aubry et al. 2007). Predicting the future range of Sierra Nevada red fox based upon its historic range and the anticipated climate changes is an important next step toward its effective conservation.

Locating additional Sierra Nevada red fox populations outside of the Lassen Peak area is critically important (Perrine et al. 2010) to better document the taxon's true spatial extent and to acquire additional specimens to refine its phylogenetic relationships with other red fox populations (e.g. Perrine et al. 2007, Aubry et al. 2009, Sacks et al. 2010). One of our primary goals for these analyses was to provide guidance for range-wide field surveys. To test our model predictions, field surveys should target areas of model discrepancy, such the Sequoia and Kings Canyon area of the southern Sierra Nevada (Fig. 5). However, surveys attempting to locate additional populations as efficiently as possible should instead target areas of model agreement, as indicated by our 2-model ensemble (Fig. 6). The highest survey priority should be to determine the extent of the newly discovered Sonora Pass population. Additional high priority areas
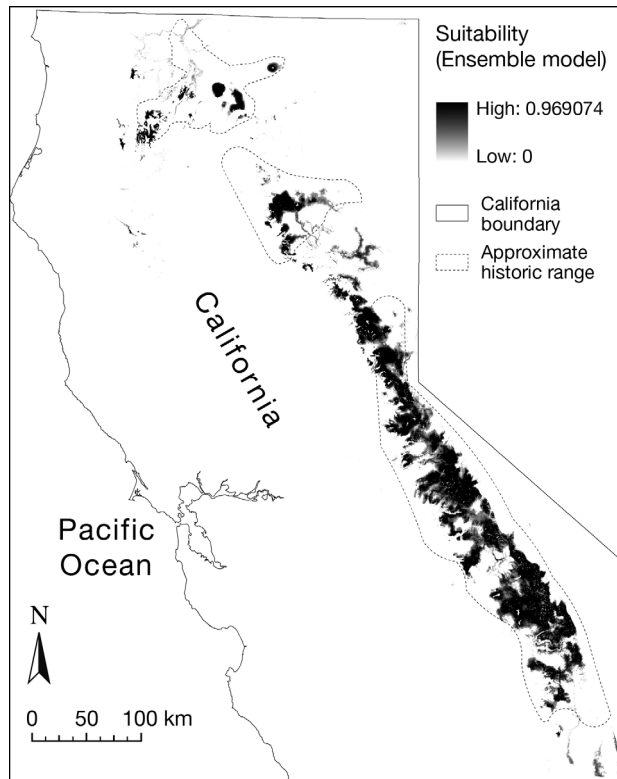
include the Mount Shasta region (Fig. 4), which is the largest predicted area north of the extant Lassen population, and the Sequoia and Kings Canyon region, which is the southernmost of the historic population centers (Grinnell et al. 1937) and an area of profound model discrepancy. Such surveys should incorporate the collection of specimens for genetic analysis because photographs alone cannot conclusively identify Sierra Nevada red foxes (Perrine et al. 2010). Additionally, the Lassen population should be closely monitored for changes in size or extent, as it remains the only known actively reproducing population of this endangered taxon (Perrine et al. 2010). If no other reproductive populations exist outside of the Lassen region, then the Sierra Nevada red fox likely warrants a higher level of state or federal protection and active management than it currently receives.

Fig. 6. *Vulpes vulpes necator.* Ensemble habitat suitability model for Sierra Nevada red fox, based on spatially-weighted logistic regression (LRW) and Maxent with subset background (MSB) models

LITERATURE CITED

Ackerly DD, Loarie SR, Cornwell WK, Weiss SB, Hamilton H, Branciforte R, Kraft NJB (2010) The geography of climate change: implications for conservation biology. Divers Distrib 16:476–487

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J Appl Ecol 43:1223–1232

Anderson RP, Raza A (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. J Biogeogr 37:1378–1393

Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. J Biogeogr 33:1677–1688

Araújo MB, New M (2007) Ensemble forecasting of species distributions. Trends Ecol Evol 22:42–47

Aubry KB, McKelvrey KS, Copeland JP (2007) Distribution and broadscale habitat relations of the wolverine in the contiguous United States. J Wildl Manag 71:2147–2158

Aubry KB, Statham MJ, Sacks BN, Perrine JD, Wisely SM (2009) Phylogeography of the North American red fox: vicariance in Pleistocene forest refugia. Mol Ecol 18: 2668–2686

Barbosa AM, Raimundo R, Vargas JM (2009) Transferability of environmental favourability models in geographic space: the case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. Ecol Model 220:747–754

Buckland ST, Burnham KP, Augustin NH (1997) Model selection: an integral part of inference. Biometrics 53:603–618

Bulluck L, Fleishman E, Betrus C, Blair R (2006) Spatial and temporal variations in species occurrence rate affect the accuracy of occurrence models. Glob Ecol Biogeogr 15: 27–38

Burnham DP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York, NY

CDFG (California Department of Fish and Game) (1996) The status of rare, threatened and endangered animals and plants of California. California Department of Fish and Game, Sacramento, CA

CDFG (2004) Special animals. California Department of Fish and Game, Sacramento, CA

Costa GC, Nogueira C, Machado RB, Colli GR (2009) Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. Biodivers Conserv 19:883–899

Daly CS, Neilson RP, Phillips DL (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J Appl Meteorol 33:140–158

Edwards TC, Cutler DR, Zimmermann NE, Geiser L, Alegria J (2005) Model-based stratifications for enhancing the detection of rare ecological events. Ecology 86:1081–1090

Elith J, Graham C, Anderson R, Dudik M and others (2006) Novel methods to 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

Ferreira de Siqueira M, Durigan G, de Marco Júnior P, Peterson AT (2009) Something from nothing: using landscape similarity and ecological niche modeling to find rare plant species. J Nat Conserv 17:25–32

Forister ML, McCall AC, Sanders NJ, Fordyce JA and others (2010) Compounded effects of climate change and habitat alteration shift patterns of butterfly diversity. Proc Natl Acad Sci USA 107:2088–2092

Gibson L, Barrett B, Burbidge A (2007) Dealing with uncertain absences in habitat modeling: a case study of a rare ground-dwelling parrot. Divers Distrib 13:704–713

Gillespie TW, Foody GM, Rocchini D, Giorgi AP, Saatchi S (2008) Measuring and modeling biodiversity from space. Prog Phys Geogr 32:203–221

Grinnell J, Dixon J, Linsdale J (1937) Fur-bearing mammals of California. University of California Press, Berkeley, CA

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135:147–186

Guisan A, Broenniamann O, Engler R, Vust M, Yoccoz N, Lehmann A, Zimmermann N (2006) Using niche-based models to improve the sampling of rare species. Conserv Biol 20:501–511

Hall ER (1981) The mammals of North America, 2nd edn. John Wiley & Sons, New York, NY

Hansen MJ, Franklin SE, Woudsma C, Peterson M (2001) Forest structure classification in the North Columbia mountains using Landsat TM tasseled cap wetness component. Can J Rem Sens 27:20–32

Hastie TJ, Pregibon D (1992) Generalized linear models. In: Chambers JM, Hastie TJ (eds) Statistical models in S. Wadsworth & Brooks/Cole, Pacific Grove, CA, p 195–248

Hayhoe K, Cayan D, Field CB, Frumhoff PC and others (2004) Emissions pathways, climate change, and impacts on California. Proc Natl Acad Sci USA 101:12422–12427

Hosmer DW, Lemeshow S (2000) Applied logistic regression. John Wiley & Sons, New York, NY

Johnson CJ, Seip DR, Boyce MS (2004) A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales. J Appl Ecol 41:238–251

Jones CC, Acker SA, Halpern CB (2010) Combining local- and large-scale models to predict the distributions of invasive plant species. Ecol Appl 20:311–325

Kapnick S, Hall A (2009) Observed changes in the Sierra Nevada snowpack: potential causes and concerns. CEC-500-2009-016-F, California Energy Commission, Sacramento, CA

Laurent E, Shi H, Gatziolis D, LeBouton J, Walters M, Liu J (2005) Using spatial and spectral precision of satellite imagery to predict wildlife occurrence patterns. Remote Sens Environ 97:249–262

Loarie SR, Carter BE, Hayhoe K, McMahon S, Moe R, Knight CA, Ackerly DD (2008) Climate change and the future of California's endemic flora. PLoS ONE 3:e2502

MacKenzie D, Nichols J, Royle JA, Pollock K, Bailey L, Hines J (2006) Occupancy estimation and modeling. Academic Press, London

Mayer KE, Laudenslayer WF Jr (eds) (1988) A guide to wildlife habitats of California. California Department of Fish and Game, Sacramento, CA

Menon S, Choudhury B, Khan ML, Peterson AT (2010) Ecological niche modeling and local knowledge predict new populations of *Gymnocladus assamicus* a critically endangered tree species. Endang Species Res 11: 175–181

Miller J, Franklin J (2002) Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. Ecol Model 157:227–247

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

Moritz C, Patton JL, Conroy CJ, Parra JL, White GC, Beissinger SR (2008) Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. Science 322:261–264

Nams VO (2001) Locate II: user's guide. Pacer, Truro, NS

Nielsen SE, Boyce MS, Stenhouse GB, Robin H, Munro M (2002) Modeling grizzly bear habitats in the Yellowhead Ecosystem of Alberta: taking autocorrelation seriously. Ursus 13:45–56

Olivier F, Wotherspoon SJ (2006) Distribution and abundance of Wilson's storm petrels *Oceanites oceanicus* at two locations in East Antarctica: testing habitat selection models. Polar Biol 29:878–892

Parra JL, Monahan WB (2008) Variability in 20th century climate change reconstructions and its consequences for predicting geographic responses of California mammals. Glob Change Biol 14:2215–2231

Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. Ecol Model 133:225–245

Perrine JD (2005) Ecology of red fox (*Vulpes vulpes*) in the Lassen Peak region of California, USA. PhD dissertation, University of California, Berkeley, CA

Perrine JD, Pollinger J, Sacks B, Barrett R, Wayne R (2007) Genetic evidence for the persistence of the critically endangered Sierra Nevada red fox in California. Conserv Genet 8:1083–1095

Perrine JD, Campbell LA, Green GA (2010) Sierra Nevada red fox (*Vulpes vulpes necator*): a conservation assessment. R5-FR-010. USDA Forest Service, Vallejo, CA

Peterson AT, Papes M, Eaton M (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography 30:550–560

Phillips SJ (2008) Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). Ecography 31:272–278

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190:231–259

Posillico M, Meriggi A, Pagnin E, Lovari S, Russo L (2004) A habitat model for brown bear conservation and land use planning in the central Apennines. Biol Conserv 118: 141–150

R Development Core Team (2005) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available at www.R-project.org

Rachlow JL, Svancara LK (2006) Prioritizing habitat for surveys of an uncommon mammal: a modeling approach applied to pygmy rabbits. J Mammal 87:827–833

Randin CF, Dirnbo T, Dullinger S, Zimmermann NE, Zappa M, Guisan A (2006) Are niche-based species distribution models transferable in space? J Biogeogr 33:1689–1703

Rebelo H, Jones G (2010) Ground validation of presence-only modelling with rare species: a case study on barbastelles *Barbastella barbastellus* (Chiroptera: Vespertilionidae). J Appl Ecol 47:410–420

Rhynsburger D (1973) Analytic delineation of Thiessen polygons. Geogr Anal 5:133–144

Sacks BN, Statham MJ, Perrine JD, Wisely SM, Aubry KB (2010) North American montane red foxes: expansion, fragmentation, and the origin of the Sacramento Valley red fox. Conserv Genet 11:1523–1539

Schempf PF, White M (1977) Status of six furbearer populations in the mountains of northern California. USDA Forest Service, California Region, San Francisco, CA

Snyder MA, Bell JL, Sloan LC, Duffy PB, Govindasamy B (2002) Climate responses to a doubling of atmospheric carbon dioxide for a climatically vulnerable region. Geophys Res Lett 29:1514 doi:10.1029/2001GL014431

St-Louis V, Pidgeon AM, Radeloff VC, Hawbaker TJ, Clayton MK (2006) High-resolution image texture as a predictor of bird species richness. Remote Sens Environ 105: 299–312

Stohlgren TJ, Ma P, Kumar S, Rocca M, Morisette JT, Jarnevich CS, Benson N (2010) Ensemble habitat mapping of invasive plant species. Risk Anal 30:224–235

Thorne JH, Kelsey R, Honig J, Morgan B (2006) The development of 70-year-old Weislander vegetation type maps and an assessment of landscape change in the central Sierra Nevada. CEC 500-2006-107, California Energy Commission, PIER Energy-related Environmental Program, Sacramento, CA

Tingley MW, Monahan WB, Beissinger SR, Moritz C (2009) Birds track their Grinnellian niche through a century of climate change. Proc Natl Acad Sci USA 106 Suppl 2: 19637–19643

Turner W, Spector S, Gardiner N, Fladeland M, Sterling E, Steininger M (2003) Remote sensing for biodiversity science and conservation. Trends Ecol Evol 18:306–314

USGS (US Geological Survey) (2000) 10 m digital elevation model. Menlo Park, CA

USDA-NRCS (US Department of Agriculture-Natural Resources Conservation Service) (2006) Major Land Resource Area (MLRA). Land resource regions and major land resource areas of the United States, the Caribbean, and the Pacific Basin. US Department of Agriculture Handbook 296

Waring RH, Running SW (1998) Forest ecosystems: analysis at multiple scales, 2nd edn. Academic Press, San Diego, CA

White J, Running S, Nemani R, Keane R, Ryan K (1997) Measurement and remote sensing of LAI in Rocky Mountain montane ecosystems. Can J Res 27:1714–1727

Wiens JA, Stralberg D, Johgsomjit D, Howell CA, Snyder MA (2009) Niches, models, and climate change: assessing the assumptions and uncertainties. Proc Natl Acad Sci USA 106 (Suppl 2):19729–19736

Wollan AK, Bakkestuen V, Kauserud H, Gulden G, Halvorsen R (2008) Modeling and predicting fungal distribution patterns using herbarium data. J Biogeogr 35: 2298–2310

Zielinski WJ, Kucera TE (1995) American marten, fisher, lynx, and wolverine: survey methods for their detection. PSW-GTR-157. USDA Forest Service, Pacific Southwest Research Station, Albany, CA

Zielinski WJ, Truex R, Schlexer F, Campbell L, Carroll C (2005) Historical and contemporary distributions of carnivores in forests of the Sierra Nevada, California, USA. J Biogeogr 32:1385–1407