Ecography 32: 907–918, 2009
doi: 10.1111/j.1600-0587.2009.05883.x

Subject Editor: Miguel Araujo. Accepted 1 September 2009

Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors

Alexandra D. Syphard and Janet Franklin

A. D. Syphard (asyphard@consbio.org), Dept of Biology, San Diego State Univ., San Diego, CA 92182-4614, USA. (Present address of A. D. S.: Conservation Biology Inst., 10423 Sierra Vista Avenue, La Mesa, CA 91941, USA.) – J. Franklin, School of Geography and Urban Planning, Arizona State Univ., Tempe, AZ 85287, USA.

Prediction maps produced by species distribution models (SDMs) influence decision-making in resource management or designation of land in conservation planning. Many studies have compared the prediction accuracy of different SDM modeling methods, but few have quantified the similarity among prediction maps. There has also been little systematic exploration of how the relative importance of different predictor variables varies among model types and affects map similarity. Our objective was to expand the evaluation of SDM performance for 45 plant species in southern California to better understand how map predictions vary among model types, and to explain what factors may affect spatial correspondence, including the selection and relative importance of different environmental variables. Four types of models were tested. Correlation among maps was highest between generalized linear models (GLMs) and generalized additive models (GAMs) and lowest between classification trees and GAMs or GLMs. Correlation between Random Forests (RFs) and GAMs was the same as between RFs and classification trees. Spatial correspondence among maps was influenced the most by model prediction accuracy (AUC) and species prevalence; map correspondence was highest when accuracy was high and prevalence was intermediate (average prevalence for all species was 0.124). Species functional type and the selection of climate variables also influenced map correspondence. For most (but not all) species, climate variables were more important than terrain or soil in predicting their distributions. Environmental variable selection varied according to modeling method, but the largest differences were between RFs and GLMs or GAMs. Although prediction accuracy was equal for GLMs, GAMs, and RFs, the differences in spatial predictions suggest that it may be important to evaluate the results of more than one model to estimate the range of spatial uncertainty before making planning decisions based on map outputs. This may be particularly important if models have low accuracy or if species prevalence is not intermediate.

Spatial prediction of species' geographic distributions has become a fundamental component of conservation planning, resource management, and environmental decision-making. Therefore, methodological issues related to species distribution models (SDMs) have been the focus of much discussion in the recent scientific literature. SDMs are quantitative, predictive models of the species–environment relationship that correlate observations of species occurrence or abundance with mapped environmental variables to make spatial predictions of habitat suitability or species occurrence (Franklin 1995, Guisan and Zimmermann 2000, Scott et al. 2002, Guisan et al. 2006). The methods are based on the assumption that species' distributions are correlated with environmental gradients represented by landscape variables that are distally or proximally related to their physiological tolerances or resource requirements, and thus, their realized niches (Austin 2002).

In recent years, a growing number of modeling methods has been applied to improve the performance and ecological validity of SDM, and the different approaches vary in terms of their complexity, assumptions, data requirements, and ease of use. The recent SDM literature has emphasized comparison of these different model types to better understand their relative differences in performance (Bio et al. 1998, Franklin 1998, Moisen and Frescino 2002, Segurado and Araújo 2004, Elith et al. 2006, Maggini et al. 2006, Guisan et al. 2007). The majority of these comparisons have focused on prediction accuracy as a measure of model performance in which one or more standard metrics are applied. For categorical prediction accuracy ("threshold-dependent"), common metrics include Kappa, Sensitivity, or Specificity. Alternatively, the area under the curve (AUC) of receiver-operating characteristic (ROC) plots (Fielding and Bell 1997) is a particularly useful metric for model comparison because it avoids the need to choose a threshold probability that separates "suitable" from "unsuitable" (or presence from absence) (i.e. it is "threshold-independent"). The AUC is also widely used because it describes the overall ability of the model to discriminate between two cases (but see Lobo et al. 2008).

While metrics like AUC are important components of model performance evaluation, there has been less emphasis in the literature on other methods of comparing and evaluating models. In particular, few studies have systematically explored the similarity among maps predicted by different model types. Yet, in many applications, the maps produced by SDMs are the key outputs that influence decision-making or designation of land, for example, for nature reserves (Mladenoff et al. 1995, Johnson et al. 2004). Therefore, measuring the amount of map overlap among predictions may provide important information about the strengths and limitations of different model types that may not be apparent from global measures such as AUC. For example, models that demonstrate equally high accuracy when assessed with test data could yield incongruent maps because the models use different assumptions, algorithms, and parameterizations.

Some studies have qualitatively compared predictive distribution maps resulting from different modeling methods and discussed their differences in the context of extent-based (non-spatial) accuracy measures, e.g. the tendency of some methods to over-predict or under-predict distributions (Loiselle et al. 2003, Elith et al. 2006). Other studies have shown large variation among projections of species' future ranges under various climate change scenarios from different models (Thuiller et al. 2004a, b, Araújo et al. 2005a, b, Araújo and New 2007), or for invasive species introduced into new regions (Crossman and Bass 2008, Kelly et al. 2008), and these studies emphasized ensemble forecasting (i.e. combining predictions across multiple models) to address the spatial uncertainty associated with these projections.

The studies that have quantitatively compared prediction maps from different models have shown that spatial predictions may vary considerably depending on the method and other modeling decisions. Prasad et al. (2006) concluded that maps produced using ensemble statistical learning methods (e.g. Random Forests), were more similar to each other (and more realistic) than to those produced using single models. In another study, global statistical models, such as generalized linear models (GLMs) and generalized additive models (GAMs), produced maps that were more similar to each other than they were to local models, that is, statistical methods that allow model parameters to vary spatially (Osborne and Suarez-Seoane 2007).

Although no studies have explicitly focused on explaining spatial divergence in map predictions, Johnson and Gillingham (2005) suggested that discrepancies in predicted maps from four different methods may have been due to differences in the predictor variable sets used to build the models. Likewise, Thuiller et al. (2004a, b) proposed that discrepancies in model projections may be related to differences in the ways that model types make predictions under different environmental conditions. Hernandez et al. (2006) also showed that the spatial prediction of suitable habitat varied depending on the number of observations available to train the model.

As with comparison of prediction maps, there has been little systematic exploration of environmental variable selection, and the relative importance of different predictor variables, among model types. The selection of environmental predictor variables in SDM is often a function of the scale of the analysis; but in general, the predictors describing the physical environment often fall into three classes: 1) climate, 2) terrain, and/or 3) substrate or landform (Franklin 1995, Mackey and Lindenmayer 2001). The predictive power of SDMs at broad scales may not be substantially improved by including variables other than climate (Thuiller et al. 2004a, b), but terrain and geological variables related to direct and resource gradients, may be more important at finer, landscape scales (Franklin 1995); in many cases, a combination of climate and edaphic factors may produce the best models (Iverson and Prasad 1998). One recent meta-analysis found that those models that included environmental predictors from multiple, hierarchical scales yielded the most accurate predictions (Meyer and Thuiller 2006).

While many individual SDM studies describe the correlations between predictors and species occurrence, i.e. the relative importance of different predictors, there is still little in the way of general guidelines about the relative importance of, e.g. climate, terrain and edaphic variables within and among model types and for different taxa. Peterson and Nakazawa (2008) showed that, when using one model type (GARP), spatial predictions of native and introduced distributions of fire ants were sensitive to the environmental data sets used to develop the models. Considering this potential influence that different environmental data sets may have on spatial predictions, the authors called for further research on the topic.

Our objective in this study was to expand the evaluation of SDM model performance to better understand how mapped predictions may vary among model types, and to explain what factors may affect spatial correspondence. Furthermore, we evaluated the selection and relative importance of different environmental variables used to predict plant species distribution for 45 species in southern California using four types of models. We asked these questions.

1) Do different SDM modeling methods produce similar spatial predictions?

   We expected map correlation to be highest between similar model types, e.g. between those that used supervised, machine learning methods (classification trees, CT, and Random Forests, RF) and those model types that are extensions of linear multiple regression models (GLMs and GAMs).

2) How does the correlation of spatial predictions from different models vary in relation to prediction accuracy, species' prevalence, species' functional type, or type of environmental variables in the model?

   We expected spatial correspondence among maps to be highest when models had greater prediction accuracy. We also expected species that occurred over smaller extents on the map to have better map correspondence because prevalence (the proportion of species' presences in the training data) may be related to habitat specificity (Stockwell and Peterson 2002, Segurado and Araújo 2004, Luoto et al. 2005, Elith et al. 2006, McPherson and Jetz 2007), which could constrain the potential area to be predicted as suitable habitat. Because the realized niche is influenced by different spatial factors for different functional types (Syphard and Franklin in press), we also expected map correlation to vary among functional types. Finally, we expected higher map correspondence to occur when climate variables were selected as important because climate varies more slowly over space than terrain or soil variables and will tend to delimit the general outline of the species range (rather than delineate occupied habitat within that range).

3) Do different modeling methods select for different types of environmental variables?

Overall, we expected climate to be more important than terrain and soil for all methods because climate directly fulfills plant species' requirements for or tolerance to heat, moisture, and light regimes. We also expected RF and CTs to select soil order (a categorical variable) more than GLMs or GAMs because categorical predictors are well handled by decision-tree methods (Breiman et al. 1984).

Methods

The species' distribution models examined in this study were developed as part of a larger project and the study area, species data, environmental data, and modeling methods are described in detail elsewhere (Syphard and Franklin in press). They will be summarized briefly here.

Study area and species data

We compiled species data for 45 plant species that dominate the foothills and coastal plain of southern California. These species are typical of the Chaparral and Sage Scrub shrubland plant communities (Westman 1981, Schoenherr 1992, Hickman 1993, Keeley 2000) and represent a range of plant functional types. Species occurrence locations (presence-absence) were acquired for 1471 southern California shrubland locations (Taylor 2004) from a database (Vegetation Type Maps (VTM), <http://vtm.berkeley.edu/>) of vegetation plots surveyed in the 1930s (Wieslander 1935, Kelly et al. 2005, Barbour et al. 2007). These species were found in at least 30 plots (prevalence >0.02). Average prevalence was 0.124 (0.03–0.53).

Environmental predictors

We used eight climate, terrain, and soil variables as predictors. Three climate variables included mean annual precipitation, mean minimum January temperature, and mean maximum July temperature interpolated to 1-km grids from 1966 to 1995 climate station data (Franklin et al. 2001). Four terrain variables included winter and summer solar radiation, Topographic Moisture Index (TMI), and slope. Terrain-distributed solar radiation (Dubayah and Rich 1995) was modeled from U.S. Geological Survey 30-m resolution digital elevation model (DEMs) using the Solar Analyst 1.0 extension for ArcView (ESRI, Redlands, CA, USA) Geographic Information System (GIS). Daily insolation was calculated for two single days, the summer and winter solstice (using site latitude of 33°N, sky size of 200, and 0.2 clear sky irradiance) and used to represent the seasonal extremes of radiation on the landscape to yield one summer and one winter radiation variable. The TMI represents relative soil moisture availability based on upslope catchment area and slope, which were derived from the DEM (Moore et al. 1991, Wilson and Gallant 2000). For the soil variable, we created a grid of soil order, a categorical variable, using the California State Soil Geographic Database (STATSGO).

Species' distribution models

We developed four models for each species using the following methods: generalized linear models (GLMs), generalized additive models (GAMs), classification trees (CTs), and Random Forests (RFs). GLMs in the form of logistic regression models are commonly used in species distribution modeling with species' presence/absence data (Guisan et al. 2002). Although GLMs allow for non-linear relationships to be accommodated using polynomial terms, they are nevertheless parametric models with distributions that do not always reflect complex species responses to the environment (Austin 2002, Austin et al. 2006). GAMs (Yee and Mitchell 1991) have been widely used in species distribution modeling as an alternative to GLMs (Lehmann et al. 2002) because global regression coefficients are replaced by local smoothing functions, allowing the structure of the data to determine the shape of the species response curves.

CTs are supervised classifiers that develop rules, based on binary recursive partitioning, that can be used to classify new observations (Breiman et al. 1984). CTs iteratively split a full data set into partitions and evaluate how well the rules that determine these splits can separate the data into homogeneous classes. Typically, CTs are partitioned until a split no longer achieves a certain level of homogeneity, and then they are "pruned" back so that the model does not over-fit the data and can provide robust predictions for new data. Classification trees easily handle categorical predictors and interactions between variables (which do not have to be specified a priori) (De'ath and Fabricius 2000). On the other hand, CTs can be unstable, that is, they may produce very different models if the inputs are slightly varied (Prasad et al. 2006). A newer ensemble modeling method, RFs, overcomes this instability by developing many (hundreds or thousands of) tree models using random subsets of the cases and the predictor variables and then averaging the predictions (Breiman 2001). Model error and variable importance for RF models are estimated via bootstrapping (Cutler et al. 2007).

Based on exploratory data analysis (Syphard and Franklin in press), we evaluated both linear and quadratic relationships for all the continuous variables in the GLMs and used three target degrees of freedom for smoothing splines in the GAMs. Backward stepwise variable selection has frequently been used in SDM (Wintle et al. 2005) and we used it here to provide a consistent and automated approach for selecting final GLMs and GAMs for all species, in spite of the acknowledged limitations of this approach (James and McCulloch 1990). We entered predictors in the following order: climate, terrain, then soil variables, based on their relative importance determined in preliminary analyses. We further refined GLMs by manually removing quadratic terms if their coefficients were positive, e.g. if the response curve was inverted. Although a response curve can theoretically be bimodal in the presence of competition (Austin and Smith 1989), we considered this fitted form (increasing probability of occurrence at extremely high and low values of a predictor) to be a poor approximation of a bimodal response, and one that produced predictions that were ecologically unrealistic (Austin 2002). Thus, we only retained the linear term for that predictor if it remained significant.

We tested spatial autocorrelation (SA) of model residuals for the GLMs because, among the model types used in this study, these global, parametric models are most susceptible to misspecification in the face of autocorrelation (Miller et al. 2007). Moran's I (Moran 1948) was calculated for lag = 4000 m. The distribution of nearest neighbor distances among the vegetation plots was 210–13 800 m (median 1600 m). Ninety percent of the plots had a nearest neighbor within 4000 m, and so 4000 m was examined as the lag distance. Monte Carlo simulation (1000 replicates) was used to estimate the significance of Moran's I because the residuals from a logistic regression are not normally distributed.

We built full classification trees for each species and then pruned using an algorithm that automatically selected the complexity parameter associated with the smallest cross-validated error. If this algorithm selected only one split, we increased the number of splits to two so our pruned tree would include at least two decision rules. For Random Forests models, we averaged the predictions from 500 trees. We evaluated three randomly selected variables for each tree based on the suggestion by Breiman (2001), that the square root of the number of variables gives optimum results.

The performance evaluation measure that we used to describe SDM prediction accuracy for each model was the area under the curve (AUC) of the receiver operating characteristic (ROC) plot (Hanley and McNeil 1982). ROC plots show the true positive predictions versus false positive predictions for all possible threshold values. Therefore, the AUC (ranging from 0 to 1) represents the probability that, for a randomly selected set of observations, the model prediction for a presence observation will be higher than the prediction for an absence observation.

We used a bootstrapping approach to estimate AUC and to determine variable importance for GLM and GAM models (Wintle et al. 2005). Five hundred bootstrapped models were created by duplicating the dataset and then randomly resampling from it so that every sample used for training or testing has the same number of observations as the original dataset. In this way, the reduced prediction accuracy expected when a model is confronted with new data could be estimated. To calculate prediction accuracy with classification trees, we used 15-fold cross-validation using the same number of splits for pruning all cross-validated models. We calculated the average AUC based on the results of the cross-validation. To calculate the AUC for RF, we used the averaged "out-of-bag" predictions (average number of cases withheld was 36%) from the models.

Modeling was carried out in the R 2.7.0 statistical programming environment (R Development Core Team 2004) using the packages gam, rpart, randomForest, ROCR, spdep, yaImpute and model_functions.R (from Wintle et al. 2005).

Ranking environmental variables

For GLMs and GAMs we estimated the relative importance of environmental variables by determining what percentage of the 500 bootstrap models retained the variable (and/or its polynomial term) as a predictor. In Random Forests, variable importance is determined by comparing the misclassification error rate of a tree with the error rate that occurs if the values of a predictor variable are randomly permuted (Cutler et al. 2007). Those variables that result in the largest difference in accuracy (averaged over the 500 model replicates) for trees developed with the true values of the variable compared to those developed with randomly values are considered to be most important. We did not assess variable importance for the classification tree models because we did not perform any bootstrapping or model averaging for this method.

Because measures of variable importance are calculated differently in Random Forests than in GLMs and GAMs, we developed a ranking system so we could compare environmental selection among the different model types. For each species in each model type, we evaluated all environmental variables and ranked them from 1 (most important) to 8. If two variables had the same importance, we assigned them both the same rank and then proceeded to rank the rest of the variables based on the order they would be in if there were no tie. We averaged the ranks together for climate, terrain, and soil variables for some analyses.

Generating prediction maps and calculating correlation

To create prediction maps for all of our models, we used the AsciiGridPredict() command in the R package yaImpute, ver. 1.0–3 (Crookston and Finley 2008). In previous studies, spatial overlap in predictions has been estimated using Kappa or Spearman rank correlations that are appropriate for categorical maps (Prasad et al. 2006, Termansen et al. 2006). Because all of the methods we used generated a likelihood of species' presence on a scale of 0–1, we used a Pearson's correlation coefficient to calculate to correlation between prediction maps for each species, pairwise between models (Termansen et al. 2006).

Analysis

To answer our questions about which variables best explain spatial correspondence among prediction maps, we averaged the pairwise map correlations among all

Table 1. Species and functional types evaluated in southern California map overlay. Prevalence is the proportion of plots in which species was present. Ranges of correlation and AUC were derived from GLM, GAM, CT, and RF model types. Functional types: shrubFac = facultative seeder shrub; shrubOS = obligate seeder shrub; shrubOR = obligate resprouters shrub; subshrFac = resprouting subshrub; subshrub S = post-fire seeding subshrub; perrherb = perennial herb.

Species scientific name	Functional type	Prevalence	Range correlation	Range AUC
Adenostoma fasciculatum	shrubFac	0.53	0.59–0.82	0.73–0.79
Adenostoma sparsifolium	shrubFac	0.06	0.74–0.92	0.85–0.93
Arctostaphylos glauca	shrubOS	0.07	0.51–0.79	0.71–0.92
Arctostaphylos pungens	shrubOS	0.06	0.55–0.85	0.79–0.91
Arctostaphylos glandulosa	shrubOR	0.14	0.55–0.91	0.78–0.84
Artemisia californica	subshrFac	0.39	0.70–0.92	0.80–0.84
Artemisia tridentata	subshrubS	0.03	0.50–0.67	0.81–0.90
Ceanothus crassifolius	shrubOS	0.12	0.48–0.81	0.75–0.84
Ceanothus cuneatus	shrubOS	0.03	0.37–0.70	0.67–0.93
Ceanothus greggii	shrubOS	0.12	0.78–0.93	0.85–0.94
Ceanothus leucadermis	shrubFac	0.12	0.70–0.90	0.77–0.89
Ceanothus tomentosus	shrubOS	0.12	0.56–0.98	0.78–0.84
Ceanothus verrucosus	shrubOS	0.03	0.61–0.84	0.74–0.92
Cercocarpus betuloides	shrubOR	0.15	0.63–0.90	0.76–0.86
Cneoridium dumosum	shrubOR	0.03	0.44–0.68	0.66–0.84
Eriophyllum confertiflorum	perrherb	0.06	0.12–0.57	0.51–0.63
Eriodictyon crassifolium	shrubFac	0.01	0.21–0.63	0.55–0.76
Eriogonum fasciculatum	subshrFac	0.46	0.45–0.96	0.58–0.66
Galium angustifolium	perrherb	0.03	0.23–0.59	0.62–0.83
Garrya veatchii	shrubFac	0.04	0.45–0.70	0.76–0.89
Gutierrezia sarothrae	subshrubS	0.05	0.29–0.83	0.60–0.80
Hazardia squarrosa	shrubOR	0.09	0.35–0.77	0.48–0.66
Heteromeles arbutifolia	shrubOR	0.12	0.53–0.81	0.64–0.77
Keckiella antirrhinoides	subshrOR	0.06	0.36–0.66	0.62–0.75
Lonicera subspicata	subshrOR	0.05	0.25–0.64	0.68–0.76
Lotus scoparius	shrubOS	0.31	0.47–0.82	0.56–0.66
Malacothamnus fasciculatus	subshrFac	0.02	0.01–0.51	0.52–0.61
Malosma laurina	shrubFac	0.3	0.78–0.93	0.79–0.83
Mimulus aurantiacus	subshrubS	0.11	0.60–0.83	0.60–0.71
Opuntia littoralis	subshrubS	0.01	0.09–0.59	0.78–0.88
Penstemon spectabilis	perrherb	0.02	0.36–0.67	0.72–0.81
Prunus ilicifolia	shrubOR	0.09	0.58–0.80	0.68–0.83
Quercus berberidifolia	shrubOR	0.37	0.77–0.97	0.76–0.81
Quercus wislizeni	shrubOR	0.04	0.52–0.75	0.79–0.93
Rhamnus ilicifolia	shrubOR	0.1	0.58–0.84	0.67–0.76
Rhamnus crocea	shrubOR	0.05	0.08–0.31	0.53–0.63
Rhus integrifolia	shrubOR	0.11	0.66–0.88	0.80–0.89
Rhus ovata	shrubFac	0.16	0.62–0.85	0.74–0.78
Salvia apiana	subshrFac	0.33	0.48–0.94	0.61–0.72
Salvia mellifera	subshrFac	0.27	0.34–0.86	0.69–0.75
Toxicodendron diversilobum	subshrOR	0.04	0.33–0.90	0.62–0.72
Trichostema lanatum	shrubFac	0.03	0.51–0.73	0.73–0.85
Viguiera laciniata	subshrOR	0.03	0.20–0.61	0.65–0.80
Xylococcus bicolor	shrubOR	0.12	0.42–0.83	0.73–0.83
Yucca whipplei	subshrOR	0.13	0.69–0.90	0.70–0.75

model types for each species to use as the dependent variable in a regression analysis. We first developed simple regression models for each explanatory variable to explore the effects of model accuracy, species prevalence, species functional type, and environmental variable importance on map correlation. After developing the simple regression analyses, we estimated a multiple regression models. We entered the explanatory variables into the model in the order of the amount of variation they explained in the simple models, and we only retained those variables that were significant at $p \leq 0.05$.

To estimate the effect of model accuracy, we averaged the AUC from the predictions of all model types for each species to use as the predictor variable. Species prevalence was calculated as the proportion of plots in which the species was present. We developed the species functional type classification (Table 1) based on natural groupings of species' life form, demographic attributes, and fire response strategy (details in Syphard and Franklin in press). For the environmental variables, we used the average importance rank for the climate, terrain, and soil variables (that were developed from GLMs, GAMs, and RFs only).

Results

There was significant ($p < 0.05$) positive spatial autocorrelation (SA) in the residuals of GLMs for only 7 of 45, or $<16\%$, of the species. There was no apparent relationship between SA in the residuals and species prevalence, AUC, or species traits. Because so few models showed significant SA in the residuals, and because the emphasis of this study

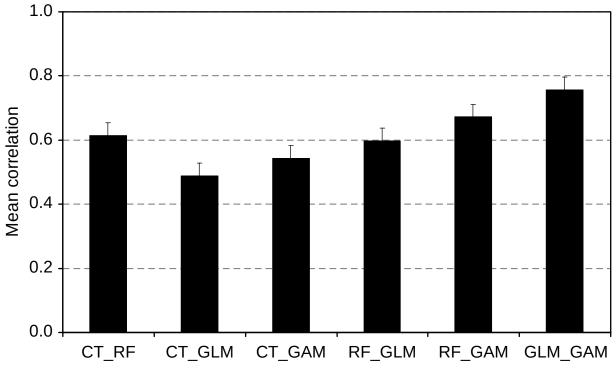

Figure 1. Pairwise correlations among prediction maps produced using classification trees (CTs), Random Forests (RFs), generalized linear models (GLMs) and generalized additive models (GAMs) for plant species in southern California. The error bars represent standard error.

was on spatial prediction and not estimation of parameters, we did not treat SA further (e.g. by fitting a spatial autoregressive error model).

Mean correlation among prediction maps varied according to the methods used to develop the models, and there was substantial variability in the correlations among species (Fig. 1). The lowest correlation between maps occurred between CTs and GAMs or GLMs, and the highest correlation occurred between GLMs and GAMs, with intermediate levels of correlation between RF and CTs, GLMs or GAMs. Analysis of variance showed that correlations were different among pairs ($F = 13.67$; $p < 0.001$) (Fig. 1).

When model accuracy and species prevalence were both low, Random Forests predicted distributions to occur over a larger extent and to be more dispersed than GLMs or GAMs (Fig. 2A, D). When different model types had reasonable accuracy (AUCs generally above 0.75), they predicted species to be distributed in the same general locations of the study area (Fig. 2B, C).

Species prevalence and model accuracy explained more variation in map correlation than the other variables, although functional type and the importance of climate in model selection were also significant (Table 2). The relationship between map correlation and prediction accuracy was positive and linear, but the relationship with prevalence was quadratic. In general, prediction maps had the lowest correspondence when both species prevalence and model accuracy were lowest. With higher prediction accuracy and species prevalence, map correlation was also much higher; but for the species with prevalence > 0.2, the relationship (between prevalence and map correlation) was negative (Fig. 3). Because only a few species (8) had prevalence values > 0.2, however, this negative relationship should be interpreted with caution.

Map correlation was higher for species that experience fire-cued germination from a dormant seed bank (facultative seeder shrubs and obligate seeder shrubs) and lowest for perennial herbs and subshrubs that respond to fire through vegetative propagation (Fig. 4). Although the importance of terrain and soil variables in the models did not influence the correspondence among prediction maps, those models for which climate was most important produced maps that had better map correlation (Table 2).

When all of the significant explanatory variables (prevalence, AUC, functional type, and climate) were included in multiple regression models, only species prevalence and AUC remained significant predictors of map correlation (Table 3). These two variables in combination explained a large amount of variation in spatial correspondence (adjusted $R^2 = 0.76$).

When averaged together across all model types (GLMs, GAMs and RFs), the climate variables were more important in the SDMs than terrain or soil, which were both similar in their relative importance (Fig. 5). The relative importance of different variables when evaluated individually, however, was different depending on the model type (Fig. 6). For the GLMs and the GAMs, the three climate variables had nearly equal importance, which was higher than the importance of the other variables. For all three model types, summer radiation was more important than the other terrain variables; and the relative difference was substantial for the GLMs and GAMs. Whereas the importance of the other three terrain variables was similar for GLMs and GAMs (although TMI was generally the lowest), winter radiation and (especially) TMI were substantially lower than slope in Random Forests. Soil order was substantially more important than the terrain variables for GLMs and GAMs. Differences in importance between terrain and soil variables were insubstantial for Random Forests.

Discussion

The use of metrics such as AUC has become standard practice in evaluating the performance of species distribution models. AUC is a very useful measure of comparative model performance because it is threshold independent, but any measure of predictive performance is limited by the data available for model evaluation. The results of this study reinforce the notion that it is also important to consider additional criteria in model evaluation, depending on the objective of the application (Austin et al. 2006, Hernandez et al. 2006). If prediction maps will be used to make conservation or resource management decisions, the spatial distribution of model uncertainty may be particularly important. While correlation among map predictions in our study significantly improved with more accurate models, there were other factors that strongly affected spatial correspondence among predictions, especially species prevalence. Map correlation also varied depending on the modeling method used, species functional traits, and type of environmental variables that were important in the models. The effect of these factors on model performance should therefore be taken into consideration for any SDM application.

One caveat with regards to map correlation is that one model may yield consistently lower (or higher) predictions than another, and yet those maps may remain highly correlated. The scaling of the predicted probabilities does tend to vary according to modeling methods or species prevalence (Real et al. 2006, Albert and Thuiller 2008) so we would expect an offset in the predicted values from different models. For this reason, correlation is an appropriate way to

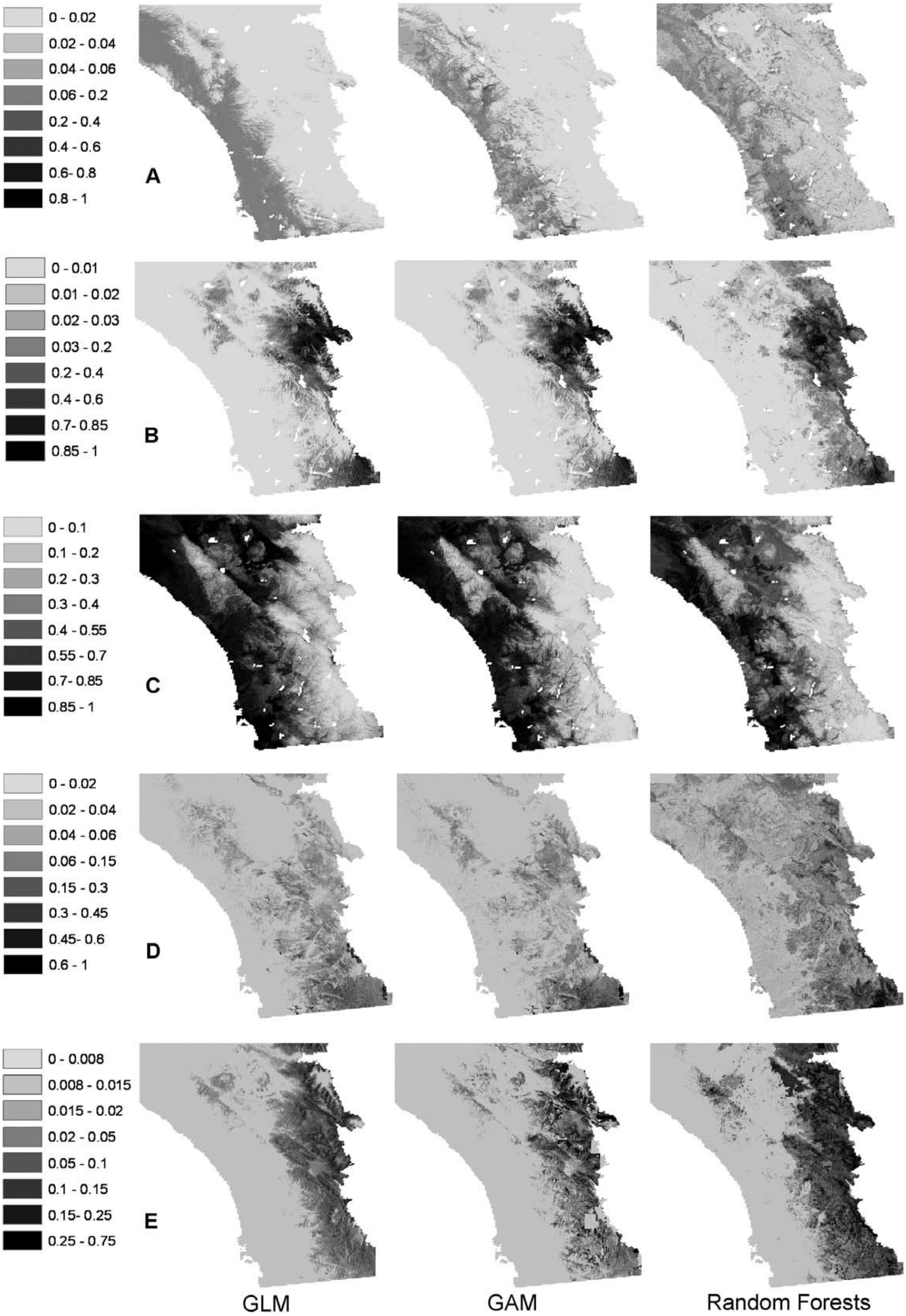

Figure 2. Maps displaying predicted probability of presence from a generalized linear model (GLM), generalized additive model (GAM), and Random Forests. A – *Viguiera laciniata* (low prevalence (0.03), low map correlation low to moderate AUC = 0.55–0.84; Table 1); B – *Adenostoma sparsifolium* (low prevalence (0.06), high map correlation, high AUC = 0.82–0.94); C – *Artemisia californica* (high prevalence (0.39), high map correlation, moderate AUC = 0.70–0.92); D – *Gutierrezia sarothrae* (low prevalence (0.05), low map correlation, low AUC = 0.60–0.85); E – *Penstemon spectabilis* (low prevalence (0.02), low map correlation, moderate AUC = 0.72–0.81).

Table 2. Model coefficients, p-values, and R^2 for individual explanatory variables in the simple regression models for map correlation in southern California.

	Variable	Model parameters		
		Coefficient	p-value	R^2
Soil order	Prevalence	2.17	<**0.001**	0.34
	Prevalence ^ 2	−3.54	**0.003**	
	AUC	0.99	<**0.001**	0.34
	Functional type	NA	**0.049**	0.16
	Climate	0.11	**0.004**	0.16
	Terrain	−0.07	0.218	0.01
	Soil	−0.01	0.646	0

establish a baseline comparison between maps. SDM predictions are often scaled or thresholded in some way for subsequent use in decision making (Freeman and Moisen 2008). If SDM predictions are used in an application where a multi-class ranked categorization of habitat quality exists (has been previously defined), then maps could be averaged and compared spatially with a multi-class kappa statistic (as in Albert and Thuiller 2008). Map similarity may diminish or intensify once these thresholds have been applied depending on the distribution of probability values predicted from each model.

With regards to the modeling methods we compared, classification trees overall had lower accuracy than the other three methods (Syphard and Franklin in press), which is likely why the pairwise comparisons of map correlation were lowest with the CTs. However, because Random Forests is essentially developed using an ensemble of trees, we were surprised that the correlation between RFs and CTs was as low as the correlation between RFs and GLMs or GAMs. The relatively low accuracy and low map correlation using single CTs is consistent with other studies that found them to be somewhat unstable (Benito Garzón et al. 2006, Prasad et al. 2006). While there are some features of CTs that may be more desirable than RFs (e.g. ability to visualize the classification rules portrayed in single trees), RFs may be a better choice for conservation practitioners trying to create the most robust predictive maps.

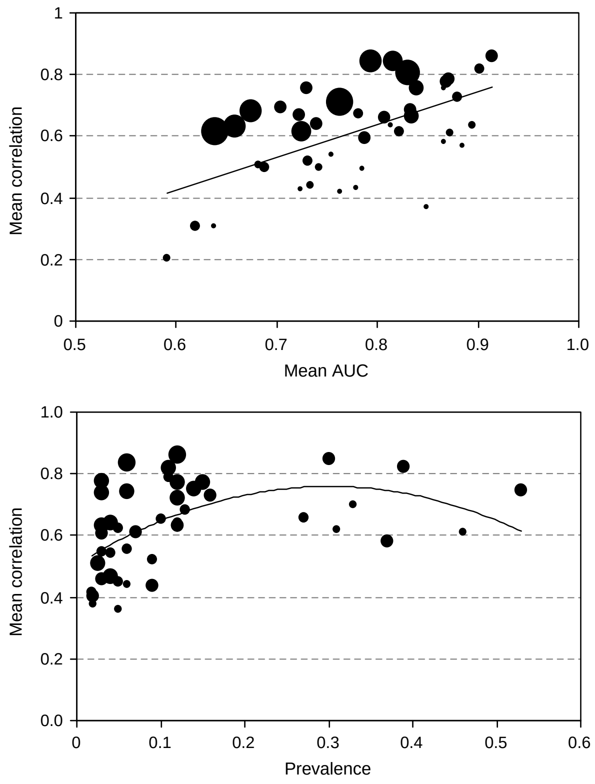

Figure 3. Mean correlation among four model types as a function of mean AUC and species' prevalence. Observations in the AUC charts (top row) are scaled by prevalence (size of circle), and observations in the charts of prevalence are scaled by AUC.

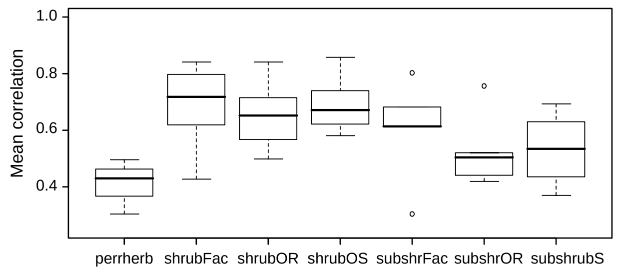

Figure 4. Boxplots for 45 plant species in southern California showing map correlation versus species' functional type. shrubFac = facultative seeder shrub; shrubOS = obligate seeder shrub; shrubOR = obligate resprouters shrub; subshrFac = resprouting subshrub; subshrubS = post-fire seeding subshrub; perrherb = perennial herb.

Although prediction accuracy was highest with Random Forests, the spatial correspondence in predictions was lower between RF and GAMs or GLMs than it was between GAMs and GLMs. Although the Random Forests models predicted greater extents of suitable habitat than GLMs or GAMs for species with low prevalence, it is unknown, based on the data we had for model evaluation, whether this low map correlation was due to true errors of commission. Alternatively, the greater predicted extent (i.e. analogous environmental conditions) may have represented areas that were truly suitable for the species, and the species may not have been sampled in that area, or it may have previously occupied the area.

While prevalence strongly affected map correlation in this study, other studies have shown that prevalence may also be significantly related to model performance. In some cases, prediction accuracy was higher when prevalence was low (Segurado and Araújo 2004, Elith et al. 2006, Hernandez et al. 2006), but McPherson et al. (2004) found that models performed best when prevalence was intermediate. In this study, prevalence had no significant effect on AUC (p = 0.46); therefore, the effect of prevalence on map correlation can not be directly attributed to the effect of prevalence on model performance (otherwise, both

Table 3. Coefficients and p-values for variables retained in the multiple regression models for map correlation in southern California.

	Variable	Coefficient	p-value
Soil order	(Intercept)	−0.51	<**0.001**
	Prevalence	2.15	<**0.001**
	Prevalence ^ 2	−3.17	<**0.001**
	AUC	1.08	<**0.001**
			R^2 −0.76

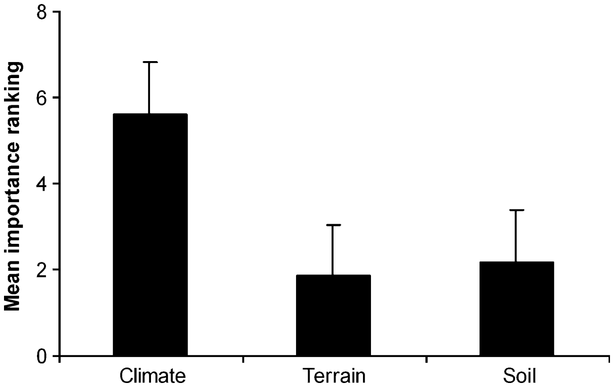

Figure 5. Mean importance ranking for climate, terrain, and soil variables using Random Forests (RFs), generalized linear models (GLMs) and generalized additive models (GAMs). The error bars represent standard error.

variables would not have been retained in our multiple regression). However, the relationship between species prevalence and map correlation may be partly related to the way that different models approximate species-response functions and how those response functions translate into predicted probabilities. If models vary in the way that their distributions of predicted probabilities reflect the prevalence of the species (based on the number of sample locations used to build the models), then presumably these model differences would be manifested more apparently in map correlations if prevalence were either very low or high.

Another reason that low prevalence may have affected map correlation is that species may have low prevalence because they are difficult to detect. One consequence of low detectability is that a species could actually be present in locations where it is predicted to be absent. This would likely affect the model's prediction accuracy, but may also affect spatial extrapolation, reflecting differences in ways that models characterize species' presences. Furthermore, if species have low prevalence, there is more room to maneuver on the map, providing a greater number of opportunities for spatial divergence in predictions.

In this study, the relationship between plant functional type and map correlation appears to result from the observed correspondence between functional type and model prediction accuracy. The functional types with the highest prediction accuracy tended to be those with high site fidelity – long-lived facultative and obligate seeders with poor dispersal and persistent seed banks (Syphard and Franklin in press). Those functional types that had higher AUC also had higher map correlation among models. Therefore, in a multiple regression model, functional type was not selected as a significant predictor of map correlation if AUC was already in the model.

A source of uncertainty in our study is the error inherent in the historic VTM data, which could also affect spatial correspondence of predictions, particularly for those models that select terrain and/or soil. Terrain and soil variables are more heterogeneous than climate at landscape scales. Because climate varies slowly over space, there is greater certainty that those variables would be accurately calculated within the 300-m range of the VTM data that had an average positional error of ~130 m (Kelly et al. 2008). Coarsening data resolution could potentially increase or decrease model performance. On one hand, performance may increase after smoothing errors in environmental or species data, but on the other hand, performance may decrease if there is a lack of spatial matching between species observations and their associated environmental predictors (Guisan et al. 2007).

The primary influence of environmental variable selection on the spatial correspondence of predicted distributions was related to climate. In other words, the more that a species distribution could be modeled through climate variables alone, the more likely the predictions were likely to overlay. Some SDMs only use climate variables (e.g. bioclimatic envelope models, Huntley et al. 2004, Kueppers

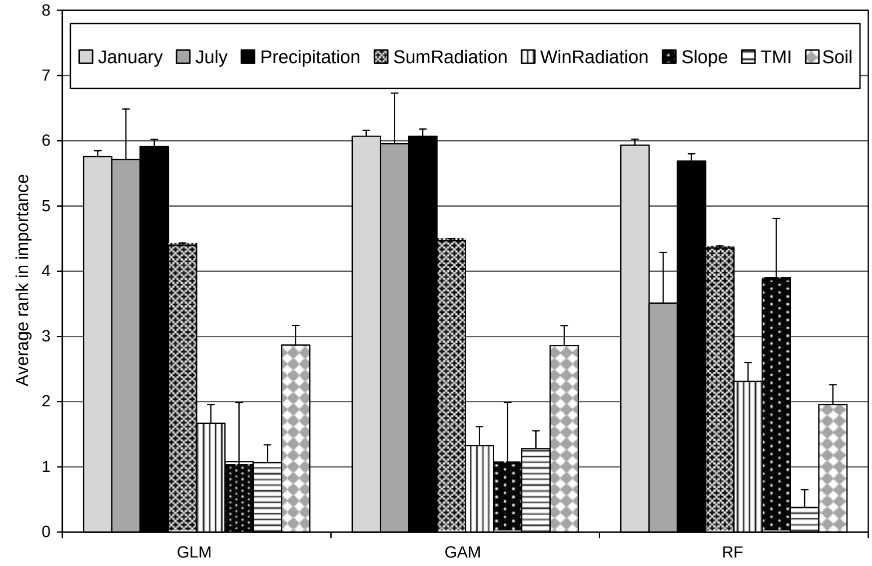

Figure 6. Mean importance ranking for all variables using Random Forests (RFs), generalized linear models (GLMs) and generalized additive models (GAMs). TMI = topographic moisture index. The error bars represent standard error.

et al. 2005, Heikkinen et al. 2006), and our results are consistent with those studies that suggest that climate tends to be the overriding driver of large-scale distributional patterns for plants. It has been shown (Segurado et al. 2006) that variable "explanatory" power (inflation of variable significance) is directly proportional to the spatial autocorrelation (SA) of the predictor, and the importance of variables found in this study was in order of increasing spatial resolution and therefore SA (terrain < soil order < climate). However, virtually all environmental drivers are spatially patterned. The general lack of SA in our model residuals suggests that the spatial pattern in the distributions is explained by the spatial pattern in the predictors. Predictors that vary slowly in space will capture broad first-order patterns in species distributions. The ranking of variable importance in this study indicates how much of the observed pattern is explained by each predictor, although it cannot show conclusively that climate imposes the most proximal limiting factor to distributions.

In addition, there was substantial variation in the importance of different environmental predictors among species, and terrain and soil were also important for explaining that variation. One potential reason that, for some species, terrain or soil variables were selected over climate variables is that their climatic range may have been greater than that which was in the study area. Thus, the model(s) sought finer-scaled variables to explain what aspects of species distribution patterns the climate variables were missing. Therefore, we suggest that both terrain and soil should be considered in any SDM study for plants at a landscape scale. As we have already noted, a meta-analysis found that models that included environmental predictors from multiple scales showed the highest predictive performance (Meyer and Thuiller 2006). Further, while there are perceived trade-offs between model parsimony and model accuracy, Drake et al. (2006) found that the most accurate models were those that included the largest number of environmental predictors, even after optimizing the models to avoid overfitting.

In conclusion, average model performance (measured by AUC) was essentially the same for the GLMs, GAMs, and Random Forests models (although CTs had lower accuracy). Yet, despite these similar accuracies, our results show that prediction maps and the environmental variables selected varied substantially among the different methods. When the goal of the SDM study is to create prediction maps, we suggest that the model evaluation process should go beyond global accuracy measures and include some evaluation of the spatial pattern of predictions. For example, in the context of climate change modeling and other applications of SDM, some authors have suggested averaging model predictions due to high variability in their projections (Thuiller et al. 2004a, b, Marmion et al. 2008). However, Araújo et al. (2005a, b) cautioned that accuracy will most likely increase only if better models are considered as opposed to more models. There are a number of approaches to ensemble forecasting in SDM and other modeling fields in addition to model averaging or consensus methods (Araújo and New 2007).

Further, it might be prudent to evaluate spatial predictions from model types that tend to be different, such as GLMs vs Random Forests, to determine a bracket of uncertainty. This might be particularly important for species that have either very low or very high prevalence. Nevertheless, it is important to consider that map correlation in this study was a function of those models that we selected to examine and the variable selection methods that we used. While we chose common methods used in SDM, the differences in predictions are ultimately a function of how the different models handle prediction.

Acknowledgements – This work was supported by a grant from the National Science Foundation (NSF BCS-0452389) to JF, J. Miller and R. Fisher. We are very grateful to N. Crookston for helping us with the prediction maps. We also thank J. Miller and R. Taylor for their contributions, J. E. Keeley for reviewing the species traits, and the anonymous reviewers for improving the manuscript with their comments.

References

Albert, C. H. and Thuiller, W. 2008. Favourability functions versus probability of presence: advantages and misuses. – Ecography 31: 417–422.

Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. – Trends Ecol. Evol. 22: 42–47.

Araújo, M. B. et al. 2005a. Validation of species-climate impact models under climate change. – Global Change Biol. 11: 1504–1513.

Araújo, M. B. et al. 2005b. Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. – Global Ecol. Biogeogr. 14: 17–30.

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

Austin, M. P. and Smith, T. M. 1989. A new model for the continuum concept. – Vegetatio 83: 35–47.

Austin, M. P. et al. 2006. Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. – Ecol. Model. 199: 197–216.

Barbour, M. et al. (eds) 2007. Terrestrial vegetation of California. – Univ. of California Press.

Benito Garzón, M. et al. 2006. Predicting habitat suitability with machine learning models: the potential area of *Pinus sylvestris* L. in the Iberian Peninsula. – Ecol. Model. 197: 383–393.

Bio, A. et al. 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. – J. Veg. Sci. 9: 5–16.

Breiman, L. 2001. Random forests. – Mach. Learn. 45: 15–32.

Breiman, L. et al. 1984. Classification and regression trees. – Wadsworth.

Crookston, N. L. and Finley, A. O. 2008. yaImpute: an R package for kNN imputation. – J. Stat. Soft. 23: 1–16.

Crossman, N. D. and Bass, D. A. 2008. Application of common predictive habitat techniques for post-border weed risk management. – Divers. Distrib. 14: 213–224.

Cutler, D. R. et al. 2007. Random forests for classification in ecology. – Ecology 88: 2783–2792.

De'ath, G. and Fabricius, K. E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. – Ecology 81: 3178–3192.

Drake, J. M. et al. 2006. Modelling ecological niches with support vector machines. – J. Appl. Ecol. 43: 424–432.

Dubayah, R. and Rich, P. M. 1995. Topographic solar radiation for GIS. – Int. J. Geogr. Inf. Syst. 9: 405–419.

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

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

Franklin, J. 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. – Prog. Phys. Geogr. 19: 474–499.

Franklin, J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. – J. Veg. Sci. 9: 733–748.

Franklin, J. et al. 2001. Stratified sampling for field survey of environmental gradients in the Mojave Desert Ecoregion. – In: Millington, A. et al. (eds), GIS and remote sensing applications in biogeography and ecology. Kluwer, pp. 229–253.

Freeman, E. A. and Moisen, G. G. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence. – Ecol. Model. 217: 48–58.

Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Model. 135: 147–186.

Guisan, A. et al. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. – Ecol. Model. 157: 89–100.

Guisan, A. et al. 2006. Making better biogeographical predictions of species' distributions. – J. Appl. Ecol. 43: 386–392.

Guisan, A. et al. 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? – Ecol. Monogr. 77: 615–630.

Hanley, J. A. and McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristics curve. – Radiology 143: 29–36.

Heikkinen, R. K. et al. 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. – Prog. Phys. Geogr. 30: 751–777.

Hernandez, P. A. et al. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. – Ecography 29: 773–785.

Hickman, J. D. (ed.) 1993. The Jepson manual: higher plants of California. – Univ. of California Press.

Huntley, B. et al. 2004. The performance of models relating species geographical distributions to climate is independent of trophic level. – Ecol. Lett. 7: 417–426.

Iverson, L. R. and Prasad, A. M. 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. – Ecol. Monogr. 68: 465–485.

James, F. C. and McCulloch, C. E. 1990. Multivariate analysis in ecology and systematics: Panacea or Pandora's box? – Annu. Rev. Ecol. Syst. 21: 129–166.

Johnson, C. J. and Gillingham, M. P. 2005. An evaluation of mapped species distribution models used for conservation planning. – Environ. Conserv. 32: 117–128.

Johnson, C. J. et al. 2004. A quantitative approach to conservation planning: using resource selection functions to identify important habitats for mountain caribou. – J. Appl. Ecol. 41: 238–251.

Keeley, J. E. 2000. Chaparral. – In: Barbour, M. G. and Billings, W. D. (eds), North American terrestrial vegetation. Cambridge Univ. Press, pp. 204–253.

Kelly, M. et al. 2005. Digitization of a historic dataset: the Wieslander California vegetation type mapping project. – Madroño 52: 191–201.

Kelly, M. et al. 2008. Considerations for ecological reconstruction of historic vegetation: analysis of the spatial uncertainties in the California Vegetation Type Map dataset. – Plant. Ecol. 194: 37–49.

Kueppers, L. M. et al. 2005. Modeled regional climate change and California endemic oak ranges. – Proc. Nat. Acad. Sci. USA 102: 16281–16286.

Lehmann, A. et al. 2002. GRASP: generalized regression analysis and spatial prediction. – Ecol. Model. 157: 189–207.

Lobo, J. M. et al. 2008. AUC: a misleading measure of the performance of predictive distribution models. – Global Ecol. Biogeogr. 17: 145–151.

Loiselle, B. A. et al. 2003. Avoiding pitfalls of using species distribution models in conservation planning. – Conserv. Biol. 17: 1591–1600.

Luoto, M. et al. 2005. Uncertainty of bioclimate envelope models based on the geographical distribution of species. – Global Ecol. Biogeogr. 14: 575–584.

Mackey, B. G. and Lindenmayer, D. B. 2001. Towards a hierarchical framework for modeling the spatial distribution of animals. – J. Biogeogr. 28: 1147–1166.

Maggini, R. et al. 2006. Improving generalized regression analysis for the spatial prediction of forest communities. – J. Biogeogr. 33: 1729–1749.

Marmion, M. et al. 2008. Evaluation of consensus methods in predictive species distribution modeling. – Divers. Distrib. 15: 59–69.

McPherson, J. M. and Jetz, W. 2007. Effects of species' ecology on the accuracy of distribution models. – Ecography 30: 135–151.

McPherson, J. M. et al. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? – J. Appl. Ecol. 41: 811–823.

Meyer, C. B. and Thuiller, W. 2006. Accuracy of resource selection functions across spatial scales. – Divers. Distrib. 12: 288–297.

Miller, J. et al. 2007. Incorporating spatial dependence in predictive vegetation models. – Ecol. Model. 202: 225–242.

Mladenoff, D. J. et al. 1995. A regional landscape analysis and prediction of favorable gray wolf habitat in the northern great lakes region. – Conserv. Biol. 9: 279–294.

Moisen, G. G. and Frescino, T. S. 2002. Comparing five modelling techniques for predicting forest characteristics. – Ecol. Model. 157: 209–225.

Moore, I. D. et al. 1991. Digital terrain modeling: a review of hydrological, geomorphologic and biological applications. – Hydrol. Process. 5: 3–30.

Moran, P. A. P. 1948. The interpretation of statistical maps. – J. Roy. Stat. Soc. B 10: 243–251.

Osborne, P. E. and Suarez-Seoane, S. 2007. Identifying core areas in a species' range using temporal suitability analysis: an example using little bustards *Tetrax tetrax* L. in Spain. – Biodivers. Conserv. 16: 3505–3518.

Peterson, A. T. and Nakazawa, Y. 2008. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. – Global Ecol. Biogeogr. 17: 135–144.

Prasad, A. M. et al. 2006. Newer classification and regression techniques: bagging and random forests for ecological prediction. – Ecosystems 9: 181–199.

R Development Core Team 2004. R: a language and environment for statistical computing. – R Foundation for Statistical Computing.

Real, R. et al. 2006. Obtaining environmental favourability functions from logistic regression. – Environ. Ecol. Stat. 13: 237–245.

Schoenherr, A. 1992. A natural history of California. – Univ. of California Press.

Scott, J. M. et al. (eds) 2002. Predicting species occurrences: issues of accuracy and scale. – Island Press.

Segurado, P. and Araújo, M. B. 2004. An evaluation of methods for modelling species distributions. – J. Biogeogr. 31: 1555–1568.

Segurado, P. et al. 2006. Consequences of spatial autocorrelation for niche based models. – J. Appl. Ecol. 43: 433–444.

Stockwell, D. and Peterson, A. T. 2002. Effects of sample size on accuracy of species distribution models. – Ecol. Model. 148: 1–13.

Syphard, A. D. and Franklin, J. in press. Species' traits affect the performance of species' distribution models for plants in southern California. – J. Veg. Sci.

Taylor, R. S. 2004. A natural history of coastal sage scrub in southern California: regional floristic patterns and relations to physical geography, how it changes over time, and how well reserves represent its biodiversity. – Geography Dept, Univ. of California.

Termansen, M. et al. 2006. The use of genetic algorithms and Bayesian classification to model species distributions. – Ecol. Model. 192: 410–424.

Thuiller, W. et al. 2004a. Biodiversity conservation – uncertainty in predictions of extinction risk. – Nature 430, doi:10.1038/nature02716.

Thuiller, W. et al. 2004b. Effects of restricting environmental range of data to project current and future species distributions. – Ecography 27: 165–172.

Westman, W. E. 1981. Diversity relations and succession in Californian coastal sage scrub. – Ecology 62: 170–184.

Wieslander, A. E. 1935. A vegetation type map of California. – Madroño 3: 140–144.

Wilson, J. and Gallant, J. 2000. Terrain analysis: principles and applications. – Wiley.

Wintle, B. A. et al. 2005. Fauna habitat modelling and mapping: a review and case study in the Lower Hunter Central Coast region of NSW. – Austral Ecol. 30: 719–738.

Yee, T. W. and Mitchell, N. D. 1991. Generalized additive models in plant ecology. – J. Veg. Sci. 2: 587–602.