

Species traits affect the performance of species distribution models for plants in southern California

Alexandra D. Syphard¹ & Janet Franklin¹

Abstract

Questions: To what extent do plant species traits, including life history, life form, and disturbance response characteristics, affect the degree to which species distributions are determined by physical environmental factors? Is the strength of the relationship between species distribution and environment stronger in some disturbance-response types than in others?

Location: California southwest ecoregion, USA.

Methods: We developed species distribution models (SDMs) for 45 plant species using three primary modeling methods (GLMs, GAMs, and Random Forests). Using AUC as a performance measure of prediction accuracy, and measure of the strength of species–environment correlations, we used regression analyses to compare the effects of fire disturbance response type, longevity, dispersal mechanism, range size, cover, species prevalence, and model type.

Results: Fire disturbance response type explained more variation in model performance than any other variable, but other species and range characteristics were also significant. Differences in prediction accuracy reflected variation in species life history, disturbance response, and rarity. AUC was significantly higher for longer-lived species, found at intermediate levels of abundance, and smaller range sizes. Models performed better for shrubs than subshrubs and perennial herbs. The disturbance response type with the highest SDM accuracy was obligate-seeding shrubs with ballistic dispersal that regenerate via fire-cued germination from a dormant seed bank.

Conclusions: The effect of species characteristics on predictability of species distributions overrides any differences in modeling technique. Prediction accuracy may be related to how a suite of species

characteristics co-varies along environmental gradients. Including disturbance response was important because SDMs predict the realized niche. Classification of plant species into disturbance response types may provide a strong framework for evaluating performance of SDMs.

Keywords: Chaparral; Coastal Sage Scrub; Disturbance response; Fire; Life history traits; Rarity; Species range.

Nomenclature: (Hickman 1993)

Abbreviations: AUC = Area under the curve; CT = Classification tree; DEM = Digital elevation model; GAM = Generalized additive model; GLM = Generalized linear model; RF = Random forests; ROC = Receiver-operating characteristic; SDM = Species distribution model; TMI = Topographic moisture index.

Introduction

Species distribution models (SDMs) estimate species responses to environmental gradients, and are used to make spatial predictions of habitat suitability or probability of species occurrence (Franklin 1995; Scott et al. 2002; Guisan et al. 2006). While the ecological underpinning of SDMs is the species–environment relationship, the models are developed using data on the actual distribution of species occurrences, which often reflect the combined influences of multiple interacting biotic and abiotic factors, including the spatial variation in disturbance regime characteristics (Pausas 1999; Pausas & Lavorel 2003; Pausas et al. 2004). Therefore, the models may account for how species environmental tolerances are convolved with their physiological adaptations to other limiting factors, such as periodic disturbance.

Because species traits may reflect their differential responses to processes that control their distribution, these traits may also affect the performance of SDMs. Some studies have explored whether species rarity (Rabinowitz 1981) affects the ability to predict their distributions from environ-

Syphard, A.D. (corresponding author: asyphard@consbio.org) & **Franklin, J.** (janet@sciences.sdsu.edu), Department of Biology, San Diego State University, San Diego, CA 92182-4614, USA.

Syphard, A.D. (asyphard@consbio.org). Current address: Conservation Biology Institute, 10423 Sierra Vista Avenue, La Mesa, CA 91941, USA

¹These authors contributed equally to this paper.

mental factors. In such studies, SDMs for rarer species, those that have smaller ranges, narrower ecological tolerances, or both, tended to have higher prediction accuracy than models developed for widespread, generalist species (e.g., Segurado & Araújo 2004; Seoane et al. 2005; Elith et al. 2006; Hernandez et al. 2006) because it is easier to separate suitable from unsuitable habitat for species with restricted environmental tolerances or very specific habitat requirements. Other studies have examined morphological, physiological, or ecological species traits with respect to SDM performance; these included migratory status, body size, and affinity with fine-scale habitat features for birds (Stockwell & Peterson 2002; McPherson & Jetz 2007), mobility and body size for butterflies (Poyry et al. 2008), and pollination type, leaf longevity and type, successional status, maximum growth rate, shade tolerance, and drought tolerance for plants (Kuhn et al. 2006; Guisan et al. 2007; Zimmermann et al. 2007). Although disturbance regime characteristics strongly affect distribution patterns in many regions, the relationship between species disturbance response and the strength of species–environment relationships has not been examined in comparison with other species traits.

The classification of species traits into disturbance response functional types has provided a framework for analysis and prediction in community and evolutionary ecology (Gillison & Carpenter 1997; Smith et al. 1997; Diaz et al. 1998; Ackerly 2003; Pausas & Lavorel 2003; McGill et al. 2006; Syphard et al. 2006). Disturbance response classifications often incorporate life history and demographic traits with disturbance response strategies, and these groups of species tend to show predictable changes along environmental as well as disturbance gradients (Noble & Gitay 1996; Rusch et al. 2003). Environmental tolerances and physiological adaptations also tend to co-vary with plant life form (Raunkiaer 1934), life history, and fire response strategies (Chapman & Crow 1981; Keeley 1981, 1998; Ackerly et al. 2002), largely due to resource allocation and life-history tradeoffs (Austin 1987; McGill et al. 2006).

Our objective was to evaluate whether species traits, including life history, life form, and disturbance response characteristics, affect the degree to which species distributions are determined by physical environmental factors by examining the relationship between these traits and model performance in SDMs developed for plant species in a fire disturbance-prone region. We classified species into disturbance response types based on unique combi-

nations of life form and fire response to evaluate their relative effect on prediction accuracy. Because the SDM literature has focused on the prediction accuracy of different modeling methods (e.g., Moisen & Frescino 2002; Segurado & Araújo 2004; Elith et al. 2006; Maggini et al. 2006), we also compared the magnitude of the effect of disturbance response and other species characteristics to the effect of model type on model performance. We developed models for 45 plant species in the southwestern ecoregion of California using four modeling methods, and compared the performance of these models to address the following predictions:

Hypothesis 1: Species traits that characterize fire disturbance response are more strongly related to SDM performance than other ecological and range characteristics because fire disturbance strongly influences plant species distributions. Fire is the predominant disturbance agent in southern California, as in other Mediterranean-type ecosystems, and many plant species have life-history traits that determine their patterns of establishment following fire.

Hypothesis 1.1: Obligate seeder shrub species have a narrower range of ecological tolerances than resprouting shrubs and other fire response types, and their distributions will be the most predictable from environmental factors. Although adult obligate seeder shrubs are killed by fire, these species produce long-lived fire-refractory seeds that only germinate in response to fire, and they typically have very short seed dispersal distances (ballistic dispersal mechanisms). Obligate resprouting shrubs are top-killed by fire, but they respond through vigorous resprouting, can recruit in gaps between fires, and have various seed dispersal syndromes. Facultative seeders regenerate using both strategies of fire-cued germination and vegetative resprouting. Some species in the region are unable to resprout or to respond to fire through fire-cued seed germination and instead rely upon seed dispersal into open areas after fire (Keeley 2000).

Hypothesis 1.2: Shrubs in general have higher site fidelity and therefore more predictable distributions than earlier-successional and shorter-lived sub-shrubs.

Hypothesis 2: In addition to fire disturbance response, other ecological and range characteristics will affect performance more than the modeling method used.

Hypothesis 2.1: Plant species that are late successional (longer-lived, woody life forms) will have distributions more predictably related to broad-scale environmental factors than pioneer species. Species with the former traits establish and persist in suitable

sites throughout their range, and therefore are detectable in species surveys (a factor that enhances accuracy in SDM). Their habitat suitability is also likely to be described by persistent features of the physical environment that are frequently available as maps for predictive SDM, such as climate, substrate and topography (Franklin 1995). Early-successional 'pioneer' species (herbaceous life forms, dependent on long-distance seed dispersal) may be more difficult to model (Guisan et al. 2007; Zimmermann et al. 2007) because they are associated with transient habitat features (e.g., light gaps) that are not always adequately captured in GIS maps of broad-scale environmental factors, and because they may not occupy all suitable habitat owing to dispersal limitations.

Hypothesis 2.2: Because species distribution patterns also reflect their relative capacity for dispersal, species with shorter dispersal distances have better site persistence and more specialized adaptations to site conditions, and thus would be more easily predicted with models.

Methods

Study area and species data

The coastal ranges and interior foothills of southern California support shrubland vegetation that is adapted to the Mediterranean-climate of the region, characterized by cool, wet winters and warm, dry summers. The most extensive vegetation type is chaparral, which is composed of dense, evergreen woody shrub species that are resilient to periodic wildfire (Keeley & Davis 2007). The distribution of chaparral species is differentially influenced by factors such as elevation, slope aspect, coastal-desert exposure, substrate, and fire regime (Hanes 1971). Sage scrub is the second most extensive vegetation type in the region and consists of drought-deciduous sub-shrubs that typically cover coastal slopes in drier locations and at lower elevations than chaparral (Westman 1982; Malanson & O'Leary 1995).

The U.S. Forest Service Wieslander Vegetation Type Map (VTM) survey was conducted from 1929 to 1934 in California, USA (Wieslander 1935a, b). The survey recorded data on abundance of trees, shrubs and herbaceous species for ~18 000 plots across California (<http://vtm.berkeley.edu/>) (Kelly et al. 2005). To develop SDMs, we used a database that included species cover and plot location for 1471 VTM shrub plots that occurred within the Natural Communities Conservation Planning area

(Taylor 2004), within California's southwest ecoregion (as defined in Hickman 1993) (Fig. 1). Because the original plot locations were drawn on 1:64 000 scale maps, the locations of the digitized points were approximated using a circle with a radius of 300 m. This imprecision limits the use of VTM data for some applications (Keeley 2004); however, VTM data have been shown to be suitable for species distribution modeling (Vayssières et al. 2000; Franklin 2002). We assumed that, because the environmental predictors used in this study vary gradually over space, uncertainty in plot locations simply added some measurement error to the data.

Of the 233 species recorded in the plot data, we selected 45 for modeling. We omitted species if there were fewer than 30 plots (prevalence <0.02) in which the species was present. Our species dataset, therefore, consisted of 1,471 locations indicating the presence and absence of 45 species, primarily woody shrubs and suffrutescent sub-shrubs, but also perennial herbs, vines and succulents, typical of chaparral or sage scrub (see supporting information Appendices S1, S2).

Environmental predictors

Based on their hypothesized relationship to the distribution of plant species in southern California (e.g., Davis & Goetz 1990; Franklin et al. 2000; Meentemeyer et al. 2001; Franklin 1998, 2002), we evaluated eight climate, terrain, and soil variables as environmental predictors (Table 1). The climate variables, including mean annual precipitation, mean minimum January temperature, and mean maximum July temperature were interpolated to 1-km resolution maps using climate station data from 1966 to 1995 and regression kriging with elevation as a covariate (Franklin et al. 2001a). Solar radiation mediates temperature and therefore available plant moisture (via evapotranspiration in this semi-arid, summer-dry environment). To develop grids of terrain-distributed solar radiation from a digital elevation model (DEM) (Dubayah & Rich 1995), we used the Solar Analyst 1.0 extension for ArcView™ GIS and U.S. Geological Survey 30-m resolution DEMs. We calculated daily insolation for the summer and winter solstice (using site latitude of 33°N, sky size of 200 cells per side, and 0.2 clear sky irradiance, the fraction of global normal radiation flux that is diffuse), and used these two variables to represent the intra-annual extremes of topographically-patterned radiation. The Topographic Moisture Index (TMI) represents relative soil moisture availability based on upslope catchment area and slope

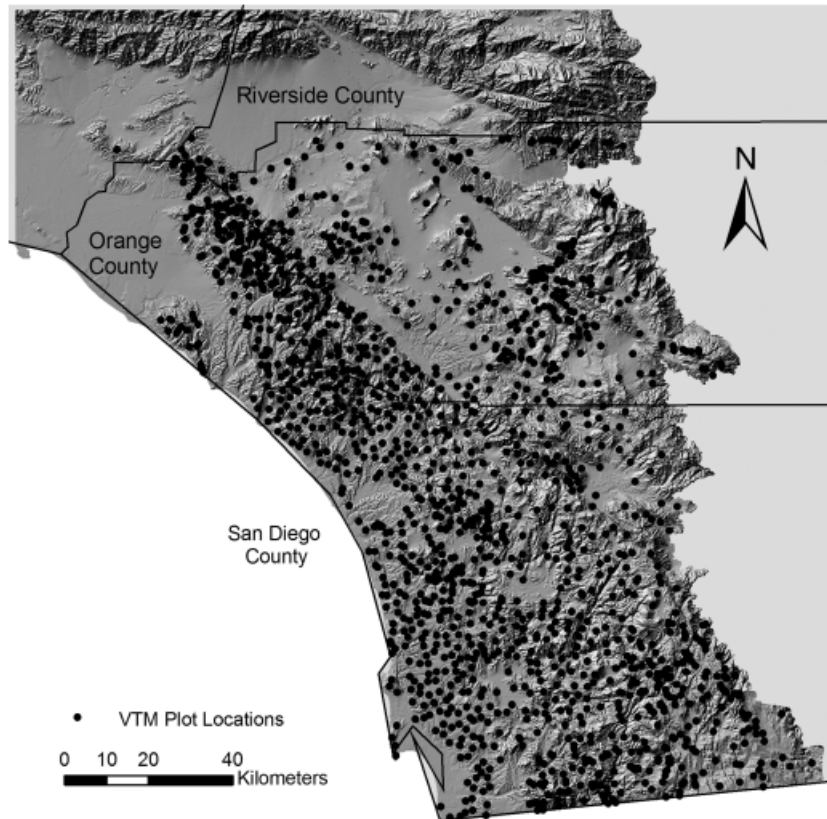


Fig. 1. VTM shrub plot locations within the Natural Communities Conservation Planning area in southern California's southwest ecoregion.

Table 1. Environmental variables used to predict plant species distributions. *<http://gis.ca.gov/catalog/BrowseRecord.epl?id=21237>

Variable	Resolution	Source	Range
Mean January minimum temperature	1 km	J. Michaelson, unpubl. data	6.10 to 7.20°C
Mean July maximum temperature	1 km	J. Michaelson, unpubl. data	24.62 to 36.39°C
Mean annual precipitation	1 km	J. Michaelson, unpubl. data	216 to 1001 mm
Mean summer solstice solar radiation	30 m	Derived from DEM (U.S. Geological Survey)	26.2 to 94.3 W h/m ²
Mean winter solstice solar radiation	30 m	Derived from the DEM	2.0 to 49.8 W h/m ²
Slope gradient	30 m	Derived from the DEM	0.04 to 66.93%
Topographic moisture index	30 m	Derived from the DEM	- 7.57 to 8.97 (unitless)
Soil type (order)	1:250 000	State Soil Geographic (STATSGO) data base for California, U.S. Department of Agriculture Natural Resources Conservation Service*	13 categories

angle, and was also calculated from the DEM (Moore et al. 1991; Wilson & Gallant 2000). We created a grid of soil type using the attribute soil order from the California State Soil Geographic Database (STATSGO).

For those environmental predictor variables whose resolution was coarser than the 300 m radius of the VTM plots (e.g., climate and soil), we assigned values through a GIS overlay technique. For variables with finer resolution (e.g., terrain), we cal-

culated the mean of those cells that fell within the extent of the 300 m plot.

Species distribution modeling procedures

We developed species distribution models using three different methods: generalized linear models (GLMs), generalized additive models (GAMs), and Random Forests (RFs) (details in Appendix S1).

Classification trees (CTs) were also tested but showed poor model performance (Appendix S1) and will not be discussed in detail. GLMs are extensions of linear regression that allow binomial distributions that are appropriate for presence/absence data to be modeled by specifying a logit link and binomial response (Venables & Ripley 1994). Residuals of GLMs were tested for spatial autocorrelation (SA). Prior to modeling, species cover was also tested for SA. Because there was minimal suggestion of SA (Appendix S3), non-spatial models were used.

GAMs allow the structure of the data to determine the shape of the response curves by replacing global regression coefficients with local smoothing functions (Yee & Mitchell 1991). CTs use binary recursive partitioning to develop classification rules that can be used to classify new observations (Breiman et al. 1984). While CTs easily handle categorical predictors and characterize variable interactions (De'ath & Fabricius 2000), they can produce very different models if the inputs are slightly varied (Edwards et al. 2006). Random Forests accounts for much of the instability of CTs by developing many (500–2000) tree models that all use random subsets of the observations and the predictor variables, and averages them (Breiman 2001).

We evaluated 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. The deviance explained in exploratory simple models developed for a subset of species was consistently highest for the three climate variables, intermediate for the terrain variables, and lowest for soil type. We used these rankings to establish the order to enter the variables and automated backward stepwise procedures to select final GLM and GAM models for each species.

We used bootstrapping (Wintle et al. 2005), resampling the data and building a series of models (500) that optimize the estimate of predictive performance, for GLMs and GAMs. The performance evaluation measure used was the area under the curve (AUC) for receiver-operating characteristic (ROC) plots (Hanley & McNeil 1982); Sensitivity

and specificity are also reported (Appendix S1). The AUC (ranging from 0.5 to 1.0) is interpreted as the probability that, for a randomly selected set of presence–absence observations, the model prediction for the presence observation will be higher than the prediction for the absence observation. For Random Forests models, we developed 500 trees and evaluated three randomly selected variables for each tree, as recommended by Breiman (2001), who suggested that the square root of the number of variables gives optimum results. To calculate the AUC, we used the averaged ‘out-of-bag’ predictions from the models. All modeling was carried out in R 2.7.0 (R Development Core Team, Vienna, Austria).

Analysis of species characteristics affecting model performance

The demographic attributes of the native shrub species in southern California are closely related to their fire disturbance response strategy (Keeley 1986). Yet, there are additional characteristics of shrubs and sub-shrubs that make them, and their distributional patterns, different from each other. In general, sub-shrubs are less shade tolerant and more sensitive to fire than shrubs; they mature early and have high rates of establishment (particularly on more xeric sites), and they recruit continuously between fires (Westman 1982; Zedler 1995; DeSimone & Zedler 2001). Like facultative seeders, these sub-shrubs respond to fire both by resprouting and seeding. However, most facultative seeder shrubs have fire-cued seed germination, whereas seed germination in sub-shrubs is rarely fire-cued.

Considering natural groupings of species life form and fire response strategy, we used a fire disturbance response classification, similar to others developed in the region (e.g., Franklin et al. 2001b; Syphard et al. 2006), to evaluate the relative effects of these unique combinations of life form and fire response on prediction accuracy (Table 2). We also included perennial herbaceous species as a category for analysis. Although the plant species in our study separate well into fire disturbance response types,

Table 2. Species disturbance response and range traits.

Traits	Description (States, Units, Range of values)
Dispersal syndrome	Ballistic, Gravity, Vertebrate-dispersed, Wind-dispersed (in order of shortest to longest dispersal distance)
Longevity	Years. 10–200.
Disturbance response	shrubFac = facultative seeder shrub; shrubOS = obligate seeder shrub; shrubOR = obligate resprouters shrub; subshrFac = resprouting sub-shrub; sub-shrub S = post-fire seeding sub-shrub; perherb = perennial herb
Range size	Number of ecoregions in which species occurs. Small (1–7 ecoregions), Medium (8–18), Large (19–29)
Prevalence	Proportion of plots where a species occurred. Percentage. 1–53%
Cover	Average cover in plots where species occurred. Percentage. 4–49%

some demographic characteristics, such as seed dispersal and longevity, may vary among species within any of these groups. Therefore, we also evaluated these traits separately in our analysis (Table 2; Appendix S2). Because there is very little information in the literature about the effective dispersal distances of our species, we grouped species according to their primary or least-distance limiting seed dispersal mechanism (i.e., we expected the dispersal distances to vary from shortest to longest by species that are ballistic, gravity, vertebrate-dispersed, or wind-dispersed, respectively). Despite these general expectations, we recognize that, in some instances, the effective dispersal distances of ballistic or gravity-dispersed species could be equally short, and the distances of animal- or wind-dispersed species could be equally long (Parker & Kelly 1989).

One common approach used when studying the effect of species range characteristics on SDMs is to classify species into different categories (Zimmermann et al. 2007; Franklin et al. 2009). In this study, we analyzed ordinal or continuous measures of geographic range, species abundance and prevalence within the study area to explore their effects independently (for example see also Brotons et al. 2004; Segurado & Araújo 2004; Luoto et al. 2005; Elith et al. 2006). We determined species abundance in plots from relative cover data recorded during the VTM survey. Species prevalence was based on the frequency of species presence (proportion of plots where a species occurred). To determine range size, we calculated the number of ecoregions in which the species had been recorded (as defined in Hickman 1993), and then classified the species as having a small (1–7 ecoregions), medium (8–18 ecoregions), or large (19–29 ecoregions) range (Table 2; Appendix S2).

The performance measure of each model for each species (i.e., the AUC) was the response variable used to address the hypotheses. To investigate the effect of species ecological and range characteristics, and model type, on prediction accuracy, we first estimated simple regression models for each explanatory variable. These included fire response type, species longevity, cover, dispersal type, range size, prevalence (Table 2), and model type. We also developed a simple model using species as the predictor for comparison purposes. Thus, the data consisted of performance measures for three models for 45 species as the response, and the associated species traits as predictors. Because explained variance, the coefficient of determination R^2 , is well established and easy to interpret, we report the results of simple linear models for this analysis.

After developing and assessing the simple models for each of the explanatory variables, we examined all predictors for multicollinearity using correlation coefficients, analysis of variance, and contingency table analysis for continuous, continuous versus categorical, and categorical versus categorical variables, respectively. We then estimated a multiple regression model based on uncorrelated predictors (and excluding species as a factor). We included the variables in the final model in order of the amount of variation they explained in the simple models, and we only retained those variables that were significant at $P < 0.05$ (Quinn & Keough 2002). The final model allowed us to further investigate the research hypotheses by identifying which factors affected the strength of species–environment correlations once other factors had been accounted for. Although the response variable (AUC) was approximately normally distributed, because it is truncated (ranges 0.5–1.0), the residuals of the regression model were examined to ensure normality, e.g., that the assumptions of the linear model were met.

Results

AUCs were low (< 0.7) for only six of the species, and generally the SDMs were able to distinguish species presence from absence based on the environmental predictors used (Table 1; Appendix S1). Residuals of the GLMs showed significant autocorrelation in only seven of 45 cases (Appendix S3). The non-parametric GAMs and decision trees would be even less sensitive to the effects of spatial autocorrelation. Fire disturbance response type explained more variation in AUC ($R^2 = 0.25$) than any

Table 3. Simple regression models of SDM performance (AUC is the response variable) as a function of each predictor variable; R^2 and P -values of the F-statistic for each simple model. $N = (45 \times 4)$ 180 models including CTs, except ‘Model Type’ is without CT models ($N = 135$). *The model including the quadratic term for cover had a significantly better fit than the simple cover model based on analysis of variance of these nested models ($F = 17.5$; $P = 5.2e-05$)

Variable	R^2	P -value
Species	0.948	< 0.001
Fire disturbance response type	0.253	< 0.001
Longevity	0.186	< 0.001
Cover	0.045	0.008
Cover + cover ² *	0.148	< 0.001
Dispersal	0.136	< 0.001
Range	0.113	< 0.001
Model type	< 0.001	0.865
Prevalence	0.035	0.017

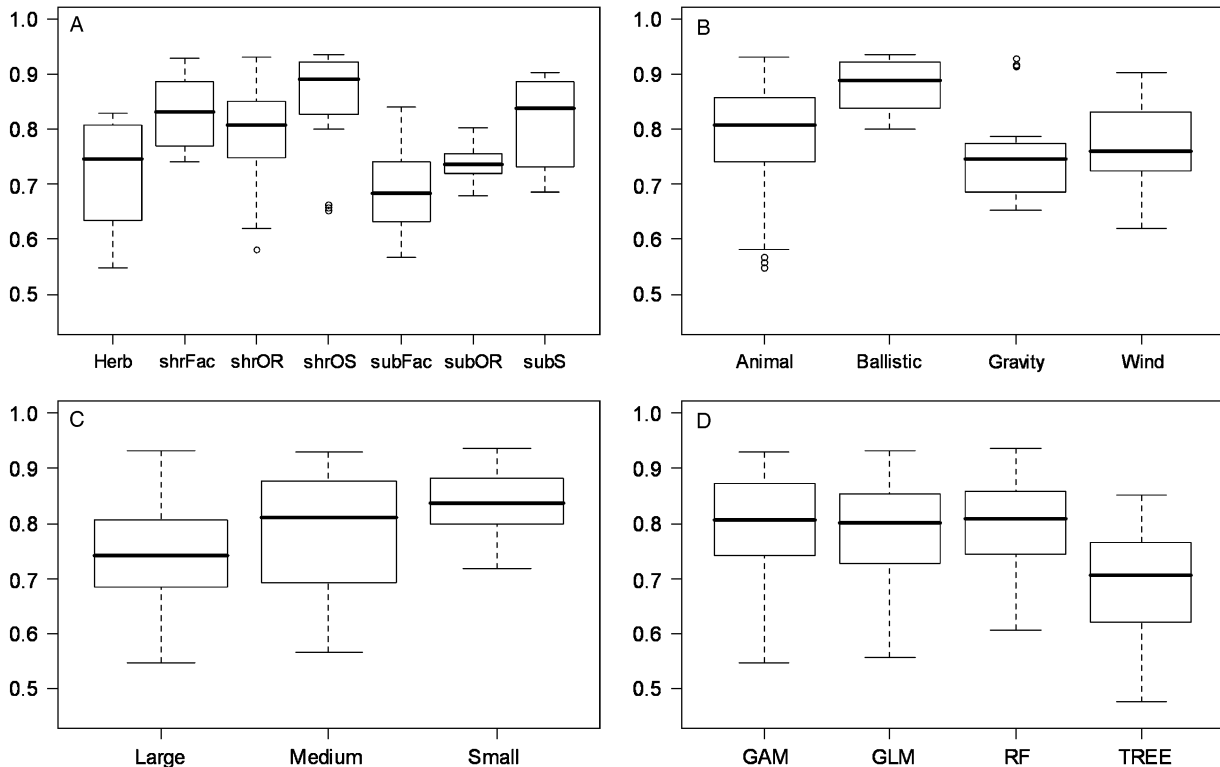


Fig. 2. Boxplots for 45 plant species and three model types showing: (a) AUC (prediction accuracy) versus disturbance response type, shrFac = facultative seeder shrub; shrOS = obligate seeder shrub; shrOR = obligate resprouters shrub; subFac = resprouting sub-shrub; sub S = post-fire seeding sub-shrub; Herb = perennial herb; (b) AUC versus dispersal mechanism; (c) AUC versus range size; and (d) AUC versus model type; GAM = generalized additive model; GLM = generalized linear model; RF = Random Forests; TREE = classification tree.

other variable (except when individual species are considered; Table 3). Overall, the obligate seeder shrub models performed best (Fig. 2a). However, there was substantial variability within some fire disturbance response types, especially herbaceous perennials and post-fire seeding sub-shrubs.

The two species demographics traits that we analyzed, longevity and dispersal mechanism, also explained substantial variation in model accuracy ($R^2 = 0.19$ and $R^2 = 0.14$, respectively). As predicted, there was a positive relationship between AUC and longevity (Fig. 3a). With regard to dispersal mechanism, AUC was substantially higher for ballistic dispersal than for the other types, and the differences between vertebrate-dispersed, gravity, and wind were minimal (Fig. 2b).

Of the range characteristics evaluated, species cover (including a quadratic term) and range size explained more variation in AUC than prevalence ($R^2 = 0.15$ and 0.11 , versus 0.03 , respectively). The relationship between species cover and AUC was nonlinear (Table 3), so low cover had a positive influence on prediction accuracy up to a point at which cover became negatively related to accuracy (Fig. 3b).

The range size of the species had an inverse relationship with model performance: the mean AUC of species with large range sizes was 0.75 , whereas that for models for small-range species was 0.83 (Fig. 2c). AUC was weakly negatively related to species prevalence (Fig. 3c). Classification trees had low mean prediction accuracy (AUC 0.69 ; Fig. 2d); when CTs are ignored, there was no significant effect of model type on accuracy, and the mean AUC for the other three model types was almost the same (0.78 – 0.79).

Correlation analysis of the predictors indicated that longevity ($F = 24.0$, $P < 0.001$), prevalence ($F = 9.56$; $P < 0.001$), and dispersal mechanism (Chi-square = 129.8 ; $P < 0.001$) differed significantly among fire disturbance response types. Further, cover, longevity, prevalence, and dispersal mechanism were intercorrelated (details not shown). Therefore, only the uncorrelated variables, fire disturbance response, cover and its squared term, and range size, were used in the multiple predictor model.

When all of the significant variables were entered in a multiple regression, the model was highly significant, with an adjusted R^2 of 0.45 . After accounting for fire disturbance response type, cover

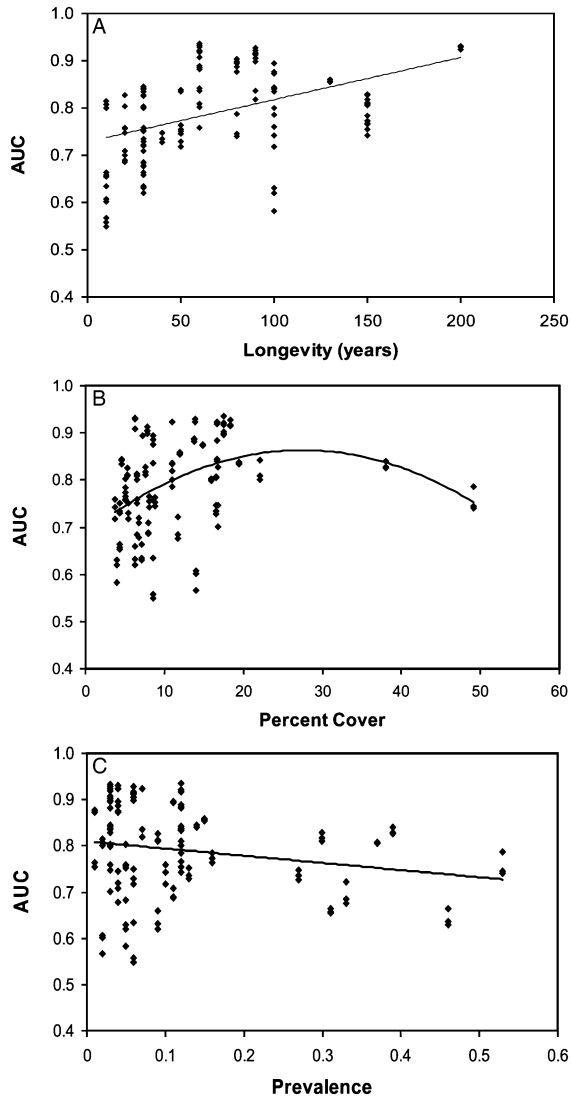


Fig. 3. Scatterplot of AUC (prediction accuracy) versus (a) species longevity; (b) percentage cover, and (c) prevalence for 45 plant species and three model types.

and range size were still significant. Obligate seeding shrubs and facultative seeding shrubs in particular had higher AUC than other functional types. Cover remained unimodally related to AUC, and species with small ranges had higher AUC (details in Appendix S4). Regression diagnostics showed that the residuals of this model were approximately normally distributed, with no indication of heteroscedasticity.

Discussion

The ability to model species–environment relationships varied widely among species in this study. The effect of individual species explained

95% of the variability in model prediction accuracy. However, classifying species into fire disturbance response types and exploring how accuracy varied according to ecological and range characteristics provided an understanding of why the species effect was so strong. Fire disturbance response type explained prediction accuracy better than any other variable (except species), supporting Hypothesis 1. Therefore, just as the classification of species into fire disturbance-related functional types has improved the prediction of community composition in dynamic landscapes (Pausas 1999), this kind of classification may also provide a strong framework for evaluating species–environment relationships and evaluating model performance in SDM.

McGill et al. (2006) raised a challenge to link species functional traits to their potential distributions on environmental gradients (their fundamental niches), and then to examine how species with different traits interact along those gradients to determine the realized niche. They proposed that this approach will make community ecology a more predictive science by identifying how communities might respond to e.g. climate change depending on whether species prefer distinct regions of niche space (distinct preference) or whether species abundance is controlled by trade-offs between dominance and tolerance. Ackerly (2003) also presented a conceptual model linking distributions in the realized niche to functional traits as a result of phenotypic plasticity, adaptive evolution, and ecological sorting.

Our study addressed, in a descriptive way, the link between species disturbance response traits and the predictability of the realized environmental niche. Specifically, we examined those plant disturbance response types that define distinct strategies for survival and reproduction in fire-prone ecosystems, and which are defined by co-varying traits such as degree of woodiness, longevity, refractory seeds, seed dormancy, potential seed dispersal distance (dispersal agents), resprouting ability, and shade tolerance.

As expected, the chaparral shrub fire disturbance response types, those most dependent on fire disturbance for reproduction, had the most predictable realized environmental niches. In particular, SDMs for the obligate and facultative seeders, which both have fire-cued seed germination, are poor long-distance dispersers, and rarely recruit in the absence of fire, outperformed those for the obligate resprouters that primarily regenerate vegetatively but can disperse seeds over long distances and recruit from seeds between fire events. Because of the higher frequency of sexual reproduction for the seeders, these species have likely adapted their

physiological and anatomical tolerances to drought and higher insolation than the resprouters (Keeley 1998); meaning they may have narrower environmental tolerances and thus, greater predictability. The seeders may also be poor competitors (Mahall & Schlesinger 1982), as well as tolerant of a narrower range of fire frequencies than the resprouters (Syphard et al. 2006).

Fire disturbance response types whose distributions are more transient in the dynamic fire-prone landscape, those that are early successional (shade intolerant), shorter lived, and less woody, had less predictable distributions than the chaparral shrubs (supporting Hypothesis 1.2). Distribution of these sub-shrubs and perennial herbs may have been less predictable because they have less specialized environmental tolerances; however, another reason for their poorer SDM performance is they may have been less detectable during the vegetation survey. Other studies have documented better SDM performance for more detectable species (Seoane et al. 2005; Poyry et al. 2008).

Although fire disturbance response type explained the most variability in model accuracy, longevity and seed dispersal mechanism were also significant. We expected accuracy to be higher for longer-lived species because they tend to persist in suitable sites throughout their range and are thus more detectable (Hypothesis 2.1), and we expected that species with the shortest potential dispersal distances (ballistic or gravity mechanisms) would have the best model performance, presumably due to better site persistence (Hypothesis 2.2). The simple regression results confirmed our expectations with longevity, and species with ballistic dispersal (the mechanism that typically results in the shortest dispersal distances) performed much better than those with other mechanisms, including gravity.

One potential reason for the effect of ballistic dispersal was that the only species that dispersed through this mechanism were obligate seeders, and the obligate seeders had better model performance than the obligate resprouters. Both longevity and seed dispersal mechanism were correlated with fire disturbance response type, and they were not included in the multiple regression model because their effects would be accounted for in the fire disturbance response classification. Thus, the significance of fire disturbance response type suggests that, at least in fire-prone regions, SDM performance may be related to the way a suite of species traits co-vary along environmental gradients. Such a classification may therefore serve as a useful frame-

work for evaluating the effect of species on SDM performance.

With regard to our second hypothesis, our results also supported our expectation that model accuracy varied more as a function of species characteristics than by model type (see also Guisan et al. 2007). Although accuracy varied somewhat among model types, the only significant difference was that classification tree models had lower accuracy than the other three methods. CTs are considered to be unstable, which is part of the reason for development of ensemble methods like Random Forests (Breiman 2001), and have also been shown in other studies to have lower prediction accuracy than GLMs, GAMs, or ensemble tree methods (Thuiller et al. 2003; Moisen et al. 2006). Although new developments in modeling methods have been instrumental in advancing the accuracy and ecological validity of SDMs, our results suggest that the effect of species on model accuracy will nevertheless override any improvements due to modeling technique.

Although fire disturbance response explained the most variation in prediction accuracy, the significance of range size and cover in this study suggest that a species geographical distribution, or rarity, can affect the variability or noise in the data used to build the models. In particular, if a species has a limited set of environmental conditions in which it can persist, and has a small range size, there may be less variability in the environmental conditions that characterize the presence observations. Although range size is not a direct measure of species environmental tolerance (Segurado & Araújo 2004; Elith et al. 2006), these factors tend to co-vary (e.g., McPherson & Jetz 2007).

Because SDM has emphasized models that predict presence versus absence, the relative proportion of such cases (the sample prevalence) has been shown to affect the resulting models. Common species usually have higher prevalence in a set of observations, unless even sampling was deliberately imposed, and therefore SDMs for these species have more 'false positive' predictions or lower specificity (Fielding & Bell 1997; Loiselle et al. 2008), while rare species with low sample prevalence have more 'false negative' predictions (Manel et al. 2001). While sample prevalence has been shown to have little effect on threshold-independent measures of prediction accuracy such as the AUC (Manel et al. 2001), other studies have suggested that the effect of sample prevalence on model performance is an artifact of sample evenness (McPherson et al. 2004). In this study, sample prevalence had a slightly negative, but not significant, effect. Therefore, the problem of

uneven sample size (low prevalence) was not apparent in our study.

Cover had a significant effect on model accuracy, but the relationship was quadratic, suggesting that at least two factors were at play. At low values, there was a positive relationship between cover and prediction accuracy, which is likely because species with very low abundance are less detectable. At higher ranges of cover, however, the relationship became negative, which is probably related to niche breadth. Cover was also correlated with longevity, dispersal mechanism, and prevalence, so in addition to detectability and niche breadth, the effect of cover may be a function of other variables as well.

In conclusion, SDM performance varied according to multiple factors. It is very useful to consider the relative importance of these factors when developing SDMs for risk assessment or resource management. While the nature of uncertainty and error in SDM is ultimately determined by model objectives, our results suggest that ecological and range characteristics of the species have a greater effect on the strength of species–environment relationships, and therefore the performance of SDMs, than the choice of model method (Barry & Elith 2006). In disturbance-prone ecosystems (regardless of the type of disturbance), disturbance response classifications that incorporate a range of species life history, life form, and disturbance response strategies can help to provide a framework for predicting which species will yield the most reliable models of species environmental response.

Acknowledgements. This work was supported by a grant from the National Science Foundation (NSF BCS-0452389) to JF, J. Miller, and R. Fisher. We thank J. Miller, L. Hierl, R. Taylor, and D. Lawson for their contributions, including comments on previous drafts of this paper, and J.E. Keeley for reviewing the species traits. We also thank F. Davis, L. Hannah, D. Stoms, and L. Ries for providing data. Comments from the anonymous reviewers greatly improved the manuscript.

References

- Ackerly, D.D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Science* 164: S165–S184.
- Ackerly, D.D., Knight, C.A., Weiss, S.B., Barton, K. & Starmer, K.P. 2002. Leaf size, specific leaf area and microhabitat distribution of chaparral woody plants: contrasting patterns in species level and community level analyses. *Oecologia* 130: 449–457.
- Austin, M.P. 1987. Models for the analysis of species' response to environmental gradients. *Vegetatio* 69: 35–45.
- Barry, S. & Elith, J. 2006. Error and uncertainty in habitat models. *Journal of Applied Ecology* 43: 413–423.
- Breiman, L. 2001. Random forests. *Machine Learning* 45: 15–32.
- Breiman, L., Friedman, J., Olshen, R. & Stone, C. 1984. *Classification and regression trees*. Wadsworth, Belmont, CA.
- Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. 2004. Presence–absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27: 437–448.
- Chapman, R.R. & Crow, G.E. 1981. Application of Raunkiaer's life form system to plant species survival after fire. *Bulletin of the Torrey Botanical Club* 108: 472–478.
- Davis, F.W. & Goetz, S. 1990. Modeling vegetation pattern using digital terrain data. *Landscape Ecology* 4: 69–80.
- De'ath, G. & Fabricius, K.E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81: 3178–3192.
- DeSimone, S.A. & Zedler, P.H. 2001. Do shrub colonizers of southern California grassland fit generalities for other woody colonizers? *Ecological Applications* 11: 1101–1111.
- Diaz, S., Cabido, M. & Casanoves, F. 1998. Plant functional traits and environmental filters at a regional scale. *Journal of Vegetation Science* 9: 113–122.
- Dubayah, R. & Rich, P.M. 1995. Topographic solar radiation for GIS. *International Journal of Geographical Information Systems* 9: 405–419.
- Edwards, T.E. Jr., Cutler, D.R., Zimmermann, N.E., Geiser, L. & Moisen, G.G. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecological Modelling* 199: 132–141.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- Fielding, A. & Bell, J. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38–49.
- Franklin, J. 2002. Enhancing a regional vegetation map with predictive models of dominant plant species in chaparral. *Applied Vegetation Science* 5: 135–146.
- Franklin, J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-

- derived variables. *Journal of Vegetation Science* 9: 733–748.
- Franklin, J. 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19: 474–499.
- Franklin, J., Keeler-Wolf, T., Thomas, K., Shaari, D.A., Stine, P., Michaelsen, J. & Miller, J. 2001a. Stratified sampling for field survey of environmental gradients in the Mojave Desert Ecoregion. In: Millington, A., Walsh, S. & Osborne, P. (eds.) *GIS and remote sensing applications in biogeography and ecology*. pp. 229–253. Kluwer Academic Publishers, Dordrecht, NL.
- Franklin, J., McCullough, P. & Gray, C. 2000. Terrain variables used for predictive mapping of vegetation communities in southern California. In: Wilson, J. & Gallant, J. (eds.) *Terrain Analysis: Principles and Applications*. pp. 331–353. Wiley & Sons, New York, NY, US.
- Franklin, J., Wejnert, K.E., Hathaway, S.A., Rochester, C.J. & Fisher, R.N. 2009. Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California. *Diversity and Distributions* 15: 167–177.
- Franklin, J., Syphard, A.D., Martin, R.M., Mladenoff, D.J., He, H.S., Simons, D.K. & Deutschman, D. 2001b. Simulating the effects of different fire regimes on plant functional types in southern California. *Ecology Modelling* 142: 261–283.
- Gillison, A.N. & Carpenter, G. 1997. A generic plant functional attribute set and grammar for dynamic vegetation description and analysis. *Functional Ecology* 11: 775–783.
- Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R. & Hastie, T. 2006. Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology* 43: 386–392.
- Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monographs* 77: 615–630.
- Hanes, T.L. 1971. Succession after fire in the chaparral of Southern California. *Ecological Monographs* 41: 27–52.
- Hanley, J.A. & McNeil, B.J. 1982. The meaning and use of the area under a receiver operating characteristics curve. *Radiology* 143: 29–36.
- Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. 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*. University of California Press, Berkeley, CA.
- Keeley, J.E. 1981. Reproductive cycles and fire regimes. In: Proceedings of the Conference Fire Regimes and Ecosystem Properties, General Technical Report WO-26, pp. 231–277. USDA, Washington, US.
- Keeley, J.E. 1986. Resilience of Mediterranean shrub communities to fire. In: Dell, B., Hopkins, A.J.M. & Lamont, B.B. (eds.) *Resilience in Mediterranean-type ecosystems*. pp. 95–112. Dr W. Junk, Dordrecht, NL.
- Keeley, J.E. 1998. Coupling demography, physiology and evolution in chaparral shrubs. In: Rundel, P.W., Montenegro, G. & Jaksic, F.M. (eds.) *Landscape degradation and biodiversity in Mediterranean-type ecosystems*. pp. 237–264. Springer-Verlag, Berlin, DE.
- Keeley, J.E. 2000. Chaparral. In: Barbour, M. & Billings, W.D. (eds) *North American terrestrial vegetation*, pp. 204–253. Cambridge University Press, Cambridge, UK.
- Keeley, J.E. 2004. VTM plots as evidence of historical change: goldmine or landmine? *Madroño* 51: 372–378.
- Keeley, J.E. & Davis, F.W. 2007. Chaparral. In: Barbour, M.G., Keeler-Wolf, T. & Schoenherr, A.A. (eds.) *Terrestrial Vegetation of California*. pp. 339–366. University of California Press, Los Angeles, CA, US.
- Kelly, M., Allen-Diaz, B. & Kobzina, N. 2005. Digitization of a historic dataset: the Wieslander California vegetation type mapping project. *Madroño* 52: 191–201.
- Kuhn, I., Bierman, S.M., Durka, W. & Klotz, S. 2006. Relating geographical variation in pollination types to environmental and spatial factors using novel statistical methods. *New Phytologist* 172: 127–139.
- Loiselle, B.A., Jorgensen, P.M., Consiglio, T., Jimenez, I., Blake, J.G., Lohmann, L.G. & Montiel, O.M. 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography* 35: 105–116.
- Luoto, M., Poyry, J., Heikkinen, R.K. & Saarinen, K. 2005. Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology & Biogeography* 14: 575–584.
- Maggini, R., Lehmann, A., Zimmermann, N.E. & Guisan, A. 2006. Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography* 33: 1729–1749.
- Mahall, B.E. & Schlesinger, W.H. 1982. Effects of irradiance on growth, photosynthesis and water use efficiency of seedlings of the chaparral shrub *Ceanothus megacarpus*. *Oecologia* 54: 291–299.
- Malanson, G.P. & O'Leary, J.F. 1995. The coastal sage scrub-chaparral boundary and response to global climate change. In: Moreno, J.M. & Oechel, W.C. (eds), *Global change and Mediterranean ecosystems*, pp. 203–224. Springer-Verlag, New York, NY, US.
- Manel, S., Williams, H.C. & Ormerod, S.J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38: 921–931.
- McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. 2006. Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution* 21: 178–185.

- McPherson, J.M. & Jetz, W. 2007. Effects of species' ecology on the accuracy of distribution models. *Ecography* 30: 135–151.
- McPherson, J.M., Jetz, W. & Rogers, D.J. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology* 41: 811–823.
- Meentemeyer, R., Moody, A. & Franklin, J. 2001. Landscape-scale patterns of shrub-species abundance in California chaparral: the role of topographically mediated resource gradients. *Plant Ecology* 156: 19–41.
- Miller, J., Franklin, J. & Aspinall, R. 2007. Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling* 202: 225–242.
- Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmerman, N.E. & Edwards Jr., T.C. 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models and tree-based methods. *Ecological Modelling* 199: 176–187.
- Moisen, G.G. & Frescino, T.S. 2002. Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling* 157: 209–225.
- Moore, I.D., Grayson, R.B. & Ladson, A.R. 1991. Digital terrain modeling: a review of hydrological, geomorphologic and biological applications. *Hydrology Processes* 5: 3–30.
- Moran, P.A.P. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society B* 10: 243–251.
- Noble, I.R. & Gitay, H. 1996. A functional classification for predicting the dynamics of landscapes. *Journal of Vegetation Science* 7: 329–336.
- Parker, V.T. & Kelly, V.R. 1989. Seed banks in California chaparral and other Mediterranean climate shrublands. In: Leck, M.A., Parker, V.T. & Simpson, R.L. (eds.) *Ecology of Soil Seed Banks*. pp. 231–255. Academic Press, New York, NY, US.
- Pausas, J.G. 1999. Response of plant functional types to changes in the fire regime in Mediterranean type ecosystems: a simulation approach. *Journal of Vegetation Science* 10: 717–722.
- Pausas, J.G., Bradstock, R.A., Keith, D.A. & Keeley, J.E. 2004. Plant functional traits in relation to fire in crown fire ecosystems. *Ecology* 85: 1085–1110.
- Pausas, J.G. & Lavorel, S. 2003. A hierarchical deductive approach for functional types in disturbed ecosystems. *Journal of Vegetation Science* 14: 409–416.
- Poyry, J., Luoto, M., Heikkinen, R.K. & Saarinen, K. 2008. Species traits are associated with the quality of bioclimatic models. *Global Ecology & Biogeography* 17: 403–414.
- Quinn, G.P. & Keough, M.J. 2002. *Experimental design and data analysis for biologists*. Cambridge University Press, Cambridge, UK.
- Rabinowitz, D. 1981. Seven forms of rarity. In: Synge, H. (ed.) *The biological aspects of rare plant conservation*. pp. 205–217. Wiley, Chichester, UK.
- Raunkiaer, C. 1934. *The life forms of plants and statistical plant geography*. Clarendon, Oxford, UK.
- Rusch, G.M., Pausas, J.G. & Leps, J. 2003. Plant functional types in relation to disturbance and land use: introduction. *Journal of Vegetation Science* 14: 307–310.
- Scott, J.M., Heglund, P.J., Morrison, M.L., Hafler, J.B., Raphael, M.G., Wall, W.A. & Samson, F.B. 2002. *Predicting species occurrences: issues of accuracy and scale*. Island Press, Covelo, CA, US.
- Segurado, P. & Araújo, M.B. 2004. An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31: 1555–1568.
- Seoane, J., Carrascal, L.M., Alonso, C.L. & Palomino, D. 2005. Species-specific traits associated to prediction errors in bird habitat suitability modelling. *Ecological Modelling* 185: 299–308.
- Smith, T.M., Shugart, H.H. & Woodward, F.I. 1997. *Plant functional types*. Cambridge University Press, Cambridge, UK.
- Stockwell, D. & Peterson, A.T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148: 1–13.
- Syphard, A.D., Franklin, J. & Keeley, J.E. 2006. Simulating the effects of frequent fire on southern California coastal shrublands. *Ecological Applications* 16: 1744–1756.
- 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. Unpublished Ph.D. thesis, Geography Department, University of California, Santa Barbara.
- Thuiller, W., Araújo, M.B. & Lavorel, S. 2003. Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science* 14: 669–680.
- Vayssières, M.P., Plant, R.E. & Allen-Diaz, B.H. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* 11: 679–694.
- Venables, W.M. & Ripley, B.D. 1994. *Modern applied statistics with S-Plus*. Springer-Verlag, New York, NY, US.
- Westman, W.E. 1982. Coastal sage scrub succession. In: Conrad, C.E. & Oechel, W.C. (eds.) *Proceedings of the symposium on dynamics and management of Mediterranean-type ecosystems*. pp. 91–99. USDA Forest Service, Pacific Southwest Forest and Range Experiment Station.
- Wieslander, A.E. 1935a. First steps of the forest survey in California. *Journal of Forestry* 33: 877–884.
- Wieslander, A.E. 1935b. A vegetation type map of California. *Madroño* 3: 140–144.
- Wilson, J. & Gallant, J. 2000. *Terrain analysis: principles and applications*. John Wiley & Sons, New York, NY, US.

- Wintle, B.A., Elith, J. & Potts, J.M. 2005. Fauna habitat modelling and mapping: a review and case study in the lower hunter central coast region of NSW. *Austral Ecology* 30: 719–738.
- Yee, T.W. & Mitchell, N.D. 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science* 2: 587–602.
- Zedler, P.H. 1995. Plant life history and dynamic specialization in the chaparral/coastal sage scrub flora in southern California. In: Arroyo, M.T.K., Zedler, P.A. & Fox, M.D. (eds.) *Ecology and Biogeography of Mediterranean Ecosystems in Chile, California, and Australia*. pp. 89–115. Springer-Verlag, New York, NY, US.
- Zimmermann, N.E., Edwards, T.C., Moisen, G.G., Frescino, T.S. & Blackard, J.A. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology* 44: 1057–1067.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Details of species distribution modeling procedures and all model performance measures reported for each species.

Appendix S2. Study species traits.

Appendix S3. Detailed analysis of spatial auto-correlations in the response variable and GLM residuals.

Appendix S4. Coefficients and *P*-values for variables in the multiple regression model for model accuracy.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received 4 August 2008;

Accepted 24 September 2009.

Co-ordinating Editor: V. Pillar.