

MACROECOLOGICAL  
METHODS



# Spatial regression methods capture prediction uncertainty in species distribution model projections through time

Alan K. Swanson<sup>1\*</sup>, Solomon Z. Dobrowski<sup>1</sup>, Andrew O. Finley<sup>2</sup>, James H. Thorne<sup>3</sup> and Michael K. Schwartz<sup>4</sup>

<sup>1</sup>Department of Forest Management, College of Forestry and Conservation, University of Montana, Missoula, MT, USA, <sup>2</sup>Department of Forestry, Michigan State University, East Lansing, MI, USA, <sup>3</sup>Information Center for the Environment, University of California, Davis, Davis, CA, USA, <sup>4</sup>USDA Forest Service, Rocky Mountain Research Station, Missoula, MT, USA

## ABSTRACT

**Aim** The uncertainty associated with species distribution model (SDM) projections is poorly characterized, despite its potential value to decision makers. Error estimates from most modelling techniques have been shown to be biased due to their failure to account for spatial autocorrelation (SAC) of residual error. Generalized linear mixed models (GLMM) have the ability to account for SAC through the inclusion of a spatially structured random intercept, interpreted to account for the effect of missing predictors. This framework promises a more realistic characterization of parameter and prediction uncertainty. Our aim is to assess the ability of GLMMs and a conventional SDM approach, generalized linear models (GLM), to produce accurate projections and estimates of prediction uncertainty.

**Innovation** We employ a unique historical dataset to assess the accuracy of projections and uncertainty estimates from GLMMs and GLMs. Models were trained using historical (1928–1940) observations for 99 woody plant species in California, USA, and assessed using temporally independent validation data (2000–2005).

**Main conclusions** GLMMs provided a closer fit to historic data, had fewer significant covariates, were better able to eliminate spatial autocorrelation of residual error, and had larger credible intervals for projections than GLMs. The accuracy of projections was similar between methods but GLMMs better quantified projection uncertainty. Additionally, GLMMs produced more conservative estimates of species range size and range size change than GLMs. We conclude that the GLMM error structure allows for a more realistic characterization of SDM uncertainty. This is critical for conservation applications that rely on honest assessments of projection uncertainty.

## Keywords

California, climate change, conservation planning, GLM, GLMM, historic data, species distribution models, transferability, uncertainty.

\*Correspondence: Alan Swanson, College of Forestry and Conservation, University of Montana, Missoula, MT 59812, USA.  
E-mail: alan.swanson@umontana.edu

## INTRODUCTION

Correlative species distribution models (SDMs) are used to predict how changes in climate will affect the spatial configuration of suitable habitat for a given species, allowing projections to be made under a variety of scenarios (Elith & Leathwick, 2009). SDMs are increasingly used for conservation planning and climate adaptation applications such as assisted migration and identifying locations suitable for reserves (Pearce & Linden-

mayer, 1998; Araújo *et al.*, 2004; Vitt *et al.*, 2009; Carroll *et al.*, 2010). Sound decisions require careful consideration of the uncertainties inherent in these projections (Burgman *et al.*, 2005; Rocchini *et al.*, 2011), yet the uncertainty associated with SDM projections, although acknowledged to be large, is poorly understood and rarely considered in applications (Elith *et al.*, 2002; Dormann, 2007a). Reasons for this include methodological issues and a lack of temporally independent data for projection validation (Dobrowski *et al.*, 2011). Repeated calls

have been made for maps of uncertainty to be presented with results (Elith *et al.*, 2002; Burgman *et al.*, 2005; Rocchini *et al.*, 2011), and their absence has led some to question the utility of SDMs for conservation planning (Heikkinen *et al.*, 2006; Dormann, 2007a). In this study we assess the ability of a spatial regression SDM method to provide a useful characterization of projection uncertainty.

The uncertainty of SDM projections is difficult to quantify given the range of contributing sources (Elith & Leathwick, 2009). Studies have shown that the choice of modelling technique introduces the greatest amount of variability in projections (Araújo *et al.*, 2005; Pearson *et al.*, 2006; Buisson *et al.*, 2010). This has led to the use of 'ensemble' methods, in which numerous models are fit using a range of methods and input data (Araújo *et al.*, 2005). Outcomes are averaged and those consistent between fitted models are deemed more reliable than those for which the models do not agree. A lack of consensus within an ensemble qualitatively suggests uncertainty, but the reasons that methods disagree are poorly understood (Burgman *et al.*, 2005).

Issues related to spatial autocorrelation (SAC) may partially explain inconsistency between methods in SDM projections. SAC arises because observations close in geographic space are generally more similar than those further apart. When a model is unable to fully explain the spatial pattern of a species' distribution, residual errors will exhibit this property, violating a key assumption of the statistical methods underlying most SDM approaches. SAC of residual error has been shown to be very common in SDM applications (Dormann, 2007b) and can easily be introduced if important covariates are missing or if a species exhibits spatial aggregation due to biotic factors. Most SDM methods are incapable of accounting for this type of error, since they consider only sampling variability and its resultant effect on the precision of parameter estimates. Although SAC has been shown not to bias parameter estimates, it has been shown to decrease their precision and lead to biased variance estimates, inflating tests of significance and thus biasing model selection procedures (Lennon, 2000; Dormann *et al.*, 2007; Beale *et al.*, 2010). This model misspecification may partially explain the disagreement between SDM methods and has been hypothesized to reduce their transferability through space and time (Randin *et al.*, 2006). Numerous methods have been proposed to correct for the adverse effects of spatial autocorrelation on SDMs (Dormann, 2007b). Generally, the focus of this research has been on methods to improve parameter estimates and tests of significance (Dormann, 2007a), and less on assessing the transferability of these models and accurately estimating projection uncertainty.

Several notable attempts have been made to quantify SDM prediction uncertainty. Buckland & Elston (1993) demonstrate a non-parametric bootstrapping approach in which numerous models are fit to permutations of the original data, resulting in maps indicating the proportion of iterations the species was predicted to be present. Hartley *et al.* (2006) present a Bayesian model averaging approach to estimating uncertainty. They fit a set of plausible models containing different covariates and

calculate uncertainty by combining between-model and within-model variability. While these approaches provide a quantitative representation of uncertainty, they do not consider the bias induced by SAC on model selection and are unable to account for uncertainty due to important covariates not considered. Other authors have presented maps of uncertainty using Bayesian spatial regression approaches (Clements *et al.*, 2006; Latimer *et al.*, 2006; Finley *et al.*, 2009a), but we are unaware of previous attempts to validate estimates of projection uncertainty using temporally independent data.

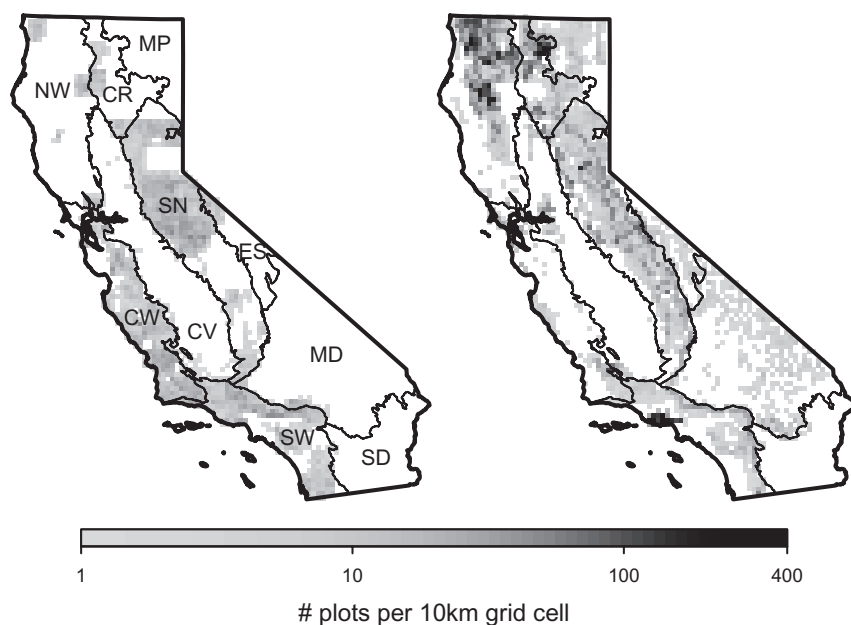
Generalized linear mixed models (GLMMs) extend generalized linear models (GLMs) to include random effects capable of accounting for additional sources of uncertainty. To account for SAC, this random effect can be specified as a spatially structured random intercept, or spatial process term, interpreted as the effects of unobserved processes with spatial structure (Diggle *et al.*, 1998). The spatially-structured random intercept has intuitive appeal in that it is able to represent the greater confidence we feel in finding a species when closer to a known presence location. The variance-covariance parameters of the random intercept control the magnitude, range and smoothness of the dependence in space, and are estimated during the model-fitting process. This avoids subjective modelling choices regarding the zone of spatial influence and allows its effect to be integrated into both parameter estimates and predictions. Spatial process GLMMs can be fit through the use of Bayesian hierarchical methods and Markov chain Monte Carlo (MCMC) techniques (Banerjee *et al.*, 2004). Although computationally intensive, this methodology provides full access to the distributions of the model's parameters given the data, i.e., posterior distributions, and the posterior predictive distributions of the response variable at unobserved locations and/or times. Latimer *et al.* (2009) and Finley *et al.* (2009a) have explored some of the utility of spatial process GLMMs (hereafter referred to as GLMMs) to model species distributions, but their projections have yet to be validated against temporally independent data. If validation shows that GLMMs are able to account for the uncertainties in modelling species distributions through time, realistic mapping of uncertainty and statistical inference on predicted range changes should be possible.

In this study we compare the ability of GLMMs with a spatially structured random intercept and non-spatial GLMs to project species distributions. We fit a suite of models for historical observations of 99 woody plant species from California, USA, and use contemporary data to assess the accuracy of projections of these models and their ability to characterize projection uncertainty through time.

## CASE STUDY

### Vegetation data

To train our models, we used presence and absence data for 99 species from 13,746 vegetation plots collected as part of the USDA Forest Service's Vegetation Type Map Project (VTM) between 1928 and 1940 (Wieslander, 1935; Thorne *et al.*, 2008)



**Figure 1** Distribution of vegetation sampling plot density (number of plots per 100 km<sup>2</sup>) for historic (left) and modern (right) periods. Text codes in the left panel are abbreviations for the ecoregions of California as defined by Hickman (1993); CR = Cascade Ranges, CV = Central Valley, CW = Central Western, ES = East of Sierras, MD = Mojave Desert, MP = Modoc Plateau, NW = Northwestern, SD = Sonora Desert, SN = Sierra Nevada, SW = Southwestern.

within the state of California, USA. Plot size was 800 m<sup>2</sup> in forests and 400 m<sup>2</sup> in other vegetation types. VTM plots were sampled in the mountainous regions of California (Fig. 1). For modern validation data, we compiled a collection of 33,596 contemporary (2000–2005) vegetation plots with presence and absence data from a variety of sources (further detail provided in Dobrowski *et al.*, 2011). Plot size in the modern data ranged from 400 m<sup>2</sup> to 800 m<sup>2</sup> in size. Vegetation plots were aggregated to 10 km by 10 km grid cells and the count of presence observations within each cell, relative to the total number of observations in that cell, was considered the response. The spatial aggregation was performed to ease computational demands and we consider this resolution adequate for a comparison between methods. Because not all species were sampled at each vegetation plot, the total number of grid cells sampled varied by species. This yielded grid cell counts for species that ranged from 825 to 1302 for the historic data and 1334 to 1929 for the modern data. Historical prevalence values ranged from 2.4% to 39.6% at the grid cell level, while modern prevalence values ranged from 0.45% to 43.7%. The historic and modern samples overlapped in 320–715 grid cells depending on species.

### Climatic data

Climatic covariates were derived from meteorological station data interpolated using the Parameter-elevation Regression on Independent Slopes Model (PRISM) (Daly *et al.*, 2008) dataset. PRISM compares favourably to other methods of climate interpolation (Daly *et al.*, 2008). PRISM data for precipitation and temperature were combined with information on geology and soils in a regional water balance model, the Basin Characteristic Model (Flint & Flint, 2007), to estimate soil water availability. Data on solar radiation, topographic shading and average cloud cover were integrated to estimate reference evapotranspiration

(ET<sub>0</sub>), actual evapotranspiration (AET), and climatic water deficit (CWD) (Flint *et al.*, unpublished data). All metrics were averaged over 30-year periods; 1911–1940 for the historic period and 1971–2000 for the modern period. For modelling purposes we selected a subset of commonly used and biologically relevant climate metrics including AET, CWD, minimum annual temperature, maximum annual temperature and annual snowfall. We removed predictors in the historic training data with correlation coefficients greater 0.85. We chose this threshold because the primary impact of collinearity is to increase variance of coefficient estimates (O'Brien, 2007), an effect that should affect both candidate models equally. The data were originally provided at a resolution of 270 m and were aggregated to 10-km resolution using a simple average.

Over the study period, the study area experienced significant changes in climate. Mean temperatures increased by approximately 1.0 °C across the state while precipitation increased in the northern half of the state resulting in spatially variable trends in climatic water balance (Dobrowski *et al.*, 2011).

### Modelling techniques

For each species we fit GLMs and GLMMs to the full historic dataset assuming a binomial distribution for the response variable and a logistic link function. We follow Latimer *et al.* (2006) in using the count of presence observations per grid cell as our response, weighted by the number of vegetation plots per grid cell. Predictions from these models reflect estimated probability of occurrence for a species within each cell, equivalent to predicted prevalence. We used quadratic functions of all five covariates to allow for nonlinear relationships between the covariates and response variables.

For the spatial models, an exponential spatial correlation function was assumed. We used a spatial *predictive process* model

to reduce the costly computations involved in estimating the spatial process (Banerjee *et al.*, 2008; Finley *et al.*, 2009b). Models were fit within a Bayesian framework using MCMC techniques. Computations were performed in R (2.10.1; R Development Core Team, 2011) using the spGLM routine in the spBayes package (Finley *et al.*, 2007). Each model required several days to complete the MCMC sampling on a quad-core server (Intel Xeon E5440 2.83 Ghz). Details about model specification and example code are included in Appendices S1 and S2 in Supporting Information.

## Model assessment

Candidate models, i.e., GLM and GLMM, were assessed using resubstituted historic training data (internal validation) and temporally independent data from the contemporary period (independent validation). For independent validation, parameter estimates from models fit to the historic data were used to make projections with the spPredict function in the spBayes library and modern climate data. The spatially varying random intercept was included in GLMM projections. For internal validation, comparisons of model fit were made using the Deviance Information Criterion (DIC; Spiegelhalter *et al.*, 2002), which is a measure of prediction accuracy with a penalty,  $p_D$ , for model complexity interpreted as the effective number of parameters. Although DIC has been criticized for a variety of theoretical and applied shortcomings (see, e.g., the discussion supplement for Spiegelhalter *et al.*, 2002), there are few alternative fit criteria suitable for hierarchical models and we feel its use for broad comparisons is reasonable. As a measure of predictive accuracy for both internal and independent validation, we used AUC (area under the receiver–operator curve), an index representing the ability of a model to discriminate between presence and absence observations (Hosmer & Lemeshow, 2000). Although AUC does not consider the calibration of predictions and required reducing our data to presence or absence within each grid cell, it remains useful for comparisons between candidate models for the same species.

To directly assess prediction uncertainty we estimated coverage rates of 90% credible intervals for probability of occurrence, derived from posterior predictive distributions for sampled grid cells. Coverage rates were calculated as the proportion of grid cells for which the observed prevalence value fell within their respective 90% credible intervals. Because a logistic link function can never return a value of zero or one, we considered intervals including 0.001 to include zero, and intervals including 0.999 to include one.

To assess both the range and significance of residual spatial dependence among the observations, we used Moran's  $I$  test based on 12 discrete distance classes. Details are given in Appendix S1.

## Range size estimates

We estimated range size as the cumulative area of cells for which the posterior predicted probability of occurrence was above a

threshold value. The threshold value was chosen to minimize the difference between sensitivity (proportion of presence observations correctly predicted) and specificity (proportion of absence observations correctly predicted) for the historic data used to fit the models. This threshold was calculated individually for each model and species. We tested the statistical significance of range size change by subtracting the posterior distributions of range size estimates for the two time periods to generate a posterior for range size change; if the 90% credible interval for this distribution excluded 0, the change was deemed significant.

In addition to estimating overall changes in range size, we identified where significant changes to the species ranges were predicted to occur. For each grid cell we compared the posterior predictive distributions in the historic period to those for projections in the modern period (see Fig. 6). From the historic posterior we calculated the probability of observing a value as extreme or more extreme than the median projected value.

## Displaying uncertainty

In order to graphically depict uncertainty in our predictions, we adapted the methods of Hengl *et al.* (2004). Median predictions for each grid cell were displayed using a colour ramp and degree of uncertainty (width of a 90% CI) was shown by increasing the whiteness of these colours.

## RESULTS

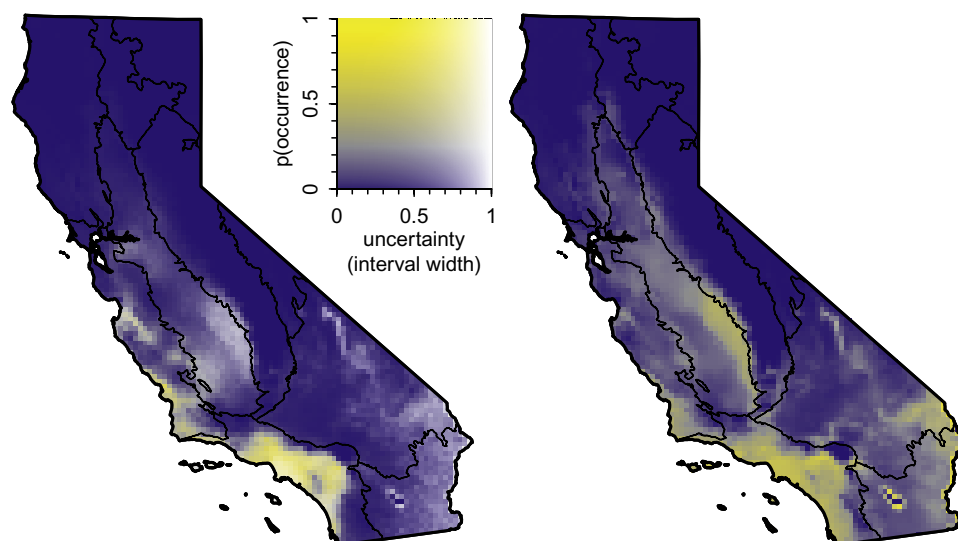
### Internal validation

Internal validation showed significant differences between model fits (Table 1 and Fig. 4). Median DIC scores dropped by 454.6 for GLMMs compared to GLMs, despite a median increase in model complexity of  $p_D = 87.5$ , suggesting a considerable improvement in fit for GLMMs over GLMs. AUC scores for GLMs had a median value of 0.88, indicating good discrimination between presence and absence observations (Swets, 1988).

**Table 1** Summary of median fit statistics on historic data (internal validation) for models fit for 99 plant species. Coverage is proportion of times a 90% credible interval for probability of occurrence contained the observed prevalence value. Range refers to the range of significant spatial autocorrelation found in binned Moran's  $I$  tests.  $p_D$  is a measure of model complexity, interpreted as the effective number of parameters in each model. DIC is the Deviance Information Criterion, lower values indicate better fit. Different letters indicate significant difference based on a matched-pairs  $t$ -test between models, adjusted for multiple comparisons following the method of Holm (1979).

|      | AUC    | Coverage | Range (km) | Moran's $I$ | $p_D$ | DIC  |
|------|--------|----------|------------|-------------|-------|------|
| GLM  | 0.88 a | 0.46 a   | 45 a       | 0.28 a      | 10.7  | 2012 |
| GLMM | 0.98 b | 0.91 b   | 0 b        | −0.02 b     | 98.2  | 1557 |

GLM, general linear model; GLMM, general linear mixed model.



**Figure 2** Example of fitted models for black sage (*Salvia mellifera*). The left panel shows predicted probability of occurrence from the spatial GLMM model. Colour indicates the prediction while the degree of whiteness indicates width of a 90% prediction interval. The right panel shows the same for the non-spatial GLM model.

GLMMs yielded a median AUC score of 0.98, indicating near-perfect discrimination between presence and absence observations. Coverage rates for GLMMs had a median value of 0.91, very close to their nominal value of 0.90, while those for GLMs had a median value of 0.46, implying overconfident predictions from the latter.

The posterior distributions of regression coefficients differed greatly between GLMMs and GLMs. Figure S1a in Appendix S1 shows an example of parameter posterior distributions for *Salvia mellifera*. Standard errors of GLMM coefficients were, on average, 2.17 times greater than that of GLM coefficients. GLMMs had fewer significant coefficients: of the 5 covariates examined, the mean number that were significant as either 1<sup>st</sup> or 2<sup>nd</sup> order (90% credible interval not including 0) was 4.5 for GLMs and 3.0 for GLMMs. GLM estimates generally fell within the 90% GLMM CI (70.4% of all parameter estimates).

The Moran statistics and range of autocorrelation given in Table 1 show that GLMMs nearly eliminated spatial autocorrelation of residual error (although 3 of the 99 species still showed significant dependence with adjacent grid cells), while all GLMs exhibited significant autocorrelation of residual error with a median range of 45 km.

### Independent validation

Temporally independent validation with modern data yielded lower mean accuracy statistics than internal validation for both GLM and GLMMs (Table 2 and Fig. 4). AUC values were slightly higher for GLMMs compared to GLMs. Coverage rates for GLMMs showed only a slight drop (compared to internal validation), remaining very close to their nominal value of 0.90 (Table 2), while those for GLMs improved but remained poor. Restricting our independent validation to those grid cells that were sampled historically had little effect on accuracy statistics but caused a slight drop in coverage rates for both candidate

**Table 2** Summary of median fit statistics on the modern data (independent validation) for models fit for 99 plant species. Coverage is proportion of times a 90% credible interval for  $p(\text{occurrence})$  contained the observed prevalence value. Letters indicate significant differences in matched-pairs  $t$ -tests, adjusted for multiple comparisons following the method of Holm (1979).

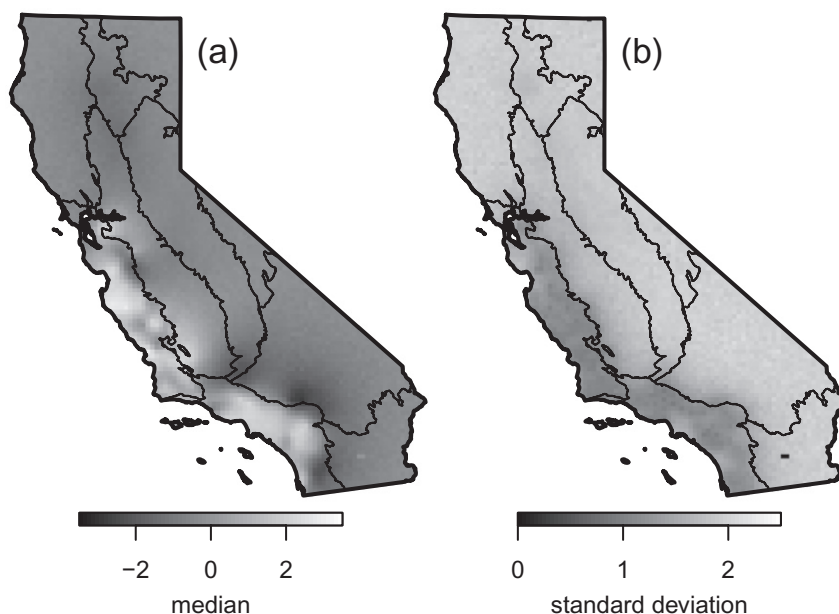
|      | AUC    | Coverage |
|------|--------|----------|
| GLM  | 0.88 a | 0.61 a   |
| GLMM | 0.89 b | 0.87 b   |

GLM, general linear model; GLMM, general linear mixed model.

models, while restricting validation to cells not sampled historically had also little effect on AUC, as was demonstrated in Dobrowski *et al.* (2011), but caused a slight increase in coverage rates for both candidate models (results not shown).

### Range size estimates and predicted changes

Mean range size estimates were correlated between time periods (Pearson correlation coefficient  $r = 0.94$  GLM,  $r = 0.99$  GLMM) and candidate models ( $r = 0.65$  historic,  $r = 0.68$  modern). Range size estimates varied by model with GLM estimates averaging c. 70% larger than GLMM estimates for both time periods (Fig. S1b in Appendix S1). Interval widths for estimated range size averaged 48.4% of range size for GLMMs vs. 25.0% for GLMs. Estimated changes in range size were also highly correlated between candidate models ( $r = 0.77$ ), but GLMM estimates predicted, on average, 50% smaller changes in range size. Figure 5 shows estimates of percentage range size change by model, highlighting estimated changes that were significant ( $\alpha = 0.10$ ). It is notable that the two models predicted similar numbers of significant changes, but in many cases failed to agree



**Figure 3** (a) Median fitted value of the GLMM spatial random intercept for *Salvia mellifera* (black sage). This can be interpreted as a latent covariate representing unobserved processes with spatial structure. Higher values indicate greater suitability than predicted by the climatic covariates included in the model. (b) Standard deviation of spatial term. This is the amount of variability added to predictions by the spatial process term.

on which species were facing these changes. Figure 6 shows an example of the spatial distribution of predicted changes in probability of occurrence for *Salvia mellifera*.

## DISCUSSION

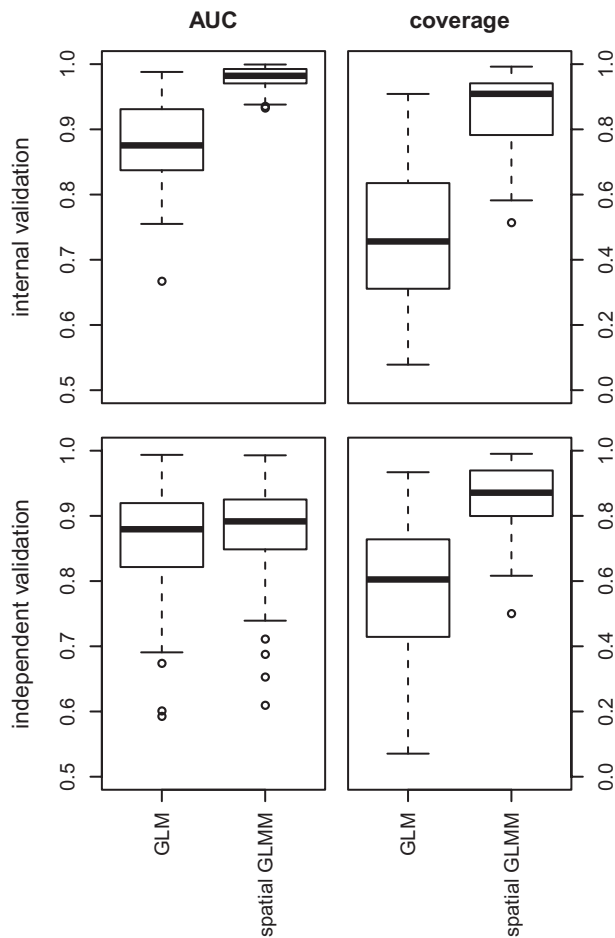
### Performance under internal vs. independent validation

GLMMs consistently outperformed GLMs under internal evaluation, but performed similarly when confronted with temporally independent data. Under internal validation, the flexibility of the spatially structured random intercept allowed it to capture spatial patterns not accounted for by our climate covariates. These patterns were smooth in space, as evidenced by the spatial autocorrelation of GLM errors and the ability of GLMMs to account for these errors. The similar performance of the candidate models under independent validation was surprising. This is apparently due to a lack of temporal persistence, for most species, of the latent effects accounted for by the spatial random intercept. In effect, many of the species' distributions shifted in ways which could not be explained by our climate covariates. From a Bayesian perspective, the spatial random intercept can be viewed as an informative prior for projections into new temporal domains – drawing the projections back toward the historic ranges when information in the covariates is lacking. If the latent effects represented by the spatial random intercept are expected to change over time, it may be desirable to specify a temporally dynamic residual spatial process, allowing the influence of the spatial random intercept to evolve over space and

time, see, e.g., Finley *et al.* (2012). To our knowledge, this methodology has not been applied to SDM projections.

### Projection uncertainty

Although the spatial random intercept did not markedly improve the projection accuracy of GLMMs, its ability to account for variability not explained by covariates yielded improved estimates of uncertainty. Including such estimates alongside mean projections gives a 'map of ignorance' as called for by Rocchini *et al.* (2011), highlighting areas where knowledge is lacking and could be improved with additional sampling effort or the inclusion of additional covariates. For instance, for *Salvia mellifera*, a historically calibrated GLM projection showed high probability of occurrence in the coastal regions of Southern California, the southern reaches of the Central Valley, and eastern portion of the Mojave desert (Fig. 2). These projections are flawed as the species does not currently occur in the latter two regions of the state. In contrast, the influence of the spatial random intercept term in the GLMM projection (Fig. 3) is readily apparent as the latter two regions of the state show lower probability of occurrence and more importantly, higher levels of uncertainty in projections to these regions (Fig. 2). In addition to improving the projections, the spatial random intercept term can provide biogeographical insights into latent covariates that can better explain the species distribution. In this case, the unobserved spatial process may be frequent disturbance from fire in the coastal sage and chaparral communities in which this species is found. *Salvia mellifera* has facultative fire adapted reproductive traits (Keeley, 1986) and although we cannot definitively prove that the spatial intercept is

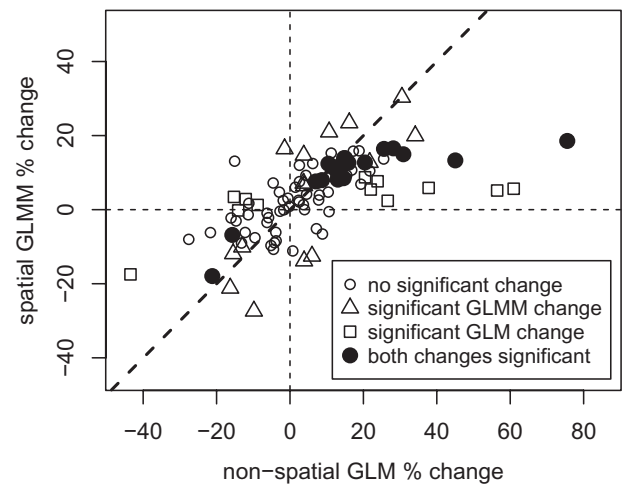


**Figure 4** Fit statistics under internal (historic data) and independent (modern data) validation. Coverage rates, shown in the right column, are the proportion of times a 90% prediction interval captured observed prevalence.

actually characterizing this latent process, this interpretation is consistent with the disturbance regime of the region and the autecology of the species.

### Conservation applications

Conservation applications of SDMs such as reserve design (Pearce & Lindenmayer, 1998; Carroll *et al.*, 2010) and assisted migration of species (Vitt *et al.*, 2009) represent costly management actions involving complex decisions for which the consequences of mistakes are high. The independently validated estimates of uncertainty we have presented have utility in this context, allowing alternatives to be assessed with regard to the confidence of projections. The results we present for *Salvia mellifera* provide a relevant hypothetical example (Fig. 2). If there were concerns over habitat loss for this species, c. 1935, then GLM results suggest the southern Central Valley and Sierra Nevada ecoregion as plausible translocation sites for assisted migration planning. However, the GLMM projection suggests that the suitability of these regions is far from certain, providing useful information to a hypothetical conservation planner.

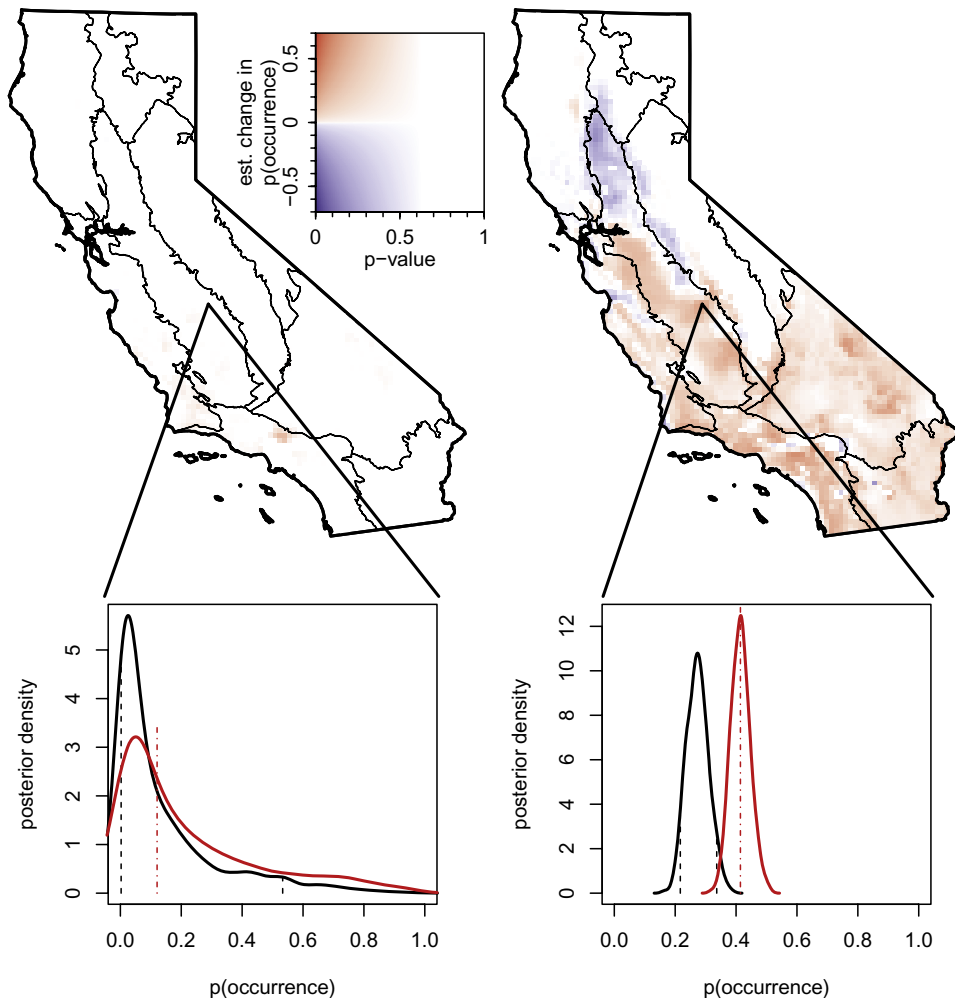


**Figure 5** Estimates of percentage change in range size over the 75-year study period for all species. Percentage change is relative to mean estimated range size for the historic period. Estimated change for GLM models is shown along the x-axis, while change for spatial GLMM models is shown on the y-axis. The thick dashed line is 1 : 1. Spatial GLMMs generally predict smaller changes in range size, and the significance of changes varies between methods.

SDMs are also used to project loss of habitat and subsequent extinction risk (Thomas *et al.*, 2004; Loarie *et al.*, 2008). Estimates of habitat loss (or gain) are driven by the shape of response curves for individual covariates, making them sensitive to model specification. In this context, spatial regression methods such as GLMMs offer a distinct advantage in that they have been shown to give more precise parameter estimates and are less likely to identify spurious covariates as significant in the presence of spatial autocorrelation (Beale *et al.*, 2010). The latter issue can be especially problematic when automated model selection techniques are used in conjunction with non-spatial SDM methods, a situation common in SDM applications. In our analysis, GLMMs yielded substantially more conservative estimates than GLMs of range size and range size change through time. This was likely due to the ability of the spatial random intercept to correctly identify areas of known absence not predicted by climate alone. Additionally, predicting a contraction or expansion of suitable habitat may be of limited use for conservation planning without regard to spatial context. We demonstrate that the posterior distributions of model projections can be used to distinguish between areas where habitat loss (or gain) is more certain compared to areas where change is less certain (Fig. 6). This type of analysis is valuable because changes occurring in areas where we have very little confidence in our original estimates should be of less concern than changes occurring in areas known to contain the focal species.

### Caveats

Numerous criticisms could be made of our methods. Weaknesses include the coarse resolution of our study, missing predictors and



**Figure 6** Estimated change in probability of occurrence over 75 years for *Salvia mellifera*. The left panel shows the spatial GLMM estimates while the right panel shows non-spatial GLM estimates. Colour ramp indicates magnitude of predicted change while degree of saturation conveys the result of a statistical test for per-pixel change in suitability over time, with darker colours indicating areas where significant change in habitat suitability. Inset plots show, for a single grid cell extracted from the central valley region, posterior distributions of predicted probability of occurrence for the two time periods and for both methods. The black lines show the posterior distribution for the historic period while the red lines show the posterior for the forecast of the historic model to the modern period. Vertical black lines show the 90% prediction intervals for the historic period, while the vertical red lines show the median value for the modern period. The width of the 90% prediction interval is analogous to that used to convey uncertainty in Fig. 2. Cases in which the modern median fell outside the 90% prediction interval for the historic period are considered significant at the 10% level. For the highlighted grid cell, the spatial GLMM did not predict a significant change while the non-spatial GLM did.

misspecification of models. We used GLMs for comparison, yet studies have shown more sophisticated methods such as generalized additive models, Random Forest and Boosted Regression Trees to produce better fitted models (e.g. Elith *et al.*, 2006). Although such methods offer many advantages, little focus has been given to their estimates of projection uncertainty, and their accuracy under spatially (Randin *et al.*, 2006) and temporally (Dobrowski *et al.*, 2011) independent validation has been questioned. The other weaknesses noted above should affect both candidate models equally, although the advantage of GLMMs would disappear under conditions in which a model is correctly specified and all relevant predictors included, conditions rarely encountered in practice (Heikkinen *et al.*, 2006; Dormann,

2007b). Finally, one might look to other approaches to assess candidate models' predictive ability, see, e.g., Gneiting & Raftery (2007) for a discussion of proper scoring rules.

## CONCLUSIONS

We found that spatial regression models, although they produced similar levels of projection accuracy under temporally independent validation, gave improved estimates of uncertainty over non-spatial methods fit to the same data. The ability of GLMMs to account for residual SAC and hence provide valid estimates of uncertainty suggests they are more suitable for drawing inference about SDM parameters and subsequent pre-

dictions. The degree of uncertainty was high in our fitted models, but their output provides valuable insight into the nature of this uncertainty and suggests ways it might be reduced. GLMM methods produced more conservative estimates of range size and range size change, and although we cannot definitely say these are more accurate than those derived from conventional methods, the statistical validity of GLMMs favours their estimates. Useful projections of species' distributions into the future require an honest assessment of projection uncertainty. GLMMs with a spatially structured random intercept offer a clear improvement over commonly used methods.

## ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation (BCS-0819430 to S.Z.D., BCS-0819493 to J.H.T., EF-1137309 and DMS-1106609 to A.O.F.), the USDA CSREES 2008-38420-19524 to A.K.S., the California Energy Commission PIER Program CEC PIR-08-006, and the USDA Forest Service Rocky Mountain Research Station (JV11221635-201). We thank Lorraine and Alan Flint for providing climate data, Jeff Braun and the Rocky Mountain Super Computing facility, and the many agencies and institutions that have collected and stewarded the historical and modern inventory data used in the analysis.

## REFERENCES

- Araújo, M.B., Cabeza, M., Thuiller, W., Hannah, L. & Williams, P.H. (2004) Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology*, **10**, 1618–1626.
- Araújo, M.B., Whittaker, R.J., Ladle, R.J. & Erhard, M. (2005) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography*, **14**, 529–538.
- Banerjee, S., Carlin, B. & Gelfand, A. (2004) *Hierarchical modeling and analysis for spatial data*. Chapman & Hall/CRC, Boca Raton, FL.
- Banerjee, S., Gelfand, A.E., Finley, A.O. & Sang, H. (2008) Gaussian predictive process models for large spatial datasets. *Journal of the Royal Statistical Society Series B*, **70**, 825–848.
- Beale, C.M., Lennon, J.J., Yearsley, J.M., Brewer, M.J. & Elston, D.A. (2010) Regression analysis of spatial data. *Ecology Letters*, **13**, 246–264.
- Buckland, S.T. & Elston, D.A. (1993) Empirical models for the spatial distribution of wildlife. *The Journal of Applied Ecology*, **30**, 478–495.
- Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2010) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145–1157.
- Burgman, M.A., Lindenmayer, D.B. & Elith, J. (2005) Managing landscapes for conservation under uncertainty. *Ecology*, **86**, 2007–2017.
- Carroll, C., Dunk, J.R. & Moilanen, A. (2010) Optimizing resiliency of reserve networks to climate change: multispecies conservation planning in the Pacific Northwest, USA. *Global Change Biology*, **16**, 891–904.
- Clements, A.C.A., Lwambo, N.J.S., Blair, L., Nyandindi, U., Kaatano, G., Kinung'hi, S., Webster, J.P., Fenwick, A. & Brooker, S. (2006) Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Tropical Medicine and International Health*, **11**, 490–503.
- Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J. & Pasteris, P.P. (2008) Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, **28**, 2031–2064.
- Diggle, P.J., Moyeed, R.A. & Tawn, J.A. (1998) Model-based geostatistics. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **47**, 299–350.
- Dobrowski, S.Z., Thorne, J.H., Greenberg, J.A., Safford, H.D., Mynsberge, A.R., Crimmins, S.M. & Swanson, A.K. (2011) Modeling plant ranges over 75 years of climate change in California, USA: relating transferability to species traits. *Ecological Monographs*, **81**, 241–257.
- Dormann, C.F. (2007a) Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, **8**, 387–397.
- Dormann, C.F. (2007b) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology & Biogeography*, **16**, 129–138.
- Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.
- Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.
- Elith, J., Burgman, M. & Regan, H. (2002) Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling*, **157**, 313–329.
- Elith, J., Graham, C., Anderson, R. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Finley, A.O., Banerjee, S. & Carlin, B.P. (2007) spBayes : an R package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software*, **19**, 1–24.
- Finley, A.O., Banerjee, S. & McRoberts, R.E. (2009a) Hierarchical spatial models for predicting tree species assemblages across large domains. *Annals of Applied Statistics*, **3**, 1052–1079.
- Finley, A.O., Sang, H., Banerjee, S. & Gelfand, A.E. (2009b) Improving the performance of predictive process modeling for large datasets. *Computational statistics and Data Analysis*, **53**, 2873–2884.

- Finley, A.O., Banerjee, S. & Gelfand, A.E. (2012) Bayesian dynamic modeling for large space-time datasets using Gaussian predictive processes. *Journal of Geographical Systems*, **14**, 29–47.
- Flint, A.L. & Flint, L.E. (2007) *Application of the basin characterization model to estimate in-place recharge and runoff potential in the Basin and Range carbonate-rock aquifer system, White Pine County, Nevada, and adjacent areas in Nevada and Utah*. US Geological Survey Scientific Investigations Report 2007–5099. Washington, DC.
- Gneiting, T. & Raftery, A.E. (2007) Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378.
- Hartley, S., Harris, R. & Lester, P.J. (2006) Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. *Ecology Letters*, **9**, 1068–1079.
- Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W. & Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **6**, 751–777.
- Hengl, T., Walvoort, D.J.J., Brown, A. & Rossiter, D.G. (2004) A double continuous approach to visualization and analysis of categorical maps. *International Journal of Geographical Information Science*, **18**, 183–202.
- Hickman, J.C. (1993) *The Jepson manual: higher plants of California*. University of California Press, Berkeley, California.
- Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.
- Hosmer, D. & Lemeshow, S. (2000) *Applied logistic regression*, 2nd edn. John Wiley & Sons, New York.
- Keeley, J.E. (1986) Seed germination patterns of *Salvia mellifera* in fire-prone environments. *Oecologia*, **71**, 1–5.
- Latimer, A.M., Wu, S., Silander, J.A. & Gelfand, A.E. (2006) Building statistical models to analyze species distributions. *Ecological applications*, **16**, 33–50.
- Latimer, A.M., Banerjee, S., Sang, H., Mosher, E.S. & Silander, J.A. (2009) Hierarchical models facilitate spatial analysis of large data sets: a case study on invasive plant species in the northeastern United States. *Ecology Letters*, **12**, 144–154.
- Lennon, J.J. (2000) Red-shifts and red herrings in geographical ecology. *Ecography*, **23**, 101–113.
- Loarie, S.R., Carter, B.E., Hayhoe, K., McMahon, S., Moe, R., Knight, C.A. & Ackerly, D.D. (2008) Climate change and the future of California's endemic flora. *PLoS ONE*, **3**, e2502.
- O'Brien, R.M. (2007) A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, **41**, 673–690.
- Pearce, J. & Lindenmayer, D. (1998) Bioclimatic analysis to enhance reintroduction biology of the endangered helmeted honeyeater (*Lichenostomus melanops cassidix*) in southeastern Australia. *Restoration Ecology*, **6**, 238–243.
- Pearson, R.G., Thuiller, W., Araújo, M.B., Martínez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P. & Lees, D.C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704–1711.
- R Development Core Team (2011) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.
- Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.
- Rocchini, D., Lobo, J.M., Jime, A., Bacaro, G. & Chiarucci, A. (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography*, **35**, 211–226.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van Der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **64**, 583–639.
- Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.
- Thomas, C., Cameron, A., Green, R., Bakkenes, M., Beaumont, L., Collingham, Y., Erasmus, B., Ferreira de Siqueira, M., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L. & Williams, S.E. (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.
- Thorne, J.H., Morgan, B.J. & Kennedy, J.A. (2008) Vegetation change over sixty years in the central Sierra Nevada, California, USA. *Madroño*, **55**, 223–237.
- Vitt, P., Havens, K. & Hoegh-guldberg, O. (2009) Assisted migration: part of an integrated conservation strategy. *Trends in Ecology and Evolution*, **24**, 473–474.
- Wieslander, A.E. (1935) A vegetation type map for California. *Madroño*, **3**, 140–144.

## SUPPORTING INFORMATION

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

**Appendix S1** Methodological details.

**Appendix S2** R code example.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

## BIOSKETCH

**Alan Swanson** is a graduate student at the University of Montana, USA. His interests lie in understanding the complex range of factors affecting the distribution of species and in statistical methods to account for uncertainty in natural systems.

Editor: José Alexandre F. Diniz-Filho