A Journal of Macroecology

Global Ecology and Biogeography

RESEARCH PAPER

Should species distribution models account for spatial autocorrelation? A test of model projections across eight millennia of climate change

Sydne Record[1]*, Matthew C. Fitzpatrick[2], Andrew O. Finley[3], Sam Veloz[4] and Aaron M. Ellison[1]

[1]*Harvard Forest, Harvard University, Petersham, MA, USA,* [2]*Appalachian Lab, University of Maryland Center for Environmental Science, Frostburg, MD, USA,* [3]*Departments of Forestry and Geography, Michigan State University, East Lansing, MI, USA,* [4]*PRBO Conservation Science, Petaluma, CA, USA*

*Correspondence: Sydne Record, Harvard Forest, Harvard University, 324 North Main Street, Petersham, MA 01366, USA.
E-mail: srecord@fas.harvard.edu

ABSTRACT

Aim The distributions of many organisms are spatially autocorrelated, but it is unclear whether including spatial terms in species distribution models (SDMs) improves projections of species distributions under climate change. We provide one of the first comparative evaluations of the ability of a purely spatial SDM, a purely non-spatial SDM and a SDM that combines spatial and environmental information to project species distributions across eight millennia of climate change.

Location Eastern North America.

Methods To distinguish between the importance of climatic versus spatial explanatory variables we fit three Bayesian SDMs to modern occurrence data for *Fagus* and *Tsuga*, two tree genera whose distributions can be reliably inferred from fossil pollen: a spatially varying intercept model, a non-spatial model with climatic variables and a spatially varying intercept plus climate model. Using palaeoclimate data with a high temporal resolution, we hindcasted the SDMs in 1000-year time steps for 8000 years, and compared model projections with palynological data for the same periods.

Results For both genera, spatial SDMs provided better fits to the calibration data, more accurate predictions of a hold-out validation dataset of modern trees and higher variance in current predictions and hindcasted projections than non-spatial SDMs. Performance of non-spatial and spatial SDMs according to the area under the receiver operating curve varied by genus. For both genera, false negative rates between non-spatial and spatial models were similar, but spatial models had lower false positive rates than non-spatial models.

Main conclusions The inclusion of computationally demanding spatial random effects in SDMs may be warranted when ecological or evolutionary processes prevent taxa from shifting their distributions or when the cost of false positives is high.

Keywords
Bayesian, eastern North America, historical validation, palaeoecology, spatial random effect.

INTRODUCTION

The last decade has witnessed a marked increase in the application of models that project the potential geographic distributions of species by linking observations of species occurrences to environmental predictor variables. These models, commonly called bioclimatic envelope, ecological niche or species distribution models (hereafter SDMs), are important

DOI: 10.1111/geb.12017
http://wileyonlinelibrary.com/journal/geb

tools for forecasting the impacts of climatic change on biological diversity and for generating conservation plans and climate change policy (Guisan & Thuiller, 2005). To project future distributions under different, plausible scenarios of climatic change, SDMs use statistical relationships between present-day distributions of species and climate (Elith *et al.*, 2010). Although generally successful at explaining and predicting current distributions of species (Franklin & Miller, 2009), impact assessments derived from SDMs have been criticized for their reliance on a number of largely untested ecological assumptions, methodological issues and statistical concerns (e.g. Pearson & Dawson, 2003; Dormann, 2007).

Chief among these issues is the failure of most SDMs to account for spatial dependence of occurrence data (Gelfand *et al.*, 2006; Bahn & McGill, 2007; Dormann, 2007; Elith *et al.*, 2010). Spatial autocorrelation arises in ecological data because nearby points tend to be more similar, in physical characteristics and/or species occurrences or abundances, than are pairs of locations that are farther apart (Legendre, 1993). When model assumptions about independent and identically distributed residuals are violated, there could be a bias in the regression parameter estimates, potentially leading to poor inference. Studies illustrate that failure to account for spatial autocorrelation can lead to misidentification of important driving variables and overly optimistic error rates (e.g. Lichstein *et al.*, 2002; Segurado *et al.*, 2006; Diez & Pulliam, 2007; Dormann, 2007), especially when small-scale patterns of explanatory variables create instability in broad-scale regression parameter estimates (Hawkins *et al.*, 2007). Further, models based solely on spatial interpolation can provide better fits to species range data than models based on explanatory environmental variables (Bahn & McGill, 2007), suggesting that spatial autocorrelation in unmeasured factors (e.g. population processes such as dispersal or underlying resources such as soil moisture) may account for most of the observed distributional patterns.

Analysis of spatial SDMs has primarily focused on predicting current or simulated species distributions using a hold-out dataset for model validation (Gelfand *et al.*, 2006; Wilson *et al.*, 2010), but projections of spatial SDMs in changing climates over long time-scales remain largely untested. Observed changes in species distributions as a result of past climatic dynamics provide a unique opportunity to compare projections of spatial and non-spatial SDMs parameterized with current conditions (Pearman *et al.*, 2008a; Nogués-Bravo, 2009; Dobrowski *et al.*, 2011; Veloz *et al.*, 2012).

Projections to environmental conditions different from those used to calibrate SDMs are subject to error (Heikkinen *et al.*, 2006) and may not be ecologically meaningful or statistically valid if there are changes in correlations between variables across time and space (Elith *et al.*, 2010) or if species–environment relationships are not conserved (e.g. Fitzpatrick *et al.*, 2007; Veloz *et al.*, 2012). It also is not known whether it is desirable to project models with spatial random effects based on the partially observed spatial distribution of a species at one time point into a new temporal domain.

In this study, we developed non-spatial and spatial SDMs for two genera of trees in eastern North America. We calibrated the models with current climate data and forest inventory and analysis (FIA) data collected by the United States Forest Service. We then projected the models back in time using palaeoclimatic simulations and extensive pollen records as independent validation data. Our approach is similar to that of Pearman *et al.* (2008a), who used fossil pollen to validate SDMs of European trees projected back to a single time in the mid-Holocene (6000 years before present). However, the availability of new palaeoclimate reconstructions, which provide millennial snapshots of historic climate for the last 21,000 years before present, allowed us to validate the models at a much finer temporal resolution.

To assess the usefulness of adding a spatial term to SDMs we consider the following: (1) a spatially varying intercept model with no climate variables; (2) a non-spatial model with climate variables; and (3) a spatially varying intercept model with climate variables. As detailed in the Methods and Appendix S2 in Supporting Information, the spatially varying intercept was introduced via spatial random effects. The rationale for choosing these candidate models is as follows. If climate variables describe a significant portion of the variability in the observed distribution, and if these variables change over time, then projections from models with climatic variables will show a conservative shift away from the observed distribution. For the spatially varying intercept model with climate variables, any projected shifts in distributions are tempered by the spatial random effects. Depending on the amount of spatial autocorrelation, spatial random effects act to draw the projected distribution back toward the observed distribution used to calibrate the model. If climate variables do not describe a significant portion of the variability in the observed distribution, then the spatial random effects will keep projected distributions close to the observed distribution, i.e. the only learning for prediction will come from the observed distribution and hence the projected probability of species occurrence will be similar to the observed probability of occurrence. With these three candidate models, we were able to tease apart differences due to spatial random effects alone, the climate variables alone and their additive effects. We parameterized and estimated model parameters following a Bayesian framework, which provided full posterior distributions for model parameters and allowed us to estimate the uncertainty in our statistical inferences. We focus on two tree genera, *Fagus* and *Tsuga*, whose distributions can be readily inferred from fossil pollen and which possess contrasting life histories.

We address three questions. (1) Do non-spatial SDMs of current distributions of *Fagus* and *Tsuga* based on climate variables exhibit residual spatial autocorrelation? (2) Do SDMs with spatial random effects that include or exclude climate variables provide better fits to the observed distributions than non-spatial SDMs with climate variables only? (3) Do hindcasted spatial SDMs better predict historic distributions than non-spatial SDMs?

METHODS

Study genera

We studied two tree genera, *Fagus* and *Tsuga*. In eastern North America, *Fagus* is represented by only one species, *Fagus grandifolia* (Ehrh.) (American beech), and *Tsuga* by only two, the widespread *Tsuga canadensis* (L.) Carr. (eastern hemlock), and the narrow endemic *Tsuga caroliniana* Engelm.) (Carolina hemlock). For both *Fagus* and *Tsuga*, the relationship between local abundance of trees and relative abundance of pollen in sediment cores has already been derived (Davis, 1981). *Tsuga* is a conifer with passively dispersed cones, whereas *Fagus* is deciduous with animal-dispersed seeds.

Occurrence data

We used FIA data to describe the current distribution of *Fagus* and *Tsuga*. In every 2428 ha of land in the United States classified as 'forested', there is one permanent FIA plot, each containing four 7.2-m fixed-radius subplots (Woudenberg *et al*., 2010). In each subplot, all trees > 12.7 cm diameter at breast height have been measured periodically since the 1940s; consistent nationwide annual inventories were initiated in 2001. We used data from the most recent full plot inventory (2003–08) to calibrate our models.

Historic distributions of *Fagus* and *Tsuga* were derived from fossil pollen data in the Neotoma Paleoecology Database (<http://www.neotomadb.org>). Palaeoclimate data (described below) were available at intervals of 1000 years before present (ka BP) from 0–21 ka BP, so we focused on millennial historic distributions of *Fagus* and *Tsuga*. Given the variation in temporal scale and spatial resolution across study sites and uncertainties associated with radiocarbon aging of pollen from sediment cores (Blaauw *et al*., 2007), we compiled pollen datasets in which *Fagus* and *Tsuga* were counted as present at a site if their pollen percentages reached threshold levels at any time within 500 years centred on each historic millennium (Appendix S1). We chose a 500-year window because cross-validation analyses of biostratigraphic ages from recently revised age models for all pollen sites suggested that 500 years is a conservative estimate of temporal uncertainty for sites in the Neotoma database (Blois *et al*., 2011). To determine the sensitivity of historic tree distributions to the pollen percentage thresholds used to define a genera's presence or absence at a site, we specified low and high thresholds for each genus (Pearman *et al*., 2008a): 0.5% or 1% for *Fagus* and 1% or 2% for *Tsuga* (Davis, 1981).

Extent and resolution

The extent of the study area was the portion of eastern North America with the highest density of pollen data (Fig. 1). This region contained 75,251 FIA sites and up to 379 Neotoma locations, depending on the time period considered. Paciorek & McLachlan (2009) found that spatial patterns relating current and past climates to abundances of pollen and trees were unre-

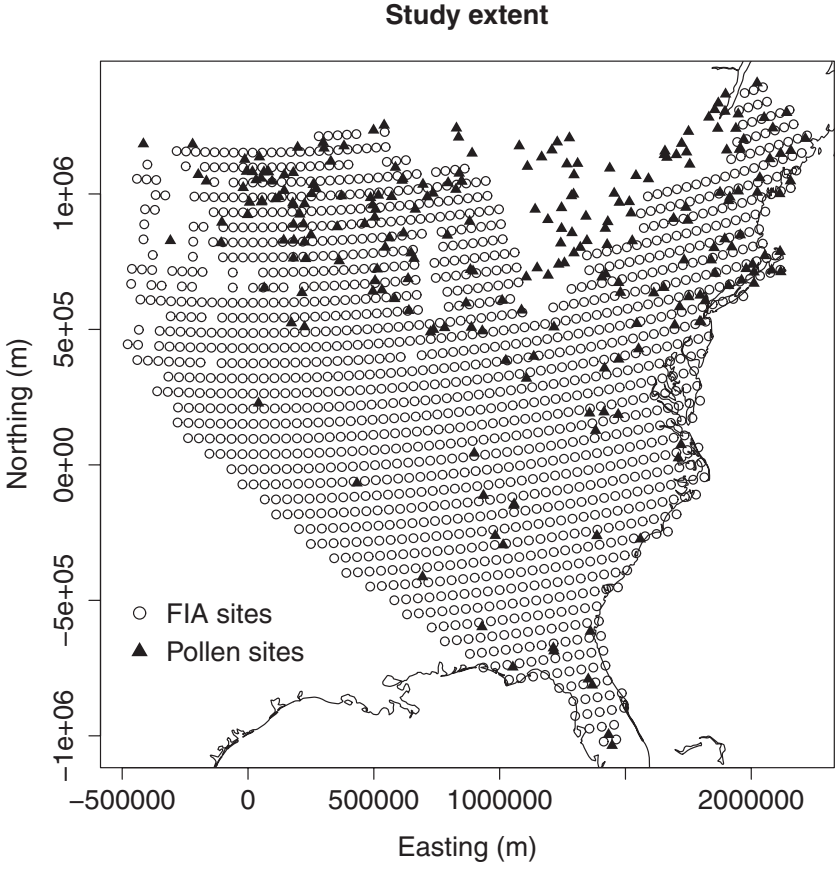

Figure 1 Map of the study extent in the eastern United States showing forest inventory and analysis (FIA) plots (hollow circles) and Neotoma pollen sites (solid triangles) snapped to a resolution of 0.5° (Alber's equal area conic projection).

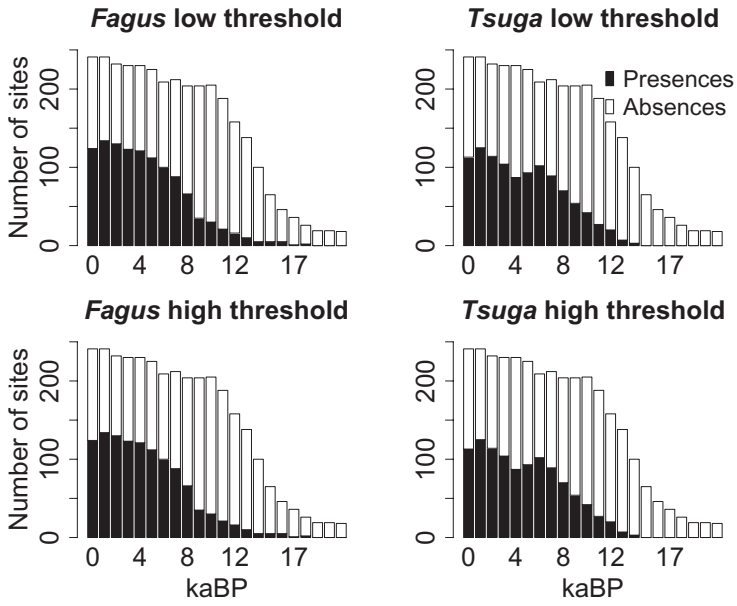

Figure 2 Numbers of sites with presences (black fill) or absences (white fill) of *Fagus* (a and c) and *Tsuga* (b and d) based on the low and high pollen thresholds from present to 21,000 years before present (ka BP) based on fossil pollen data from the Neotoma database. Data extending beyond 8 ka BP were not used in the analyses due to the low number of presences of *Fagus* and *Tsuga* beyond that time.

liable at resolutions below *c.* 50 km, so the climatic predictors for our model (see below) were downscaled to a resolution of 0.5° (*c.* 50–80 km depending on latitude). We upscaled the current tree occurrence data for each grid cell in the climate spatial data layers, keeping track of the number of FIA sites per 0.5° cell to be used as weights in the models (Appendix S2). Following this aggregation there were a total of 1419 FIA observations with presence/absence ratios for *Fagus* and *Tsuga* of 706/713 and 380/1039, respectively. The number of aggregated pollen observations varied for each 1 ka BP time period (Fig. 2). Although both palaeoclimatic and pollen data extended back 21 ka BP, the total sample size and the number of pollen grains of each genus declined rapidly beyond 8 ka BP (Fig. 2). Thus, our hindcast projections only extend from 1 to 8 ka BP, which allowed us to validate the models using a minimum of 200 grid cells containing observations, at least 50 of which contain presences for each genus.

Climate data

Modern climate data came from the observed dataset of the Climate Research Unit (CRU) at the University of East Anglia (Brohan *et al.*, 2006). Palaeoclimate data for this study came from a recent transient simulation of the CCSM3 global circulation model (GCM) (Liu *et al.*, 2009). The standard change-factor approach was employed to statistically downscale and reduce bias in the climate data (Wilby *et al.*, 2004). For each climate variable at each millennial interval, the difference between the modelled palaeoclimate and modelled modern climate was calculated and then resampled to a $0.5° \times 0.5°$ grid to match the resolution of the CRU observed climate dataset (Mitchell & Jones, 2005).

Decadal averages of seasonal variables were the highest temporal resolution data available from the archived CCSM3 simulations. To get a 'snapshot' of climatic conditions at each millennial time point, decadal averages of seasonal climate variables from the CRU or CCSM3 simulations were calculated for the first 100 years of each millennium (e.g. 8.0 to 7.9 ka BP). Because summaries of modern observed climate are available at centennial scales, these same centennial summaries of palaeoclimate were derived to aid comparisons between palaeo and modern SDMs. Bioclimatic variables that captured precipitation and temperature averages and seasonalities were used because response surface analyses for *Fagus* and *Tsuga* have shown that climatic annual averages, annual ranges and seasonality were important factors controlling the Holocene migrations of these genera (Bartlein *et al.*, 1986). Specifically, we calculated six bioclimatic variables (Hijmans *et al.*, 2005): annual mean temperature (BIO1), mean diurnal range (BIO2), temperature seasonality (BIO4), temperature annual range (BIO7), annual precipitation (BIO12) and precipitation seasonality (BIO15).

Two of the six calculated bioclimatic variables, temperature seasonality and temperature annual range, had within-time correlations with the other bioclimatic variables ≥ 0.7, so they were not included as explanatory variables in the models that included environmental predictors (see Appendix S3). The correlations between mean diurnal range and annual precipitation varied between modern and historic times (see Appendix S3), and such changing correlation structures between times could be problematic when projecting models beyond the present (Elith *et al.*, 2010). To determine if sufficient variance in the current distribution was explained by the two remaining variables with stable correlation structures over time (i.e. annual mean temperature and precipitation seasonality), we compared a model with annual mean temperature, precipitation seasonality, mean diurnal range and annual precipitation with another that included only annual mean temperature and precipitation seasonality.

Model calibration

We used Bayesian generalized linear models (GLMs) to model the occurrence of genera. While approaches such as neural networks and genetic algorithms have been used for SDMs, and although model projections can be sensitive to the type of statistical model employed (Elith *et al.*, 2010), classical approaches do not provide the statistical inferences we sought. Even though GLMs describe a central tendency and not a limiting effect (for example of temperature or precipitation extremes), Bayesian spatial GLMs provide exact inference for the random model parameters, including spatial random effects, by estimating entire posterior distributions at both observed and unobserved geographic locations (Gelfand *et al.*, 2006). Because our goal was to compare consistently SDMs with three different specifications [i.e. spatially varying intercept only (SVI), climate only, and spatially varying intercept plus climate], we adopted a Bayesian approach in fitting all of the models. Model structure is detailed in Appendix S2; model code is provided in Appendix S4.

Including the SVI has a potential for overfitting as it allows variable intercepts for every location and thus a very flexible spatial fit to the FIA data. As a null model, we also fit a multilevel B-spline to the FIA data (Lee *et al.*, 1997) using the 'MBA' package of 'R' statistical software to determine whether our hindcasting test for the inclusion of a SVI in the Bayesian models was sufficient. As an exploratory analysis into the strength of the residual spatial dependence in the FIA data, we calculated Moran's *I* from the residuals of the non-spatial GLMs. This latter analysis was conducted using the Spatial Analyst Tool in ArcMap10 (ESRI, 2011).

Model fit to calibration data

We fitted the Bayesian models to 90% of the FIA data ($n = 1277$) and randomly selected a 10% hold-out dataset ($n = 142$) to assess predictive performance. We also used the deviance information criterion (DIC) to rank the fit of the Bayesian models to the calibration data (Spiegelhalter *et al.*, 2002) – the DIC is the sum of the Bayesian deviance (a measure of model fit) and the effective number of parameters (a penalty for model complexity). Lower DIC values indicate better model fit. Models are compared using ΔDIC:

$$\Delta\mathrm{DIC}_i = \mathrm{DIC}_i - \min(\mathrm{DIC})$$

where min(DIC) is the DIC value for the model with the best fit (i.e. the lowest DIC value). In general, $\Delta\mathrm{DIC} < 2$ indicates weak evidence, $5 < \Delta\mathrm{DIC} < 10$ indicates strong evidence, and $\Delta\mathrm{DIC} > 10$ indicates very strong evidence that one model is preferred over another (Spiegelhalter *et al.*, 2002).

FIA hold-out dataset and pollen validations

When projecting the spatial models back in time for the pollen validation, the random effects serve to draw the projected distributions for each genus back toward that of the observed distribution used for model calibration (i.e. the FIA data) in the new time period (Appendix S2). To compare the performance of the models in predicting current and projecting past distributions, three measures were calculated using the 'ROCR' package of 'R' statistical software: the area under the curve (AUC) of the receiver operating curve (ROC), false negative rates (FNR) and false positive rates (FPR). The calculation of FNRs and FPRs requires the continuous outputs to be converted to a binary form using a threshold, in this case the value that maximizes the sum of sensitivity and specificity (Liu *et al.*, 2005; Lobo *et al.*, 2008).

Differences in AUC, FNR and FPR between models, genera, pollen percentage thresholds, time and the model × genus interaction were tested with three GLMs. To normalize residuals and reduce heteroskedasticity, AUC, FNR and FPR were all arcsin transformed. Model, genera, pollen percentage threshold and the model × genus interaction entered the GLM as fixed factors, and time entered as a covariate. The model × genus interaction was of particular interest as it tested whether or not different models performed better or worse in hindcasting the presence–absence of the two genera. The data were analysed with separate GLMs for AUC, FNR and FPR to facilitate the interpretation of Tukey's honestly significant differences post hoc comparisons at the expense of increasing Type II error rates. Bonferroni corrections of the *P*-values from the tests did not alter the significance of any of the effects.

RESULTS

Parameter estimates and model fit to calibration data

In non-spatial models with two climatic variables (i.e. annual mean temperature and precipitation seasonality) or four climatic variables (i.e. annual mean temperature, mean diurnal range, annual precipitation and precipitation seasonality), all climatic variables were significant predictors of presence–absence: none of the 95% credible intervals of the parameter estimates included zero (Tables 1 & 2). In contrast, in the spatial models some of the climatic explanatory variables were not significant predictors of presence–absence (e.g. annual mean temperature in the *Tsuga* models with two climatic variables and mean diurnal range in the *Fagus* model with four climatic variables; Tables 1 & 2). Changes in the magnitude and sign of parameter estimates between non-spatial and spatial models suggested that non-spatial models violated the assumption of independent identically distributed residuals. The residuals of the non-spatial models for both *Fagus* and *Tsuga* also exhibited significant positive spatial autocorrelation (Moran's $I = 0.604$, $P < 1 \times 10^{-7}$ for *Fagus*; Moran's $I = 0.761$, $P < 1 \times 10^{-7}$ for *Tsuga*), supporting the conclusion that non-spatial models were inappropriate for these data.

For *Fagus*, the SVI plus climate model with annual mean temperature and precipitation seasonality had the lowest DIC value and $\Delta\mathrm{DIC} > 10$ relative to all other *Fagus* models (Table 3, Fig. 3). In contrast, for *Tsuga*, the SVI model with no bioclimatic

Table 1 Parameter credible intervals (2.5%, 50.0% and 97.5% percentiles) for the *Fagus* spatially varying intercept (SVI), non-spatial (NS2 and NS4) and SVI plus climate (SVI2 and SVI4) models. The numbers two and four in the acronyms for the non-spatial and SVI plus climate models indicate the number of bioclimatic explanatory variables included in the models. The two climatic variables models included annual mean temperature (BIO1) and precipitation seasonality (BIO15). The four climatic variables models included annual mean temperature (BIO1), mean diurnal range (BIO2), annual precipitation (BIO12) and precipitation seasonality (BIO15). For models with spatial random effects, the spatial random effect variance and spatial decay parameter are denoted σ^2 and φ, respectively.

Model	β parameter	2.5%	50.0%	97.5%
SVI	Intercept	−7.23	−5.28	−2.72
SVI	σ^2	8.11	12.90	20.24
SVI	φ	1.09×10^{-6}	1.62×10^{-6}	2.63×10^{-6}
NS2	Intercept	−3.06	−3.01	−2.96
NS2	BIO1	−0.48	−0.46	−0.43
NS2	BIO15	−1.83	−1.78	−1.72
NS4	Intercept	−3.11	−3.06	−3.01
NS4	BIO1	−0.62	−0.58	−0.54
NS4	BIO2	0.33	0.37	0.40
NS4	BIO12	−0.20	−0.14	−0.09
NS4	BIO15	−2.03	−1.96	−1.90
SVI2	Intercept	−7.49	−5.77	−4.41
SVI2	BIO1	−1.57	−1.25	−0.89
SVI2	BIO15	−0.97	−0.47	−0.08
SVI2	σ^2	6.35	10.32	17.25
SVI 2	φ	1.15×10^{-6}	1.90×10^{-6}	3.20×10^{-6}
SVI 4	Intercept	−8.27	−5.47	−3.13
SVI 4	BIO1	−1.37	−0.83	−0.25
SVI 4	BIO2	−0.16	−0.03	−0.11
SVI 4	BIO12	−0.15	−0.47	0.80
SVI 4	BIO15	−0.89	−0.36	−0.12
SVI 4	σ^2	5.53	10.50	17.78
SVI 4	φ	1.14×10^{-6}	1.91×10^{-6}	3.69×10^{-6}

Table 2 Parameter credible intervals (2.5%, 50% and 97.5% percentiles) for the *Tsuga* spatially varying intercept (SVI), non-spatial (NS2 and NS4) and SVI plus climate (SVI 2 and SVI 4) models. The numbers two and four in the acronyms for the non-spatial and SVI plus climate models indicate the number of bioclimatic explanatory variables included in the models. The two climatic variables models included annual mean temperature (BIO1) and precipitation seasonality (BIO15). The four climatic variables models included annual mean temperature (BIO1), mean diurnal range (BIO2), annual precipitation (BIO12) and precipitation seasonality (BIO15). For models with spatial random effects, the spatial random effect variance and spatial decay parameter are denoted σ^2 and φ, respectively.

Model	β parameter	2.5%	50%	97.5%
SVI	Intercept	−9.10	−7.68	−4.15
SVI	σ^2	12.6	22.3	36.4
SVI	φ	1.09×10^{-6}	2.23×10^{-6}	2.74×10^{-6}
NS2	Intercept	−3.50	−3.45	−3.40
NS2	BIO1	−1.14	−1.11	−1.07
NS2	BIO15	−1.20	−1.16	−1.12
NS4	Intercept	−3.55	−3.50	−3.45
NS4	BIO1	−1.34	−1.30	−1.25
NS4	BIO2	0.31	0.35	0.40
NS4	BIO12	0.07	0.14	0.21
NS4	BIO15	−1.25	−1.21	−1.12
SVI2	Intercept	−10.18	−8.38	−3.45
SVI 2	BIO1	0.07	0.48	0.89
SVI 2	BIO15	−1.09	−0.55	−0.05
SVI 2	σ^2	10.86	18.57	32.11
SVI 2	φ	1.09×10^{-6}	1.68×10^{-6}	2.96×10^{-6}
SVI 4	Intercept	−8.28	−5.73	−4.00
SVI 4	BIO1	−1.28	−0.85	−0.26
SVI 4	BIO2	−0.16	−0.03	0.11
SVI 4	BIO12	−0.15	0.47	0.80
SVI 4	BIO15	−0.81	−0.36	0.12
SVI 4	σ^2	5.94	10.58	17.86
SVI 4	φ	1.14×10^{-6}	1.89×10^{-6}	3.43×10^{-6}

predictors had the lowest DIC value and ΔDIC > 10 relative to all other *Tsuga* models (Table 3, Fig. 4).

The non-spatial SDMs for both *Fagus* and *Tsuga* that included only annual mean temperature and precipitation seasonality had ΔDIC values > 10 relative to the non-spatial models that included annual mean temperature, precipitation seasonality, mean diurnal range and annual precipitation (Table 3). Given that the correlative relationship between mean diurnal range and annual precipitation was unstable between modern and historic times (see Appendix S3) and that the inclusion of them did not provide a large decrease in the ΔDIC, these two climatic variables were excluded from the models used for prediction that were validated with the 10% hold-out FIA dataset and fossil pollen record.

FIA hold-out dataset and pollen validations

For the present-day 10% hold-out FIA dataset for both genera, the non-spatial model performed worse than the SVI, SVI plus climate or multilevel B-spline models in terms of AUC, FNR and FPR (Table 4, Appendix S5). However, the same was not true when models were hindcasted. Based on AUC, there were significant main effects of model type (non-spatial, SVI, SVI plus climate, FIA B-spline; $F_{3,118} = 32.4$, $P = 2.4 \times 10^{-15}$), and a significant genus × model interaction ($F_{3,118} = 13.8$, $P = 8.8 \times 10^{-8}$) (Table 4, Appendix S5) on model performance. For the *Fagus* hindcasts, on average the non-spatial model had higher AUC values than the spatial models (i.e. SVI and SVI plus climate) and FIA multilevel B-spline models, but the opposite was true for *Tsuga*. The FNRs in the hindcasting validation varied by model ($F_{3,118} = 8.1$, $P = 6.2 \times 10^{-5}$). The FIA data multilevel B-spline model had the highest FNR and post hoc comparisons showed that there were no significant differences between the non-spatial and spatial models in FNRs (Table 4, Appendix S5). Similar to the FNRs, the FPRs also varied by model ($F_{3,118} = 9.0$, $P = 1.95 \times 10^{-5}$) (Table 4, Appendix S5). The FIA data multilevel B-spline and the non-spatial models had higher FPRs than the

Table 3 Fits of the spatially-varying intercept (SVI), non-spatial and SVI plus climate SDMs to the modern forest inventory and analysis (FIA) occurrence data for *Fagus* and *Tsuga.* Bioclimatic variables included in the models with climatic predictors were: annual mean temperature (BIO1), mean diurnal range (BIO2), annual precipitation (BIO12) and precipitation seasonality (BIO15). Model fit was evaluated with the deviance information criterion (DIC), which is the sum of P_D (the effective number of parameters) and the posterior mean of the deviance. To facilitate model comparison, ΔDIC was also calculated, where the model with the lowest DIC has a value of zero and all other models are compared to it.

Model	Bioclimatic variable	Genus	P_D	DIC	ΔDIC
SVI	None	*Fagus*	247	35,893	81
Non-spatial	1, 15	*Fagus*	3	41,497	5685
Non-spatial	1, 2, 12, 15	*Fagus*	5	41,125	5313
SVI-climate	1, 15	*Fagus*	248	35,812	0
SVI-climate	1, 2, 12, 15	*Fagus*	251	35,826	14
SVI-climate	None	*Tsuga*	170	23,685	0
Non-spatial	1, 15	*Tsuga*	3	30,025	6340
Non-spatial	1, 2, 12, 15	*Tsuga*	5	29,715	6030
SVI-climate	1, 15	*Tsuga*	164	23,708	23
SVI-climate	1, 2, 12, 15	*Tsuga*	160	23,727	42

spatial models. There were no significant genus × model interactions for FNRs ($F_{3,118} = 2.3$, $P = 0.08$) and FPRs ($F_{3,118} = 1.7$, $P = 0.18$). Overall for the three measures, model performance worsened as models were projected further back in time (AUC: $F_{1,118} = 118$, $P = 2.0 \times 10^{-6}$; FNR: $F_{1,118} = 98.7$, $P = 2.0 \times 10^{-16}$; FPR: $F_{1,118} = 109$, $P = 2.0 \times 10^{-16}$). Also, model performance was better (i.e. higher AUC and lower FNR and FPR) for *Tsuga* than for *Fagus* (AUC: $F_{1,118} = 10.0$, $P = 0.002$; FNR: $F_{1,118} = 65.5$, $P = 5.8 \times 10^{-13}$; FPR: $F_{1,118} = 88$, $P = 6.3 \times 10^{-16}$) and for low pollen percentage thresholds than for high pollen percentage thresholds (AUC: $F_{1,118} = 14.0$, $P = 2.8 \times 10^{-4}$; FNR: $F_{1,118} = 15.3$, $P = 1.5 \times 10^{-4}$; FPR: $F_{1,118} = 24.9, 2.13 \times 10^{-16}$). For all three test metrics (i.e. AUC, FNR, FPR), the multilevel B-spline fit to the FIA data, which we used as a 'perfectly fitted' model to assess whether or not the spatial models were overfitted to the calibration data, performed the worst. This assured us that the pollen validation test was stringent enough.

DISCUSSION

A key question regarding the application of SDMs to predicting the response of species to climate change is whether the failure to include ecological and evolutionary processes (e.g. dispersal, biotic interactions, readjustment lags) will prove to be problematic (reviewed by Pearson & Dawson, 2003). Depending on the species and its life history, ecological and evolutionary processes may (or may not) lead to its inability to track changes in climate. While there is evidence that vagile organisms (e.g. butterflies) can track rapid climate change (Warren *et al.*, 2001), sessile organisms (e.g. trees) may not readily disperse to newly suitable habitat, resulting in limited niche space filling (Svenning & Skov, 2004; Meier *et al.*, 2012). Species undergoing climate-driven range expansions coupled with enemy release are hypothesized to be more capable of realizing their potential niche (Hellman *et al.*, 2012), whereas species limited by a particular resource (e.g. host availability) can be constrained to the spatial distribution of the resource (Merrill *et al.*, 2007). There is evidence that shorter-lived taxa (e.g. insects and herbaceous plants; Woodward, 1990; Thomas *et al.*, 2001) can evolve in response to rapid climate change, but longer-lived taxa that cannot evolve as quickly may experience readjustment lags (Pearson & Dawson, 2003).

For those taxa whose distributions do not shift over time as a result of ecological and evolutionary processes, the inclusion of spatial random effects in SDMs could improve projections by providing a more conservative prediction of distributional shifts, especially when climatic variables do not explain much variability in their observed distributions. Alternatively, when climatic variables explain most of the variability in a taxon's observed distribution and the taxon is capable of tracking climate, then accounting for spatial autocorrelation in SDMs won't provide better projections. In other words, the spatial random effects keep the projected distribution similar to the data used for model calibration, unless the covariates (e.g. climatic variables) suggest otherwise. Further, if the climate variables do not explain much of the variability in the observed distribution and the genera's distribution shifts far from the observed distribution over time, then none of the models defined here will perform well. The predictive abilities of non-spatial and spatial SDMs have rarely been compared with temporally varying validation datasets to test these assertions (Gelfand *et al.*, 2006).

In this study we tested the predictive abilities of non-spatial and spatial SDMs across eight millennia using data from the pollen record (Appendix S1). We found that spatial SDMs had better fits to the calibration data, higher predictive accuracy for a modern hold-out validation dataset and greater variance in their outputs than non-spatial SDMs (see also Gelfand *et al.*, 2006; Bahn & McGill, 2007). For *Fagus*, the SVI plus climate model provided a better fit to the calibration data than the SVI model, but the opposite was true for *Tsuga*. Also for the two climatic variable models, for *Fagus* there was no change in the sign of the climatic regression coefficients between the non-spatial and spatial models (Table 1), but with *Tsuga* there was a sign change in the regression coefficient for annual mean temperature between the non-spatial and SVI plus climate models (Table 2). This result suggests that for *Tsuga* the spatial random effect could be accounting for dependence in the model's residuals across space as several other studies have found that parameter estimates are affected by spatial autocorrelation (Dormann, 2007; Kühn, 2007; Bini *et al.*, 2009; Hodges & Reich, 2010).

In the hindcasting analyses, the SVI and SVI plus climate models performed similarly. This suggests that the climatic variables do not contribute much to explaining the variability of occurrence relative to that explained by the spatial random

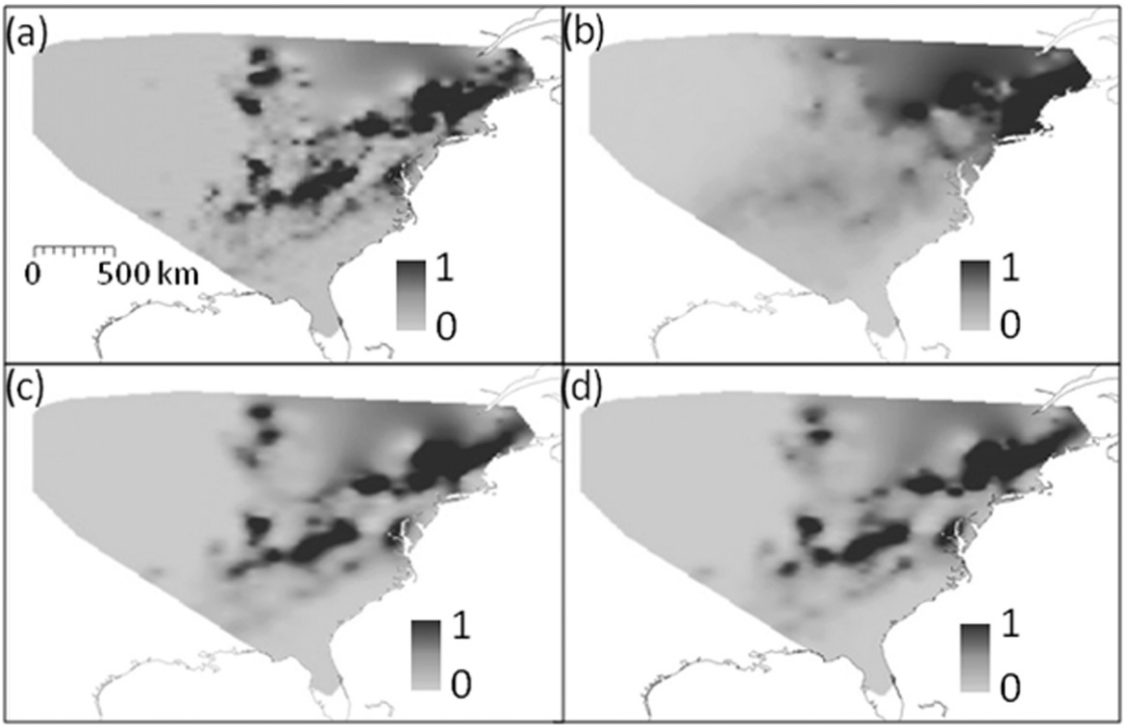

Figure 3 Maps of (a) a surface approximation of the probability of occurrence of *Fagus* generated by a multilevel B-spline fit to the raw forest inventory and analysis (FIA) data and the predicted probability of presence of the (b) non-spatial, (c) spatially varying intercept, and (d) spatially varying intercept plus climate species distribution models to modern *Fagus* FIA data (Alber's equal area conic projection). The surface approximation in (a) was calculated with the MBA package in R.

effects. AUC values based on fossil pollen indicated that the non-spatial model performed better for *Fagus* than either of the two spatial models, but the opposite was true for *Tsuga*. However, FNR values did not differ among the models for either genus, and FPR values were greater for non-spatial models for both genera. We have more confidence in FNR and FPR values than in AUC values because the latter describes portions of the ROC curve that are rarely encountered and weights omission and commission errors equally (Lobo *et al.*, 2008). With the pollen record, equal weighting of omission and commission errors may not be ideal; we have much more confidence in the presence of pollen grains than in their absence (Blaauw *et al.*, 2007; Blois *et al.*, 2011) and false negatives in the pollen record are more problematic than false positives. The lack of differences in false negative rates between models shows that the non-spatial and spatial models have similar FNRs.

Although we have less confidence in actual absences in the pollen data, the FPRs are interesting when considering the ecological and evolutionary processes leading to conserved spatial structure in the distributions of species. The greater FPRs of non-spatial models for both genera suggest that spatial effects may account for smaller-scale climatic spatial structure that is not otherwise estimated in large-scale or averaged temperature and precipitation values (Gelfand *et al.*, 2006; Hawkins *et al.*, 2007). Evidence from the fossil pollen and palaeoclimate records suggests that climatic shifts can result in abrupt ecological changes in vegetation that are driven by internal dynamics, such as site-specific environmental characteristics (e.g. soil moisture) or biotic interactions (e.g. competition) that create geographically localized variation in vegetation composition (Williams *et al.*, 2011). Taxon-specific responses to climate forcing also could explain why the SVI model had the lowest DIC for *Tsuga* and why the two spatial models performed better with regard to both AUC and FPR for *Tsuga* but not for *Fagus.* Approximately 5.5 ka BP *Tsuga* experienced a range contraction known as the 'hemlock decline' potentially due to an abrupt change in climate, a phytophagous insect infestation or both (Bhiry & Filion, 1996; Foster *et al.*, 2006). If the hemlock decline was due to an abrupt change in climate, then localized ecological changes could have resulted in stronger spatial structure in its distribution. However, decoupling changes in distributions due to climate and spatial structure due to biotic interactions or site-specific abiotic characteristics is difficult because observed spatial structure is (or was) inherently linked to abrupt climate change.

Alternatively, the spatial random effects may have captured a missing covariate, such as an ecological process that generates spatial structure (Clayton *et al.*, 1993; Paciorek, 2010). Such processes could include dispersal, competitive interactions, land-use history or underlying features of the terrain. For example, if dispersal limitation prevents distributional shifts, then we might expect that spatial SDMs would perform better

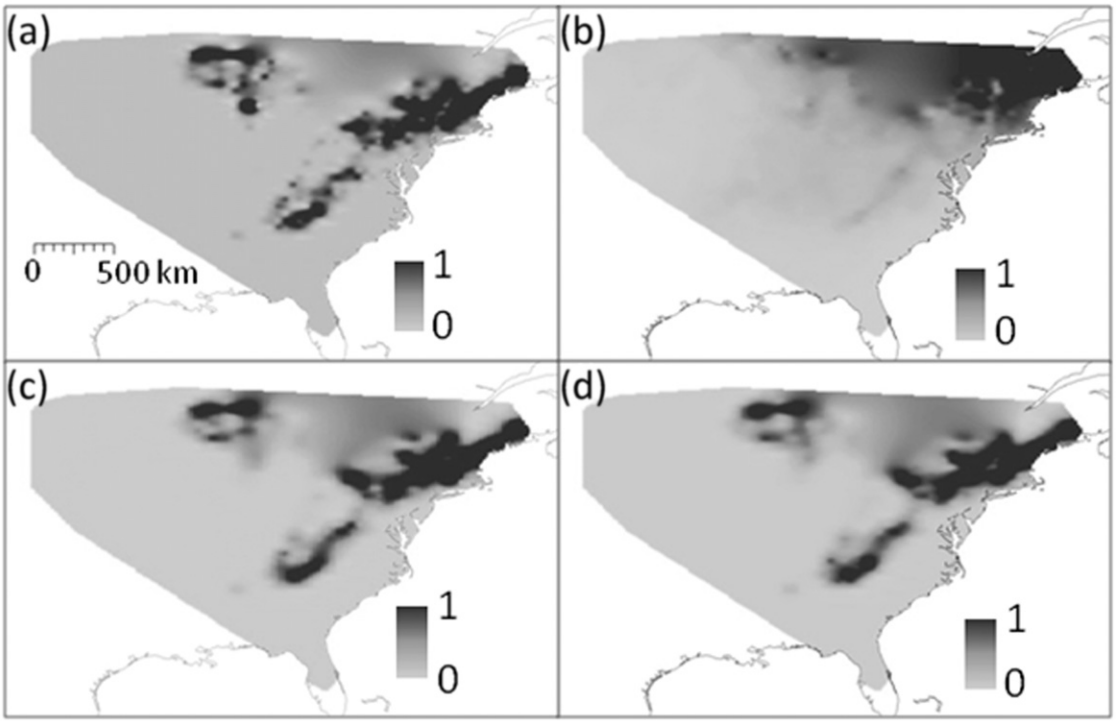

Figure 4 Maps of (a) a surface approximation of the probability of occurrence of *Tsuga* generated by a multilevel B-spline fit to the raw forest inventory and analysis (FIA) data and the predicted probability of presence of the (b) non-spatial, (c) spatially varying intercept, and (d) spatially varying intercept plus climate species distribution models to modern *Tsuga* FIA data (Alber's equal area conic projection). The surface approximation in (a) was calculated with the MBA package in R.

for dispersal-limited taxa (e.g. *Tsuga*) that cannot track changes in climate, but not necessarily for taxa with effective dispersal vectors (e.g. *Fagus*) that can gain dominance by migrating faster to climatically favourable sites (Pearman *et al.*, 2008b). These taxon-specific differences in dispersal mode and degree of dominance could explain why *Tsuga* seemed to be less responsive to climate over the past eight millennia than *Fagus* as evidenced by the better performance over time of the two spatial models with regard to both AUC and FPR for *Tsuga*, but not for *Fagus*. Simulation experiments for European trees with spatially explicit process models accounting for changing macroclimate, competition and habitat connectivity showed that some of the spatial autocorrelation between two time periods may be due to very slow migration rates resulting in severe time lags that are not accounted for in non-dynamic and non-spatial SDMs (Meier *et al.*, 2012). Also, Dobrowski *et al.* (2011) found that non-spatial SDMs fitted to widespread plants with more effective dispersal mechanisms had higher predictive accuracy over 75 years of climate change in California than non-spatial SDMs fitted to dispersal-limited plants.

Given the results of this study, should researchers include spatial random effects in SDMs? We found that for two long-lived eastern North American trees, spatial models provided better fits to calibration data and lower FPRs, but not necessarily improvements in AUC or the FNR. The better fits of the spatial SDMs may have resulted from the richness of the FIA data used to calibrate the models. Large samples of evenly dispersed data will probably capture any spatial structure; consequently a spatial SDM should fit well. However, when sample sizes are small, there is less of a chance that the spatial structure will be represented adequately. Ultimately, whether to include spatial random effects in SDMs will depend on the taxon being modelled, the cost of false positives and the quality of the data.

ACKNOWLEDGEMENTS

This research was supported by the US Department of Energy's National Institute for Climate Change Research, through subaward 3892-HU-DOE-4157 to A.M.E. and M.C.F., NSF grant DBI 10-03938 to A.M.E., and NSF grants DMS-1106609 and EF-1137309 to A.O.F. FIA data were provided by Brett Butler and Elizabeth LaPoint (both with the US Forest Service) pursuant to a Memorandum of Understanding 09MU11242305123 between the US Forest Service and Harvard University. Eliza Ledwell and Elisabete Baker-Vail assisted with initial data organization and programming. The Harvard Forest Lab Group, J. Williams, N. Zimmerman and three anonymous referees provided valuable insights.

Table 4 Model performance as measured by the area under the receiver operating curve (AUC), false negative rates (FNR) and false positive rates (FPR) for the non-spatial model, spatially varying intercept (SVI) model, SVI plus climate and multilevel B-spline fit to modern *Fagus* and *Tsuga* occurrence data from the Forest Inventory and Analysis (FIA) data. Predictions of the models for modern time were validated with a 10% hold-out dataset from the FIA data. Hindcasts were validated with data from the fossil pollen record provided by the Neotoma database using the 'high' pollen thresholds for both genera. The numbers behind the AUC, FNR and FPR values in parentheses for the Bayesian models represent the standard error calculated from 1000 random draws from the post burn-in Markov chain Monte Carlo (MCMC) iterations. For the FIA multilevel B-spline approximation there is no standard error as there were no MCMC iterations to draw from.

Genus	Performance measure	Time (ka BP)	Non-spatial	SVI	SVI-climate	FIA
Fagus	AUC	0	0.87 (4×10^{-4})	0.91 (0.01)	0.92 (0.01)	0.91
		1	0.89 (5×10^{-4})	0.87 (0.02)	0.87 (0.02)	0.86
		2	0.90 (4×10^{-4})	0.88 (0.02)	0.88 (0.02)	0.86
		3	0.89 (6×10^{-4})	0.88 (0.01)	0.88 (0.02)	0.86
		4	0.88 (6×10^{-4})	0.87 (0.02)	0.87 (0.02)	0.84
		5	0.85 (1×10^{-3})	0.85 (0.02)	0.84 (0.02)	0.83
		6	0.84 (2×10^{-3})	0.84 (0.02)	0.83 (0.03)	0.83
		7	0.81 (1×10^{-3})	0.80 (0.02)	0.80 (0.03)	0.78
		8	0.73 (2×10^{-3})	0.76 (0.01)	0.74 (0.02)	0.71
Fagus	FNR	0	0.22 (0.01)	0.14 (0.04)	0.14 (0.03)	0.11
		1	0.20 (0.01)	0.23 (0.04)	0.22 (0.09)	0.26
		2	0.19 (0.02)	0.21 (0.05)	0.20 (0.09)	0.24
		3	0.19 (0.01)	0.19 (0.04)	0.21 (0.09)	0.23
		4	0.22 (0.01)	0.20 (0.04)	0.22 (0.09)	0.23
		5	0.28 (0.02)	0.24 (0.04)	0.25 (0.10)	0.26
		6	0.26 (0.01)	0.25 (0.05)	0.27 (0.10)	0.24
		7	0.30 (0.01)	0.31 (0.05)	0.32 (0.10)	0.31
		8	0.34 (0.03)	0.33 (0.04)	0.35 (0.07)	0.38
Fagus	FPR	0	0.23 (0.01)	0.14 (0.02)	0.14 (0.02)	0.12
		1	0.21 (0.02)	0.23 (0.04)	0.22 (0.05)	0.23
		2	0.20 (0.02)	0.21 (0.04)	0.20 (0.06)	0.22
		3	0.20 (0.01)	0.19 (0.03)	0.21 (0.06)	0.22
		4	0.24 (0.02)	0.19 (0.04)	0.22 (0.07)	0.25
		5	0.28 (0.03)	0.24 (0.03)	0.25 (0.07)	0.26
		6	0.27 (0.02)	0.25 (0.04)	0.26 (0.07)	0.28
		7	0.26 (0.02)	0.31 (0.04)	0.30 (0.07)	0.29
		8	0.35 (0.01)	0.33 (0.04)	0.34 (0.07)	0.40
Tsuga	AUC	0	0.85 (3×10^{-3})	0.95 (0.02)	0.95 (8×10^{-3})	0.97
1			0.85 (3×10^{-3})	0.91 (0.01)	0.91 (0.02)	0.82
		2	0.86 (4×10^{-4})	0.89 (0.01)	0.89 (0.02)	0.81
		3	0.87 (4×10^{-4})	0.88 (0.01)	0.87 (0.02)	0.80
		4	0.83 (3×10^{-3})	0.86 (0.02)	0.85 (0.02)	0.80
		5	0.84 (3×10^{-3})	0.90 (0.02)	0.89 (0.02)	0.84
		6	0.86 (2×10^{-3})	0.91 (0.02)	0.90 (0.02)	0.80
		7	0.85 (5×10^{-3})	0.88 (0.02)	0.87 (0.02)	0.80
		8	0.76 (5×10^{-3})	0.89 (0.02)	0.88 (0.02)	0.79
Tsuga	FNR	0	0.20 (0.03)	0.11 (0.04)	0.11 (0.04)	0.05
		1	0.16 (0.07)	0.16 (0.04)	0.18 (0.03)	0.20
		2	0.19 (0.02)	0.18 (0.04)	0.20 (0.03)	0.21
		3	0.19 (0.03)	0.18 (0.04)	0.20 (0.04)	0.20
		4	0.21 (3×10^{-3})	0.20 (0.05)	0.21 (0.04)	0.21
		5	0.25 (0.02)	0.17 (0.04)	0.18 (0.04)	0.20
		6	0.20 (0.02)	0.16 (0.04)	0.18 (0.03)	0.20
		7	0.25 (0.01)	0.18 (0.04)	0.19 (0.03)	0.24
		8	0.30 (0.01)	0.19 (0.05)	0.19 (0.04)	0.33
Tsuga	FPR	0	0.22 (0.01)	0.11 (0.03)	0.11 (0.03)	0.09
		1	0.19 (0.03)	0.16 (0.04)	0.17 (0.04)	0.23
		2	0.16 (0.01)	0.18 (0.04)	0.19 (0.03)	0.20
		3	0.19 (1×10^{-3})	0.18 (0.04)	0.19 (0.03)	0.22
		4	0.23 (0.01)	0.20 (0.03)	0.21 (0.03)	0.26
		5	0.24 (0.02)	0.17 (0.04)	0.18 (0.04)	0.23
		6	0.19 (0.01)	0.16 (0.05)	0.17 (0.04)	0.18
		7	0.23 (0.01)	0.18 (0.04)	0.18 (0.03)	0.22
		8	0.32 (0.01)	0.20 (0.04)	0.21 (0.03)	0.31

ka BP, thousand years before present.

REFERENCES

Bahn, V. & McGill, B.J. (2007) Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, **16**, 733–742.

Bartlein, P.J., Prentice, I.C. & Webb, T. (1986) Climatic response surfaces from pollen data for some eastern North American taxa. *Journal of Biogeography*, **13**, 35–57.

Bhiry, N. & Filion, L. (1996) Mid-Holocene hemlock decline in eastern North America linked with phytophagous insect activity. *Quaternary Research*, **45**, 312–320.

Bini, L.M., Diniz-Filho, J.A.F., Rangel, T.F.L.V.B. *et al.* (2009) Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography*, **32**, 193–204.

Blaauw, M., Christen, J.A., Mauquoy, D., van der Plicht, J. & Bennett, K.D. (2007) Testing the timing of radiocarbon-dated events between proxy archives. *The Holocene*, **17**, 283–288.

Blois, J.L., Williams, J.W., Grimm, E.C., Jackson, S.T. & Graham, R.W. (2011) A methodological framework for assessing and reducing temporal uncertainty in paleovegetation mapping from late-Quaternary pollen records. *Quaternary Science Reviews*, **30**, 1926–1939.

Brohan, P., Kennedy, J.J., Harris, I., Tett, S.F.B. & Jones, P.D. (2006) Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *Journal of Geophysical Research*, **111**, D12106.

Clayton, J.S., Carlin, B.P. & Montomoli, C. (1993) Spatial correlation in ecological analysis. *Journal of Epidemiology*, **22**, 1193–1201.

Davis, M.B. (1981) Quaternary history and the stability of forest communities. *Forest succession* (ed. by D.C. West, H.H. Shugart and D.B. Botkin), pp. 132–153. Springer, New York.

Diez, J.M. & Pulliam, H.R. (2007) Hierarchical analysis of species distributions and abundance across environmental gradients. *Ecology*, **88**, 3144–3152.

Dobrowski, S.Z., Thorne, J.H., Greenberg, J.A., Safford, H.D., Mynsberge, A.R., Crimmins, S.M. & Swanson, A.K. (2011) Modeling plant distributions over 75 years of measured climate change in California, USA: relating temporal transferability to species traits. *Ecological Monographs*, **81**, 241–257.

Dormann, C.F. (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, **16**, 129–138.

Elith, J., Kearney, M. & Phillips, S. (2010) The art of modeling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

ESRI (2011) ArcGIS desktop: Release 10. Environmental Systems Research Institute, Redlands, CA.

Fitzpatrick, M.C., Weltzin, J.F., Sanders, N.J. & Dunn, R.R. (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography*, **16**, 24–33.

Foster, D.R., Oswald, W.W., Faison, E.K., Doughty, E.D. & Hansen, B.C.S. (2006) A climatic driver for abrupt mid-Holocene vegetation dynamics and the hemlock decline in New England. *Ecology*, **87**, 2959–2966.

Franklin, J. & Miller, J.A. (2009) *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge.

Gelfand, A.E., Latimer, A., Wu, S. & Silander, J.A. (2006) Building statistical models to analyze species distributions. *Hierarchical modelling for the environmental sciences: statistical methods and applications* (ed. by J.S. Clark and A.E. Gelfand), pp. 77–07. Oxford University Press, Oxford.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Hawkins, B.A., Diniz-Filho, J.A.F., Bini, L.M., De Marco, P. & Blackburn, T.M. (2007) Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography*, **30**, 375–384.

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W. & Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modeling under climate change. *Progress in Physical Geography*, **30**, 751–777.

Hellman, J.J., Prior, K.M. & Pelini, S.L. (2012) The influence of species interactions on geographic range change under climate change. *Annals of the New York Academy of Sciences*, **1249**, 18–28.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *Journal of Climatology*, **25**, 1965–1978.

Hodges, J.S. & Reich, B.J. (2010) Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, **64**, 325–334.

Kühn, I. (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, **16**, 129–138.

Lee, S., Wolberg, G. & Shin, S.Y. (1997) Scattered data interpolation with multilevel B-splines. *IEEE Transactions on Visualization and Computer Graphics*, **3**, 229–244.

Legendre, P. (1993) Spatial autocorrelation – trouble or new paradigm. *Ecology*, **74**, 1659–1673.

Lichstein, J.W., Simons, T.R., Shriner, S.A. & Franzreb, K.E. (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, **72**, 445–463.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.

Liu, Z., Otto-Bleisner, B.L., He, F., Brady, E., Tomas, R., Clark, P.U., Carlson, A.E., Lynch-Stieglitz, J., Curry, W., Brook, E., Erickson, D., Jacob, R., Kutzbach, J. & Cheng, J. (2009) Transient simulation of last deglaciation with a new mechanism for Bølling–Allerød warming. *Science*, **325**, 310–314.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.

Meier, E.S., Lischke, H., Schmatz, D.R. & Zimmerman, N.E. (2012) Climate, competition and connectivity affect future

migration and ranges of European trees. *Global Ecology and Biogeography*, **21**, 164–178.

Merrill, R.M., Gutierrez, D., Lewis, O.T., Gutierrez, J., Diez, S.B. & Wilson, R.J. (2007) Combined effects of climate and biotic interactions on the elevational range of a phytophagous insect. *Journal of Animal Ecology*, **77**, 145–155.

Mitchell, T.D. & Jones, P.D. (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, **25**, 693–712.

Nogués-Bravo, D. (2009) Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography*, **18**, 521–531.

Paciorek, C.J. (2010) The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*, **25**, 107–125.

Paciorek, C.J. & McLachlan, J.S. (2009) Mapping ancient forests: Bayesian inference for spatio-temporal trends in forest composition using the fossil pollen proxy record. *Journal of the American Statistical Society*, **104**, 608–622.

Pearman, P.B., Randin, C.F., Broennimann, O., Vittoz, P., van der Knapp, W.O., Engler, R., Le Lay, G., Zimmermann, N.E. & Guisan, A. (2008a) Prediction of plant species distributions across six millennia. *Ecology Letters*, **11**, 357–369.

Pearman, P.B., Guisan, A., Broennimann, O. & Randin, C.F. (2008b) Niche dynamics in space and time. *Trends in Ecology and Evolution*, **23**, 149–158.

Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.

Segurado, P., Araújo, M.B. & Kunin, W.E. (2006) Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, **43**, 433–444.

Spiegelhalter, D.J., Best, N., Carlin, B.P. & van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B*, **64**, 583–639.

Svenning, J.-C. & Skov, F. (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, **7**, 565–573.

Thomas, C.D., Bodsworth, E.J., Wilson, R.J., Simmons, A.D., Davis, Z.G., Musche, M. & Conradt, L. (2001) Ecological and evolutionary processes at expanding range margins. *Nature*, **411**, 577–581.

Veloz, S.D., Williams, J.W., Blois, J.L., He, F., Otto-Bliesner, B. & Liu, Z. (2012) No-analog climates and shifting realized niches during the late Quaternary: implications for 21st-century predictions by species distribution models. *Global Change Biology*, **18**, 1698–1713.

Warren, M.S., Hill, J.K., Asher, T.J., Fox, R., Huntley, B., Roy, D.B., Telfer, M.G., Jeffcoate, S., Harding, P., Jeffcoate, G., Willis, S.G., Greatorex-Davies, J.N., Moss, D. & Thomas, C. (2001) Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature*, **414**, 65–69.

Wilby, R.L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P. & Mearns, L.O. (2004) *Guidelines for use of climate scenarios developed from statistical downscaling methods*. Data Distribution Center of the Intergovernmental Panel on Climate Change, Norwich, UK.

Williams, J.W., Blois, J.L. & Shuman, B.N. (2011) Extrinsic and intrinsic forcing of abrupt ecological change: case studies from the late Quaternary. *Journal of Ecology*, **99**, 664–677.

Wilson, T.L., Odei, J.B., Hooten, M.B. & Edwards, T.C. (2010) Hierarchical spatial models for predicting pygmy rabbit distribution and relative abundance. *Journal of Applied Ecology*, **47**, 401–409.

Woodward, F.I. (1990) The impact of low temperatures in controlling the geographical distribution of plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **326**, 585–593.

Woudenberg, S.W., Conkling, B.L., O'Connell, B.M., LaPoint, E.B., Turner, J.A. & Waddell, K.L. (2010) *The forest inventory and analysis database: database description and user's manual version 4.0 for phase 2*. General Technical Report RMRS-GTR-245. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Appendix S1 Presence–absence plots of historic pollen distributions.

Appendix S2 Detailed description of the models.

Appendix S3 Plots of between- and within-time correlations of palaeoclimate data.

Appendix S4 Code for analyses programmed in R and C++.

Appendix S5 Results of the *Fagus* and *Tsuga* low pollen threshold analysis.

BIOSKETCH

Sydne Record is a post-doctoral research fellow at Harvard University – Harvard Forest with broad research interests in validating ecological models and Bayesian statistics.

Author contributions: M.C.F., S.R. and A.M.E. conceived the ideas for the study. S.R., M.C.F. and A.O.F. conducted statistical analyses. S.R. led the writing of the manuscript with critical comments from all co-authors. S.V. provided the downscaled climate data.

Editor: Niklaus Zimmermann