Ecological Applications, 24(5), 2014, pp. 990–999
© 2014 by the Ecological Society of America

More than the sum of the parts: forest climate response from joint species distribution models

James S. Clark,[1,2,4] Alan E. Gelfand,[2] Christopher W. Woodall,[3] and Kai Zhu[1]

[1]*Nicholas School of the Environment, Duke University, Durham, North Carolina\ 27708 USA*
[2]*Department of Statistical Science, Duke University, Durham, North Carolina 27708 USA*
[3]*USDA Forest Service, Northern Research Station, St. Paul, Minnesota 55108 USA*

Abstract. The perceived threat of climate change is often evaluated from species distribution models that are fitted to many species independently and then added together. This approach ignores the fact that species are jointly distributed and limit one another. Species respond to the same underlying climatic variables, and the abundance of any one species can be constrained by competition; a large increase in one is inevitably linked to declines of others. Omitting this basic relationship explains why responses modeled independently do not agree with the species richness or basal areas of actual forests.

We introduce a joint species distribution modeling approach (JSDM), which is unique in three ways, and apply it to forests of eastern North America. First, it accommodates the joint distribution of species. Second, this joint distribution includes both abundance and presence–absence data. We solve the common issue of large numbers of zeros in abundance data by accommodating zeros in both stem counts and basal area data, i.e., a new approach to zero inflation. Finally, inverse prediction can be applied to the joint distribution of predictions to integrate the role of climate risks across all species and identify geographic areas where communities will change most (in terms of changes in abundance) with climate change. Application to forests in the eastern United States shows that climate can have greatest impact in the Northeast, due to temperature, and in the Upper Midwest, due to temperature and precipitation. Thus, these are the regions experiencing the fastest warming and are also identified as most responsive at this scale.

Key words: biodiversity; climate change; species abundance; species distributions; species occurrence; species presence/absence; zero inflation.

Introduction

Where is biodiversity most vulnerable to contemporary climate change? Biodiversity risk is often evaluated as the loss of species predicted to occur under a future climate scenario (Rodríguez-Castañeda et al. 2012). Species distribution models (SDMs) calibrate the distribution of each species to contemporary climate independent of other species (Iverson et al. 2008, Guisan and Thuiller 2005, Gelfand et al. 2006, Botkin et al. 2007, Fischlin et al. 2007, McMahon et al. 2011, Thuiller et al. 2011). Sometimes one species is included as a predictor of another (Araújo and Luoto 2007, Barbet-Massin and Jiguet 2011). In other cases, ordinations are applied to multiple species to identify environmental gradients that are combinations of input variables, such as climate. However, even where more than one species is included in a model, their combined responses are not treated as a joint distribution (Yee 2004; but see Ovaskainen et al. 2010 for presence–absence data). This is problematic, because loss of biodiversity is estimated as the sum of losses predicted by independent models for each species, and estimates are inaccurate (Baselga and Araújo 2010). We identify how problems arise and present an approach to resolve this problem, providing uncertainty on predictions for current and future climate using a single model for the joint distribution of all species simultaneously.

Evaluating the effects of climate on biodiversity typically begins with fitting a model for a species' occurrence or abundance to climate and other environmental variables (Guisan and Zimmermann 2000, Thuiller 2003, Latimer et al. 2006, Elith and Leathwick 2009, Marmion et al. 2009, Chakraborty et al. 2010, Benito et al. 2013). Models might be fitted to each of S species. A variable-selection criterion is often used to evaluate the importance of each input, or explanatory variable, for the species under consideration. The result is a collection of S models, each with different combinations of inputs. This collection of models cannot represent how the response is damped by competition when species tend to benefit from similar conditions, yet compete for limited space/resources. Baselga and Araújo (2010) found that different ways of clustering species in models did not correct inaccuracies in predictions made from independently fitted models.

Manuscript received 30 May 2013; revised 1 October 2013; accepted 15 October 2013. Corresponding Editor: D. S. Schimel.

[4] E-mail: jimclark@duke.edu

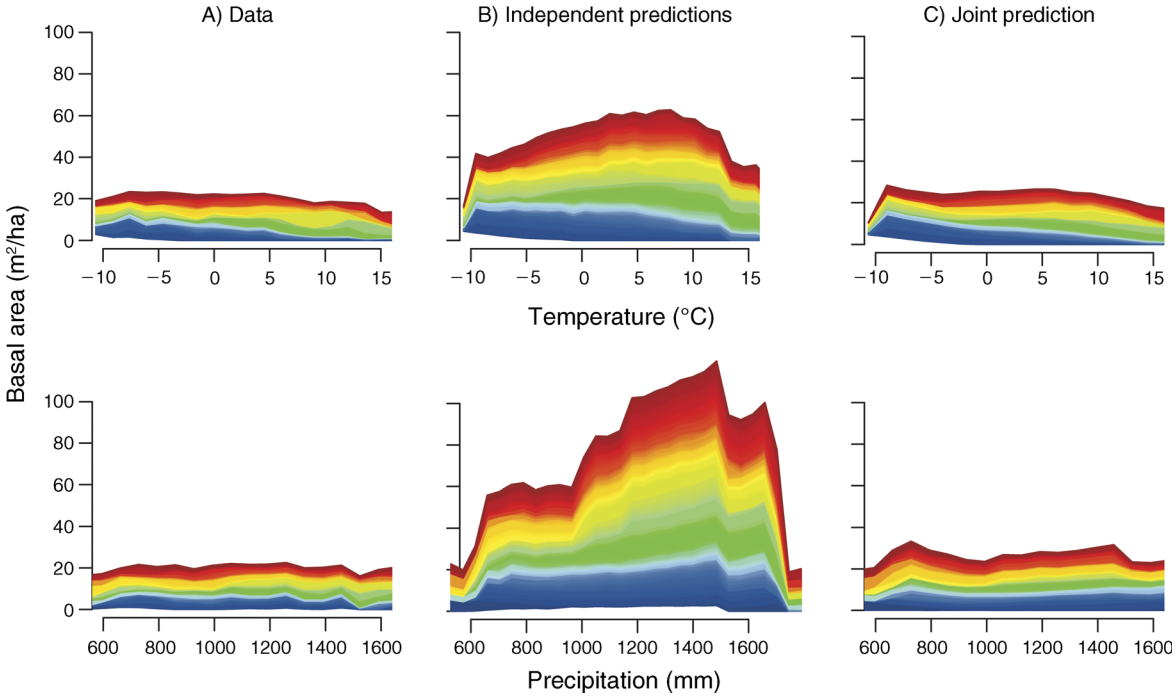

FIG. 1. Cumulative basal area by species (one color per species) for (A) the raw FIA data, (B) as predicted by models calibrated independently for each species, and (C) predicted by a joint model for all species. The latter predicts the data well, whereas the sum of independent models (panel B) is unrealistically high.

For example, independent models for clusters of species predict as many as five different forest types in a given location, where, by definition, there can be only one. Guisan and Rahbek (2011) found that independent models collectively imply too many species per location. A growing number of models that include multiple species (e.g., Ferrier et al. 2002, Elith et al. 2006, Mokany and Ferrier 2011, Mokany et al. 2011, 2012) do not model the joint distribution itself, but rather, rely on ad hoc methods to combine results from independent models. Ordination based regressions that include multiple species, but not their joint distribution (e.g., canonical Gaussian ordination [CGO]), did not perform better than independent models for each species (Yee 2004). In CGO, multiple species are used to identify a reduced set of predictor variables, rather than to estimate a joint distribution of species abundances. Yee (2004) suggests that species dependence might possibly be handled through generalized estimating equations, but we are unaware of implementations.

We describe an alternative approach that recognizes a joint distribution of the occurrence and the abundances of species competing for space. When species compete, an increase in the abundance of one species is attended by decreases in others. This fundamental property of communities results from competition, but is missing in SDMs. This space constraint is almost certainly violated when modeling is done on a species-by-species basis. In addition, independent models cannot exploit information contained in the joint relationships, such as the tendency for species to occupy similar sites for reasons other than those taken up by predictor variables (Clark et al. 2011, Ovaskainen and Soininen 2011). To appreciate how important the joint distribution is, consider that most ecologists could more accurately predict abundance of a species based on knowledge of the other species present at a site than they could based on, say, mean temperature or precipitation. We exploit the information inherent in the joint distribution of species with a joint model.

Problems that result from modeling species independently are easy to illustrate. For example, independently fitting 96 tree species in the USDA Forest Service's Forest Inventory and Analysis (FIA) database (Woudenberg et al. 2010) against climate variables (Fig. 1B) predicts basal area more than twice as high as observed (Fig. 1A). Published results that appear inconsistent with one another could result, in part, from the missing joint distribution. For example, one analysis of FIA data does not show climate sensitivity of species abundance across wide gradients in temperature and precipitation (Canham and Thomas 2010), while others suggests substantial response (Iverson and Prasad 2008, Zhu et al. 2014).

Given that species distribution modeling is usually concerned with impacts on multiple species (e.g., species richness, the number species at risk of extinction, the number migrating north or south, changes in total

NPP), current approaches could benefit from a joint modeling approach. Such a model would be more challenging to implement than individual models for each species, but it could ultimately be simpler than the collection of independent models.

Here we describe joint species distribution modeling (JSDM), demonstrate some of the challenges, and provide ways to address them. JSDM can be used to identify the important climatic variables that influence the species distributions across a full range of species in communities. It begins with a joint distribution of species (both occurrence and abundance) and builds on inverse prediction (Clark et al. 2012, 2013) to help identify which inputs (e.g., climatic variables) are most important in driving a high dimensional response (e.g., the distribution and abundance of co-occurring and competing species). The presence–absence component of our model is treated similar to that of Ovaskainen et al. (2010) and Ovaskainen and Soininen (2011). It is combined with abundance data in a single coherent model. The approach yields predictions of vulnerability for the full community of species, showing greatest potential for impacts of changing moisture in the Upper Midwest and temperature in the Upper Midwest and Northeast. We begin by summarizing five challenges for species distribution modeling in general that motivate our approach. We then describe the model and application.

Motivating Challenges

Much has been written about the challenges with species distribution modeling (Pearson and Dawson 2003, Ibañez et al. 2006, Latimer et al. 2009, Meier et al. 2011, Schurr et al. 2012) that guide the approach that follows. We do not repeat previous reviews (e.g., lack of dispersal dynamics or missing predictors, such as CO_2, fire, or soils), but focus instead on specific issues especially relevant for the multivariate setting.

The missing joint distribution.—A SDM for one species quantifies its relationship with climate variables. That relationship is complicated by other species. If there are 100 species, there can be 100 different models, each (implicitly) conditional on the abundances of the other 99. Our model for the joint distribution allows for dependence between species not accounted for by climate variables, and it yields uncertainty estimates that accommodate the joint response. We do not claim to quantify species interactions, only that the model allows for the fact that species abundances do covary. Our approach also differs from ordination, in which multiple species define gradients, but are not modeled as a joint distribution (e.g., Yee 2004).

We emphasize that the joint distribution does not quantify or predict "species interactions"; rather it accommodates the patterns that result from them. Although there are a growing number of species distribution models that are believed to quantify species interactions, observational data do not allow us to isolate effects of climate from competition. First is a scale mismatch. Individuals interact, plot- or region-scale abundances do not. Relationships at the plot and region scale result from interactions at the individual scale, but the aggregation of competing individuals into plot or region abundances introduces Simpson's paradox, including loss of information and distortion of relationships (Clark et al. 2011). A tendency for two species to occur together (positive correlation) can suggest mutualism (they benefit one another) or just the opposite, strong competition (high niche overlap). Likewise, negative correlations in species abundance can result from strong competition (competitive exclusion) or no competition (absence of niche overlap). Second, the incorporation of "species interactions" in SDMs also causes confusion when one species is assumed to predict another. Abundance of the predictor is fixed in the model, whereas the response is treated as stochastic (uncertain). We are unaware of biological or modeling justification for deciding which is which, unless one is observed without error, and it is uninfluenced by the other. JSDM discussed here assumes all species respond to climate and to one another, without prescribing some as predictors of others.

The important predictors.—Because species distribution modeling yields a different model and, thus, a different set of predictors for each species, there are as many sets of predictors as there are species. Some approaches use the same predictors for all species (e.g., Huntley et al. 2008). Alternatively, model selection is employed to pare down the number of predictors of specific species in the interest of parsimony and when overfitting is deemed undesirable. On the basis of a model-selection criterion, temperature might be selected for one species, precipitation for another, quadratic and interaction terms for others, and so forth. Given that the species are not responding independently, the approach appears both unnecessarily complex and potentially misleading. For example, a predictor may be significant for one species but not for its competitors. If species are modeled jointly, we can evaluate the role of inputs in the context of the full response. We apply inverse prediction, an adaptation of dynamic inverse prediction (Clark et al. 2011, 2013; J. Brynjarsdottir and A. E. Gelfand, *unpublished manuscript*), to reduce the dimensionality of sensitivity analysis. There is one model, and it can be inverted to determine the importance of predictors for species jointly.

The problem with an overabundance of absences.—Beyond the problems posed by presence-only data (e.g., Elith et al. 2006), there are problems with high-zero incidence in presence–absence and abundance data that are not widely appreciated. Zero-inflated models (ZIMs) assume that abundance is conditional on presence. A Bernoulli model for zeros is combined with a second model for abundance, which also admits zero values. The zero model "inflates" the zero class of the abundance model, in recognition of the fact that absence

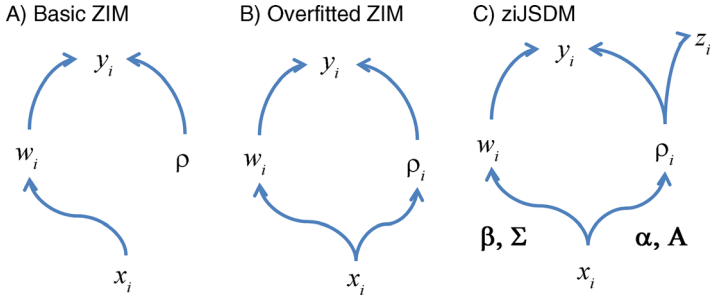

FIG. 2. Graphs for zero-inflated models (ZIM) with climate x_i influencing basal area y_i, allowing for zeros in a latent variable w_i and from a Bernoulli variable for absence having probability ρ. Climate may enter the model (A) for basal area alone or (B) for both basal area and absence. (C) The absence model can be informed by count data z_i, where $\boldsymbol{\alpha}$ is the parameter matrix, $\mathbf{A}$ is the species covariance matrix, $\boldsymbol{\beta}$ is the parameter matrix for basal area, and $\boldsymbol{\Sigma}$ is a covariance matrix.

might occur in at least two ways: (1) the sample is outside the species range or (2) the sample is within the range, but sample plots do not include it for other reasons. ZIMs are distinct from two-component models that do not have zero inflation, such as a Bernoulli-gamma mixture of Canham and Thomas ([2010]; the gamma distribution does not have point mass at zero).

However, severe identifiability issues arise with ZIMs when both the presence and the abundance of a species depend on the same predictor variables (Fig. 2B; Ghosh et al. 2012); sparse occurrence means that zeros are compatible with both components. Resolving the identifiability problem would appear to require that the same predictors not be included in both model components (Fig. 2A). The problem with this solution to non-identifiability is that it precludes use of the model for prediction. For example, if a model assumes that temperature affects abundance, then it could not also be included in the model for presence. Such a model could not predict how the species distribution is influenced by temperature. It could only predict abundance given temperature. Alternatively, if the climate variable is included in the zero model, but not the abundance model, then climate is assumed to have no effect on abundance throughout the species range. We resolve the identifiability problem by jointly modeling two types of abundance data that both admit zeros, basal area and stem counts, to anchor the two components of our ZIM (Fig. 2C). Basal area is modeled as a multivariate probit model, with an explicit zero class. Stem counts also have an explicit zero class that is the basis for presence–absence and, thus, the distribution of the species. This differs from a standard ZIP model, where zero inflation would come from a binomial distribution. In other words, both components admit zeros, but neither is binomial. This is a zero-inflated JSDM (ziJSDM).

Evaluating climate vulnerability from multiple species.—Climate vulnerability cannot be considered in isolation from interactions within a forest stand. Responses to drought and warming winters interact with light availability and local moisture (Clark et al. 2012). When models are fitted to species separately results cannot be objectively combined to yield a joint prediction of climate impacts. Joint modeling makes this prediction possible. We show how inverse prediction (Clark et al. 2011, 2013) can be applied to a joint distribution of species, thus providing a coherent prediction from the full species information. We emphasize again that we are not specifically modeling competition, but rather the species dependence that results from species interactions but that are not taken up by the mean structure of the model (see *The missing joint distribution*).

After summarizing the data set, we describe our approach that addresses each of these considerations.

MATERIAL AND METHODS

Data

FIA data were extracted for 43 396 inventory plots from annual inventories (2003–2009) in 31 states in the eastern US from the FIA database (FIADB version 4.0; Woudenberg et al. 2010). As FIA is authorized by the U.S. Congress to conduct a systematic and consistent inventory of all US forests (USDA Forest Service 2012), inventory methods are documented (Bechtold and Patterson 2005) with all relevant data available online and summarized in national resource reports (Smith et al. 2009). For this analysis, we use stem counts and basal areas from all plots aggregated into the 427 FIA ecoregions (McNab et al. 2007). Use of ecoregions, rather than individual plots, helps bring the response variables (stem counts and basal areas) into alignment with available climate data, which are highly smoothed. Plots are not instrumented; there is no plot-scale climate data. Analyses of responses at the plot scale and climate at the region scale cannot predict species abundance (Zhu et al. 2014). We analyzed the 96 most abundant species, which included 1 651 958 trees. Basal area is used here as the index of abundance. Stem counts from seedling and tree plots are the basis for modeling presence–absence.

Climate data are the parameter-elevation regressions on independent slopes model (PRISM) data set (*available online*).[5] Data are averages for 1990–2000. Data are provided at 800-m resolution, but observations are much coarser than this. The interpolated grid incorporates some information on landscape features. The data set is highly spatially smoothed relative to actual landscape climate variation. To maximize transparency,

[5] http://www.prism.oregonstate.edu/

we focus on two climate variables of known importance, winter temperature (January, February, March) and annual precipitation, although they do not emerge as important in all studies. We select these variables as basis for this demonstration of the joint modeling approach because they have been widely used in the past. Standard model-selection techniques are available for our approach and could be used to examine any available variables.

Model development

We develop a model for joint distributions of both presence/absence and abundance, both in response to climate, for all species, as outlined in Fig. 2C. Our goals here are to model both distribution and abundance, the latter represented by basal area, and to use count data to anchor the zero class. Together this is a zero-inflated JSDM (ziJSDM) model (Appendix). It applies a latent variable approach to basal area modeling.

There is a pair of length-S vectors of species at location i, one for counts ($\mathbf{z}_i$) and another for abundance (basal area; $\mathbf{y}_i$). Both have mean vectors that are multivariate normal at the first stage, predicted by temperature and precipitation, with coefficient vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and residual covariance matrices $\mathbf{A}$ and $\boldsymbol{\Sigma}$, respectively. Both models (presence–absence and abundance) contain the same linear, quadratic, and interaction terms for these two climate variables. This redundancy is responsible for overfitting in standard ZIMs (Ghosh et al. 2012). Specifically, the model components compete to explain the zeros. This is recognized when a predictor that occurs in both model components tends to have positive coefficients in one and negative coefficients in the other.

Our approach departs from a standard ZIM in key ways. First, is the treatment of zeros. Rather than a mixture for two types of zeros in a single response variable (e.g., one interpreted as absence from the ecoregion and another as presence in the ecoregion, but absence from the sample), we model ecoregion absence conditionally from the stem count $z_{is} \sim \text{Poisson}(A_i\gamma_{is})$, where A_i is the sample area for ecoregion i, and γ_{is} is the density of species s in i. The absence model is the zero class of the Poisson distribution

$$\Pr(z_{is} = 0) = e^{-A_i\gamma_{is}}$$

where $z_{is} = 0$, $y_{is} = 0$ (basal area requires stems). However, z_{is} and y_{is} contain different information about zeros, because, unlike a standard ZIM, where $P(y_{is} = 0)$ is a binary event, z_{is} is not binary. The joint distribution for log density is a multivariate GLM $\ln \boldsymbol{\gamma}_i \sim N_S(\mathbf{x}_i\boldsymbol{\alpha}, \mathbf{A})$, where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{iS})$ is the vector of species' densities in ecoregion i, $\boldsymbol{\alpha}$ is the $p \times S$ parameter matrix, and $\mathbf{A}$ is the species covariance matrix.

Second, zero inflation enters as a multivariate probit model for a latent variable w_{is}, equal to basal area y_{is} when y_{is} is non-zero (positive), but taking negative values otherwise. The joint distribution of basal area $\mathbf{w}_i \sim \text{TN}_S(\mathbf{x}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{h}_i)$ is a truncated normal distribution with mean, variance, and truncation vector $\mathbf{h}_i$ that depends on predictor, Pi (Fig. 2c, Appendix), $\boldsymbol{\beta}$ is the $p \times S$ parameter matrix for basal area, and $\boldsymbol{\Sigma}$ is a covariance matrix. The zero class from the abundance model is multivariate probit.

In summary, the model allows for joint distributions of basal area $\mathbf{Y}$ and density $\mathbf{Z}$, both of which depend on climate. We have two sources of zeros (zero-inflation), but neither relies solely on presence–absence data. Zero inflation modeled in this way allows us to quantify climate effects on both distribution and abundance without overfitting. Prior distributions and posterior simulation are described in the Appendix. To summarize the Appendix, prior distributions are non-informative for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and informative for $\boldsymbol{\Sigma}$ and $\mathbf{A}$.

Inverse prediction

Climate vulnerability is both a sensitivity problem and a prediction problem: how much will the distribution, abundance, or both change with climate? The answer is different for each species, but we cannot simply add together sensitivity coefficients from models fitted to each species separately. First, there are simply too many sensitivity coefficients. If there are S species and p predictors, there are as many as $p \times S$ sensitivity coefficients. Second, they are not independent, and we have no objective way of combining them. One species can benefit from ameliorating climate only at the expense of others. A species insensitive to climate will nonetheless respond if competitors respond. The joint model described above captures these relationships among species, but, by itself, does not provide an answer to the question "Where is the greater vulnerability of the joint distribution?" The answer to this question is buried in $>10^2$ coefficients, $p \times S$ coefficients in $\boldsymbol{\alpha}$ and $p \times S$ coefficients in $\boldsymbol{\beta}$. We require a sensitivity analysis that combines the information from all coefficients in the context of the full fitted model.

Inverse prediction combines information in a multivariate response vector (e.g., species abundance or presence) and the full fitted model to predict the climate variables having most impact (Clark et al. 2011, 2013; J. Brynjarsdottir and A. E. Gelfand, *unpublished manuscript*). Inverse prediction comes from inverting the fitted model to predict the inputs (climate) from the outputs (distribution and abundance). Accurate and precise predictions of a climate variable are possible where that variable exerts strong controls on the multivariate response vector: in this case, the joint distribution of species. Conversely, poor predictions obtain where a climate variable has no impact on the response. The advantage of inverse prediction over standard sensitivity analysis comes from the fact that it combines all responses into a single sensitivity value weighted by the fitted model itself (Clark et al. 2011, 2013). There is a predictive distribution for the input q in region i ranked by the proper scoring rule of Gneiting

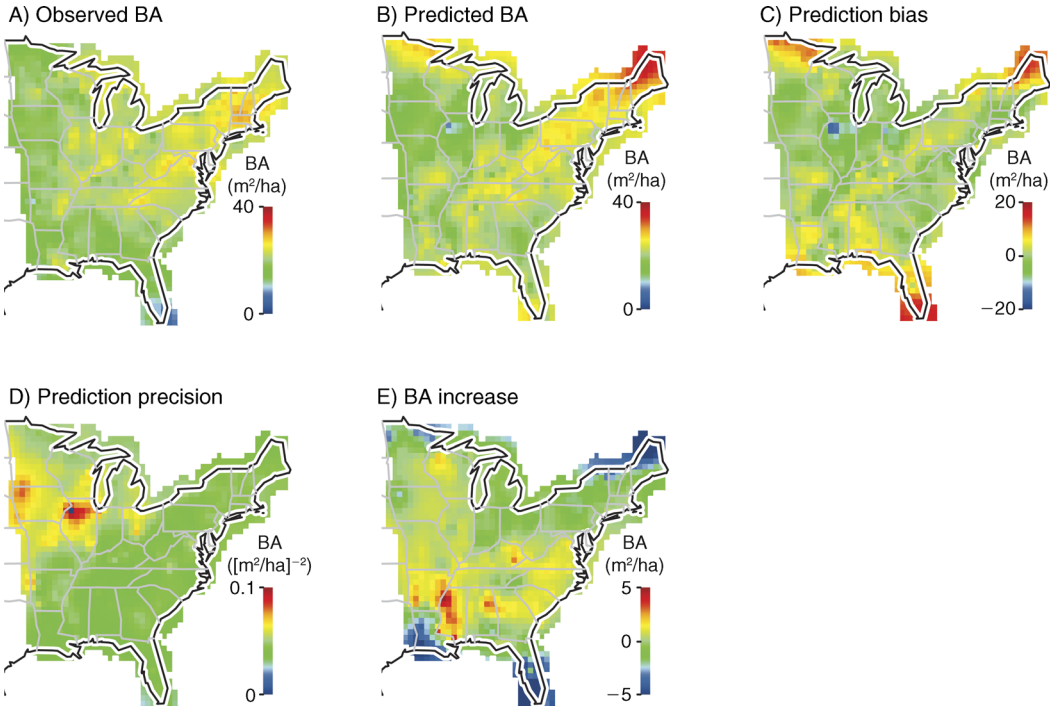

FIG. 3. Current and prediction (A, B) mean basal area (BA), (C, D) prediction bias and precision, and (E) predicted change in basal area for the A2 scenario.

and Raftery (2007). High scores result from accurate predictions (predicted inputs close to observed inputs), and confident (but not overconfident) predictions.

RESULTS

Our first concern was to establish that data have updated prior distributions of parameters. Because there are a large number of parameter estimates ($p \times S$ in both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$), not all might learn from data. In fact, we obtained concentrated posteriors for both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ for most of the 96 species in our analysis. Posterior summaries are provided in the Appendix.

We found that the sum of independent SDMs does not produce sensible predictions for total basal area (Fig. 1B), and we wanted to know if the ziJSDM model is an improvement. We found that predictions for total basal area for our model (Fig. 1C) are comparable to observed values (Fig. 1A). This contrasts with the predictions of unrealistically high values from independently fitted SDMs (Fig. 1B). A mapped comparison based on all 96 species (Fig. 3A, B) shows that a tendency to overpredict basal area occurs in northern New England and northern Minnesota (Fig. 3C), but the qualitative patterns are in agreement (Fig. 3A, B). Predictions are most confident in the Midwest (Fig. 3D).

Like total basal area, species richness predictions (Fig. 4B) are quite close to observed (Fig. 4A). These are shown at a value of 0.6 probability of presence. Note that 0.5 is the midway point, but it also represents maximum uncertainty. We display values for 0.6 in Fig. 4B to reflect "more certain than not" that the species is present.

The tendency for lower prediction errors for basal area in the ziJSDM than for independently fitted models for each species (Appendix) supports the use of count data to allow for geographic variation in presence/absence estimates. The ziJSDM model does not suffer from overfitting problems. Ghosh et al. (2012) found that redundant predictors in the two components tended to offset one another, with high values in one component of the mixture offset by low values in the other. In our example, this would be recognized as a negative parameter estimate for temperature in $\boldsymbol{\beta}$ paired with a positive estimate in $\boldsymbol{\alpha}$. This offsetting relationship apparently results from the fact that zeros can be assigned to either component of the model, when the same predictors are included in both. Instead, we find a tendency for a positive relationship between the two in this analysis (Appendix), which is consistent with biological expectations and not with overfitting: a species is likely to be abundant where it is likely to be present, and not vice versa.

Inverse prediction from the joint model provides a synthesis of regions where temperature and precipitation have most impact on the full community of species. Scores for temperature are higher than those for precipitation (Fig. 5, right panels). Temperature scores high in regions with the lowest and highest temperatures (Fig. 5, upper left). Precipitation scores high in regions with the lowest precipitation. Results suggest that

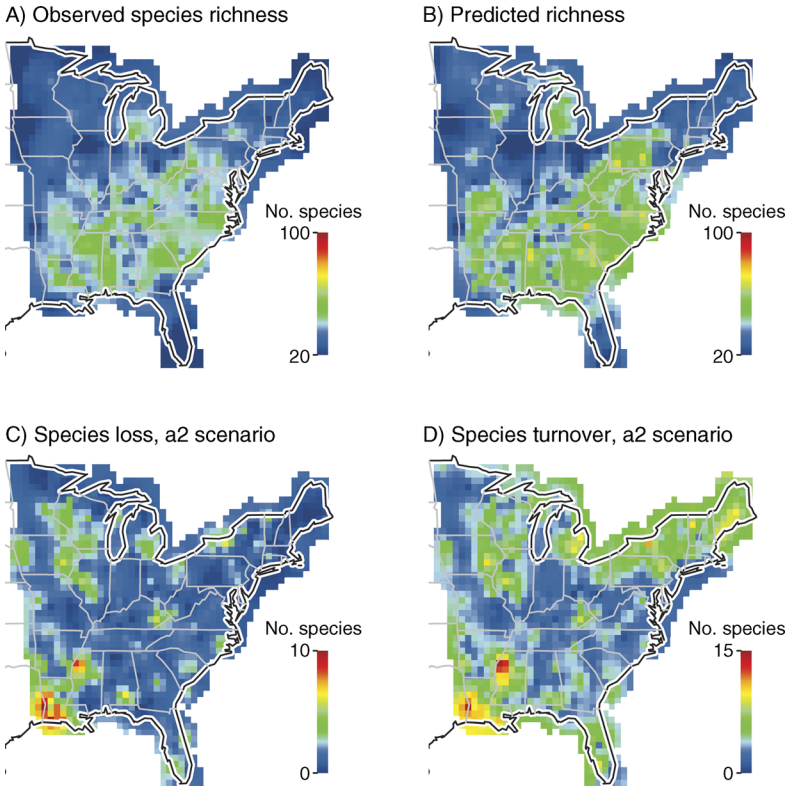

FIG. 4. (A) Observed and (B) zero-inflated joint species distribution modeling (ziJSDM) prediction of species richness at 60% confidence (those present with probability > 0.6) for current climate. (C) Species losses predicted for the A2 scenario and (D) turnover (gains plus losses) at 60% confidence .

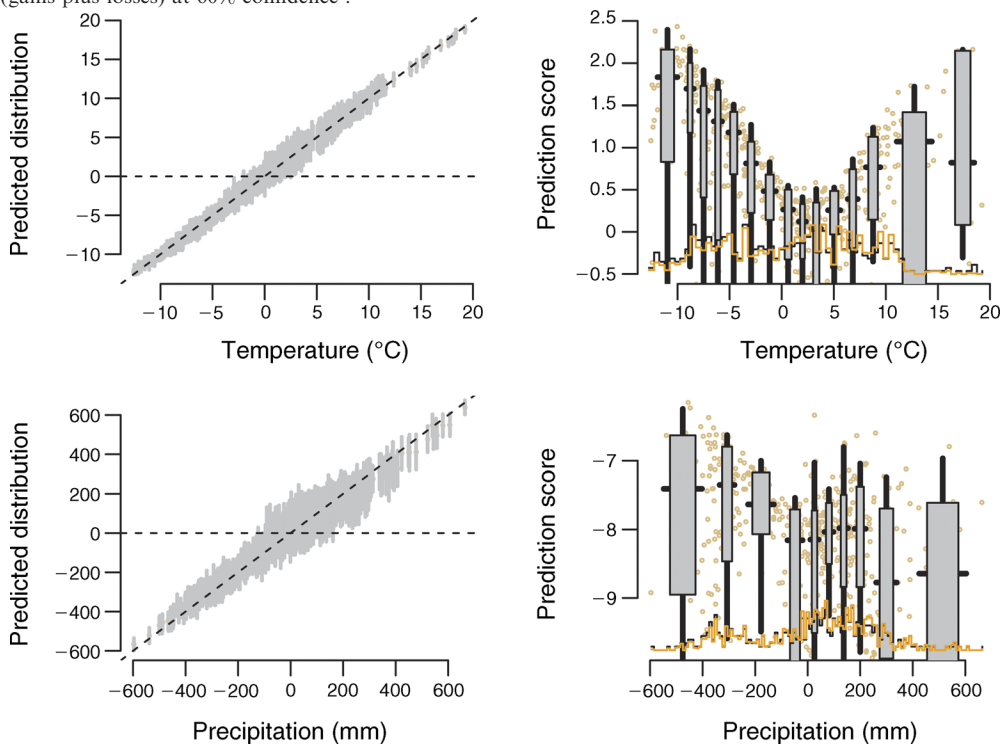

FIG. 5. Temperature and precipitation on horizontal axes compared with predictive distributions (95% as shaded gray bars at left) and prediction scores (points and bars for 68% and 95% at right). Also shown with predictive scores are distributions of data (black histograms) and distributions of non-zero values (orange histograms). Both variables are centered on the mean value.

temperature is more important than precipitation (overall higher scores), temperature is most important at extremes, and precipitation is least important where it is highest. The latter trend could result if precipitation is less often limiting in the wettest regions. A mapping of prediction scores shows largest temperature effects in the north and south (Fig. 6, top). Precipitation effects are greatest in the upper Midwest (Fig. 6, bottom). Unlike temperature, sensitivity to precipitation is especially low where precipitation is highest, i.e., the Gulf Coast. The eastern central region emerges as least sensitive to both variables.

Applying the fitted model to the A2 scenario (Nakicenovic and Swart 2000, Solomon et al. 2007) we find that the difference between predicted basal area for current conditions (Fig. 3B) and for the A2 scenario show gains in basal area in the South and Midwest, the largest in the southwestern portion of the map (Fig. 3E). Despite the gains in basal area, species losses are largest in the South and Midwest (Fig. 4C). Total turnover (sum of losses and gains) is greatest in the northern and southwestern portions of the map (Fig. 4D).

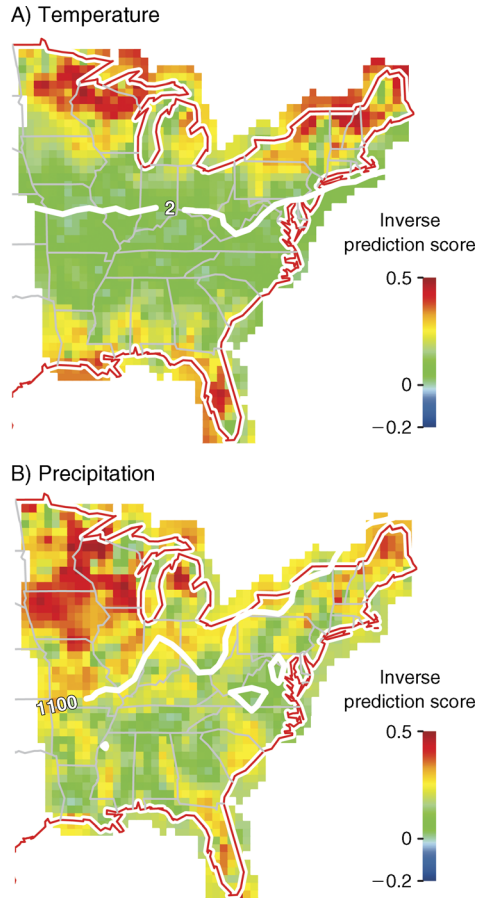

FIG. 6. Inverse prediction scores for (A) temperature and (B) precipitation. Mean values are shown as white lines.

DISCUSSION

Despite continuing debate over SDMs, climate envelopes, and niche modeling generally, these methods remain one of the few ways to evaluate potential change in species' distribution and abundance at the sub-continental scale. Our approach addresses some of these concerns, by synthesizing presence–absence and abundance jointly across species, by resolving the over-fitting problem of zero-inflated models, and offering prediction of communities as opposed to individual species. The advantages of a single model for all species, rather than independent models for each, extend beyond convenience. Because the model learns from the joint relationships in data, it can provide accurate predictions (Figs. 1, 3, 4). Both distribution and abundance are contained in the same model, and the fitted model predicts both (Figs. 3B, 4B).

Inverse prediction reduces the high-dimensional distribution down to one dimension per predictor, identifying where sensitivity is high based on all species (Fig. 6). In contrast to models that seek to identify vulnerable regions by summing the results from independently fitted models, our approach does not rely on ad hoc rules for putting independent models together. It applies only accepted distribution theory and a coherent treatment of the joint distribution, grounded on the constraint that species do not respond independently. Abundance of one species increases at the expense of another.

Joint species distribution modeling (JSDM) exploits information in the joint distribution to synthesize and improve prediction. In many situations, including most applications involving SDMs, there is limited explanatory information in input variables. Regional climate variables are not nearly sufficient to predict the rich variation in diversity across landscapes. The fact that species abundances are among the best predictors of one another stimulated a tradition of ordination and classification of plant communities (Bray and Curtis 1957, Whittaker 1967, ter Braak and Prentice 1988). Exploiting this information in a JSDM makes the difference between useful and non-useful predictions (e.g., Fig. 1).

Should joint species distribution modeling (JSDM) be the sole approach for species distribution modeling, or is there still a place for independent modeling of species? JSDM provides no advantage for species that provide no information on one another and when there is no interaction between them. Independent models of species can be as effective as JSDM when considering species that bear no relationship to one another. Whenever SDMs involve species that interact or that respond to similar inputs, JSDM is expected to provide large improvements.

By itself JSDM does not solve the problem of quantifying fine-scale processes, such as species interactions. As discussed above, the JSDM accommodates outcomes of fine-scale processes as they control patterns

in the aggregate scale. Although the improvement in predictions is substantial, we cannot misinterpret this as the "effect" of one species on another. We do not parameterize competition, but rather allow for the fact that the responses of species are interrelated beyond what can be explained by predictor variables. The advantages of joint approach for predicting change in distribution and abundance does not mean that all sources of uncertainty can be quantified. An important benefit of this joint specification within a hierarchical framework is the capacity to include error from parameters, the limited capacity of climate to predict distribution and abundance, and the correlation structure of density and basal area. Prediction bias and precision in abundance (Fig. 3) is available directly from the model, as is uncertainty in presences (0.6 is used in Fig. 4). These sources of uncertainty are coherently integrated with inverse prediction.

Inverse prediction highlights the upper Midwest and Northeast as regions most closely controlled by temperature and precipitation (Fig. 6): the coldest and driest regions. Application to climate change scenarios identifies the same areas as places of high species turnover (Fig. 4D). In the case of these northern regions, they also coincide with the region where climate change is most rapid (Zhu et al. 2012). The capacity to evaluate climate relationships across subcontinental scale regions can complement efforts to evaluate biodiversity threats of climate change.

Acknowledgments

This study was supported by NSF grants EF-1137364, CDI 0940671, and the Coweeta LTER. We appreciate comments on the manuscript from Janneke HilleRisLambers and three anonymous reviewers.

Literature Cited

Araújo, M. B., and M. Luoto. 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecology and Biogeography 16:743–753.

Barbet-Massin, M., and F. Jiguet. 2011. Back from a predicted climatic extinction of an island endemic: a future for the Corsican Nuthatch. PLoS ONE 6(3):e18228.

Baselga, A., and M. B. Araújo. 2010. Do community models fail to project community variation effectively? Journal of Biogeography 37:1842–1850.

Bechtold, W. A., and P. L. Patterson 2005. The enhanced Forest Inventory and Analysis Program: national sampling design and estimation procedures. USDA Forest Service, Southern Research Station, Asheville, North Carolina, USA.

Benito, B. M., L. Cayuela, and F. S. Albuquerque. 2013. The impact of modelling choices in the predictive performance of richness maps derived from species-distribution models: guidelines to build better diversity models. Methods in Ecology and Evolution 4:327–335.

Botkin, D. B., et al. 2007. Forecasting the effects of global warming on biodiversity. BioScience 57:227–236.

Bray, J. R., and J. T. Curtis. 1957. An ordination of upland forest communities of southern Wisconsin. Ecological Monographs 27:325–349.

Canham, C. D., and R. Q. Thomas. 2010. Frequency, not relative abundance, of temperate tree species varies along climate gradients in eastern North America. Ecology 91:3433–3440.

Chakraborty, A., A. E. Gelfand, J. A. Silander, Jr., A. M. Latimer, and A. M. Wilson. 2010. Modeling large scale species abundance through latent spatial processes. Annals of Applied Statistics 4:1403–1429.

Clark, J. S., D. M. Bell, M. H. Hersh, M. Kwit, E. Moran, C. Salk, A. Stine, D. Valle, and K. Zhu. 2011. Individual-scale variation, species-scale differences: inference needed to understand diversity. Ecology Letters 14:1273–1287.

Clark, J. S., D. M. Bell, M. Kwit, A. Powell, R. Roper, A. Stine, B. Vierra, and K. Zhu. 2012. Individual-scale inference to anticipate climate-change vulnerability of biodiversity. Philosophical Transactions of the Royal Society B 367:236–246.

Clark, J. S., D. M. Bell, M. Kwit, A. Powell, and K. Zhu. 2013. Dynamic inverse prediction and sensitivity analysis with high-dimensional responses: application to climate-change vulnerability of biodiversity. Journal of Biological, Environmental, and Agricultural Statistics 18:376–404.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology and Systematics 40:677–697.

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

Ferrier, S., M. Drielsma, G. Manion, and G. Watson. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in north-east New South Wales: II. Community-level modelling. Biodiversity and Conservation 11:2309–2338.

Fischlin, A., et al. 2007. Ecosystems, their properties, goods and services. Page 102 *in* M. Parry, O. Canziani, and J. Palutikof, editors. Climate change 2007: impacts, adaptation and vulnerability. Contribution of Working Group II to the Assessment Report Four of the Intergovernmental Panel of Climate Change (IPCC). Cambridge University Press, Cambridge, UK.

Gelfand, A. E., J. A. Silander, S. Wu, A. Latimer, P. O. Lewis, A. G. Rebelo, and M. Holder. 2006. Explaining species distribution patterns through hierarchical modeling. Bayesian Analysis 1:41–92.

Ghosh, S., A. E. Gelfand, K. Zhu, and J. S. Clark. 2012. The k-ZIG: flexible modeling for zero-inflated counts. Biometrics 68:878–885.

Gneiting, T., and A. E. Raftery. 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102:359–378.

Guisan, A., and C. Rahbek. 2011. SESAM—a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. Journal of Biogeography 38:1433–1444.

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

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

Huntley, B., Y. C. Collingham, S. G. Willis, and R. E. Green. 2008. Potential impacts of climatic change on European breeding birds. PLoS ONE 3:e1439.

Ibáñez, I., J. S. Clark, M. C. Dietze, K. Feeley, M. Hersh, S. LaDeau, A. McBride. N. E. Welch, and M. S. Wolosin. 2006. Predicting biodiversity change: outside the climate envelope, beyond the species–area curve. Ecology 87:1896–1906.

Iverson, L. R., A. M. Prasad, S. N. Matthews, and M. Peters. 2008. Estimating potential habitat for 134 eastern U.S. tree species under six climate scenarios. Forest Ecology and Management 254:390–401.

Latimer, A. M., S. Banerjee, H. Sang, E. Mosher, and J. A. Silander. 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.

Latimer, A. M., S. S. Wu, A. E. Gelfand, and J. A. Silander. 2006. Building statistical models to analyze species distributions. Ecological Applications 16:33–50.

Marmion, M., M. Parviainen, M. Luoto, R. K. Heikkinen, and W. Thuiller. 2009. Evaluation of consensus methods in predictive species distribution modelling. Diversity and Distributions 15:59–69.

McMahon, S. M., S. P. Harrison, W. S. Armbruster, P. J. Bartlein, C. M. Beale, M. E. Edwards, J. Kattge, G. Midgley, X. Morin, and I. C. Prentice. 2011. Improving assessment and modelling of climate change impacts on global terrestrial biodiversity. Trends in Ecology and Evolution 26:249–259.

McNab, W. H., D. T. Cleland, J. A. Freeouf, J. J. E. Keys, G. J. Nowacki, and C. A. Carpenter. 2007. Description of ecological subregions: sections of the conterminous United States. General Technical Report WO-76B edition. USDA Forest Service, Washington, D.C., USA.

Meier, E. S., T. C. Edwards, F. Kienast, M. Dobbertin, and N. E. Zimmermann. 2011. Competition patterns of trees along macro-climatic gradients and their potential influence on the present and future distribution of *Fagus sylvatica*. Journal of Biogeography 38:371–382.

Mokany, K., and S. Ferrier. 2011. Predicting impacts of climate change on biodiversity: a role for semi-mechanistic community-level modelling. Diversity and Distributions 17:374–380.

Mokany, K., T. D. Harwood J. M. Overton G. M. Barker, and S. Ferrier. 2011. Combining alpha- and beta-diversity models to fill gaps in our knowledge of biodiversity. Ecology Letters 14:1043–1051.

Mokany, K., T. D. Harwood, K. J. Williams, and S. Ferrier. 2012. Dynamic macroecology and the future for biodiversity. Global Change Biology 18:3149–3159.

Nakicenovic, N., and R. Swart, editors. 2000. Intergovernmental Panel on Climate Change special report on emissions scenarios. Cambridge University Press, Cambridge, UK.

Ovaskainen, O., J. Hottola, and J. Siitonen. 2010. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. Ecology 91:2514–2521.

Ovaskainen, O., and J. Soininen. 2011. Making more out of sparse data: hierarchical modeling of species communities. Ecology 92:289–295.

Pearson, R. G., and T. P. Dawson. 2003. "Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?" Global Ecology and Biogeography 12(5):361–371.

Rodríguez-Castañeda, G., A. R. Hof, R. Jansson, and L. E. Harding. 2012. Predicting the fate of biodiversity using species' distribution models: enhancing model comparability and repeatability. PLoS ONE 7(9):e44402.

Schurr, F. M., J. Pagel, J. S. Cabral, J. Groeneveld, O. Bykova, R. B. O'Hara, F. Hartig, W. D. Kissling, H. P. Linder, and G. F. Midgley. 2012. How to understand species' niches and range dynamics: a demographic research agenda for biogeography. Journal of Biogeography.

Smith, W. B., P. D. Miles, C. H. Perry, and S. A. Pugh. 2009. Forest resources of the United States, 2007. USDA Forest Service, Washington Office, Washington, D.C., USA.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, editors. 2007. Climate change 2007. The physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK.

ter Braak, C. J. F., and I. C. Prentice. 1988. A theory of gradient analysis. Advances in Ecological Research 18:271–317.

Thuiller, W. 2003. BIOMOD—optimizing predictions of species distributions and projecting potential future shifts under global change. Global Change Biology 9:1353–1362.

Thuiller, W., S. Lavergne, C. Roquet, I. Boulangeat, B. Lafourcade, and M. B. Araújo. 2011. Consequences of climate change on the Tree of Life in Europe. Nature 470:531–534.

USDA Forest Service. 2012. Forest inventory and analysis: fiscal year 2011 business report. WO-FS-999. U.S. Department of Agriculture, Forest Service, Washington, D.C., USA.

Whittaker, R. H. 1967. Gradient analysis of vegetation. Biology Reviews 42:207–264.

Woudenberg, S. W., et al. 2010. The forest inventory and analysis database: database description and users manual version 4.0 for Phase 2. General Technical Report RMRS-GTR-245. USDA Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, USA.

Yee, Thomas W. 2004. A new technique for maximum-likelihood canonical Gaussian ordination. Ecological Monographs 74:685–701.

Zhu, K., C. W. Woodall, and J. S. Clark. 2012. Failure to migrate: lack of tree range expansion in response to climate change. Global Change Biology 18:1042–1052.

Zhu, K., C. W. Woodall, S. Ghosh, A. E. Gelfand, and J. S. Clark. 2014. Dual impacts of climate change: forest migration and turnover through life history. Global Change Biology 20:251–264.

Supplemental Material

Appendix

Detailed description of the model, prior distributions, diagnostics, and inverse prediction (*Ecological Archives* A024-058-A1).