


Joint species distribution modelling for spatio-temporal occurrence and ordinal abundance data

Erin M. Schliep¹  | Nina K. Lany^{2,3} | Phoebe L. Zarnetske^{2,3} |
Robert N. Schaeffer⁴ | Colin M. Orians⁴ | David A. Orwig⁵ | Evan L. Preisser⁶

¹Department of Statistics, University of Missouri, Columbia, Missouri

²Department of Forestry, Michigan State University, East Lansing, Michigan

³Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, East Lansing, Michigan

⁴Department of Biology, Tufts University, Medford, Massachusetts

⁵Harvard Forest, Harvard University, Petersham, Massachusetts

⁶Department of Biological Sciences, University of Rhode Island, Kingston, Rhode Island

Correspondence

Erin M. Schliep, Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211 USA.
Email: schliepe@missouri.edu

Funding information

Arnold and Mabel Beckman Foundation; Michigan State University; National Institute of Food and Agriculture, Grant/Award Number: 2011-67013-30142; NSF DEB, Grant/Award Number: 0715504, 1256769, 1256826, 0620443

Editor: Antoine Guisan

Abstract

Aim: Species distribution models are important tools used to study the distribution and abundance of organisms relative to abiotic variables. Dynamic local interactions among species in a community can affect abundance. The abundance of a single species may not be at equilibrium with the environment for spreading invasive species and species that are range shifting because of climate change.

Innovation: We develop methods for incorporating temporal processes into a spatial joint species distribution model for presence/absence and ordinal abundance data. We model non-equilibrium conditions via a temporal random effect and temporal dynamics with a vector-autoregressive process allowing for intra- and interspecific dependence between co-occurring species. The autoregressive term captures how the abundance of each species can enhance or inhibit its own subsequent abundance or the subsequent abundance of other species in the community and is well suited for a 'community modules' approach of strongly interacting species within a food web. R code is provided for fitting multispecies models within a Bayesian framework for ordinal data with any number of locations, time points, covariates and ordinal categories.

Main conclusions: We model ordinal abundance data of two invasive insects (hemlock woolly adelgid and elongate hemlock scale) that share a host tree and were undergoing northwards range expansion in the eastern U.S.A. during the period 1997–2011. Accounting for range expansion and high inter-annual variability in abundance led to improved estimation of the species–environment relationships. We would have erroneously concluded that winter temperatures did not affect scale abundance had we not accounted for the range expansion of scale. The autoregressive component revealed weak evidence for commensalism, in which adelgid may have predisposed hemlock stands for subsequent infestation by scale. Residual spatial dependence indicated that an unmeasured variable additionally affected scale abundance. Our robust modelling approach could provide similar insights for other community modules of co-occurring species.

KEYWORDS

biotic interactions, coregionalization, invasive species, Markov chain Monte Carlo, rank probability scores, vector autoregression

1 | INTRODUCTION

Species distribution models are commonly used in basic and applied ecological research to study the factors that define the distribution and abundance of organisms. They are used to quantify species' relationships with abiotic conditions, to predict species' response to land-use

and climatic change and to identify potential conservation areas (Guisan & Zimmermann, 2000). Traditionally, species distribution models correlate static observations of the occurrence (presence and absence) or abundance of a species with abiotic variables. Occurrence and abundance of a species, however, can also change through time and across space through colonization and may not be at equilibrium

with the environment (Bolker & Pacala, 1997; Ives, Dennis, Cottingham, & Carpenter, 2003). Non-equilibrium is expected for invasive species and species affected by climate change (Elith, Kearney, & Phillips, 2010; Peterson, 2003). Additionally, dynamic local interactions among species in a community can affect abundance and scale up to affect distributions at broader spatial scales (Araújo & Rozenfeld, 2014; Gotelli, Graves, & Rahbek, 2010). Capturing these spatio-temporal ecological processes requires the use of more robust modelling techniques that account for temporal dynamics and joint dependencies among co-occurring species (reviewed by Ehrlén & Morris, 2015; Guisan & Thuiller, 2005).

Joint dependencies among co-occurring species in a community have been modelled in multivariate generalized linear model frameworks (Clark, Gelfand, Woodall, & Zhu, 2014; Clark, Nemerbut, Seyednasrollah, Turner, & Zhang, 2017; Hui, Warton, Foster, & Dunstan, 2013; Kissling et al., 2012; Ovaskainen, Hottola, & Siitonen, 2010; Pollock et al., 2014; Thorson, Scheuerell et al., 2015). In joint species distribution models, residual dependence in occurrence or abundance can arise from biotic interactions and as correlated responses to an unmeasured covariate. These generalized linear modelling approaches rely on the assumption that species are at equilibrium with their environment because the models are static; they integrate data over a fixed time interval, without modelling changes through time.

Incorporating a time series of observations can improve species distribution models. For example, including vector autoregressive terms in a joint species distribution model can identify the temporal effects of biotic interactions, such as competition between co-occurring species (Mutshinda, O'Hara, & Woivod, 2009, 2011) or heterospecific attraction (Sebastián-González, Sánchez-Zapata, Botella, & Ovaskainen, 2010). Using temporal autocorrelation to incorporate dynamic biotic interactions into joint species distribution models is especially well suited for a 'community modules' approach that focuses on smaller subsets of strongly interacting species within a community (Gilman, Urban, Tewksbury, Gilchrist, & Holt, 2010; Holt, 1997). Incorporating these dynamic processes when time-series data are gathered over multiple years at biologically relevant intervals is an important step towards incorporating demography and biotic interactions into distribution models (Schurr et al., 2012). Additionally, temporal random effects could be used to address violations of the assumption that species distributions and abundance are at equilibrium with abiotic factors. This is important because violations of the assumption that species' ranges are at equilibrium can affect parameter estimation and interpretation of model results (Elith et al., 2010).

In species distribution models, it is also important to account for the spatial patterns often present in occurrence or abundance data collected at multiple locations. Residual spatial patterns can arise from ecological processes, such as movement or dispersal, source-sink dynamics, aggregation or social structure, or in response to unmeasured environmental covariates (Keitt, Bjornstad, Dixon, & Citron-Pousty, 2002). Ignoring spatial patterns violates model assumptions and potentially introduces bias in parameter estimates that describe the species-environment relationship (Dormann et al., 2007; Keitt et al., 2002). The ability of a species distribution model to predict

abundance at unobserved locations (Carroll, Johnson, Dunk, & Zielinski, 2010; Dray et al., 2012) or under climate-change scenarios (Crane, Liedloff, Vesk, Fukuda, & Wintle, 2014; Record, Fitzpatrick, Finley, Veloz, & Ellison, 2013) is often improved by using a spatially explicit model structure where information can be shared across both species and space (Ovaskainen, Roy, Fox, & Anderson, 2016; Thorson, Iannelli, Munch, Ono, & Spencer, 2015).

Abundance data are likely to be more informative than binary occurrence data for species distribution modelling, even if abundance is measured coarsely (Howard, Stephens, Pearce-Higgins, Gregory, & Willis, 2014). Abundance data are better indicators of the effects of a species on an ecosystem, the probability of extinction of rare organisms and the potential of one species to affect other species within a local community through biotic interactions (Ehrlén & Morris, 2015). However, additional challenges arise when abundance data are observed imperfectly. Ordinal data are one type of abundance data with imperfect detection where abundance is recorded in ordered categories, and binary occurrence data are a special case of ordinal data with only two abundance categories (present and absent). Ordinal scales for abundance data are commonly used in ecology when it is not possible or practical to record count data, but information on abundance beyond simple presence/absence is desired (examples in Table 1). Other ordinal responses, such as severity or intensity of disturbance, disease or damage, are also common.

Bayesian hierarchical models can effectively capture these dependencies and dynamics that are inherent in multivariate spatio-temporal datasets, and Markov chain Monte Carlo methods make estimating the complex posterior distributions of these models computationally feasible (Banerjee, Carlin, & Gelfand, 2014). We present a spatial joint species distribution model for multivariate ordinal abundance data with extensions that incorporate temporal random effects to address the assumption that species are at equilibrium with the environment, and vector autoregression to model temporal dynamics in abundance within and among species. We provide an additional model extension that accommodates replicated measurements at each location and time point. Explicitly modelling replicated measures enables inference on the variability in ordinal abundance within species and the covariance between species at the level of the observational unit when replicated measurements are made at each location and time point. We include annotated R functions for fitting multispecies models within a Bayesian framework for ordinal data with any number of abundance categories, time points, spatial locations and covariates. The functions permit the user to apply each individual model extension (i.e., temporal random effects, temporal dynamics and replicated measurements) separately or in combination. We developed custom Metropolis-within-Gibbs algorithms for fitting the models, and we evaluate their efficiency and accuracy.

We illustrate the model and each extension with a spatially explicit time series of the ordinal abundance of two invasive insects [hemlock woolly adelgid (*Adelges tsugae*) and elongate hemlock scale (*Fiorinia externa*)] that share a host plant and were undergoing northwards range expansion in the eastern U.S.A. during the period 1997–2011. We evaluate how violation of the assumption of equilibrium with the environment

TABLE 1 Some examples of ordinal data in ecology that could be analysed using this multivariate spatio-temporal distribution model

Example (reference)	Data
Occurrence (binary)	$Y = \begin{cases} 0 & \text{absent} \\ 1 & \text{present} \end{cases}$
Vegetation cover class (Wikum & Shanholtzer, 1978)	$Y = \begin{cases} 1 & <5\% \\ 2 & 5-25\% \\ 3 & 25-50\% \\ 4 & 50-75\% \\ 5 & 75-100\% \end{cases}$
Insect abundance (density) (Gómez et al., 2015)	$Y = \begin{cases} 0 & 0 \text{ individuals per metre} \\ 1 & 1-10 \text{ individuals per metre} \\ 2 & 11-100 \text{ individuals per metre} \\ 3 & >100 \text{ individuals per metre} \end{cases}$
Ontogeny of fungal diseases of trees (Garnas, Houston, Ayres, & Evans, 2012)	$Y = \begin{cases} 0 & \text{absent} \\ 1 & \text{sparse; few localized fruiting structures} \\ 2 & \text{light; scattered; moderate fruiting} \\ 3 & \text{moderate; many isolated infections with abundant fruiting bodies} \\ 4 & \text{heavy; large areas covered with fruiting bodies} \end{cases}$
Insect damage to leaves (Pocock & Evans, 2014)	$Y = \begin{cases} 0 & \text{no evidence of damage} \\ 1 & \text{just a couple damaged patches} \\ 2 & \text{more green than damaged} \\ 3 & \text{cannot decide whether green or damaged dominates} \\ 4 & \text{damaged patches definitely cover more than half of the leaf} \end{cases}$
Wildfire severity class (Bigler, Kulakowski, & Veblen, 2005)	$Y = \begin{cases} 1 & \text{unburned} \\ 2 & \text{low; surface fire, overstorey largely not scorched} \\ 3 & \text{moderate; many canopy crowns scorched, some green crowns remain} \\ 4 & \text{high; all canopy trees killed} \end{cases}$
Aerial surveys of beetle-attacked trees (Franklin, Wulder, Skakun, & Carroll, 2003)	$Y = \begin{cases} 1 & <10 \text{ red-attacked trees per } 50\text{-m-diameter plot} \\ 2 & 10-20 \text{ red-attacked trees per } 50\text{-m-diameter plot} \\ 3 & 21-50 \text{ red-attacked trees per } 50\text{-m-diameter plot} \end{cases}$
Seagrass standing biomass (g m^{-2}) (Mumby et al., 1997)	$Y \in \{1, 2, 3, 4, 5, 6\}$, where 1 is the minimum and 6 is the maximum

can affect parameter estimation and demonstrate the types of ecological inference that can be made about biotic interactions by quantifying dependence between species through time and across space.

2 | METHODS

We begin with a multivariate generalized linear model with probit link function (as described by Pollock et al., 2014) as a 'baseline' joint species distribution model. We extend the link function to accommodate ordinal abundance categories and add a spatial random effect to account for residual spatial dependence (Section 2.1), making it a multivariate generalized linear mixed model. Then, we describe three model extensions that can be applied individually or all together to incorporate non-equilibrium conditions (Section 2.2), temporal dynamics (Section 2.3) and replicated observations (Section 2.4). We present the general model for data observable for any number of species S , but emphasize that these methods are intended for testing specific

hypotheses for a small number of strongly interacting species within a community module. Annotated R code for fitting these models and extensions is included as online supporting information.

2.1 | General description of a spatio-temporal joint species distribution model

The multivariate generalized linear model with probit link function for binary data (Pollock et al., 2014) forms the foundation of our model. Let $Y_{it}^{(s)} \in \{0, 1, \dots, L-1\}$ denote the observable ordinal abundance for species s at plot i and time t . Here, L is the number of ordinal categories, and $L = 2$ for binary occurrence data because binary data are a special case of ordinal data with only two categories (presence and absence). Although responses other than ordinal abundance also apply here (e.g., severity of disturbance, disease or damage; see Table 1), we refer to the response variable as 'abundance' for the purposes of illustrating this approach.

To extend the link function to accommodate ordinal abundance categories, we assume that each ordinal response variable is the result of thresholding an underlying latent, or true, abundance. We apply the latent variable approach with probit link function, where the ordinal response $Y_{i,t}^{(s)}$ is the result of a truncation or thresholding process applied to a latent Gaussian variable (Albert & Chib, 1993). Clark et al. (2017) use a similar latent variable approach in their generalized joint attribute model for modelling multiple variable types, including continuous, ordinal, composition, zero-inflated and censored data.

The relationship between $Y_{i,t}^{(s)}$ and the latent continuous response $Z_{i,t}^{(s)}$ is

$$Y_{i,t}^{(s)} = \begin{cases} 0 & Z_{i,t}^{(s)} < \lambda_1^{(s)} \\ 1 & \lambda_1^{(s)} \leq Z_{i,t}^{(s)} < \lambda_2^{(s)} \\ \vdots & \\ L-1 & \lambda_{L-1}^{(s)} \leq Z_{i,t}^{(s)} \end{cases} \quad (1)$$

where $\lambda^{(s)} = (\lambda_1^{(s)}, \dots, \lambda_{L-1}^{(s)})$ is a species-specific vector of threshold parameters bounding each ordinal category such that $\lambda_1^{(s)} = 0$ and $\lambda_l^{(s)} \leq \lambda_{l+1}^{(s)}$ for $l = 1, \dots, L-2$ and all s . Here, we assume that L is the same for all species; however, this can be easily modified to allow for species-specific ordinal categories. Combining response variables from different probability distributions is also possible in this latent multivariate framework (Schliep & Hoeting, 2013).

In the multivariate setting, latent abundance is modelled using a multivariate normal distribution, with one dimension for each species $s = 1, \dots, S$ in the study. The latent continuous responses for location (i.e., plot) i and time t , denoted $(Z_{i,t}^{(1)}, \dots, Z_{i,t}^{(S)})'$, are conditionally independent given $\mathbf{K}_{i,t}$. That is, letting $\mathbf{Z}_{i,t} = (Z_{i,t}^{(1)}, \dots, Z_{i,t}^{(S)})'$,

$$\mathbf{Z}_{i,t} \sim \text{multivariate normal}(\mathbf{K}_{i,t}, \mathbf{I}_S) \quad (2)$$

where $\mathbf{K}_{i,t} = (K_{i,t}^{(1)}, \dots, K_{i,t}^{(S)})'$ is a multivariate latent process of abundance for plot i and time t , and \mathbf{I}_S is an $S \times S$ identity matrix for parameter identifiability under the probit model.

The multivariate latent abundance is defined using fixed effect covariates specific to each plot-year and spatially correlated errors. Fixed covariates represent climate variables capturing the relationship between abiotic conditions and abundance unique to each species (i.e., the abiotic niche, or climate envelope). We specify latent abundance for $t = 1, \dots, T$ as

$$\mathbf{K}_{i,t} = \beta \mathbf{X}_{i,t} + \eta_{i,t} \quad (3)$$

where β is an $S \times P$ matrix of species-specific coefficients describing the species-environment relationship, and $\mathbf{X}_{i,t}$ is a vector containing an intercept and $P - 1$ time- and plot-specific covariates where

$$\beta = \begin{bmatrix} \beta_0^{(1)} & \beta_1^{(1)} & \dots & \beta_{P-1}^{(1)} \\ \vdots & & & \\ \beta_0^{(S)} & \beta_1^{(S)} & \dots & \beta_{P-1}^{(S)} \end{bmatrix} \quad \text{and} \quad \mathbf{X}_{i,t} = (1, X_{i,t,1}, \dots, X_{i,t,P-1})'. \quad (4)$$

The spatial random effect, $\eta_{i,t}$, is modelled using a linear model of coregionalization (Gelfand, Schmidt, Banerjee, & Sirmans, 2004) to capture the dependence between species and across space not accounted for by the covariates. Letting \mathbf{A} be an $S \times S$ lower triangular matrix, the linear model of coregionalization is defined as

$$\eta_{i,t} = \mathbf{A} \begin{pmatrix} W_{i,t}^{(1)} \\ \vdots \\ W_{i,t}^{(S)} \end{pmatrix}$$

where each $W_{i,t}^{(s)}$ is an independent Gaussian process with spatial covariance matrix $\Sigma^{(s)}$. This model specification assumes that the random effects, $\eta_{i,t}$ are independent in time. We discuss extension of the model to include temporal random effects and temporal dynamics in Sections 2.2 and 2.3, respectively. Here, $\mathbf{A}\mathbf{A}'$ can be interpreted as the covariance of $\mathbf{K}_{i,t}$. This implies that

$$\text{Cov}(\mathbf{K}_{i,t}, \mathbf{K}_{i',t}) = \mathbf{A} \begin{bmatrix} \Sigma_{i,i'}^{(1)} & & 0 \\ & \ddots & \\ 0 & & \Sigma_{i,i'}^{(S)} \end{bmatrix} \mathbf{A}'.$$

We use an exponential covariance function where the correlation between $W_{i,t}^{(s)}$ and $W_{i',t}^{(s)}$ is

$$\text{Corr}(W_{i,t}^{(s)}, W_{i',t}^{(s)}) = \Sigma_{i,i'}^{(s)} = \exp(-d_{i,i'} \phi^{(s)}), \quad (5)$$

where $d_{i,i'}$ is the distance between plots i and i' , and $\phi^{(s)}$ is the spatial decay parameter for process s . Note that when the number of columns of \mathbf{A} is less than the number of rows, this model is similar to a factor model (Thorson et al., 2016; Warton et al., 2015). This reduced rank specification of the multispecies model is beneficial as the number of species being considered increases (Taylor-Rodríguez, Kaufeld, Schliep, Clark, & Gelfand, 2016).

The spatial decay parameters $\phi^{(s)}$ permit estimation of the effective range, the distance at which the residual spatial correlation drops below 0.05, for each species. Note that $\phi^{(s)}$ is not the spatial decay parameter for species s unless all off-diagonal elements $A_{1,1}, \dots, A_{1,s-1}$ are zero. Therefore, the effective range for each species will be a function of \mathbf{A} and the spatial decay parameters $\phi = \{\phi^{(1)}, \dots, \phi^{(S)}\}$ (Banerjee et al., 2014, Chapter 9). For a two-species model, the effective range for the average latent abundance process of species 1 solves $\exp(-d\phi^{(1)}) = 0.05$, whereas the range of the average latent abundance process of species 2 solves

$$\frac{A_{2,1}^2 \exp(-d\phi^{(1)}) + A_{2,2}^2 \exp(-d\phi^{(2)})}{A_{2,1}^2 + A_{2,2}^2} = 0.05. \quad (6)$$

Each posterior sample of \mathbf{A} , $\phi^{(1)}$ and $\phi^{(2)}$ enables a solution for a corresponding d resulting in posterior samples of the range for both species (Banerjee et al., 2014, Chapter 9). The magnitude of the effective range indicates whether ecological process such as dispersal, response to unmeasured environmental covariates or interactions with other species strongly affect the distribution and abundance of the study species (Dormann et al., 2007; Keitt et al., 2002). A short effective range indicates that the model and chosen covariates are capturing the variability in the process.

2.2 | Adding temporal random effects to account for non-equilibrium

Species-specific temporal random effects can be added to this baseline model to address the ways in which spreading invasive species or species that are range shifting because of climate change can violate the

assumption of equilibrium with the environment. The temporal random effects also capture the inter-annual variability in overall abundance across survey years for each species. To accomplish this, the model for latent abundance in equation (3) becomes

$$\mathbf{K}_{i,t} = \alpha_t + \beta \mathbf{X}_{i,t} + \eta_{i,t} \quad (7)$$

where $\alpha_t = (\alpha_t^{(1)}, \dots, \alpha_t^{(S)})'$ denotes the species-specific random effects for time t . For identifiability of the species-specific intercepts $\beta_0^{(s)}$, the last year of the species-specific temporal random effects $\alpha_T^{(s)}$ are set to zero for all s . We assume that $\alpha_t^{(s)}$ for all t and s are independent a priori as they are capturing the temporal variation in each species. Although there are many possible model specifications for having species-specific temporal random effects, this specification was chosen for ease in comparing inference for the model versus the submodel that does not contain temporal random effects.

The R function `Multivariate.Ordinal.Spatial.Model` within the included annotated R code allows a user to fit this model for S species with any number of locations, time points, covariates and ordered abundance categories (along with the worked example described in Section 2.5). The function allows the user to specify a model with or without temporal random effects for comparison. Inference on the model parameters and latent variables is obtained within the Bayesian framework. Details regarding the prior distributions and Metropolis-within-Gibbs Markov chain Monte Carlo (MCMC) sampling algorithm are given in Supporting Information Appendices S1 and S2.

2.3 | Adding temporal dynamics

Next, we extend the spatial model with temporal random effects outlined in Section 2.2 to incorporate additionally the temporal dependence within and among species via a vector autoregressive term that allows inter- and intraspecific processes to affect abundance. This term captures how the abundance of each species can enhance or inhibit its own abundance or the abundance of other species in the community in the next time step. This autoregressive approach is similar to dynamic range models for univariate (Pagel & Schurr, 2012; Schurr et al., 2012) or multivariate (Mutshinda et al., 2009, 2011; Thorson et al., 2016; Thorson, Munch, & Swain, 2017) count data that explicitly include per-capita population demographic models. Additionally, one could incorporate temporal dependence by specifying a spatio-temporal covariance function for the random effects, η_{it} (for examples of such functions, see Cressie & Wikle, 2011).

Latent abundance is modelled using a lag 1 vector autoregressive model. For $t = 1$, we specify latent abundance $\mathbf{K}_{i,t}$ analogous to Equation 7 for all i . Then, for $t = 2, \dots, T$, we specify

$$\mathbf{K}_{i,t} = \alpha_t + \beta \mathbf{X}_{i,t} + \rho \mathbf{K}_{i,t-1} + \eta_{i,t}. \quad (8)$$

The vector autoregressive parameter ρ is an $S \times S$ matrix where

$$\rho = \begin{bmatrix} \rho_{1,1} & \rho_{1,2} & \dots & \rho_{1,S} \\ \rho_{2,1} & \rho_{2,2} & \dots & \rho_{2,S} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{S,1} & \rho_{S,2} & \dots & \rho_{S,S} \end{bmatrix}. \quad (9)$$

Note that this matrix is not necessarily symmetrical; the effect of species 1 on the abundance of species 2 in the subsequent time step can be distinguished from the effect of species 2 on the abundance of species 1 in the subsequent time step. As a vector autoregressive model simplification, by setting all off diagonal elements of ρ to zero, the temporal dynamics in the latent abundance process would be only within species, not between species. The terms α_t , $\beta \mathbf{X}_{i,t}$ and $\eta_{i,t}$ are as defined in Sections 2.1 and 2.2.

2.4 | Adding replicated observations within a location

The final extension to the model allows for multiple observations, or replicates, of a species' abundance for a given location and time. For example, cover class may be observed at several quadrats within a site, occurrence may be recorded at several points along a transect, or a species' abundance may be observed on multiple trees within a forest stand. In these examples, the quadrat, point and tree are the units of observation, whereas the site, transect or forest stand are the units of inference. Such 'pseudoreplication' is common in ecological data (Hurlbert, 1984; Steel, Kennedy, Cunningham, & Stanovick, 2013) and may be especially prevalent in observational data. Additionally, observational units may not be uniquely labelled (e.g., if quadrats are not marked with a permanent pin, volunteers do not stop at the same exact locations each time they walk a transect, or individual trees are not given a permanent identification tag). In these cases, the same observational units are not resampled in each year of study. For data with replicated observations, making inference on the variance within species and covariance between species at the level of the observational unit can provide insight into intra- and interspecific interactions at the local scale.

To accomplish this, we relax the assumption of conditional independence between $Z_{i,t}^{(s)}$ and $Z_{i,t}^{(s')}$ given average latent abundance and allow for dependence between species at the level of the observational unit. Let $Y_{i,t,j}^{(s)} \in \{0, 1, \dots, L-1\}$ denote the ordinal abundance response for the j th replicate observation of species s at plot i and time t . Let $\mathbf{Z}_{i,t,j} = (Z_{i,t,j}^{(1)}, \dots, Z_{i,t,j}^{(S)})'$ denote the vector of latent continuous responses for observation j at plot i and time t . Now, for $j = 1, \dots, J_{i,t}$, where $J_{i,t}$ denotes the number of observations for plot i and time t ,

$$\mathbf{Z}_{i,t,j} \stackrel{i.i.d.}{\sim} \text{multivariate normal}(\mathbf{K}_{i,t}, \Omega_i). \quad (10)$$

Here, $\mathbf{K}_{i,t} = (K_{i,t}^{(1)}, \dots, K_{i,t}^{(S)})'$ is the multivariate latent process of abundance for plot i and time t as before, but now the identity matrix in Equation 2 is replaced by Ω_i , an $S \times S$ covariance matrix capturing the dependence between species for plot i at the level of the observational unit. Modelling $\mathbf{Z}_{i,t,j}$ given $\mathbf{K}_{i,t}$ and Ω_i as *i.i.d.* assumes that for plot i and time t , each $\mathbf{Z}_{i,t,j}$ for $j = 1, \dots, J_{i,t}$ is replicate measure of some multivariate latent process of abundance.

The plot-specific covariances, Ω_i , for the multivariate continuous latent measure of abundance cannot be time varying for identifiability of $\lambda_t^{(s)}$ and therefore represent the overall dependence among species across all study years. For $i = 1, \dots, n$, $\Omega_i \sim i.i.d.$ inverse-Wishart $(v+1, 2v \text{diag}(\frac{1}{\delta^{(1)}}, \dots, \frac{1}{\delta^{(S)}}))$, and therefore $\frac{1}{\delta^{(s)}}$ represents the variability of ordinal abundance categories on samples within a quadrat for species s .

The function `Multivariate.Ordinal.Spatial.ModelX` included in the annotated R code allows a user to fit this full model (with temporal random effects, temporal dynamics and replicate observations) for S species with any number of locations, time points and covariates and up to four ordered abundance categories. The code permits the user to specify whether temporal random effects, temporal dynamics and/or replicated observations should be included in the model. The user can further specify whether to include both within-species and interspecific temporal dynamics, or to set the interspecific 'cross-dynamics' to zero and permit only within-species temporal dependence. Likewise, the user can force the spatial processes to be independent for each species by setting the off-diagonal elements of the \mathbf{A} matrix to zero. The flexibility to turn model components 'on and off' allows a researcher to build a model specific to the study system and facilitates comparison of competing submodels. This function uses C++ code in obtaining samples from the joint posterior distribution, increasing the computational efficiency of the MCMC algorithm. We also include a worked example of this full model using the data described in Section 2.5. We conducted a simulation study for the full model by simulating three datasets using the estimates of the parameters obtained using the methods described in Section 2.5, each with a different specification of the autocorrelation parameter matrix, ρ . Details for this simulation are given in Supporting Information Appendix S2.

2.5 | Example system: hemlock woolly adelgid and elongate hemlock scale in New England, U.S.A.

We applied the above models to a spatially explicit time series of multivariate ordinal data on the abundance of two invasive insects, hemlock woolly adelgid (*Adelges tsugae*) and elongate hemlock scale (*Fiorinia externa*), that share the host tree eastern hemlock (*Tsuga canadensis*) in the eastern U.S.A. Adelgid has caused widespread decline and mortality of eastern hemlock across the eastern U.S.A. since its introduction from Asia in the early 1950s and threatens to eliminate eastern hemlock across its range. Scale was introduced from Asia in 1908 and is less harmful to hemlocks than adelgid. Both invasive insects were undergoing northwards range expansion in the eastern U.S.A. during the study period, 1997–2011 (Gómez et al., 2015).

Adelgid and scale abundances were visually assessed in 142 eastern hemlock stands across Massachusetts and Connecticut. Initial assessment of stands occurred in 1997–1998 (Connecticut only; Orwig, Foster, & Mausel, 2002) and 2002–2004 (Massachusetts only; Orwig et al., 2012). All stands were subsequently assessed in 2005, 2007, 2009 and 2011 (Gómez et al., 2015, and references within). Ordinal abundance was assigned to one of four ordered categories as described by Gómez et al. (2015) and Table 1. In this example, $Y_{i,t,j}^{(s)}$ denoted the ordinal measure of abundance for stand i , tree j , in year t of species s , where $s = 1$ for adelgid and $s = 2$ for scale. At the initial sampling of stands, one abundance category representing the average insect abundance category across all eastern hemlock trees in the stand was recorded such that $J = 1$. For the remaining four years, $J = 50$ trees unless 50 eastern hemlock trees could not be located in the stand. Individual trees were not marked. Figure 1 shows the location of study

sites and the proportion of trees in each stand that were infested with adelgid and scale in 2005–2011; the distance between stands ranged from 0.5 to 165 km, with an interquartile range of 30–93 km.

We included mean winter temperature before the growing season (from 1 December to 31 March) for each plot in each year (1997, 2003, 2005, 2007, 2009 and 2011) as the fixed covariate in the model because it has been shown to affect the abundance of both species strongly (McClure, 1989; Paradis, Elkinton, Hayhoe, & Buonaccorsi, 2008; Preisser, Elkinton, & Abell, 2008). Mean winter temperatures were interpolated to the centroid of each eastern hemlock stand using PRISM data at 4 km resolution (PRISM Climate Group & Oregon State University, 2015). The covariate was centred within each year to enable comparison of the temporal random effects, α_t .

To illustrate the spatial model with temporal random effects (described in Sections 2.1 and 2.2), we simplified the full dataset. First, the repeated observations of ordinal abundance on individual hemlock trees within stands were collapsed into a single value for each stand in each year. We used the mode of observations within a stand. To explore the value of using data on ordinal abundance ($L = 4$ categories) versus binary occurrence ($L = 2$ categories), as well as to demonstrate how these methods can be used for both occurrence and ordinal abundance data, we assigned a single binary occurrence category (mode = 0 versus mode > 0) for each species. The included R function `Bivariate.Ordinal.Spatial.Model` fits models for these worked binary and ordinal examples as well as for user-specified data.

We fitted the more complex model described in Sections 2.3 and 2.4 that contains dynamic temporal processes and replicated measurements to the full dataset. The included R function `Bivariate.Ordinal.Spatial.ModelX` fits the full extension model with temporal random effect, temporal dynamics and replicated measurements for this worked ordinal example as well as for user-specified data. We use this full model to evaluate model fit and prediction, and for ecological inference. The model was run for 50,000 MCMC iterations. The first 10,000 samples were discarded as burn in, and Monte Carlo standard errors for each parameter were computed. To evaluate the effect of violating the equilibrium assumption, we compared these results with those from a model that did not include temporal random effects (α_t) describing average abundance in the study region for each species.

2.6 | Model fit and prediction

To assess model fit, we computed marginal rank probability scores (RPS). Marginal RPS is a probabilistic method for assessing prediction accuracy that describes the equality of predicted and actual data (Gneiting, Balabdaoui, & Raftery, 2007). Using the posterior estimates, we generated predictions of the ordinal response for each location, time and species. Then, marginally for each species, RPS was computed as

$$\frac{1}{3} \sum_{k=0}^3 (F_{i,t}^{(s)}(k) - \hat{F}_{i,t}^{(s)}(k))^2,$$

where $F_{i,t}^{(s)}(k)$ and $\hat{F}_{i,t}^{(s)}(k)$ are the empirical CDFs of the observed and generated ordinal response data, respectively. For example, the empirical CDF for plot i , time t and species s is computed as

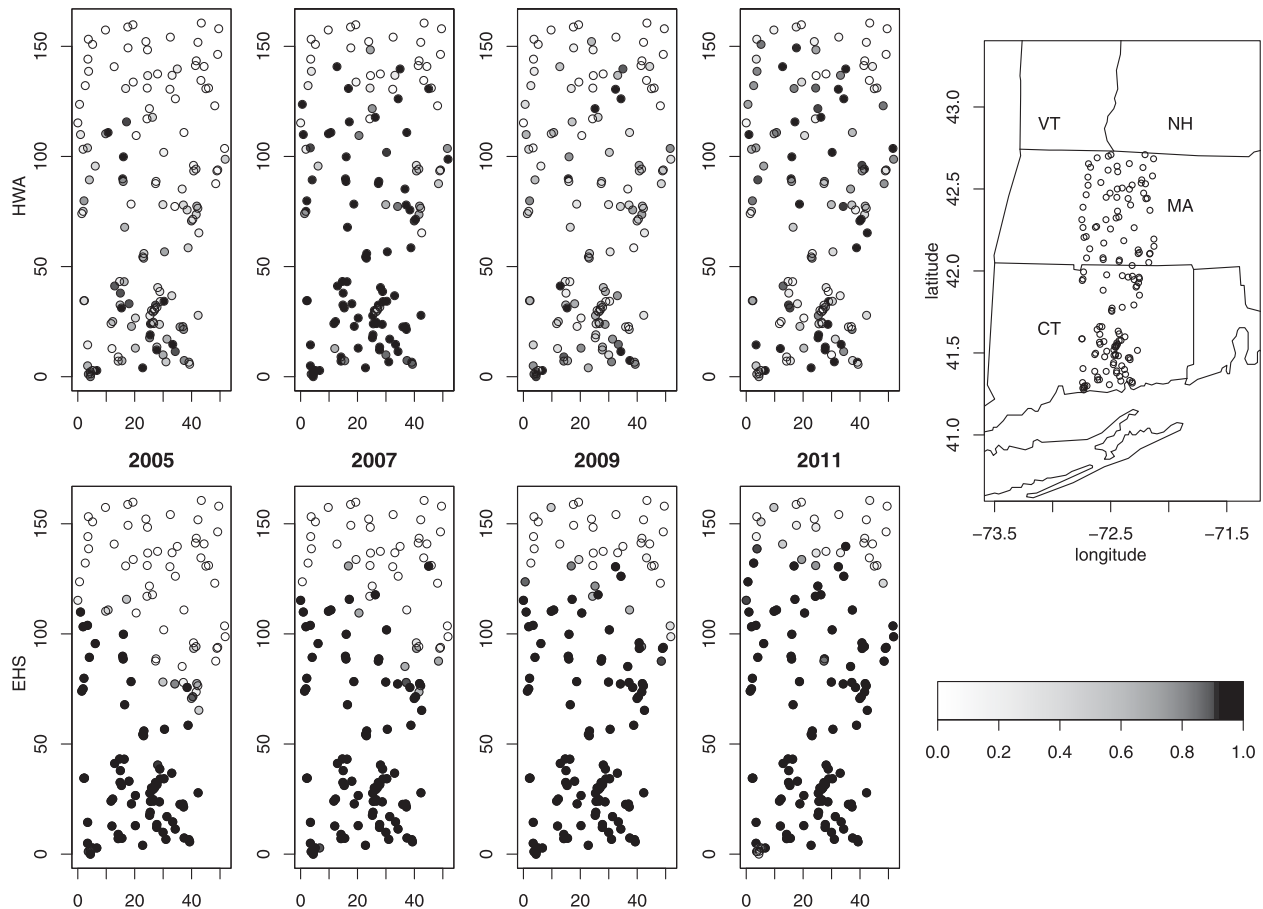


FIGURE 1 The proportion of observed trees on each stand that had one or more hemlock woolly adelgid (HWA, top) and elongate hemlock scale (EHS, bottom) across each year. Eastings and northings on each axis are given in kilometres. Both species showed northward range expansion during this time period. The right panel shows where the 142 surveyed hemlock stands are located across Connecticut and Massachusetts, U.S.A

$$F_{i,t}^{(s)}(k) = \frac{1}{J_{i,t}} \sum_{j=1}^{J_{i,t}} I_{|Z_{i,t,j}^{(s)}| \leq k}$$

Rank probability score is a particularly attractive method for assessing model fit for ordinal response data with replicated observations because it enables comparison of the distributions as opposed to individual observed and predicted ordinal values. Small values of marginal RPS indicate that the distribution of the data generated from the fitted model closely resembles the distribution of the observed data for that location, time and species. Perfect matching between predicted and actual data would yield an RPS score of zero. Rank probability score provides an alternative to information theoretical approaches to model selection that focuses on predictive ability, and can be combined with k-fold cross-validation to perform out-of-sample prediction (Gneiting et al., 2007). The R function RPS included in the annotated R code calculates RPS using samples from the posterior distribution.

A common goal of species distribution models is to predict the occurrence or abundance of species at unobserved locations. To assess prediction accuracy under the model, we conducted 10-fold cross-validation where we partitioned the locations into 10 disjoint sets. The model was then fitted 10 times, each time using a different dataset as

the testing data and the remaining nine sets as the training data. We computed the marginal RPS for all out-of-sample prediction locations using the posterior predictive distributions of the 10-fold cross-validation runs.

3 | RESULTS

3.1 | Simulations

In both the ordinal ($L = 4$ categories) and binary ($L = 2$ categories) simulations of the spatial joint species distribution model with temporal random effects described in Sections 2.1 and 2.2, the model recovered the true parameter values well, and no issues of convergence were detected (Supporting Information Tables S1 and S2). A significant positive relationship between the latent abundance and the covariate (mean winter temperature) resulted from both models. The credible intervals of $\beta_0^{(1)}$ and $\beta_1^{(1)}$ for the model fitted to the binary data were slightly above the true value (Supporting Information Table S2), which is attributable to the lack of information in the binary data. This simulation study adds to the growing body of evidence that using abundance data in species distribution models is better than using data on

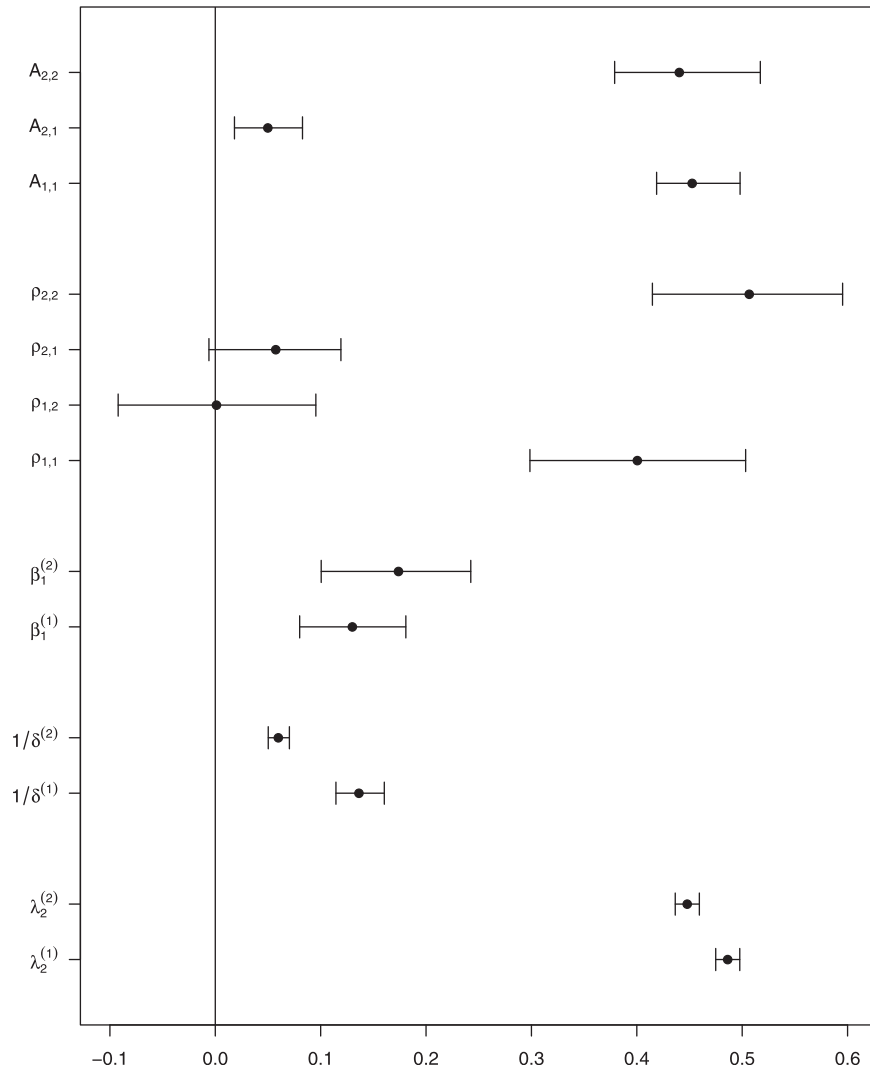


FIGURE 2 Mean and 95% credible intervals of the posterior distributions of the parameters from the spatio-temporal joint species distribution model for adelgid (species 1) and scale (species 2). Parameters describe the latent thresholding of ordinal abundance categories (λ), variability of abundance on individual trees within a hemlock stand for each species (δ), the effect of the environment (mean winter temperature) on the abundance of each species (β_1), temporal autocorrelation within and between species (ρ), and spatial autocorrelation within and shared between species (A)

occurrence, even if coarsely measured (Howard et al., 2014). Simulations of the full model that additionally include temporal dynamics and replicated observations, described in Sections 2.3 and 2.4, also recovered the parameter values well, with no convergence issues (Supporting Information Table S3). In particular, the parameter values of the autocorrelation matrix were recovered, indicating that the model can distinguish different types of vector autoregressive structures. The computation time for obtaining posterior samples from the spatial model with temporal random effects described in the model in Sections 2.1 and 2.2 was c. 5,000 iterations per hour. The computation time for obtaining posterior samples from the full extensions model that additionally includes temporal dynamics and replicated measures of each observational unit (described in Sections 2.3 and 2.4) was c. 5,000 iterations per day. These computing times were for $n = 142$ spatial locations, $T = 6$ time points, and calibrated on an iMac with 4 GHz Intel Core i7 CPU and 32 GB 1867 MHz DDR3 RAM.

3.2 | Inference for invasive insect example

Most of the posterior distributions of the parameters of the dynamic spatio-temporal joint species distribution model fitted to the full adelgid and scale dataset were significantly different from zero, according to the 95% credible intervals (Figure 2). The β_1 coefficients for both species were significantly positive, indicating that abundance increased with mean winter temperature for both species. The parameter $A_{2,1}$ in the lower triangular matrix of the linear model of coregionalization was positive, indicating that the average latent spatial processes for the two species exhibited some dependence. The posterior mean estimate of the effective range for adelgid was 3.76 km, and the posterior mean estimate for scale was 24.18 km (Figure 3).

The temporal random effects, $\alpha_t^{(s)}$, indicated high inter-annual variability in adelgid abundance and generally increasing scale abundance over time (Figure 4). The model without temporal random effects overestimated the strength of the species–environment relationship between adelgid

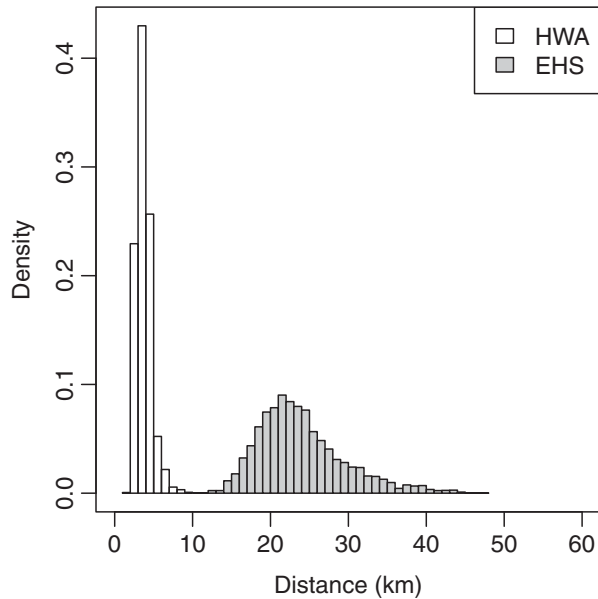


FIGURE 3 Posterior distributions of the effective range (in kilometres) of residual spatial autocorrelation for adelgid (HWA) and scale (EHS). Effective range indicates the distance at which residual spatial autocorrelation drops below 0.05, after accounting for weather-related covariates, temporal dynamics and dependence between species. The greater effective range for scale suggests that an additional, unmeasured factor affects its abundance

abundance and mean winter temperature (β_1 coefficient), compared with the model with the temporal random effect, while underestimating the same relationship for scale abundance (Figure 5). Importantly, the β_1 coefficient for scale was significantly positive in the model that included temporal random effects to account for non-equilibrium, whereas the 95% credible intervals included zero in the model without the temporal random effects. The estimates and credible intervals for $\rho_{1,1}$ and $\rho_{2,2}$ indicated that adelgid and scale have significant, positive within-species temporal autocorrelation. Neither $\rho_{1,2}$ nor $\rho_{2,1}$ was significantly different from zero according to the

credible intervals, but the posterior mean estimate of the cross-species autocorrelation parameter, $\rho_{2,1}$, was 0.06, suggesting that high average abundance of adelgid at time $t - 1$ may have led to an increase in average abundance of scale at time t . Lastly, the estimates of $1/\delta^{(1)}$ and $1/\delta^{(2)}$ indicated that, in general, the distribution of ordinal responses across trees in a stand was more variable for adelgid than for scale.

3.3 | Model validation

Root RPS scores did not indicate lack of model fit (Figure 6, top). For each year and species, the median root RPS across all sites was between 0.02 and 0.04, and was typically lower for scale than for adelgid. An exception occurred in 2009, the year of an unexplained dip in landscape-level abundance of scale (Figure 4). Out-of-sample prediction was also assessed using RPS (Figure 6, bottom). For adelgid, the years with higher abundance on average (2007 and 2011; Figure 4) also had higher predicted root RPS. Predicted root RPS for scale, in contrast, was fairly constant over the 4 years. As a benchmark of comparison for out-of-sample prediction of the model, we also computed RPS using the empirical distribution of the observed data for each species (Figure 6, bottom); that is, for each species, we computed the empirical density of the ordinal response variable across all years and used it as our predictive distribution. The median predicted root RPS from our model is less than that obtained from the empirical distribution for each species and year, indicating that the model explained some of the variability in the response, although the gain is much greater for scale than adelgid.

4 | DISCUSSION

4.1 | Spatial dependence within and between species

We found positive spatial dependence between the latent abundance of the hemlock woolly adelgid and elongate hemlock scale as indicated by the posterior distribution of $A_{2,1}$ (Figure 2). Therefore, even with the

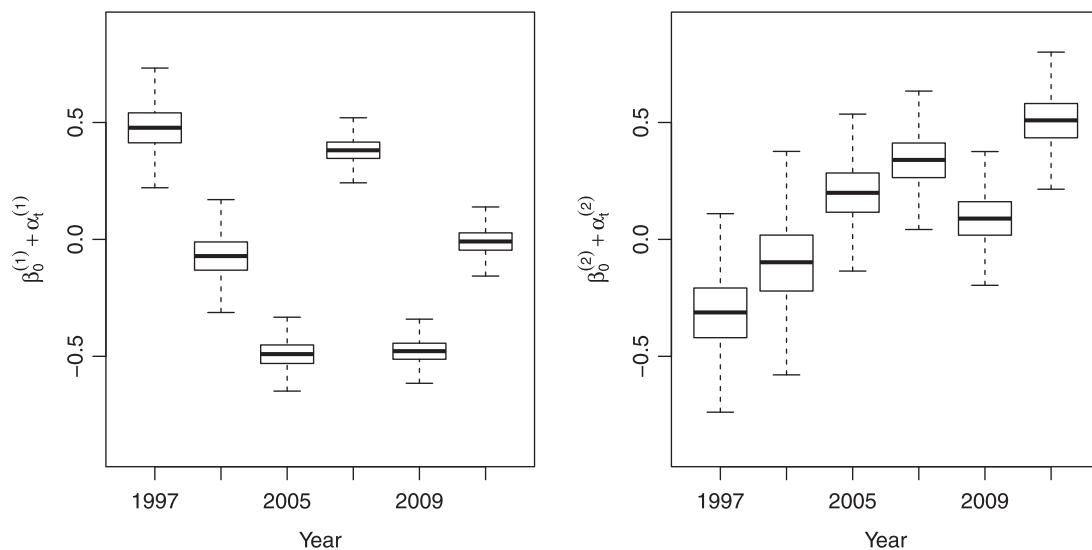


FIGURE 4 Boxplots of the posterior distribution of the time-varying random intercepts, $\beta_0^{(s)} + \alpha_t^{(s)}$, indicated high inter-annual variation in abundance for adelgid (left) and generally increasing abundance of scale during the study (right) at the landscape scale

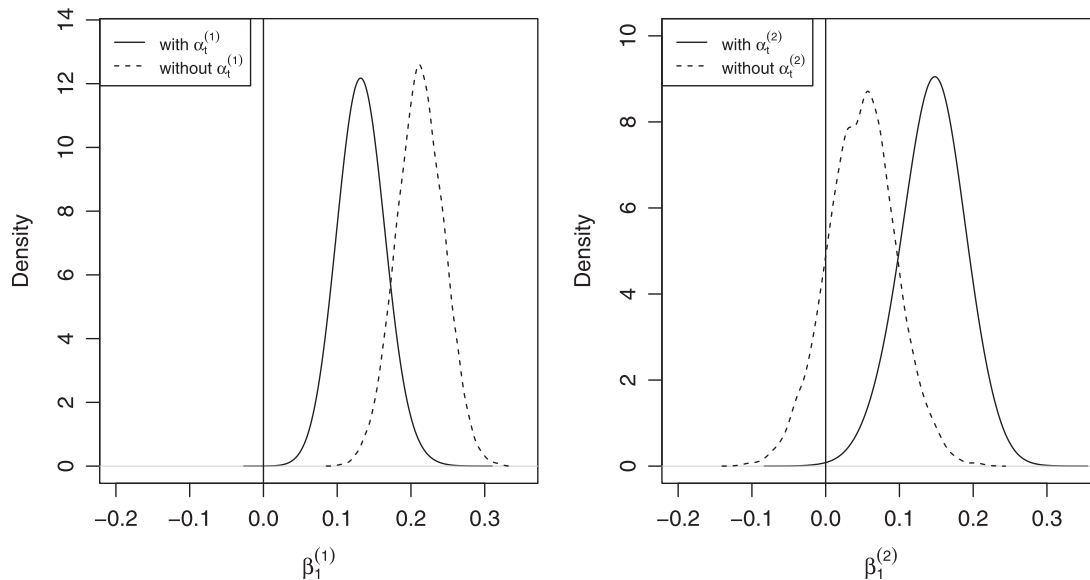


FIGURE 5 Posterior probability of β coefficients representing the species–environment relationship for adelgid (left) and scale (right) in the northeastern U.S.A., with versus without accounting for non-equilibrium of range-shifting species and inter-annual variation in abundance with a species-specific temporal random effect, $\alpha_t^{(s)}$

other components in the model (i.e., temporal random effects, covariates and temporal dynamics) there was remaining spatial structure in average, stand-level latent abundance shared between the two species. The parameter estimates of the spatial random effects also indicated that the effective range varied between the two species. We do not expect that an unmeasured covariate that acts at the regional scale (e.g., a climate-related variable) strongly affected average adelgid abundance because residual average latent abundance was not spatially correlated at distances > 4 km after accounting for mean winter temperature, temporal processes and dependence among species (Figure 3). However, the moderate effective range of spatial correlation for scale (c. 24 km) indicates that additional covariates, possibly weather related, could affect the abundance of this species.

Computation time, although reasonable for this example system with 142 locations, can quickly become prohibitive when a large number of sampling locations are observed. Possible dimension reduction techniques using predictive processes (Banerjee, Gelfand, Finley, & Sang, 2008) or Gaussian Markov random fields to approximate the Gaussian field (Lindgren, Rue, & Lindström, 2011) could be considered.

4.2 | Benefits of temporal random effects to account for non-equilibrium

The species-specific temporal random effects, $\alpha_t^{(s)}$, permitted estimation of the effect of abiotic conditions on the abundance of each insect species while accounting for potential violations of the assumption that species are at equilibrium distribution or abundance with the environment. Average latent abundance varied greatly from year to year for both species (Figure 4), indicating that including temporal random effects was appropriate. The estimates of the time-varying random intercepts for adelgid generally show an alternating pattern of abundance in the study region between adjacent years of observation. The

time-varying random intercepts for scale generally increased over the study period (Figure 4). In 2009, however, average scale abundance decreased in a way that was not accounted for by mean winter temperature or autoregressive processes. This suggests that an unmeasured regional-scale covariate, such as a summer heat wave or drought, might have affected scale abundance in that year.

Failure to account for violations of the equilibrium assumption with the temporal random effects would have led to different conclusions about the effect of abiotic conditions on each species (Figure 5). If we had not accounted for the northward range expansion and general increase in overall scale over the course of the study, we would have erroneously concluded that winter temperatures do not affect scale abundance, despite numerous studies that have clearly documented the negative effects of cold winter temperatures on scale survival and abundance (e.g., McClure, 1989; Preisser et al., 2008).

Non-equilibrium dynamics were modelled using temporal random effects for each species that were assumed independent a priori. Temporal random effects explicitly capture the variability in abundance across years in the study period and improve estimation of the species–environment relationship. Although the temporal random effects do not benefit forecasting abundance at future time periods, the model is valuable for spatial prediction; that is, for predicting species abundance at unobserved spatial locations within observed time periods. Other autoregressive latent abundance random variable specifications could be used to forecast abundance, and uncertainty, properly at both observed and unobserved spatial locations.

4.3 | Intra- and interspecific temporal dynamics

We found weak evidence for interspecific temporal autocorrelation. The posterior mean estimate of $\rho_{2,1}$ was positive (although the 95% credible interval contained zero), indicating that scale may be more abundant

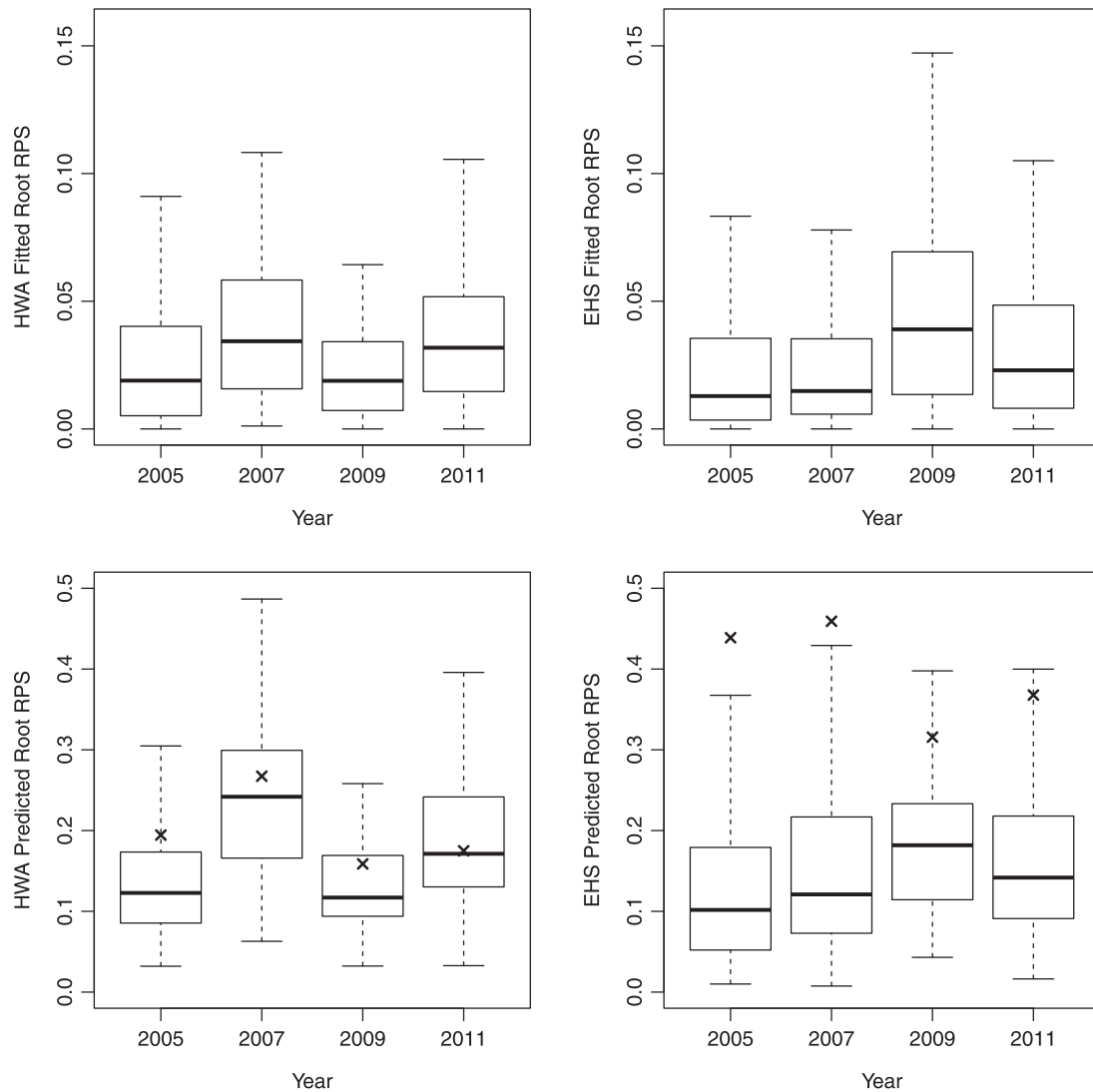


FIGURE 6 Square root of the rank probability score for in-sample plots (top) and out-of-sample plots (bottom) for adelgid (left) and scale (right). The median predicted root rank probability scores (RPS) obtained for each species and year when using the empirical distribution as the predictive distribution are denoted by \times . The distance between the \times and the median predicted root RPS represents the improvement in prediction provided by the model versus the empirical distribution

when adelgid abundance was high in the previous time step, but $\rho_{1,2}$ was clearly indistinguishable from zero (Figure 2). A positive posterior estimate of $\rho_{2,1}$ combined with an estimate of $\rho_{1,2}$ that was indistinguishable from zero would imply a commensalism, in which adelgid predisposed stands to future infestation by scale, but not vice versa. Such a commensalism could occur if adelgid manipulates host plant defensive chemistry (Pezet et al., 2013) or the nitrogen content of foliage (Gómez, Orians, & Preisser, 2012; Soltis, Gómez, Gonda-King, Preisser, & Orians, 2015) in ways that benefit scale. Of all the model terms that describe dependence between species at different temporal or spatial scales, only $\rho_{1,2}$ and $\rho_{2,1}$ can indicate directionality, because the ρ matrix is not necessarily symmetric. Significant positive within-species temporal autocorrelation was captured by the posterior estimates of $\rho_{1,1}$ and $\rho_{2,2}$, indicating that for both adelgid and scale, stands with higher than average abundance tended also to have higher than average abundance of that same species at the subsequent sampling occasion.

4.4 | Replicated observations within a location

The estimates of $1/\delta^{(1)}$ and $1/\delta^{(2)}$ indicated that, in general, the ordinal abundance category on individual trees was more variable within a stand for adelgid than for scale. Although the reason for this difference is not known for certain, adelgid preferentially feeds on new hemlock shoots and can deplete the available feeding sites on a tree, which could contribute to this pattern (McClure, 1979, 1991). Having Ω_i unstructured greatly increases the number of parameters in the model. Two possible simplifications would be either to assume conditional independence of $Z_{i,t,j}^{(1)}$ and $Z_{i,t,j}^{(2)}$ given $\mathbf{K}_{i,t}$ or to assume a global covariance matrix Ω .

4.5 | Generality and connection to other approaches

Temporal dynamics, spatial dependence and the influence of interactions with other species are inherent in the distribution and abundance

of species. We show that failure to account for range shifting or other non-equilibrium of abundance with the environment can affect estimation of the species–environment relationship, and demonstrate how including temporal dependence between species can indicate biotic interactions. Given that the model is defined generally for S species, as the number of species increases, so does the number of parameters that need estimating. In particular, the $S(S-1)/2$ parameters of \mathbf{A} and the S^2 parameters of ρ may be difficult to identify from the data as S increases. Although our current model assumes fully regulated dynamics, Thorson et al. (2017) propose investigating the number of regulatory relationships that are identifiable from the data in order to reduce the dimension of ρ . Dimension reduction techniques, such as clustering across species or ordination, could also be used for large S (Hui, 2016; Ovaskainen et al., 2016; Thorson et al., 2016). These clustering techniques are especially appropriate when the goal is to map species diversity or to improve prediction for rare species by borrowing strength from more common species in the community (Warton et al., 2015). Clustering techniques are not used here because our goal was to test a specific hypothesis about how fine-scale biotic interactions affect distribution and abundance for a smaller subset of strongly interacting species within a community. Recently, Ovaskainen et al. (2017) proposed a general hierarchical model for species communities that directly models dependence between species' environmental niches in addition to species dependence at the level of the response variable. As the realized niche of a species encompasses the range of conditions in which a species can exist in the presence of interactions with other species (Hutchinson, 1957), and the distribution of a species can be interpreted as projecting the realized niche onto geographical space (Wiens, 2011), incorporating such interactions among smaller subsets of strongly interacting species into distribution models is generally applicable to ecological systems (Gilman et al., 2010). The methods we present can be adapted to a wide range of ecological data and sampling schemes, providing a flexible approach for inferring ecological process from pattern and making predictions for a specific conservation or management aim.

DATA ACCESSIBILITY

Data are available via the LTER Network Data Portal at <https://doi.org/10.6073/pasta/55236414e515e94f5866d0b1e91475e0> and are also accessible through the Harvard Forest Data Archive (HF289). R scripts are included as online supporting information.

ACKNOWLEDGMENTS

N.K.L. was supported by the Arnold and Mabel Beckman Foundation and Michigan State University, and P.L.Z. was supported by Michigan State University and by the USDA National Institute of Food and Agriculture, Hatch project 1010055. We would like to thank Alan Gelfand for insightful conversations and James Clark for efficient C++ code. This project was funded by the following grants: NSF DEB-0715504, NSF DEB-1256769, NSF DEB-1256826, NIFA 2011-67013-30142 and is a contribution of the Harvard Forest Long-Term Ecological Research Program (NSF DEB-0620443).

ORCID

Erin M. Schliep  <http://orcid.org/0000-0002-2803-3467>

REFERENCES

- Albert, J., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*, 669–679.
- Araújo, M. B., & Rozenfeld, A. (2014). The geographic scaling of biotic interactions. *Ecography*, *37*, 406–415.
- Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Boca Raton, FL, USA: CRC Press.
- Banerjee, S., Gelfand, A. E., Finley, A. O., & Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*, 825–848.
- Bigler, C., Kulakowski, D., & Veblen, T. T. (2005). Multiple disturbance interactions and drought influence fire severity in Rocky Mountain subalpine forests. *Ecology*, *86*, 3018–3029.
- Bolker, B., & Pacala, S. W. (1997). Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology*, *52*, 179–197.
- Carroll, C., Johnson, D. S., Dunk, J. R., & Zielinski, W. J. (2010). Hierarchical Bayesian spatial models for multispecies conservation planning and monitoring. *Conservation Biology*, *24*, 1538–1548.
- Clark, J. S., Gelfand, A. E., Woodall, C. W., & Zhu, K. (2014). More than the sum of the parts: Forest climate response from joint species distribution models. *Ecological Applications*, *24*, 990–999.
- Clark, J. S., Nemergut, D., Seyednasrollah, B., Turner, P. J., & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecological Monographs*, *87*, 34–56.
- Crase, B., Liedloff, A., Vesik, P. A., Fukuda, Y., & Wintle, B. A. (2014). Incorporating spatial autocorrelation into species distribution models alters forecasts of climate-mediated range shifts. *Global Change Biology*, *20*, 2566–2579.
- Cressie, N., & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Hoboken, NJ, USA: Wiley.
- Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., ... Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, *30*, 609–628.
- Dray, S., Pélissier, R., Couteron, P., Fortin, M. J., Legendre, P., Peres-Neto, P. R., ... Wagner, H. H. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, *82*, 257–275.
- Ehrlén, J., & Morris, W. F. (2015). Predicting changes in the distribution and abundance of species under environmental change. *Ecology Letters*, *18*, 303–314.
- Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, *1*, 330–342.
- Franklin, S. E., Wulder, M. A., Skakun, R. S., & Carroll, A. L. (2003). Mountain pine beetle red-attack forest damage classification using stratified Landsat TM data in British Columbia, Canada. *Photogrammetric Engineering and Remote Sensing*, *69*, 283–288.
- Garnas, J. R., Houston, D. R., Ayres, M. P., & Evans, C. (2012). Disease ontogeny overshadows effects of climate and species interactions on population dynamics in a nonnative forest disease complex. *Ecography*, *35*, 412–421.
- Gelfand, A. E., Schmidt, A. M., Banerjee, S., & Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, *13*, 263–312.

- Gilman, S. E., Urban, M. C., Tewksbury, J., Gilchrist, G. W., & Holt, R. D. (2010). A framework for community interactions under climate change. *Trends in Ecology and Evolution*, 25, 325–331.
- Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 243–268.
- Gómez, S., Gonda-King, L., Orians, C. M., Orwig, D. A., Panko, R., Radville, L., ... Preisser, E. L. (2015). Interactions between invasive herbivores and their long-term impact on New England hemlock forests. *Biological Invasions*, 17, 661–673.
- Gómez, S., Orians, C. M., & Preisser, E. L. (2012). Exotic herbivores on a shared native host: Tissue quality after individual, simultaneous, and sequential attack. *Oecologia*, 169, 1015–1024.
- Gotelli, N. J., Graves, G. R., & Rahbek, C. (2010). Macroecological signals of species interactions in the Danish avifauna. *Proceedings of the National Academy of Sciences USA*, 107, 5030–5035.
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8, 993–1009.
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186.
- Holt, R. (1997). Community modules. In A. Gange & V. Brown (Eds.), *Multitrophic Interactions in Terrestrial Systems* (pp. 333–350). New York, USA: Blackwell Science.
- Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D., & Willis, S. G. (2014). Improving species distribution models: The value of data on abundance. *Methods in Ecology and Evolution*, 5, 506–513.
- Hui, F. K. C. (2016). boral – Bayesian ordination and regression analysis of multivariate abundance data in R. *Methods in Ecology and Evolution*, 7, 744–750.
- Hui, F. K. C., Warton, D. I., Foster, S. D., & Dunstan, P. K. (2013). To mix or not to mix: Comparing the predictive performance of mixture models vs. separate species distribution models. *Ecology*, 94, 1913–1919.
- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54, 187–211.
- Hutchinson, G. (1957). Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415–427.
- Ives, A., Dennis, B., Cottingham, K., & Carpenter, S. (2003). Estimating community stability and ecological interactions from time-series data. *Ecological Monographs*, 73, 301–330.
- Keitt, T. H., Bjørnstad, O. N., Dixon, P. M., & Citron-Pousty, S. (2002). Accounting for spatial pattern when modeling organism-environment interactions. *Ecography*, 25, 616–625.
- Kissling, W. D., Dormann, C. F., Groeneveld, J., Hickler, T., Kühn, I., McNerny, G. J., ... O'Hara, R. B. (2012). Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, 39, 2163–2178.
- Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 423–498.
- McClure, M. S. (1979). Self-regulation in populations of the elongate hemlock scale, *Fiorinia externa* (Homoptera, Diaspididae). *Oecologia*, 39, 25–36.
- McClure, M. S. (1989). Importance of weather to the distribution and abundance of introduced adelgid and scale insects. *Agricultural and Forest Meteorology*, 47, 291–302.
- McClure, M. S. (1991). Density-dependent feedback and population-cycles in *Adelges tsugae* (Homoptera, Adelgidae) on *Tsuga canadensis*. *Environmental Entomology*, 20, 258–264.
- Mumby, P. J., Edwards, A. J., Green, E. P., Anderson, C. W., Ellis, A. C., & Clark, C. D. (1997). A visual assessment technique for estimating seagrass standing crop. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 7, 239–251.
- Mutshinda, C. M., O'Hara, R. B., & Woiwod, I. P. (2009). What drives community dynamics? *Proceedings of the Royal Society B: Biological Sciences*, 276, 2923–2929.
- Mutshinda, C. M., O'Hara, R. B., & Woiwod, I. P. (2011). A multispecies perspective on ecological impacts of climatic forcing. *Journal of Animal Ecology*, 80, 101–107.
- Orwig, D. A., Foster, D. R., & Mausel, D. L. (2002). Landscape patterns of hemlock decline in New England due to the introduced hemlock woolly adelgid. *Journal of Biogeography*, 29, 1475–1487.
- Orwig, D. A., Thompson, J. R., Povak, N. A., Manner, M., Niebyl, D., & Foster, D. R. (2012). A foundation tree at the precipice: *Tsuga canadensis* health after the arrival of *Adelges tsugae* in central New England. *Ecosphere*, 3, 1–16.
- Ovaskainen, O., Hottola, J., & Siitonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, 91, 2514–2521.
- Ovaskainen, O., Roy, D. B., Fox, R., & Anderson, B. J. (2016). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, 7, 428–436.
- Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., ... Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20, 561–576.
- Pagel, J., & Schurr, F. M. (2012). Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, 21, 293–304.
- Paradis, A., Elkinton, J., Hayhoe, K., & Buonaccorsi, J. (2008). Role of winter temperature and climate change on the survival and future range expansion of the hemlock woolly adelgid (*Adelges tsugae*) in eastern North America. *Mitigation and Adaptation Strategies for Global Change*, 13, 541–554.
- Peterson, A. T. (2003). Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology*, 78, 419–433.
- Pezet, J., Elkinton, J., Gomez, S., McKenzie, E. A., Lavine, M., & Preisser, E. (2013). Hemlock woolly adelgid and elongate hemlock scale induce changes in foliar and twig volatiles of eastern hemlock. *Journal of Chemical Ecology*, 39, 1090–1100.
- Pocock, M. J. O., & Evans, D. M. (2014). The success of the horse-chestnut leaf-miner, *Cameraria ohridella*, in the UK revealed with hypothesis-led citizen science. *PLoS One*, 9, e86226.
- Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., ... McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5, 397–406.
- Preisser, E. L., Elkinton, J. S., & Abell, K. (2008). Evolution of increased cold tolerance during range expansion of the elongate hemlock scale *Fiorinia externa* Ferris (Hemiptera: Diaspididae). *Ecological Entomology*, 33, 709–715.
- PRISM Climate Group, & Oregon State University. (2015). Retrieved from <http://www.prism.oregonstate.edu/explorer/>
- Record, S., Fitzpatrick, M. C., Finley, A. O., Veloz, S., & Ellison, A. M. (2013). Should species distribution models account for spatial autocorrelation? A test of model projections across eight millennia of climate change. *Global Ecology and Biogeography*, 22, 760–771.
- Schliep, E. M., & Hoeting, J. A. (2013). Multilevel latent Gaussian process model for mixed discrete and continuous multivariate response data.

- Journal of Agricultural, Biological, and Environmental Statistics*, 18, 492–513.
- Schurr, F. M., Pagel, J., Cabral, J. S., Groeneveld, J., Bykova, O., O'Hara, R. B., ... Zimmermann, N. E. (2012). How to understand species' niches and range dynamics: A demographic research agenda for biogeography. *Journal of Biogeography*, 39, 2146–2162.
- Sebastián-González, E., Sánchez-Zapata, J. A., Botella, F., & Ovaskainen, O. (2010). Testing the heterospecific attraction hypothesis with time-series data on species co-occurrence. *Proceedings of the Royal Society B: Biological Sciences*, 277, 2983–2990.
- Soltis, N. E., Gómez, S., Gonda-King, L., Preisser, E. L., & Orians, C. M. (2015). Contrasting effects of two exotic invasive hemipterans on whole-plant resource allocation in a declining conifer. *Entomologia Experimentalis Et Applicata*, 157, 86–97.
- Steel, E. A., Kennedy, M. C., Cunningham, P. G., & Stanovick, J. S. (2013). Applied statistics in ecology: Common pitfalls and simple solutions. *Ecosphere*, 4, 1–13.
- Taylor-Rodríguez, D., Kaufeld, K., Schliep, E. M., Clark, J. S., & Gelfand, A. E. (2016). Joint species distribution modeling: Dimension reduction using Dirichlet processes. *Bayesian Analysis*, 1–29.
- Thorson, J. T., Ianelli, J. N., Larsen, E. A., Ries, L., Scheuerell, M. D., Szuwalski, C., & Zipkin, E. F. (2016). Joint dynamic species distribution models: A tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography*, 25, 1144–1158.
- Thorson, J. T., Ianelli, J. N., Munch, S. B., Ono, K., & Spencer, P. D. (2015). Spatial delay-difference models for estimating spatiotemporal variation in juvenile production and population abundance. *Canadian Journal of Fisheries and Aquatic Sciences*, 72, 1897–1915.
- Thorson, J. T., Munch, S. B., & Swain, D. P. (2017). Estimating partial regulation in spatio-temporal models of community dynamics. *Ecology*, 98, 1277–1289.
- Thorson, J. T., Scheuerell, M. D., Shelton, A. O., See, K. E., Skaug, H. J., & Kristensen, K. (2015). Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution*, 6, 627–637.
- Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. K. C. (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology and Evolution*, 30, 766–779.
- Wiens, J. J. (2011). The niche, biogeography and species interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366, 2336–2350.
- Wikum, D., & Shanholtzer, G. F. (1978). Application of the Braun-Blanquet cover-abundance scale for vegetation analysis in land development studies. *Environmental Management*, 2, 323–329.

BIOSKETCH

ERIN M. SCHLIEP's research interests include spatio-temporal models, spatial point patterns and autoregressive models for environmental processes within the Bayesian paradigm. Specifically, she focuses on developing new statistical models, computational algorithms for parameter estimation and prediction, as well as theory within the field of spatial and spatio-temporal statistics. She collaborates with researchers in varying disciplines at the interface of statistics and environmental science (<http://www.stat.missouri.edu/schliepe/>).

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Schliep EM, Lany NK, Zarnetske PL, et al. Joint species distribution modelling for spatio-temporal occurrence and ordinal abundance data. *Global Ecol Biogeogr.* 2018;27:142–155. <https://doi.org/10.1111/geb.12666>