



Modeling Community Dynamics Through Environmental Effects, Species Interactions and Movement

Becky TANG[✉], James S. CLARK, Peter P. MARRA, and Alan E. GELFAND

Understanding how communities respond to environmental change is frustrated by the fact that both species interactions and movement affect biodiversity in unseen ways. To evaluate the contributions of species interactions on community growth, dynamic models that can capture nonlinear responses to the environment and the redistribution of species across a spatial range are required. We develop a time-series framework that models the effects of environment–species interactions as well as species–species interactions on population growth within a community. Novel aspects of our model include allowing for species redistribution across a spatial region, and addressing the issue of zero inflation. We adopt a hierarchical Bayesian approach, enabling probabilistic uncertainty quantification in the model parameters. To evaluate the impacts of interactions and movement on population growth, we apply our model using data from eBird, a global citizen science database. To do so, we also present a novel method of aggregating the spatially biased eBird data collected at point-level. Using an illustrative region in North Carolina, we model communities of six bird species. The results provide evidence of nonlinear responses to interactions with the environment and other species and demonstrate a pattern of strong intraspecific competition coupled with many weak interspecific species interactions.

Supplementary materials accompanying this paper appear online.

Key Words: Competition; Dispersal; eBird; Interaction strength; Markov chain Monte Carlo.

1. INTRODUCTION

Understanding the role of species interactions is challenged by the fact that they are constantly shifting—models assume fixed effects of one species on another, while movement

B. Tang (✉), Department of Mathematics, Middlebury College, Middlebury, VT 05753, USA
(E-mail: btang@middlebury.edu).

B. Tang · A. E. Gelfand, Department of Statistical Science, Duke University, Durham, NC 27708, USA.

J. S. Clark, Nicholas School of the Environment, Duke University, Durham, NC 27708, USA.

P. P. Marra, Department of Biology and McCourt School of Public Policy, Georgetown University, Washington, DC, USA.

© 2022 International Biometric Society

Journal of Agricultural, Biological, and Environmental Statistics, Volume 28, Number 1, Pages 178–195

<https://doi.org/10.1007/s13253-022-00520-3>

assures that influences are dynamic (Paine et al. 1998). Quantifying species interactions is further complicated by environment–species interactions, the fact that species responding to the environment are, at the same time, responding to other species that are also responding to the environment (Clark et al. 2021). A framework that enables learning about environment–species interactions has to allow that they occur when species encounter one another, which varies in space and time. In this paper, we develop a framework that leverages dynamic data to determine the potential to learn about environment–species interactions with movement. Simulation studies confirm that both interactions and movement can be estimated from spatio-temporal data. Application to eBird data (eBird, 2017), a citizen-science compilation of bird observations, provides evidence that the relatively few strong competitive interactions are predominantly intraspecific.

We use the term environment–species interaction (ESI) to combine the biotic and abiotic effects of environment that both depend on population sizes of interacting species (Clark et al. 2020). Competing species can be positively correlated in observations, because similar environmental responses bring them into direct conflict, or negatively correlated, because they tend to displace one another (Clark 2010). To clarify and quantify the effects of ESI, it requires dynamic data. Further, dynamic abundance data enable learning about the effect of species abundances on the growth rates of others in the context of a variable environment. Dynamic models incorporating species–species interactions provide important insights (e.g., Ives et al. 2003; Schliep et al. 2018), including ESI (Clark et al. 2021), but they have not been combined with movement.

To evaluate environment–species interactions, we allow for redistribution across spatial regions over time. The absence from an area occurs when the habitat is unsuitable or when excluded by competitors or natural enemies that share habitat requirements. Populations respond to these ESI not just through births and deaths, but, more commonly through redistribution. Models for movement include diffusion processes (Wikle 2003), cellular automata (Engler and Guisan 2009), and integrating particle systems (Smolik et al. 2010). We implement a combination of dispersal scales within a framework that admits flexible assumptions about movement.

We apply our model to eBird data, a global citizen science database dedicated to avifauna observation. We evaluate the model in a 450 km² region around the Research Triangle in North Carolina. Our application addresses two data challenges. Like many ecological data sets, eBird data are dominated by zeros. However, rather than a typical zero-inflated model, we are interested in learning about interactions in addition to species abundances. Therefore, observed zeros are accommodated in order to better isolate the relationships between species and their environment. Second, eBird data pose the added challenge of spatial bias; certain areas are sampled heavily, while other areas may not be visited at all (Tang et al. 2021). Our approach allows for aggregating point-level eBird data within spatial or areal units in order to obtain counts per area (CPA) abundances. Results from the application show that there is strong ESI with forested land cover, while species interactions reveal few strong intraspecific ones and mostly weak interspecific ones. This pattern aligns with studies that explore how species interactions impact community stability.

2. STUDY AREA, SPECIES, AND DATA

2.1. EBIRD

Citizen scientists using eBird report the type and number of bird species detected as a checklist. We use only “complete” checklists that report all individuals identified. Each checklist contains starting location, date, unique observer ID, and unique checklist ID. Optional information includes the duration and start time of the observation. Data were further filtered for quality control: checklists missing duration and/or time were removed. Additionally, checklists that recorded a duration length longer than 8 h were removed. eBird data were cleaned using the *auk* package in R (Strimas-Mackey et al. 2018), and all analyses were conducted using R (Version 3.6.1) (R Core Team 2013). While eBird data arise at point level, our proposed model operates on abundances observed within areal units, e.g., counts per area. As we discuss later in Sect. 6.1, we discretize the selected study region of the Research Triangle in North Carolina into areal units based on sampling effort and land cover. In this region, fractions of land cover types of agriculture, developed, forest, and water are (0.011, 0.454, 0.483, 0.052).

Our analysis incorporates both seasonal migrants as well as non-migrant species, so we focus on data from breeding-season observations. Migrant species that winter in Central and South America are absent from our region only in winter months (November to March). Checklists recorded during May and June can be used to model change in species that have migrated or remained as residents. We focus on data from 2011 to 2019 in the Research Triangle, retaining 5571 complete checklists from the Research Triangle. Sampling effort increases over time and is spatially aggregated near preferred observation areas (Fig. 2).

To compare effects across potential competitors of both migrant and non-migrant species, we fitted our model to three resident species [tufted titmouse (*Baeolophus bicolor*), carolina chickadee (*Poecile carolinensis*), carolina wren (*Thryothorus ludovicianus*)] and three migrant species [gray catbird (*Dumetella carolinensis*), chipping sparrow (*Spizella passerina*), and eastern towhee (*Pipilo erythrophthalmus*)]. These species are abundant and have the potential to interact through shared resources, as they are predominantly insectivores and omnivores at this time of year.

2.2. SPATIAL COVARIATES

Associated with each areal unit are the categorical variable of (dominant) land cover (National Land Cover Database 2016; see Table S6 for land cover aggregations) (Dewitz 2019) and the continuous variable Enhanced Vegetation Index (EVI), a measure of vegetation greenness (Vermote et al. 2016). Our region is dominated by two land cover types, forest and developed, that are constant across the years. EVI is taken as the average for the areal unit and year.

3. MODEL DESCRIPTION

Consider a community of bird species that co-occur in the same area and may interact through trophic and spatial relationships. We assume births occur during the breeding season,

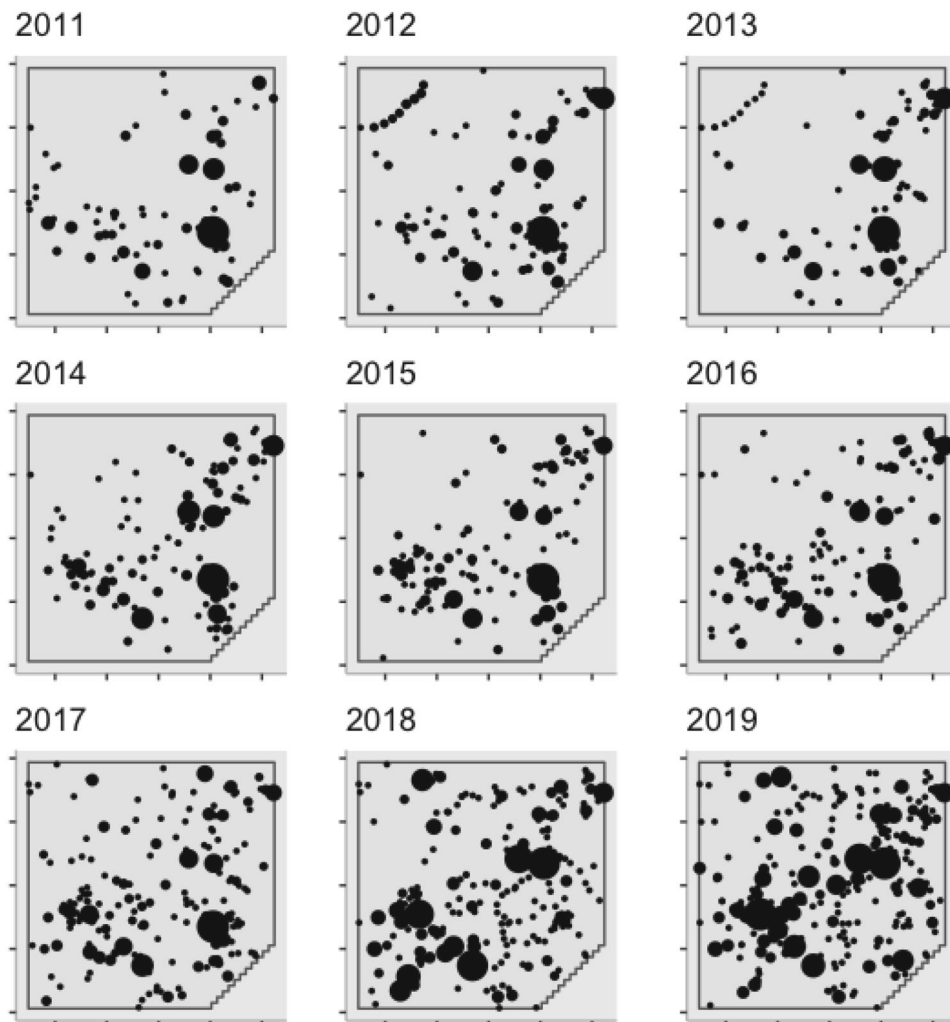


Figure 1. Unique visited locations in the Research Triangle by year. Symbol size corresponds to total number of visits during May and June.

individuals interact through competition for resources and/or nesting sites, and there is overwintering prior to the next breeding season. Species differ in migration status (migrants vs. year-round residents), the extent to which they move between breeding seasons (even residents move), their habitat affinities (simplified here to landcover type and EVI), and the extent to which they interact with others of the same and other species (intra- and inter-specific interaction, respectively). There are environment–species interactions (ESI), in the sense that the effects of species on each other are context-dependent (Clark et al. 2020).

Dynamics are described by an expanded multi-species Lotka–Volterra (LV) model (Takeuchi 1996) for the effects of density-independent growth potential, environment, and other species. To allow model fitting, the LV system of differential equations is discretized (annual time step) and stochasticized (process, observation, and parameter error), adapted

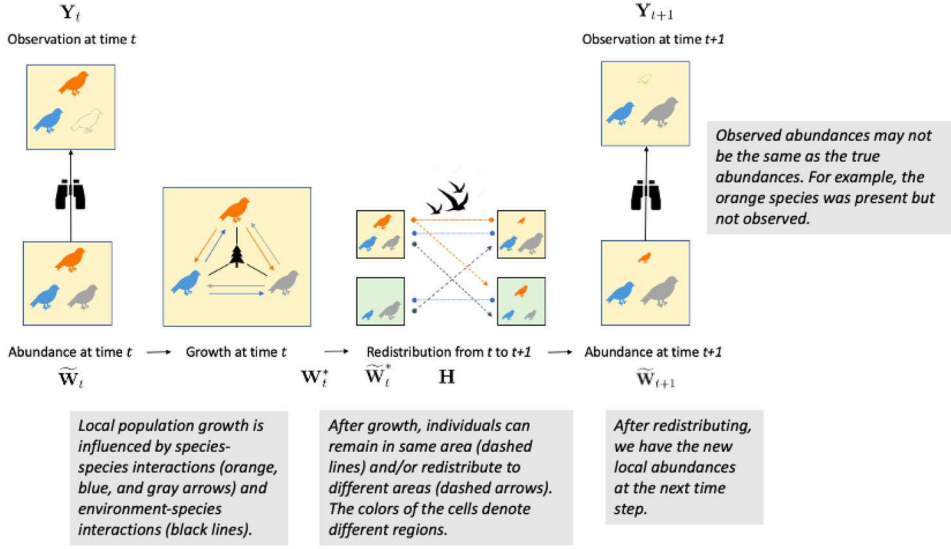


Figure 2. Visual representation of our modeling framework, from time t to time $t+1$.

from Clark et al. (2020). The principal innovation here implements temporal dynamics across a landscape of varying cover types, where movement can be local, simple diffusion, or long-distance, depending on species. The following summarizes the spatial domain in terms of subregions and the notion of local absence. We then discuss model structure, the redistribution process, and model fitting.

3.1. OVERVIEW

Assume that we have a spatial region of interest that is divided into n areal units. These units may be grid cells or more general units of uneven shape and area. Let i denote the i -th areal unit, $j \in (1, \dots, J)$ the j -th species, and let $t \in (1, \dots, T)$ index time. Our data $y_{ijt} \in [0, \infty)$ are comprised of counts per area of species j in the i -th areal unit at time t . For every (i, j, t) , we consider the observed data as arising jointly driven by a corresponding latent true $\tilde{w}_{ijt}$. Our goal is to estimate these $\tilde{w}_{ijt}$ and learn about the associated parameters that impact the growth of a species over time.

We present the framework visually in (Fig. 1). At a high level, our framework models abundances that iterate between growth and redistribution of populations, with the data y_{ijt} being observed after redistribution and before growth. Because simultaneously modeling redistribution and growth within the same process would lead to difficulties in computation, we choose to split the two processes into separate steps. Specifically at time t , population dynamics/growth are regulated by environment-species interactions (ESI) locally within each areal unit. We model growth using a discretized extension of the Lotka–Volterra (LV) model that incorporates uncertainty. After population growth, the individuals in each cell redistribute across the spatial region. This is represented using a species-specific redistribution matrix. In our framework, we assume that redistribution is independent across species.

Table 1. Model variables and parameters, with t indexing time

Name	Dimension	Definitions
<i>Process</i>		
$\mathbf{Y}_t$	$n \times J$	Observed abundances $[y_{1t} y_{2t} \cdots y_{Jt}]$
$\tilde{\mathbf{W}}_t$	$n \times J$	Latent species abundances before growth $[\tilde{\mathbf{w}}_{1t} \tilde{\mathbf{w}}_{2t} \cdots \tilde{\mathbf{w}}_{Jt}]$
$\mathbf{W}_t^*$	$n \times J$	Latent species abundances after growth $[\mathbf{w}_{1t}^* \mathbf{w}_{2t}^* \cdots \mathbf{w}_{Jt}^*]$
$\tilde{\mathbf{W}}_t^*$	$n \times J$	Left-censored latent species abundances after growth $[\tilde{\mathbf{w}}_{1t}^* \tilde{\mathbf{w}}_{2t}^* \cdots \tilde{\mathbf{w}}_{Jt}^*]$ with element $\tilde{w}_{ijt}^* = \max(0, w_{ijt}^*)$
$\mathbf{H}_j$	$n \times n$	Redistribution matrix for species j ; forms $nJ \times nJ$ $\mathbf{H} = \text{diag}(\mathbf{H}_j)$
$\mathbf{X}_t$	$n \times p_x$	Design matrix for growth; first column holds 1s
$\mathbf{U}_t$	$U \times n$	i -th column $\tilde{\mathbf{w}}_{it} \otimes \tilde{\mathbf{w}}_{it}$ Containing all unique combinations $\tilde{w}_{ijt} \cdot \tilde{w}_{ijt}$
$\mathbf{V}_t$	$p_x J \times n$	i -th column $\tilde{\mathbf{w}}_{it} \otimes \tilde{\mathbf{x}}_{it}$ Containing all combinations $\tilde{w}_{ijt} \cdot \tilde{x}_{it}$
$\sigma_{\eta,j}^2$	—	Process error for species j
$\sigma_{\nu,j}^2$	—	Observation error for species j
$\mathbf{P}$	$p_x J \times J$	ρ Reorganized for model fitting; multiplies $\mathbf{V}_t$
$\mathbf{A}$	$U \times J$	α Reorganized for model fitting; multiplies $\mathbf{U}_t$
<i>Structural zero</i>		
$\mathbf{Q}_t$	$n \times p_q$	Design matrix for zeros due to unsuitability
δ	$p_q \times J$	Coefficients for zeros due to unsuitability $[\delta_1 \cdots \delta_J]$; multiplies $\mathbf{Q}_t$
β	$2 \times J$	Coefficients for zeros due to chance $[\beta_1 \cdots \beta_J]$; β_j multiplies $(1, \tilde{w}_{ijt})$

n —number of areal units; J —number of species; U —number of unique pairwise species combinations; p_x —number of environmental predictors for ESI, plus one for intercept; p_q —number of predictors for zeros due to unsuitability plus one

Once individuals settle, we have the post-redistribution abundances at time $t + 1$, and we assume observation takes place following redistribution.

We model the population growth and redistribution processes as operating on the latent w_{ijt} . In our framework, these values are restricted to be non-negative to ensure proper behavior in both the growth model and the redistribution process, as we elaborate upon in Sect. 4.1. In the following sections, we use tildes and stars on the w_{ijt} to clearly denote each stage of the framework (i.e., growth, redistribution, or observation; see Table 1). We note that an observed $y_{ijt} = 0$ can arise due to one of two reasons: (1) the species j is truly absent or (2) the species present is but unobserved. More details are provided in Sect. 4.3.

4. EXPLICIT MODELING DETAILS

Our model extends the dynamic Generalized Joint Attribute Modeling (Clark et al. 2020) framework by embedding ESI and species–species interactions into a dynamic framework that incorporates dispersal across a landscape. The dynamic process is constructed using a multivariate first-order autoregressive model (MAR(1)) (Ives et al. 2003; Schliep et al. 2018; Ovaskainen et al. 1855). In univariate autoregressive (1) models, the growth-rate parameter ρ links the population abundance $\tilde{w}_{ijt}$ of species j in cell i at time t to the abundance at time $t + 1$, $\tilde{w}_{ij,t+1}$. Ives et al. (2003) extended the AR(1) model to whole communities using the MAR(1) model as an approximation to the nonlinear ecological process. In particular, the MAR(1) process assumes that the effect of interactions on the growth rates of each species is linear. See (Fig. 1) for a graphic visualization of our modeling framework.

4.1. GROWTH MODEL

The traditional LV model of community ecology describes how the abundance of one species in the community affects the growth rates of another species. Thus, the population growth component of our model framework is the LV model, which is then extended to incorporate the effects of ESI on population growth rates. Specifically, the dynamics of species j is described by two terms: (1) the autoregressive density-independent growth rate ρ_j and (2) the impact from interactions with the other species that depends on their abundances $\alpha_{jj'}$. We parameterize the LV model to enable learning about both types of terms through the following system of J equations for each cell i :

$$\frac{d\tilde{\mathbf{w}}_{it}}{dt} = \text{diag}(\tilde{\mathbf{w}}_{it}) [\boldsymbol{\rho} + \boldsymbol{\alpha}\tilde{\mathbf{w}}_{it}], \quad (1)$$

where $\tilde{\mathbf{w}}_{it}$ is a J -vector of species abundances, $\boldsymbol{\rho}$ are the density-independent growth coefficients, and $\boldsymbol{\alpha}$ is the $J \times J$ species interaction matrix with coefficients $\alpha_{jj'}$, i.e.,

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1J} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{J1} & \alpha_{J2} & \cdots & \alpha_{JJ} \end{bmatrix}$$

The $\alpha_{jj'}$ describe the effect of the abundance of species j' on the growth rate of species j . In particular, in the context of competition, $\alpha_{jj'} < 0$. We expect asymmetry ($\alpha_{jj'} \neq \alpha_{j'j}$) as it is highly unlikely that two species have the same exact effect on each other. The interpretation of the LV model requires the $\tilde{w}_{ijt} \geq 0$. Interactions involving negative abundances do not have an ecological interpretation. Additionally, the signs of the species–species interaction coefficients α_{jj} have specific interpretations, and allowing for negative $\tilde{w}_{ijt}$ would inhibit the interpretation of these relationships.

We extend the LV model by introducing ESI through the density-independent coefficients $\boldsymbol{\rho}$. Typically, $\boldsymbol{\rho}$ is the J -vector $(\rho_1, \dots, \rho_J)$. In our model, the abiotic environment can have effects on population growth rates. Thus assuming we have p_x environmental variables of interest, $\boldsymbol{\rho}$ is a $J \times p_x$ matrix of density-independent growth coefficients. The first column holds the intercepts for density-independent growth, and the remaining columns hold the coefficients for ESI:

$$\boldsymbol{\rho} = \begin{bmatrix} \rho_{10} & \rho_{11} & \rho_{12} & \cdots & \rho_{1,p_x-1} \\ \rho_{20} & \rho_{21} & \rho_{22} & \cdots & \rho_{2,p_x-1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \rho_{J0} & \rho_{J1} & \rho_{J2} & \cdots & \rho_{J,p_x-1} \end{bmatrix}$$

Due to mismatch in dimensions, we cannot use $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ directly in the model. Therefore, we re-organize the matrices into sparse matrices $\mathbf{A}$ and $\mathbf{P}$ that maintain the linear relationship between the ESI and growth. In the case of $J = 2$, this looks like:

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & 0 \\ \alpha_{12} & \alpha_{21} \\ 0 & \alpha_{22} \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \rho_{10} & 0 \\ \vdots & \vdots \\ \rho_{1,p_x-1} & 0 \\ 0 & \rho_{20} \\ \vdots & \vdots \\ 0 & \rho_{2,p_x-1} \end{bmatrix}$$

$\mathbf{P}$ is the matrix of coefficients for the ESI terms $\mathbf{v}_{it}$, which are defined as the interactions between the abundances $\tilde{\mathbf{w}}_{it}$ and the environmental variables plus intercept $\mathbf{x}_{it}$. For $J = 2$ species, $\mathbf{v}_{it} = (\tilde{w}_{i1t}, \tilde{w}_{i1t}x_{it,1}, \dots, \tilde{w}_{i1t}x_{it,p_x}, \tilde{w}_{i2t}, \tilde{w}_{i2t}x_{it,1}, \dots, \tilde{w}_{i2t}x_{it,p_x})'$. The re-ordered $\mathbf{A}$ matrix operates on the unique species–species interaction pairs $\mathbf{u}_{it}$, i.e., $\tilde{\mathbf{w}}_{it} \otimes \tilde{\mathbf{w}}_{it}$ minus replicates. For $J = 2$, $\mathbf{u}_{it} = (\tilde{w}_{i1t}\tilde{w}_{i1t}, \tilde{w}_{i1t}\tilde{w}_{i2t}, \tilde{w}_{i2t}\tilde{w}_{i2t})'$. Using this notation, the discrete-time version of Eq. 1 incorporating ESI is:

$$\tilde{\mathbf{w}}_{i,t+1} = \tilde{\mathbf{w}}_{it} + \mathbf{P}'\mathbf{v}_{it} + \mathbf{A}'\mathbf{u}_{it} \quad (2)$$

Thus, the growth of each species j within a given areal unit is impacted by both species–species interactions as well as ESI. For our framework, we interpret the quantity on the left of Eq. 2 to be the abundances post-growth at time t , as opposed to the abundances at time $t + 1$. For clarity, we replace the left-hand side of Eq. 2 with $\mathbf{w}_{it}^*$ to clearly show that growth operates within the same time t .

Lastly, we introduce uncertainty into population growth via species-specific growth error $\sigma_{\gamma,j}^2$. Then, the overall model for growth for all species across all areal units at time t is:

$$\mathbf{W}_t^* \sim MVN_{n \times J}(\tilde{\mathbf{W}}_t + \mathbf{V}_t'\mathbf{P} + \mathbf{U}_t'\mathbf{A}, \mathbf{I}_n, \boldsymbol{\Sigma}), \quad (3)$$

where $MVN_{n \times J}$ refers to the matrix-variate Normal distribution of dimension $n \times J$. $\boldsymbol{\Sigma} = \text{diag}(\sigma_{\gamma,1}^2, \dots, \sigma_{\gamma,J}^2)$ specifies independent process error across species, and the n -dimensional identity matrix $\mathbf{I}_n$ specifies that population growth is independent across the areal units (local growth). We remark that (3) can result in negative w_{ijt}^* . In order to prevent redistributing negative abundances in the next stage, which would lead to unexplained loss, we left-censor to obtain $\tilde{w}_{ijt}^* = \max(0, w_{ijt}^*)$. Thus after growth and censoring, we have the matrix $\tilde{\mathbf{W}}_t^*$ of latent species abundance at time t .

4.2. REDISTRIBUTION

After growth, we assume that populations move and redistribute across the spatial region. That is, some birds that were located in cell i at time t are likely to be in a different cell i' at time $t + 1$. For each species, we have an $n \times n$ redistribution matrix $\mathbf{H}_j$ that specifies the behavior of grid-based dispersal across the map. The value in the i -th row and i' -th column, $\mathbf{H}_{j,[i,i']}$, represents the dispersal of the population into cell i from cell i' . The columns of $\mathbf{H}_j$ are normalized to sum to one, as we want to conserve the total abundance of each species. Then, we simply redistribute the post-growth abundances $\tilde{\mathbf{W}}_{t+1}^*$ to obtain the abundances at

time $t + 1$ (less some nuances addressed in the next subsection): $\tilde{\mathbf{W}}_{j,t+1} = \mathbf{H}_j \tilde{\mathbf{W}}_{j,t+1}^*$. As mentioned in the Introduction, there are several ways that dispersal can be defined, such as diffusion, local dispersal kernels, and long-distance movement. Our notation here is general so any suitable choice of grid-based dispersal can be easily adopted.

4.3. DATA MODEL

The observations y_{ijt} are related to the latent $\tilde{w}_{ijt}$ through an appropriate data model. Focusing first on the observed positive abundances $y_{ijt} > 0$, we use the following model to link the observations to the latent abundances: $\log(y_{ijt}) \sim N(\log(\tilde{w}_{ijt}), \sigma_{\eta,j}^2)$. The observation errors $\sigma_{\eta,j}^2$ are independent and species-specific.

An observed $y_{ijt} = 0$ may occur due to one of two reasons: (1) the true latent abundance associated with that observation is truly zero, or (2) the species was present but not observed. In the first case, the associated $\tilde{w}_{ijt} = 0$. The marginal probability $\pi_{ijt} = P(\tilde{w}_{ijt} = 0)$ is induced by the model and not available in closed-form, but can be estimated during model fitting. The conditional probability of an observed zero given true presence, $q_{ijt} = P(y_{ijt} = 0 | \tilde{w}_{ijt} > 0)$, can also be estimated. We model this conditional probability using a probit link: $\hat{q}_{ijt} = \Phi(\delta_{0j} + \delta_{1j}\tilde{w}_{ijt})$. It follows that the observation model for $y_{ijt} > 0$ given $\tilde{w}_{ijt} > 0$ has total mass $1 - q_{ijt}$, and the marginal probability of an observed zero is $P(y_{ijt} = 0) = \pi_{ijt} + (1 - \pi_{ijt})q_{ijt}$.

5. MODEL FITTING DETAILS

5.1. PRIOR CHOICE AND MODEL ASSESSMENT

Due to the large number of parameters associated with the species–species interactions and ESI, model inference focuses on the most important interactions. We can only hope to identify those that have consequential effects, so it is important to exploit ecological understanding. In order to introduce sparsity, pairs of species that are suspected to not directly interact have their corresponding interaction coefficients set to zero a priori in $\mathbf{A}$ (Kissling et al. 2012; Ives et al. 2003). Often, the type of interaction (e.g., competition, predation) is known *a priori*, which can be incorporated by setting the sign of the corresponding coefficient to be positive or negative, as suggested by Ives et al. (2003). This can be achieved using priors with bounds $[-c_{jj'}, 0]$ or $[0, c_{jj'}]$ (Clark et al. 2020). Experiences with model fitting reveal difficulty in capturing the intercepts for density-independent growth in $\mathbf{P}$ (ρ) when using uninformative priors. Therefore, we suggest informative priors for these parameters. For the latent w_{ijt}^* , we use uniform priors with liberal bounds. Lastly, we take inverse gamma priors for the error parameters $\sigma_{\eta,j}^2$ and $\sigma_{\gamma,j}^2$, and diffuse, normal priors for δ and β .

Our model is fitted with a Metropolis–Hastings within Gibbs sampler to estimate the coefficients and latent abundances of interest. We provide the closed-form full conditionals for the growth interaction coefficients in Appendix A.1 and present the remaining sampling steps in Supplement S1. In the case of simulated data, we can assess our model’s performance by examining parameter recovery (Sect. 5.2). For real data, we suggest fitting the model on the first $T - 1$ time points of data and then, predicting for the held-out, final T -th time

point. Then, model checks such as posterior predictive checks (e.g., graphical or Bayesian p -values) or sampled predictive p -values can be performed (Conn et al. 2018). We also suggest comparing the posterior and prior densities for the growth coefficients α and ρ to assess if there is Bayesian learning and to interrogate if the priors dominate inference. For example, if the posteriors of many $\alpha_{jj'}$ are tightly concentrated on one of the prior bounds, re-evaluation of modeling assumptions and prior choice is recommended.

5.2. SIMULATION

We focus on examining how well we can recover the α and ρ parameters related to species growth, as these coefficients have ecological interpretations. We explore the effects of altering the amount of redistribution, the proportion of zeros in the observed data, the number of areal units, or the number of species in the community. We simulated data using the areal units of equal size (i.e., a grid-system) across $T = 20$ time points. Process and observation errors were the same for all simulations, but varied across species. Additionally, each species had the same redistribution matrix, constant across time for simplicity. Redistribution was defined according to Gaussian dispersal based on the distance between grid centers.

Simulation results are presented in Supplement S2, and we summarize the main findings here. We assess the recovery of the growth parameters by comparing nominal and empirical coverage. The largest factor impacting our ability to recover the α coefficients is the proportion of observed 0s in the data (Table S3a). It is particularly the intraspecific competition coefficients α_{jj} where empirical coverage falls below nominal. This is also the case when the number of areal units increases (Table S2a). Recovery of the interspecific species–species interaction coefficients $\alpha_{jj'}$ is generally quite robust, with empirical coverage at or above nominal. When the data have few observed zeros, the model is quite robust to increasing the number of species J (Tables S4, S5).

In the event of large amounts of redistribution, empirical coverage of the density-independent growth intercepts in ρ_{j0} falls below nominal (Table S1b). This may due to the fact that little in the model informs whether the loss or gain in abundance of a particular species in a given areal unit is due to growth or redistribution.

Taken together, our simulation encourages application of the model to data comprising relatively abundant species that do not redistribute across long-distances at high rates. Given these conditions, our model appears to be robust to the number of species in the community.

6. APPLICATION: EBIRD DATA

As mentioned in Sect. 2, we consider a community of six species. Three are resident species in the region of interest (tufted titmouse, carolina chickadee, carolina wren), whereas the remaining three are migrants that return to the study region during the breeding season (gray catbird, chipping sparrow, eastern towhee). Therefore, we focus on data collected during the combined months of May and June in each year, resulting in an annual time step.

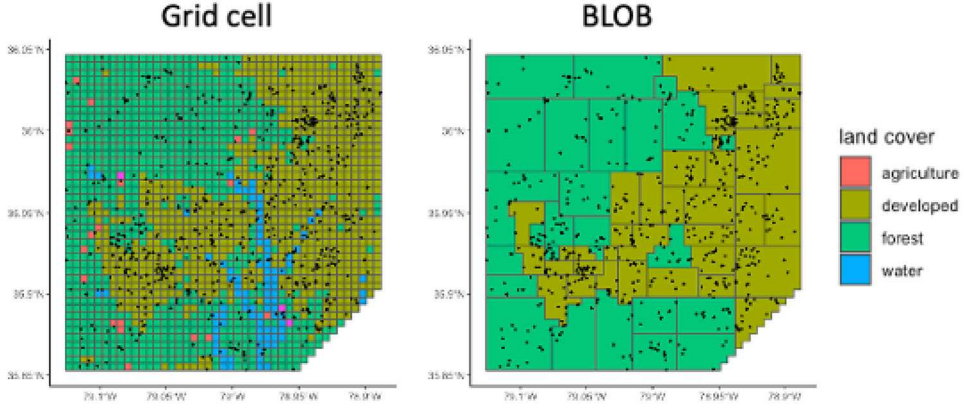


Figure 3. Spatial region colored by land cover type in 500×500 m grid cells (left) and our choice of BLOBs (right). Points are unique eBird observation locations in May–June during the years of interest.

6.1. DATA FORMATTING AND REDISTRIBUTION

In order to apply our model to eBird data, a key challenge arises to determine how to discretize the spatial region. Due to biases in sampling location (Johnston et al. 2020; Tang et al. 2021), using equal-sized grid cells/areal units of reasonable area over a large spatial region would result in the majority of the cells having no observations in a given year. To work around this issue, we create areal units of varying area that are irregularly shaped but approximately uniform in land cover type. The intention is that each resulting unit will have at least one observation for the majority of years in our analysis while preserving the type of land cover, as interactions may be habitat dependent. We refer to these single-land cover dominated areal units as ‘bulk land-occupancy blocks’, or BLOBs. See Fig. 3 for the 62 BLOBs we adopted considering only two dominant land cover types: forest and developed.

The next step in order to apply our model to eBird data requires obtaining the response variable y_{ijt} for each BLOB i , species j , and time t . Borrowing from the literature, we use the idea of fixed-radius point count (Hutto et al. 1986) to obtain a notion of sampling effort for each observation. In particular, we associate a circle of radius r to each observation, which we take to be the effective sampling area for that observation. Then, we take our observed count per area y_{ijt} to be the total sum of the counts of species j within BLOB i divided by the total effective sampling area:

$$y_{ijt} = \frac{\sum_{m=1}^{n_{it}} y_{ijt,m}}{\sum_{m=1}^{n_{it}} \pi r^2} = \frac{\sum_{m=1}^{n_{it}} y_{ijt,m}}{n_{it} \pi r^2}. \quad (4)$$

Upon forming the BLOBs, we use the dominant land cover type as a covariate for ESI.

Irregularly shaped and sized BLOBs require a nuanced redistribution matrix. In particular, it becomes inappropriate to use a redistribution kernel defined solely through the distance between centroids. Instead, we allow redistribution to occur both locally via Gaussian dispersal and at long-distances with some constant propensity for far-range movement. Additionally, to be realistic, redistribution matrices differ by species. Details are given in Appendix A.2.

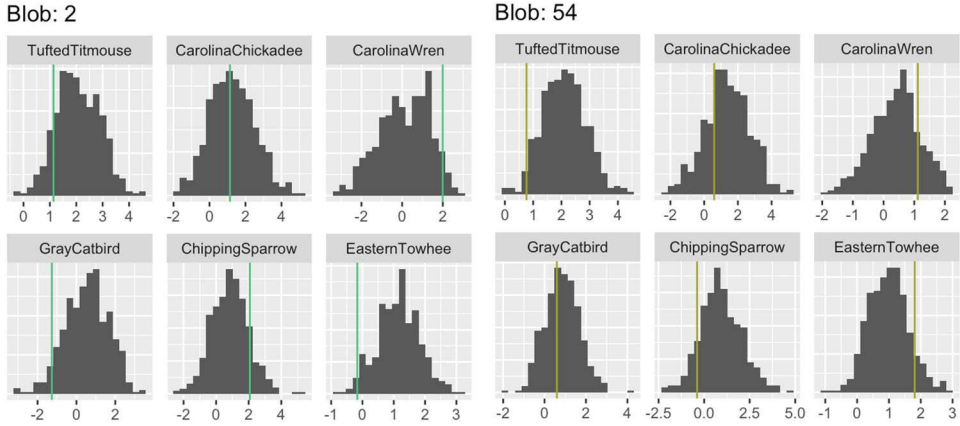


Figure 4. Posterior predictive densities of species abundances (log scale) in selected areal units in held-out year of data. The vertical line denotes observed log count-per-area, where green is forest land cover and yellow is developed (Color figure online). .

6.2. PRIORS

Based on feeding guilds of the six species, a priori we set the $\alpha_{jj'}$ interaction coefficients between the three lower canopy cleaner insectivores (tufted titmouse, carolina chickadee, carolina wren) and the two ground forager omnivores (chipping sparrow, eastern towhee) to 0. The remaining priors for the $\alpha_{jj'}$ specify competitive interactions. Additionally, informative priors for the intercepts in ρ were obtained using available information concerning clutch size, number of broods, and survival rate (see Section S3, Table S7).

7. RESULTS

We examine the effect of varying the different parameters in the redistribution matrix to obtain the best fitting model. Tables of these parameters and corresponding RMSPEs are provided in Tables S10, S11. The following results are reported for the values of parameters in the matrices $\mathbf{H}_j$ that yield the overall superior performance in terms of RMSPE for the held-out 2019 CPAs. Figure 4 plots histograms of posterior predictive distributions for the CPAs of all species in a few BLOBs in the 2019 North Carolina test data, along with the true observed CPA. We see that the posterior predictive distributions capture the true CPA quite well. In Figure S2, we provide spatial maps of posterior mean estimated latent log-abundance of carolina chickadee across time. The growth process variances are estimated much larger for the more abundant, resident species, and carolina chickadee has the largest estimated observation error (Table S14). Posterior summaries of the β coefficients in the regressions to estimate the conditional probability of observed zero given presence are provided in Table S15. The majority of species have an estimated significant negative $\beta_{1,j}$ coefficient, suggesting that as the abundances of species j increases, the probability of zero due to chance decreases. This is expected, as a species that is highly abundant in an area should be easily observed.

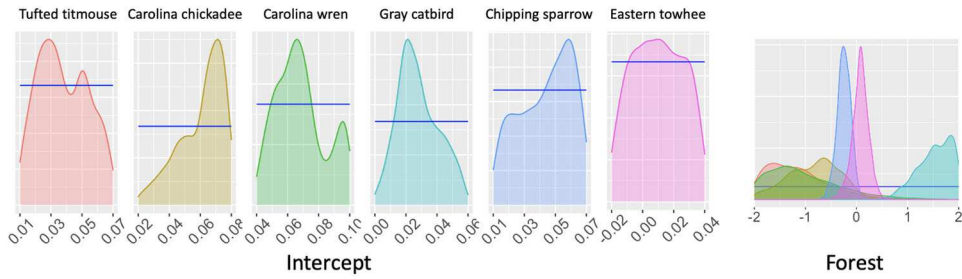


Figure 5. Prior (blue) and posterior densities for ρ coefficients by species. Posteriors for ESI (forest) reveal strong Bayesian learning for chipping sparrow and eastern towhee (Color figure online). .

7.1. ENVIRONMENT–SPECIES INTERACTIONS

The posterior densities for the ESI parameters plotted in Fig. 5 reveal strong Bayesian learning (posterior summaries provided in Table S13). The coefficient for forest is significantly positive for the gray catbird and significantly negative for chipping sparrow. Chipping sparrow are quite commonly found in gardens and parks in many areas (Dunn 2017), which may help explain the negative coefficient for forest land cover compared to the baseline of developed. In contrast, gray catbirds are less commonly found in residential areas.

7.2. SPECIES–SPECIES INTERACTIONS

For clarity, the coefficients α mathematically represent the species–species interactions in the model, but do not necessarily indicate the presence/absence of an interaction in nature. In the following, we interpret the coefficients within the scope of the statistical model. The majority of the interspecific competition parameters in α were estimated to be close to 0, corresponding to few direct pairwise interspecific interactions, whereas the intraspecific competition coefficients tended to be more strongly negative (Table S12). The prior and posterior densities are plotted in Fig. 6. We see that many of the intraspecific interactions (carolina chickadee, gray catbird, and chipping sparrow) are estimated to be large and negative. In many cases, the posterior densities of the α coefficients are different from the prior, demonstrating evidence of Bayesian learning. Cases where the posterior resembles the prior are most common for interactions involving gray catbird. This may be a result of a priori setting the gray catbird to interact with all the species when the data may not be able to inform about all these interactions.

8. DISCUSSION

Our framework captures nonlinear responses of a species to interactions with the environment and other species in the community, and our modeling enables inference on these important effects. We incorporate the notion of movement to more accurately represent the dynamics of an ecological system. Allowing populations to redistribute across a spatial region can lead to more accurate representation of the relationships between interactions

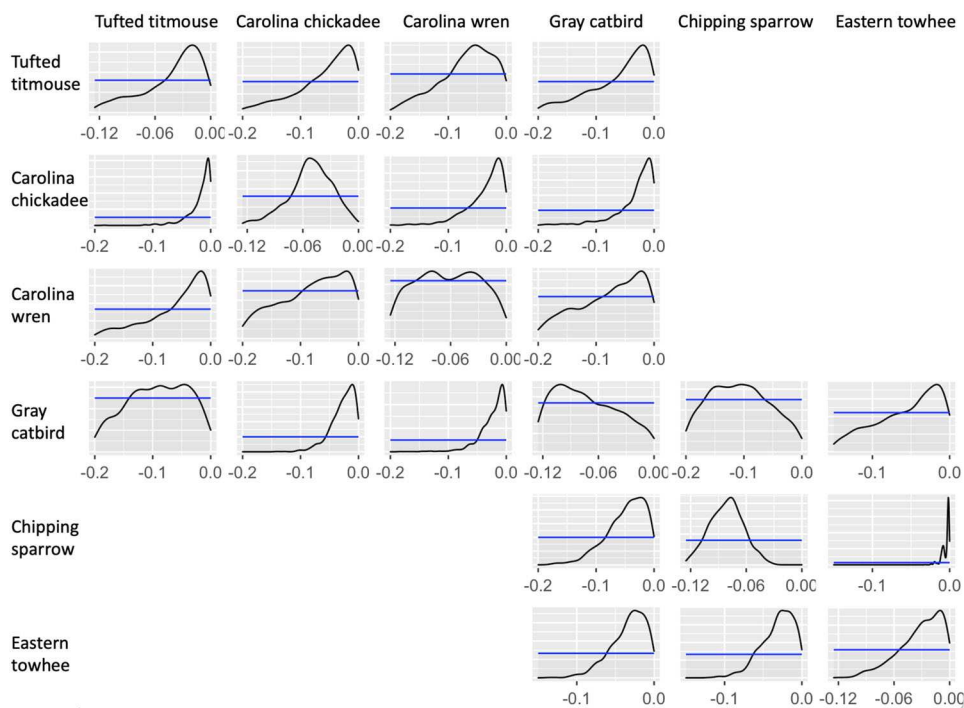


Figure 6. Prior (blue) and posterior (black) densities for α coefficients, where column species affect row species. Intraspecific competition is represented by the plots on the main diagonal. The white spaces represent interactions that are zero a priori. Posteriors that look different from the prior indicate Bayesian learning (Color figure online).

and population growth by disentangling the two components. Simulation studies reveal difficulty in recovering some species–species interaction parameters $\alpha_{jj'}$ when the data consist of a high proportion of observed zeros. These findings are not surprising; we expect difficulty in learning about species–species interactions when a species is not present and/or highly mobile. Additionally, it is well understood that fitting mechanistic models require large amounts of high-quality data (Urban et al. 2016). Therefore, it seems that our model is best suited for applications where species are observed to be relatively abundant.

The application using eBird data suggests a pattern of few strong and many weak species interactions. We found evidence of strong ESI with forested land cover for a few species, which is expected as land cover type affects nesting sites, cover, and food. We acknowledge that our application could benefit from incorporating additional environmental covariates. However, the small areas analyzed for this demonstration do not admit meaningful variation in climate variables such as temperature.

Many of the intraspecific coefficients (α_{jj}) were estimated to be large and negative. The majority of the interspecific ($\alpha_{jj'}$, $j \neq j'$) coefficients were either estimated to be close to 0 or not informed by the data; the model estimates few strong and many weak interactions. This trend is in agreement with the few studies that have examined interaction strengths, though these studies tend to focus on predator–prey interactions developed using food-web theory, rather than competitive interactions. For example, when examining population-level interactions through the Jacobian community matrices of soil food webs, de Ruiter et al.

(1995) found evidence of skew toward weak interaction strengths for both the negative effects of predators on prey, as well as for positive effects of prey on predators. The skew toward weak interaction strengths may be an important factor when examining community stability, as Allesina and Tang (2012) found that many weak interactions can help stabilize competitive networks, and Haydon (2000) demonstrated that the effects of intraspecific interactions may override the effects of interspecific interactions.

Here, we hypothesize why estimated parameters in α did not reflect the interspecific competition between pairs of species that we expected or have been documented to interact. This is most likely a result of our large-scale model-based approach, rather than analyzing data collected from a smaller scale experiment. The mathematical interpretation of the species-interaction coefficients α do not directly parallel the ecological interpretation of interaction. Our framework is developed for the population-level, and so unless there are repeated competitive effects that species j' exerts onto j that impact its growth rate, our model cannot learn about interactions that take place between individuals. Another factor to consider is spatial scale; closer individuals are more likely to interact with each other than distant ones (Wootton and Emmerson 2005), and interaction strength may change in heterogeneous habitats as the habitat structure can alter consumer–resource interactions (Werner et al. 1983). Additionally, allowing the species to redistribute could lead to fewer localized interactions and therefore, potentially weaker interaction strength. Lastly, the duration of each time interval for observing the potential interactions must be appropriate. If measurements are taken over too long of a time interval, indirect effects and density-dependent feedbacks can develop, which can influence the estimates of pairwise interactions (Berlow et al. 1999).

A limitation of our large model-based approach is that we cannot conclude the existence of interactions between species. For example, subtle preferences in microhabitat preference (ex. tall trees vs scrub) and foraging behavior could be explaining the patterns in the data, rather than a true competitive interaction. However, these findings could be useful for further research studies examining the potential competitive interactions between species. Altogether, our findings highlight the importance of developing realistic models that may capture both the biotic and abiotic processes that influence community dynamics. Future work includes developing a richer notion of redistribution that incorporates covariate information (e.g., temperature) to better represent the biological process, as well as applying the model to communities with difference dynamics (e.g., predatory/prey).

ACKNOWLEDGEMENTS

This manuscript has been developed based upon work supported by the National Science Foundation Graduate Research Fellowship Program (1644868) to Becky Tang, and the National Science Foundation (DEB-1754443) and the Programme d'Investissement d'Avenir under project FORBIC (18-MPGA-0004) (*Make Our Planet Great Again*) to James Clark.

Declarations

Conflict of interest The authors declare no conflict of interest.

Data Availability Raw and derived data analyzed are included as supplementary material.

[Received May 2022. Revised October 2022. Accepted October 2022. Published Online November 2022.]

A APPENDIX

A.1 POSTERIOR UPDATE FOR GROWTH COEFFICIENTS

Sampling the ESI coefficient matrix $\mathbf{P}$ can be achieved in a similar fashion, so here we provide details about sampling $\mathbf{A}$. Let $\Sigma = \text{diag}(\sigma_{\gamma,j}^2)$. The full conditional posterior distribution for $\text{vec}(\mathbf{A})$ is UJ multivariate Normal with mean $\mathbf{m}$ and covariance $\mathbf{M}$, where $\mathbf{M} = (\sum_{t=1}^{T-1} C_t)^{-1}$, $\mathbf{m} = \mathbf{M} \sum_{t=1}^{T-1} c_t$, and

$$\begin{aligned} C_t &= S^{-1} \otimes (\mathbf{U}_t' \mathbf{U}_t) \\ c_t &= C_t \otimes \text{vec}((\mathbf{U}_t' \mathbf{U}_t)^{-1} \mathbf{U}_t' (\mathbf{W}_t^* - (\tilde{\mathbf{W}}_t + \mathbf{V}' \mathbf{P}))) \end{aligned} \quad (5)$$

Recall that $\mathbf{A}$ is a sparse matrix, structured such that most elements are fixed at 0. Thus, we only require updating a subset of the elements of $\mathbf{A}$, or equivalently $\text{vec}(\mathbf{A})$. Define $\mathbf{A}_u$ to be a vector of this subset of elements of $\mathbf{A}$ to be updated, and vector $\mathbf{A}_{u'}$ the elements of $\mathbf{A}$ fixed at 0. Here, u holds indices of nonzero elements in $\text{vec}(\mathbf{A})$, and u' the indices of the zero elements in $\text{vec}(\mathbf{A})$. Rather than sampling all of $\text{vec}(\mathbf{A})$, we reduce dimensionality by conditioning on $\mathbf{A}_{u'} = \mathbf{0}$. We reorganize $\text{vec}(\mathbf{A})$, $\mathbf{m}$, and $\mathbf{M}$ such that the elements u are ordered before u' :

$$\text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{A}_u \\ \mathbf{A}_{u'} \end{bmatrix} \quad \mathbf{m} = \begin{bmatrix} \mathbf{m}_u \\ \mathbf{m}_{u'} \end{bmatrix} \quad \mathbf{M} = \begin{bmatrix} \mathbf{M}_{uu} & \mathbf{M}_{uu'} \\ \mathbf{M}_{u'u} & \mathbf{M}_{u'u'} \end{bmatrix}$$

Multivariate normal theory yields the conditional distribution from which we sample:

$$\mathbf{A}_u | \mathbf{A}_{u'}, \mathbf{P}, \Sigma \sim MVN_{|u|J}(\mathbf{m}_u - \mathbf{M}_{uu'} \mathbf{M}_{u'u}^{-1} \mathbf{m}_{u'}, \mathbf{M}_{uu} - \mathbf{M}_{uu'} \mathbf{M}_{u'u}^{-1} \mathbf{M}_{u'u}) \quad (6)$$

A.2 REDISTRIBUTION MATRIX

We first define local redistribution from BLOB B_i to B_k . Using a kernel $f(\cdot; \theta)$ and uniform distribution over B_i , the probability r_{ik}^* of moving from i to k is:

$$r_{ik}^* = \frac{1}{B_i} \int_{B_k} \int_{B_i} f(x - y; \theta) dx dy \quad (7)$$

for $x \in B_i$, $y \in B_k$. Then, we take the probability r_{ik} of local dispersal from B_i to B_k as

$$r_{ik} = \frac{r_{ik}^*}{\sum_{k'} r_{ik'}^*} \quad (8)$$

We also allow for long distance redistribution from B_i . That is, it is possible for birds located in BLOB i at time t to redistribute to any other BLOB on the map at time $t + 1$. The

probability of long distance redistribution to B_k from B_i , s_{ik} , is proportional to its area:

$$s_{ik} = \frac{|B_k|}{\sum_{k' \neq i} |B_{k'}|} \quad (9)$$

Letting l denote the proportion of dispersal that could occur due to long-distance, redistribution from B_i to any receiving B_k takes the form $\mathbf{H}_{[k,i]} = (1 - l)r_{ik} + ls_{ik}$, which is column-normalized to preserve abundances.

REFERENCES

- Allesina S, Tang S (2012) Stability criteria for complex ecosystems. *Nature* 483(7388):205–208
- Berlow EL, Navarrete SA, Briggs CJ, Power ME, Menge BA (1999) Quantifying variation in the strengths of species interactions. *Ecology* 80(7):2206–2224
- Clark JS (2010) Individuals and the variation needed for high species diversity in forest trees. *Science* 327(5969):1129–1132
- Clark JS, Scher CL, Swift M (2020) The emergent interactions that govern biodiversity change. *Proceedings of the National Academy of Sciences* 117(29):17074–17083
- Clark JS, Andrus R, Aubry-Kientz M, Bergeron Y, Bogdziewicz M, Bragg DC, Brockway D, Cleavitt NL, Cohen S, Courbaud B et al (2021) Continent-wide tree fecundity driven by indirect climate effects. *Nature communications* 12(1):1–11
- Conn PB, Johnson DS, Williams PJ, Melin SR, Hooten MB (2018) A guide to bayesian model checking for ecologists. *Ecological Monographs* 88(4):526–542
- de Ruiter PC, Neutel A-M, Moore JC (1995) Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* 269(5228):1257–1260
- Dewitz J (2019) National land cover dataset (nlcd) 2016 products (ver. 2.0, July 2020). U.S. Geological Survey data release
- Dunn JL (2017) Field guide to the birds of North America. National Geographic Books, Berlin
- eBird (2017) eBird: an online database of bird distribution and abundance [web application]. Cornell Lab of Ornithology, Ithaca
- Engler R, Guisan A (2009) Migclim: predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions* 15(4):590–601
- Haydon DT (2000) Maximally stable model ecosystems can be highly connected. *Ecology* 81(9):2631–2636
- Hutto RL, Pletschet SM, Hendricks P (1986) A fixed-radius point count method for nonbreeding and breeding season use. *The Auk* 103(3):593–602
- Ives AR, Dennis B, Cottingham K, Carpenter S (2003) Estimating community stability and ecological interactions from time-series data. *Ecological monographs* 73(2):301–330
- Johnston A, Moran N, Musgrove A, Fink D, Baillie SR (2020) Estimating species distributions from spatially biased citizen science data. *Ecol Model* 422:108927
- Kissling W, Dormann CF, Groeneveld J, Hickler T, Kühn I, McInerney GJ, Montoya JM, Römermann C, Schiffrers K, Schurr FM et al (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography* 39(12):2163–2178
- Ovaskainen O, Tikhonov G, Dunson D, Grøtan V, Engen S, Sæther B-E, Abrego N (2017) How are species interactions structured in species-rich communities? a new method for analysing time-series data. *Proc R Soc B Biol Sci* 284(1855):20170768
- Paine RT, Tegner MJ, Johnson EA (1998) Compounded perturbations yield ecological surprises. *Ecosystems* 1(6):535–545

- R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Schliep EM, Lany NK, Zarnetske PL, Schaeffer RN, Orians CM, Orwig DA, Preisser EL (2018) Joint species distribution modelling for spatio-temporal occurrence and ordinal abundance data. *Global Ecology and Biogeography* 27(1):142–155
- Smolik M, Dullinger S, Essl F, Kleinbauer I, Leitner M, Peterseil J, Stadler L-M, Vogl G (2010) Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *Journal of Biogeography* 37(3):411–422
- Strimas-Mackey M, Miller E, Hochachka W (2018) auk: eBird data extraction and processing with AWK. R package version 0.3.0. <https://cornelllabofornithology.github.io/auk/>
- Takeuchi Y (1996) Global dynamical properties of Lotka-Volterra systems. World Scientific
- Tang B, Clark JS, Gelfand AE (2021) Modeling spatially biased citizen science effort through the ebird database. *Environ Ecol Stat* 28:1–22
- Urban MC, Bocedi G, Hendry AP, Mihoub J-B, Pe'er G, Singer A, Bridle J, Crozier L, De Meester L, Godsoe W et al (2016) Improving the forecast for biodiversity under climate change. *Science* 353(6304):aad8466
- Vermote E, Justice C, Claverie M, Franch B (2016) Preliminary analysis of the performance of the landsat 8/oli land surface reflectance product. *Remote Sensing of Environment* 185:46–56
- Werner EE, Gilliam JF, Hall DJ, Mittelbach GG (1983) An experimental test of the effects of predation risk on habitat use in fish. *Ecology* 64(6):1540–1548
- Wikle CK (2003) Hierarchical bayesian models for predicting the spread of ecological processes. *Ecology* 84(6):1382–1394
- Wootton JT, Emmerson M (2005) Measurement of interaction strength in nature. *Annu. Rev. Ecol. Evol. Syst.* 36:419–444

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.