ARTICLE

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

Analyzing dynamic species abundance distributions using generalized linear mixed models

Erik Blystad Solbu[1] [iD] | Bert van der Veen[1,2,3] [iD] | Ivar Herfindal[3] [iD] |
Knut Anders Hovstad[3,4] [iD]

[1]Department of Landscape and
Biodiversity, Norwegian Institute of
Bioeconomy Research (NIBIO),
Trondheim, Norway

[2]Department of Mathematics, Norwegian
University of Science and Technology
(NTNU), Trondheim, Norway

[3]Centre for Biodiversity Dynamics,
Department of Biology, Norwegian
University of Science and Technology
(NTNU), Trondheim, Norway

[4]Norwegian Biodiversity Information
Centre, Trondheim, Norway

Correspondence
Erik Blystad Solbu
Email: erik.solbu@nibio.no

Funding information
Norges Forskningsråd, Grant/Award
Numbers: 223257, 244608, 272408/F40

Handling Editor: Paul B. Conn

Abstract

Understanding the mechanisms of ecological community dynamics and how they could be affected by environmental changes is important. Population dynamic models have well known ecological parameters that describe key characteristics of species such as the effect of environmental noise and demographic variance on the dynamics, the long-term growth rate, and strength of density regulation. These parameters are also central for detecting and understanding changes in communities of species; however, incorporating such vital parameters into models of community dynamics is challenging. In this paper, we demonstrate how generalized linear mixed models specified as intercept-only models with different random effects can be used to fit dynamic species abundance distributions. Each random effect has an ecologically meaningful interpretation either describing general and species-specific responses to environmental stochasticity in time or space, or variation in growth rate and carrying capacity among species. We use simulations to show that the accuracy of the estimation depends on the strength of density regulation in discrete population dynamics. The estimation of different covariance and population dynamic parameters, with corresponding statistical uncertainties, is demonstrated for case studies of fish and bat communities. We find that species heterogeneity is the main factor of spatial and temporal community similarity for both case studies.

KEYWORDS
environmental variance, generalized linear mixed model, Poisson lognormal, population dynamics, spatial and temporal correlation, species abundance distribution, species heterogeneity, variance partitioning

INTRODUCTION

Human activities are causing biodiversity declines at unprecedented rates (Brondizio et al., 2019). The human factors affecting biodiversity patterns, which vary in time and space, combined with complex dynamics of species communities, make it challenging to acquire knowledge on the ecological mechanisms involved in biodiversity change. However, it is vital to obtain a better understanding of the natural and anthropogenic processes causing biodiversity change to initiate actions to reduce and reverse biodiversity loss (Davison et al., 2021; Naeem et al., 2012; Sutherland

et al., 2004). Ecological models that quantitatively relate biodiversity changes to ecological processes will improve our understanding of changes in species communities.

One approach to studying community dynamics by Engen and Lande (1996b) showed how population dynamic models can generate species abundance distributions. Species abundance distributions describe the differences in abundance of species in a community (McGill, 2011), which could now be explained by population dynamical processes. Specifically, Engen and Lande (1996a, 1996b) and later Diserud and Engen (2000) showed how the species' growth rates, shapes and strengths of density regulation, carrying capacities, responses to environmental and demographic stochasticity, all determine the mean and variance of the species abundance distribution. The spatio-temporal analysis of communities using the dynamic species abundance distribution was explained in Engen et al. (2002).

The variance of the dynamic species abundance distribution can be partitioned into several different well known components of population dynamics (Engen et al., 2002). The variance partitioning consists of environmental noise, demographic stochasticity, species heterogeneity, expressed as variance in log carrying capacity between species, and sampling variation. For instance, Bellier et al. (2014) found that the spatio-temporal environmental variance accounted for 75% of the variation in a freshwater zooplankton community, whereas in a tropical butterfly community the species heterogeneity accounted for 81% of the variance (Engen et al., 2002). This is critical information for conservation, as large variation in species heterogeneity indicates that many species have low levels of carrying capacity, which increases their extinction risk (Henle et al., 2004).

The purpose of this paper was to describe how standard general linear mixed models (GLMMs) (Bolker et al., 2009; Harrison et al., 2018) can be used to model dynamic species abundance distributions, and to partition the variance of the abundance distribution into several components with a well defined ecological meaning. By doing so, we can relate population dynamic theory to a modeling framework that is available, accessible, and familiar to many ecologists. First we revisit the theory of dynamic species abundance distributions, and show how the population dynamic model generates different components of variation in relative and mean log abundance of species. We then define a GLMM with variance components that correspond to the variance of the dynamic species abundance distribution. With the connection between dynamic species abundance distribution and GLMM established, we investigate how different community dynamics can affect our estimation with a simulation study. Finally, we illustrate how community dynamics and spatial and temporal correlation is analyzed for two community data sets found in an open-access database.

METHODS

Dynamic species abundance distributions

Let X_i be the log abundance of species i, we assume the dynamics of X_i can be described by a stochastic differential equation

$$dX_i = [r_i - \gamma X_i]dt + \sigma_s dB_i(t) + \sigma_E dE(t), \qquad (1)$$

where we have ignored among-species density dependence and demographic variance for brevity (please refer to equation 8.1b in Lande et al. (2003) for more details regarding the extended process). The density independent long-run growth rate r_i of species i is assumed to be normally distributed among species with mean μ_r and variance σ_r^2. The species-specific strength of density dependence γ is assumed to be the same for all species and has a Gompertz (log–linear) form. The general response to environmental variability, i.e., the response shared for all species, is described by σ_E^2 and the species-specific response by σ_s^2. Finally, $dB_i(t)$ and $dE(t)$ are Brownian motions that are independent, identically distributed, and follow a normal distribution with mean zero and variance dt. All notation introduced above and in the following paragraphs is summarized in Appendix S1.

The continuous time process in Equation (1) is known as a diffusion process, and has two properties describing the population dynamics. The first property can be thought of as expected change in abundance, describing deterministic dynamics, and is denoted by $r_i - \gamma X_i$ for species i. The second property describes random fluctuations, denoted by σ_s^2 and σ_E^2 for species i. Both properties are described over a very short time interval and are known as infinitesimal mean and infinitesimal variance, respectively (Czuppon & Traulsen, 2021). If the infinitesimal mean is linear, e.g., $M(dX) = a + bX$, and the variance is constant, $V(dX) = c$, the process has a stationary distribution that describes the probability distribution of $X_i(t)$ when the effect of its initial value has diminished. This stationary distribution is normal with mean $\mu_X = a/b$ and variance $\sigma_X^2 = c/2b$. It can be shown that the mean log abundance $\overline{X} = \frac{1}{S}\sum_{i=1}^{s} X_i$ and the relative log abundance $x_i = X_i - \overline{X}$ are also diffusion processes (Lande et al., 2003). Assuming that the number of species S is large so that $\frac{1}{S} \approx 0$, the diffusion processes of x_i and $\overline{X}$ have linear infinitesimal means, and constant infinitesimal variances.

The stationary distribution of relative abundance x_i has mean $(r_i - \mu_r)/\gamma$ and variance $\sigma_s^2/2\gamma$, and the mean log abundance $\overline{X}$ has mean $\overline{r}/\gamma$ and variance $\sigma_E^2/2\gamma$, where $\overline{r} = \frac{1}{S}\sum_{i=1}^{s} r_i$. Because r_i is also normally distributed, the species abundance distribution of relative log abundances is normal with mean zero and variance

$$\mathrm{var}[x_i(t)] = \frac{\sigma_s^2}{2\gamma} + \frac{\sigma_r^2}{\gamma^2}. \tag{2}$$

From the description above, the distribution of log abundances X_i, can be obtained by reformulating it as $X_i = X_i - \overline{X} + \overline{X} = x_i + \overline{X}$, which from the diffusion processes has a normal distribution with mean $\overline{r}/\gamma$ and variance $\mathrm{var}[X_i(t)] = \sigma_r^2/\gamma^2 + \sigma_s^2/2\gamma + \sigma_E^2/2\gamma$. The temporal covariation of time interval u, follows from properties of the diffusion processes and is

$$\mathrm{cov}[X_i(t), X_i(t+u)] = \frac{\sigma_r^2}{\gamma^2} + \frac{\sigma_s^2}{2\gamma}e^{-\gamma u} + \frac{\sigma_E^2}{2\gamma}e^{-\gamma u}. \tag{3}$$

In the next section we show how the same covariance function can be defined in a GLMM framework.

GLMM formulation

In a model that describes spatial and temporal dynamics in an ecological community, the response $y_{ij}(t,z)$ is typically the abundance of a species i, at time t and location z, and observation replicate j (if applicable). The abundance is measured as counts, and the observational distribution is assumed to be $\mathrm{Poisson}\left[\mu_{ij}(t,z)\right]$, with log-link $\eta_{ij}(t,z) = \log\left[\mu_{ij}(t,z)\right]$. We simplify our analysis by assuming that temporal (t) and spatial (z) dynamics can be modeled separately. This separation of dimensions means that we fit *two* models if both dimensions are represented in the data set. In the following we consider temporal dynamics, but the same results can be used for spatial dynamics.

For a community studied over time, the linear predictor $\boldsymbol{\eta} = \left[\eta_{ij}(t)\right]$ is

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} = \mathbf{1}\beta_0 + [\mathbf{Z}_h \ \ \mathbf{Z}_e \ \ \mathbf{Z}_c \ \ \mathbf{Z}_o]\begin{bmatrix}\mathbf{b}_h \\ \mathbf{b}_e \\ \mathbf{b}_c \\ \mathbf{b}_o\end{bmatrix}, \tag{4}$$

where $\boldsymbol{X}$ is the design-matrix for fixed effects, and $\boldsymbol{\beta}$ is the vector of fixed effects. In all cases considered here, we assume $\mathbf{X} = \mathbf{1}_N$, and $\boldsymbol{\beta} = \beta_0$, where $\mathbf{1}_N$ is an $N \times 1$ vector of 1 s and β_0 is a scalar. The total number of observations N is the product of the number of species S, time points T and repeated observations J.

The design-matrix $\mathbf{Z}$ for random effects is split into four different components (their ecological interpretation is explained in the next section). These are most compactly described using Kronecker products, '$\otimes$', meaning that we multiply each element of the matrix to the left of the product with the whole matrix to the right. The first matrix, $\mathbf{Z}_h = \mathbf{I}_S \otimes \mathbf{1}_T \otimes \mathbf{1}_J$, where $\mathbf{I}_S$ is a diagonal matrix of order S, assigns one random effect to each species. The second matrix, $\mathbf{Z}_e = \mathbf{I}_S \otimes \mathbf{I}_T \otimes \mathbf{1}_J$, assigns one random effect to each species at each time point. The third matrix, $\mathbf{Z}_c = \mathbf{1}_S \otimes \mathbf{I}_T \otimes \mathbf{1}_J$ assigns one random effect to each time point, and finally $\mathbf{Z}_o = \mathbf{I}_S \otimes \mathbf{I}_T \otimes \mathbf{I}_J$ assigns one random effect to each observation. Similarly, the vector of random effects $\boldsymbol{b}$ has four components: $\mathbf{b}_h = [h_i]$ is a $S \times 1$ vector of among-species effects, $\mathbf{b}_e = [e_i(t)]$ is a $ST \times 1$ vector of within-species effects, $\mathbf{b}_c = [c(t)]$ is a $T \times 1$ vector of among-time effects, and $\mathbf{b}_o = \left[\varepsilon_{ij}(t)\right]$ is a $STJ \times 1$ vector of observation-level effects.

The distribution of the random effects is

$$\mathbf{b} = \begin{bmatrix}\mathbf{b}_h \\ \mathbf{b}_e \\ \mathbf{b}_c \\ \mathbf{b}_o\end{bmatrix} \sim N\left(\begin{bmatrix}\mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\sigma_h^2\mathbf{I}_S & & & \\ & \sigma_e^2\mathbf{I}_S \otimes \boldsymbol{\rho} & & \\ & & \sigma_c^2\boldsymbol{\rho} & \\ & & & \sigma_o^2\mathbf{I}_S \otimes \mathbf{I}_T \otimes \mathbf{I}_J\end{bmatrix}\right), \tag{5}$$

where σ_h^2 is the among-species variation, σ_e^2 is the within-species variation, and $\boldsymbol{\rho}$ is a $T \times T$ symmetric matrix with elements $\rho_{kl} = e^{-\gamma|t_l - t_k|}$, for any $k, l \in 1, ..., T$ and '$|\ |$' is the absolute value. Furthermore, σ_c^2 is the among-time variation, and σ_o^2 is the observation-level variation. From the definitions above, we can write the elements of the linear predictor $\boldsymbol{\eta}$ as

$$\eta_{ij}(t) = \beta_0 + h_i + c(t) + e_i(t) + \varepsilon_{ij}(t), \tag{6}$$

with expectation $\mathrm{E}\left[\eta_{ij}(t)\right] = \beta_0$, and covariance

$$\mathrm{cov}\left[\eta_{ij}(t), \eta_{kl}(t+u)\right] = \begin{cases} \sigma_h^2 + \sigma_e^2 + \sigma_c^2 + \sigma_o^2 & \text{for } i = k, j = l \text{ and } u = 0 \\ \sigma_h^2 + \sigma_e^2 e^{-\gamma u} + \sigma_c^2 e^{-\gamma u} & \text{for } i = k \text{ and } j \neq l \\ \sigma_c^2 e^{-\gamma u} & \text{for } i \neq k. \end{cases} \tag{7}$$

We see that the covariance between linear predictors of the same species at different time points (second line in Equation 7) has the same form as the covariance of log abundances in Equation (3), with $\sigma_h^2 = \sigma_r^2/\gamma^2$, $\sigma_e^2 = \sigma_s^2/2\gamma$, and $\sigma_c^2 = \sigma_E^2/2\gamma$. The correspondence between the dynamic species abundance distribution (Equation 3) and the GLMM model (Equation 7) shows that the different random effects have direct relations to population dynamics, which we will cover in the next section.

We stated above that we analyzed spatial and temporal dynamics separately. When having only spatial data, the results above remain the same by replacing "t" with "z," "T" with the number of locations "L" and the correlation $\boldsymbol{\rho}$ has elements $e^{-\alpha d(z_l, z_k)}$, where α is the inverse spatial scaling, $1/\alpha$, and $d(z_l, z_k)$ is the distance between z_k and z_l for any $k, l \in 1, ..., L$. If we have both time points and locations, the separation of dimensions into two models mean that we group observations by location when fitting the temporal model, and group by time points when fitting the spatial model. We return to the separate models in the case studies, where we have both space and time as dimensions.

Interpretation of intercept and random effects

Because the observational unit is counts and we use the log-link function, the intercept β_0 is the expected log abundance of species in the community. In most ecological studies, only a fraction of the true abundance is sampled so that the intercept β_0 is confounded with the sampling intensity, ν, and on the log scale we estimate $\widetilde{\beta}_0 = \ln\nu + \beta_0$. If the proportion of the community that has been sampled is known approximately, the estimated intercept can be corrected accordingly. Under the assumption of density regulation, the intercept is equal or proportional to the average log carrying capacity $r/\gamma = \ln K$ over all species, locations and times (Engen et al., 2002).

The among-species random effect, $\mathbf{b}_h \sim N(0, \sigma_h^2 \mathbf{I}_S)$ describes the variation in mean abundance among species. Because the mean abundance is proportional to the carrying capacity, the species level random effect can be interpreted as variation in carrying capacity among species. From Equations (3) and (7) we have that the species level random effect can be transformed to variation in density independent growth rate through $\sigma_r^2 = \sigma_h^2 \gamma^2$. In Figure 1a we have illustrated the effect of among-species variation. Each species has a unique carrying capacity sampled from the whole distribution of carrying capacities (horizontal lines [sample] and green square [distribution], respectively, in Figure 1a). The among-species variation is also called species heterogeneity in the community (Engen & Lande, 1996b), or ecological heterogeneity among species (Lande et al., 2003). From Equation (7) we see that the among-species variation adds a constant term, σ_h^2, to the covariance among-time points within species (Figure 1e).

The among-time random effect, $\mathbf{b}_c \sim N(0, \sigma_c^2 \boldsymbol{\rho})$, describes variation in mean abundance among-time points common to all species. The common environmental stochasticity, quantified by the variance σ_c^2, describes the environmental effects that have a synchronizing effect on species abundances. The larger σ_c^2 is, the more similar abundances will fluctuate over time, as illustrated in Figure 1b. Due to density dependence, the common environmental variance is temporally correlated, generating a non-negative covariation among species (Figure 1f). From Equations (3) and (7) we have that the temporal level random effect can be transformed to general responses to environmental variation through $\sigma_E^2 = 2\gamma\sigma_c^2$.

The species-temporal random effect, $\mathbf{b}_e \sim N(0, \sigma_e^2 \mathbf{I}_S \otimes \boldsymbol{\rho})$, describes the within-species variation due to environmental stochasticity. The within-species variation σ_e^2 reduces the synchrony among species within years (please refer to Figure 1c). However, within-species environmental variance is temporally correlated, determined by γ in $\rho_{kl} = e^{-\gamma|t_l - t_k|}$, so a small γ gives a larger temporal synchrony in the fluctuations over time (Figure 1c,g). From Equations (3) and (7) we have that the species-temporal level random effect can be transformed to species-specific responses to environmental variation through $\sigma_s^2 = 2\gamma\sigma_e^2$.

The observation-level random effect, $\mathbf{b}_o \sim N(0, \sigma_o^2 \mathbf{I}_S \otimes \mathbf{I}_T \otimes \mathbf{I}_O)$, describes sampling variation due to uncertainty in the measurements or differences in observer skills or any other factor that might result in different abundances of the same species at the same time point. In addition, the observation-level random effect can account for overdispersion in the Poisson distribution (Browne et al., 2005). However, not all sources of variation can be separated individually, so overdispersion can also include effects such as demographic stochasticity, i.e., variation in individual survival and fecundity, in small local populations (Lande et al., 2003). The extra noise due to observation differences adds a final layer of variation to the observed dynamics (Figure 1d), expressed as an element on the diagonal of the covariance matrix of linear predictors (Figure 1h).

Simulated example and case studies

In the simulation example, we look at our ability to estimate different variance components and the population dynamic parameters, for different sampling intensities ν,
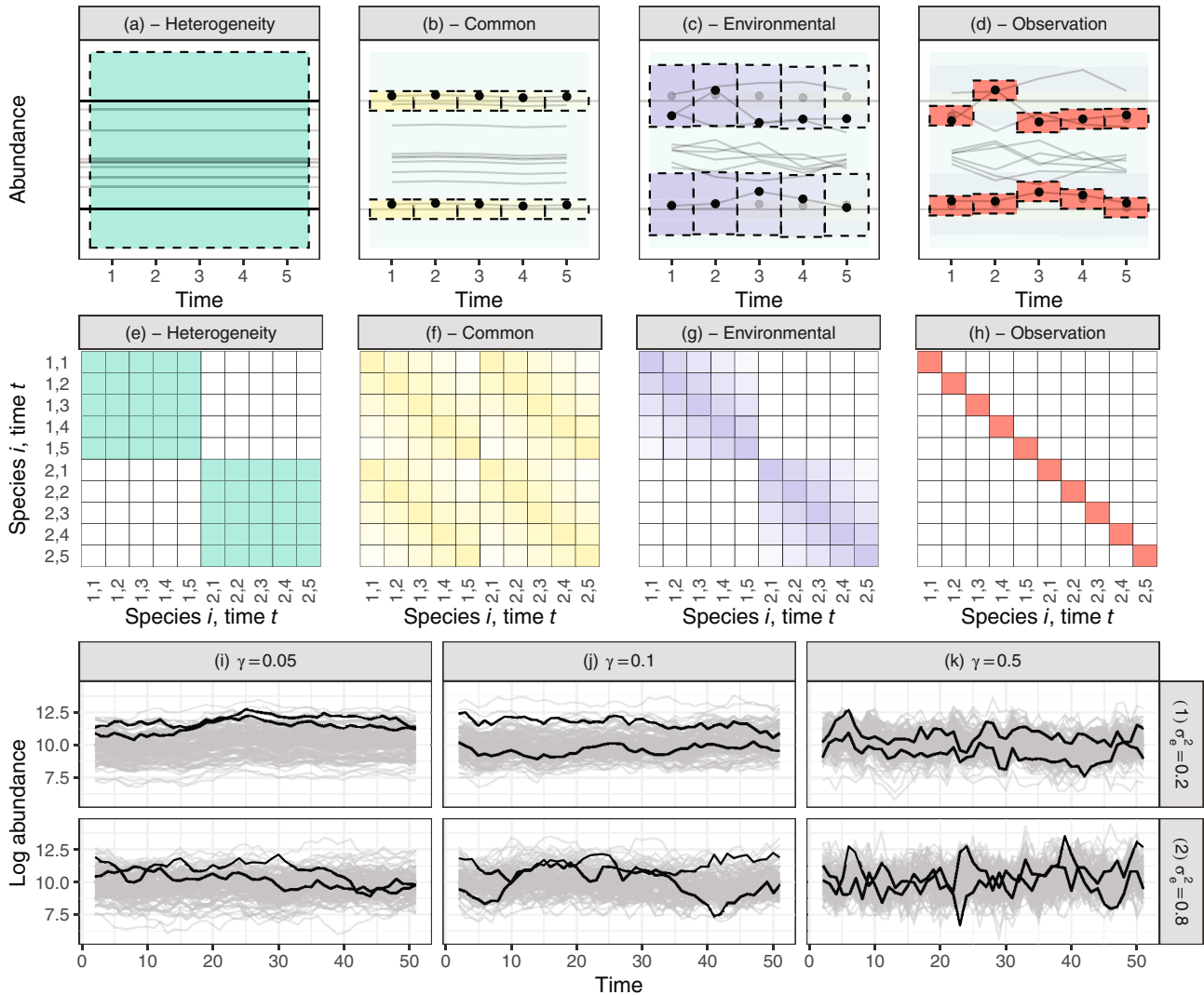
FIGURE 1 Population dynamics and covariance structure of random effects. (a–d) Schematic illustration of population dynamic interpretation of random effects. Two highlighted (black) species are shown changing throughout. Each square marked by dashed lines is a random effect, indicating a 95% probability interval around the mean. Going from (a) to (d), a new random effect is added from the previous figure. (a) Among-species variation or species heterogeneity, σ_h^2. The horizontal lines are 10 species' carrying capacities. (b) Common environmental variance σ_c^2, synchronizing fluctuations among species. A small random effect, correlated among years (yellow shading), is added to the sampled carrying capacities from (a). (c) Within-species environmental variation σ_e^2. For each species, a random effect, correlated among years (purple shading), is added to each time point. The gray dots are the previous values of the highlighted species. (d) Observation-level random noise σ_o^2. This is similar to (b), but now an independent effect is added to each species within each year. (e–h) Covariance matrix for linear predictors of two species. (e) Species heterogeneity generating covariation over time within each species. (f) Common environmental variance is correlated, gradually declining, over time among species. (g) Within-species environmental variance is correlated, gradually declining, over time within species. (h) Observation variance adds variance, reducing the overall correlation. (i–k) Sample realizations of community dynamics used in the simulation example. Strengths of density regulation, γ, from 0.05 (i) to 0.5 (k). (i–k) (1) has within-species variation σ_e^2 equal to 0.2, and (i–k) (2) has within-species variation equal to 0.8. Each simulation consists of 100 species, with two species highlighted in black.

and different strengths of density regulation γ. The simulations are based on the discrete version of the dynamic model presented in Equation (1) (please refer to the simulation procedure in Appendix S2). We simulate communities with log carrying capacity equal to 10, and strength of density regulation either 0.01, 0.1 or 0.5, so

the mean growth rate is accordingly $\mu_r = 10\gamma$. The within-species variation σ_e^2 is ranging from $0.2, 0.4, ..., 0.8$, whereas the among-species variation is $\sigma_h^2 = 1 - \sigma_e^2$, i.e., within- and among-species variations account for 20% to 80% each. The species-specific response to environmental stochasticity is $\sigma_s^2 = 2\gamma\sigma_e^2$, and the ecological

heterogeneity among species is $\sigma_r^2 = \gamma^2 \sigma_h^2$, expressed as variation in growth rate among species. General effect of environmental stochasticity, σ_c^2, is either 0.01, or 0.1, and we have $\sigma_E^2 = 2\gamma\sigma_c^2$. The sampling intensity, expressed as proportion of community sampled ν, is either 0.01, 0.1, or 0.5. The number of species simulated is 100 that all have the same initial log abundance. The dynamics are simulated for 370 time steps and the first 320 are removed before estimation to ensure that the populations have reached their stationary distributions. Although the assumption of stationary environment is unrealistic, especially over long time periods, the simulations can generate quite different dynamics, as illustrated in Figure 1i–k.

We consider two case studies with data from the open-access database BioTIME (Dornelas et al., 2018). The first data set is of a fish community in Ria de Aveiro, Portugal. The fish data contained observations from 4 years, and nine locations. We therefore fitted two models, one for modeling temporal dynamics and one for spatial dynamics. The distribution of random effects for the temporal model is

$$
\mathbf{b}_{\text{fish,time}} = \begin{bmatrix} \mathbf{b}_h \\ \mathbf{b}_e \\ \mathbf{b}_c \end{bmatrix}
$$
$$
\sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_h^2 \mathbf{I}_S & & \\ & \sigma_e^2 \mathbf{I}_L \otimes \mathbf{I}_S \otimes \boldsymbol{\rho} & \\ & & \sigma_c^2 \mathbf{I}_L \otimes \boldsymbol{\rho}_c \end{bmatrix} \right),
$$

(8)

where we estimate the following random effects: the among-species variation σ_h^2, the within-species variation σ_e^2 within locations, and the common environmental variation σ_c^2 among locations.

Although the model formulations in Equations (1) and (4) have assumed the same temporal scaling for species-specific and general response to environmental variation, this might not always be the case, for instance due to competition among species for common resources. To account for this difference, we estimate different correlation parameters for the temporal within-species and common environmental variation, $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_c$, with inverse temporal scales $1/\gamma$ and $1/\gamma_c$, respectively. We have no repeated observations, so we cannot estimate an observation-level random effect in this example. A similar formulation as in Equation (8) is done for the spatial model (Appendix S3: Equation S1), replacing the species-specific and general environmental correlation with $e^{-\alpha d(z_l, z_k)}$ and $e^{-\alpha_c d(z_l, z_k)}$, respectively.

The second data set is of a bat community in the Amazon near Manaus, Brazil. The data set available through BioTIME has 5 years of observations, ranging from 1997 to 2013 at 12 locations. We use additional habitat information from the data set to compare community dynamics for the community as a whole, and split into two data sets, where one is from locations with "continuous forest" habitats, whereas the other is locations from "fragmented interior" habitats. Because there are multiple dates with observations within a single year, we use "date" as repeated observations to illustrate the effect of overdispersion. For general environmental variation we also estimate uncorrelated noise due to sampling units (dates), in which the random effect is defined as $\mathbf{b}_u \sim N\left(\mathbf{0}, \sigma_u^2 \mathbf{I}_L \otimes \mathbf{I}_T \otimes \mathbf{I}_J \right)$ and J is the number of dates. The exact formulation of random effects for the temporal and spatial models are summarized in Appendix S3: Equations S2 and S3.

When analyzing community dynamics such as the fish and bats data, we estimate several variance and correlation parameters. We can summarize their effect in terms of community similarity by computing the temporal and spatial correlation in relative and mean log abundance. The correlation in relative log abundance consists of all within- and among-species variation components (please refer to Figure 1e,g,h). For example the temporal correlation in relative log abundance within a location in the fish community is $\rho_x(u) = \left(\sigma_e^2 e^{-\gamma u} + \sigma_h^2 \right) / \left(\sigma_e^2 + \sigma_h^2 \right)$. The correlation in mean log abundance is computed using all within- and among-time or location variation components (please refer to Figure 1f), for example the spatial correlation in mean log abundance for the bat community is $\rho_{\overline{X}}(v) = \sigma_c^2 e^{-\alpha_c v} / \left(\sigma_c^2 + \sigma_u^2 \right)$. All correlation functions for the case studies are summarized in Appendix S3: Equations S4–S11.

We assume that for each recorded combination of location, time point and observation replicate, any species without an abundance is a zero in the data set. The models where fitted in R (R Core Team, 2021) using the *glmmTMB* R-package (Brooks et al., 2017) and *DHARMa* R-package (Hartig, 2021) for residual diagnostics, but note that many GLMM R-packages share the syntax used for demonstration of fitting models here. Parameter uncertainty is estimated using parametric bootstrap (Efron & Tibshirani, 1994) with 1000 replicates in each case. All code used in this paper can be found online at https://doi.org/10.5281/zenodo.6401148 (Solbu et al., 2022), and a detailed guide for the fish data set in Appendix S6.

RESULTS

Simulation example

The estimation of the GLMM (Equation 4) on data generated by the discrete version of the population dynamic
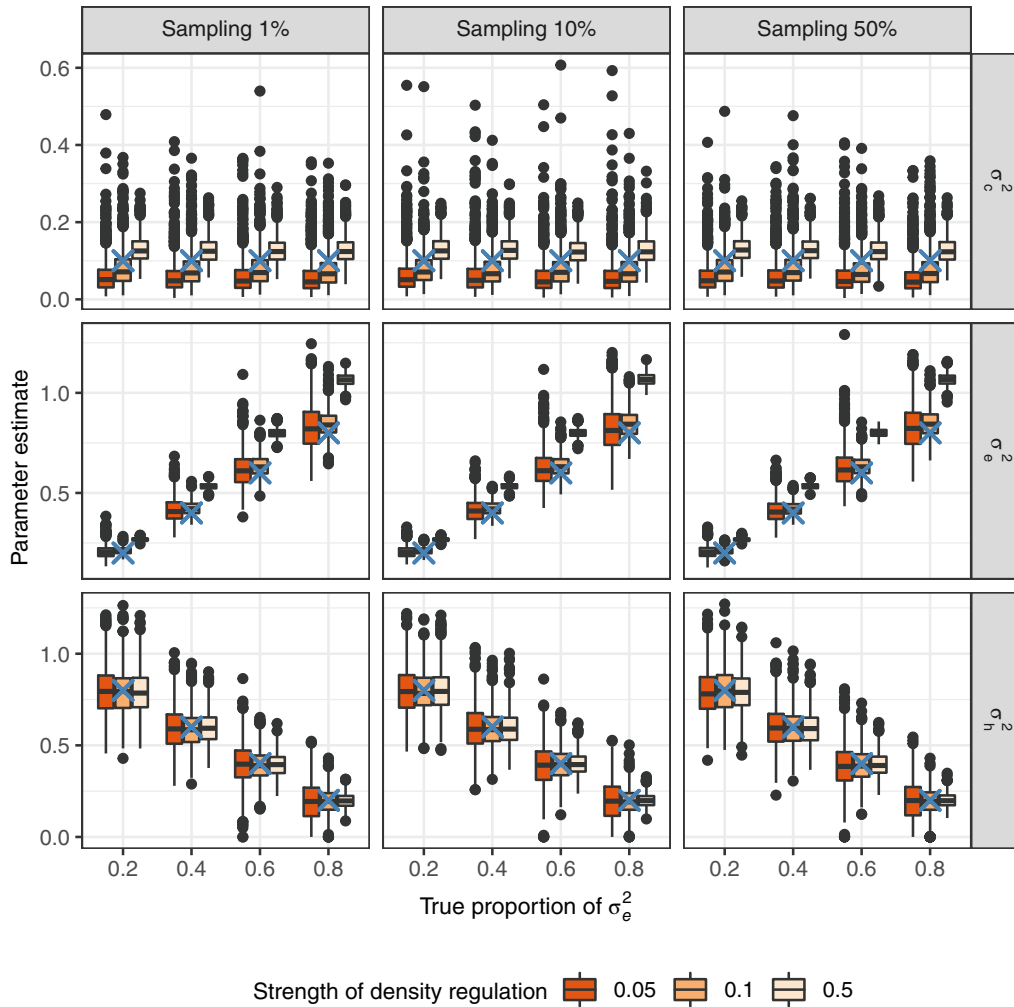
FIGURE 2 Estimates of common environmental effect (first row), within-species variation (second row), and among-species variation (third row), from fitting a generalized linear mixed model to simulated discrete population data. The columns are different sampling intensities, whereas the *x*-axis is the true proportion of variance due to within-species variation. The boxplots are point estimates from 1000 simulations, and the blue 'X' indicates the true values. Results for common environmental effect $\sigma_c^2 = 0.1$ are shown, while $\sigma_c^2 = 0.01$ is found in Appendix S5: Figure S3.

model (Equation 1) is summarized in Figures 2 and 3. The estimated among-species variation, σ_h^2 is accurate, regardless of sampling intensity and strength of density regulation (Figure 2, third row). The within-species variation, σ_e^2 is overestimated as strength of density regulation increases (Figure 2, second row). The common environmental variance, σ_c^2, ranges from under- to overestimated as strength of density regulation increases (Figure 2, first row). The results for $\sigma_c^2 = 0.01$ were similar, and can be viewed in Appendix S1: Figures S1–S3.

The estimated strength of density regulation is accurate for populations with weak density regulation (Figure 3a,b), and γ is overestimated for strong density regulation (Figure 3c). As a consequence of the estimates of γ, the species-specific response to environmental stochasticity σ_s^2, and ecological heterogeneity among

species, expressed as variation in growth rate among species σ_r^2, is overestimated for strong density regulation as $\sigma_s^2 = 2\gamma\sigma_e^2$ and $\sigma_r^2 = \gamma^2\sigma_h^2$ (Figure 3i,l). Temporal autocorrelation in general environmental effects, γ_c is overestimated (please refer to Appendix S1: Figure S2), whereas general response to environmental fluctuations σ_E^2 is accurate for weak density regulation (Figure 3d,e).

Fish community

In the fish community case study, the estimated total variance of the species abundance distribution, the variance in relative log abundance (Equation 2), is 14.1 based on the temporal GLMM (Equation 4, Figure 4e; Appendix S4: Table S1). Our approach estimates that 64% of this variation
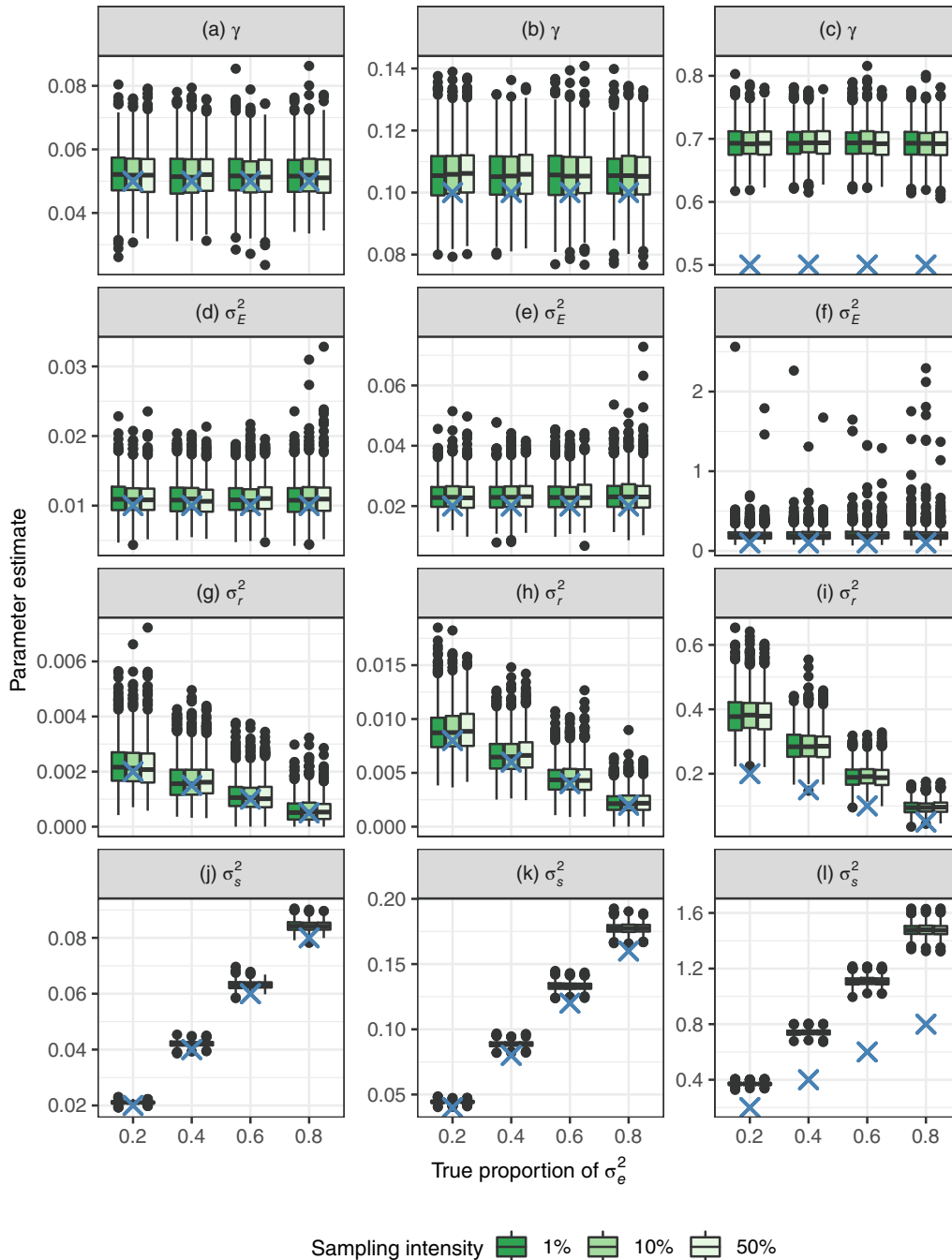
FIGURE 3 Estimates of strength of density regulation (first row), general response to environmental stochasticity (second row), ecological heterogeneity among species (third row), expressed as variation in growth rate, and species-specific response to environmental stochasticity (fourth row). The columns have different (true) strengths of density regulation: $\gamma = 0.05$, 0.1 and 0.5, respectively, whereas the x-axis is the true proportion of variance due to within-species variation. The boxplots are point estimates from 1000 simulations, and the blue 'X' indicates the true values. Results for common environmental effect $\sigma_c^2 = 0.1$ are shown, whereas $\sigma_c^2 = 0.01$ is found in Appendix S5: Figure S4.

is due to ecological heterogeneity among species, whereas the remaining proportion is attributed to species-specific responses to environmental stochasticity (Figure 4f). Variation in the mean log abundance, due to general environmental stochasticity, is roughly one for both the temporal and spatial model. The estimated strength of density regulation is

$\hat{\gamma} = 0.17\,\text{year}^{-1}$ (the hat symbol '⌢' indicates estimate), so the corresponding temporal scale $\widehat{1/\gamma}$ is 5.8 years. In Figure 4c we see how the uncertainty in temporal correlation for the general environment, the dynamic of mean log abundance, is much larger than the species-specific in Figure 4a, which is the dynamic of relative log abundance.
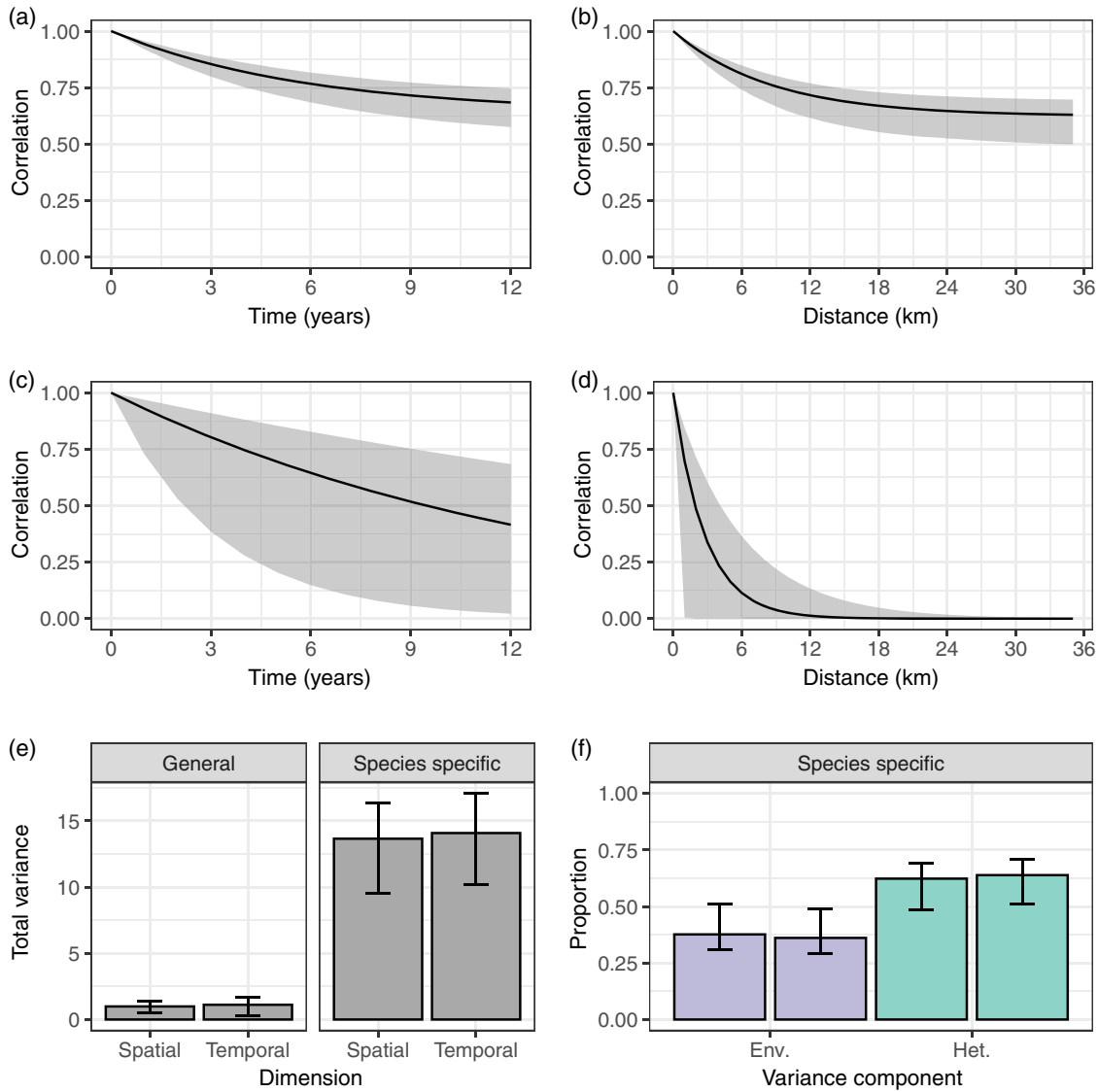
FIGURE 4 Analysis of fish community data from Ria de Aveiro, Portugal. Temporal (a) and spatial (b) correlations of relative log abundance with gray bands as 95% confidence intervals. Temporal (c) and spatial (d) correlations of mean log abundance with gray bands as 95% confidence intervals. (e) Estimated total variance for mean log abundance (general environmental effects) and relative log abundance (species-specific effects) with 95% confidence intervals. (f) Proportion of the total variance attributed to temporal environmental variation within-species (σ_e^2, "Env."), and among-species variation or ecological heterogeneity, (σ_h^2,"Het.") with 95% confidence intervals. For each variance component the columns are from the spatial (left) and temporal (right) models. All confidence intervals are based on 1000 bootstrap replicates.

The difference in uncertainty is because the general environment is estimated with only four time points over nine levels (locations), whereas the species-specific has four time points over 648 levels (species and locations).

Bat community

In the bat community case study, the among-species variation, or ecological heterogeneity among species, accounts for 74% of the total variation in the species abundance distribution when analyzing all habitats together (Figure 5d and Appendix S4: Table S2). Species-specific response to environmental stochasticity accounts for another 12% of the variance, whereas the final 14% is due to overdispersion. The large amount of overdispersion is not surprising because we have used samples from different dates as "repeated observations" of the community. Seasonal variation in species composition will affect this estimate, in addition to demographic stochasticity and intraspecific aggregation (Engen et al., 2002; Lande et al., 2003). The partitioning of the variance is similar for the "continuous" and
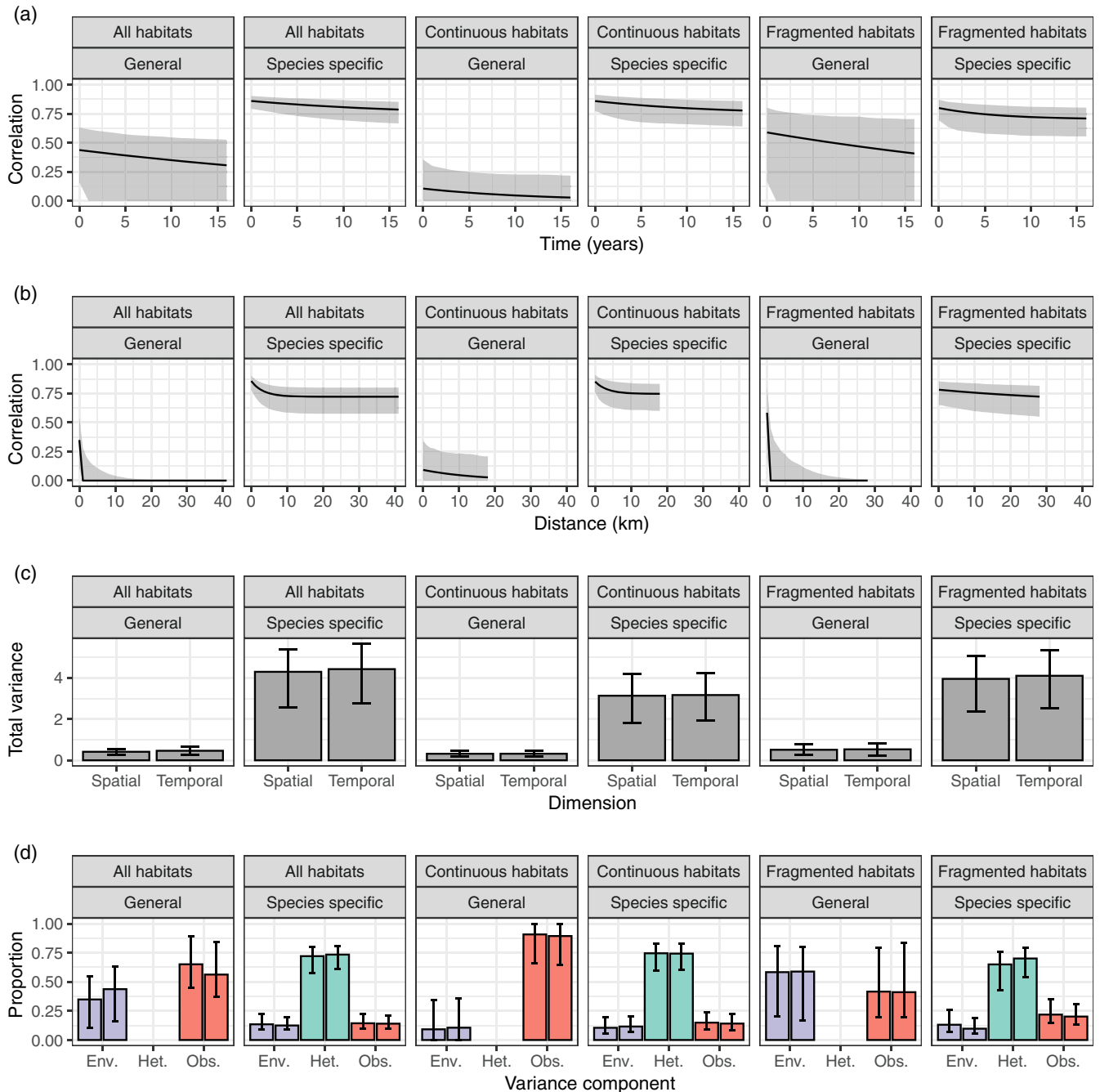
FIGURE 5 Analysis of bat community data from Manaus, Brazil. Temporal (a) and spatial (b) correlation in "General" mean log abundance and "Species specific" relative log abundance, with gray bands as 95% confidence intervals. (c) Estimated total variance for mean log abundance (general environmental effects) and relative log abundance (species-specific effects) with 95% confidence intervals. (d) Proportion of the total variance attributed to temporal environmental variation within-species σ_e^2, among-species variation or ecological heterogeneity σ_h^2, and observation-level random effects or overdispersion σ_o^2 with 95% confidence intervals for species-specific effects. For the general effects we have environmental variation (σ_c^2, "Env.") and uncorrelated random effects (σ_u^2, "Obs."). For each variance component the columns are from the spatial (left) and temporal (right) models. All confidence intervals are based on 1000 bootstrap replicates.

"fragmented" habitats analyzed separately, but "continuous" habitats have a smaller total variance, 3.17 compared with 4.10 in "fragmented" habitats. Although the proportions of variance are almost identical (74-12-14% "continuous," 70-10-20% "fragmented"), the spatial and

temporal scaling differ considerably. The "continuous" habitats have a much weaker strength of density regulation $\widehat{\gamma} = 0.08\,\text{year}^{-1}$, half that of the "fragmented" habitats $\widehat{\gamma} = 0.16\,\text{year}^{-1}$. This means that the temporal scaling, also known as the mean return time to equilibrium, is

twice as long for "continuous" habitats compared with fragmented ones. The spatial scale is shorter for "continuous" habitats than "fragmented" ones, with $\widehat{1/\alpha} = 3.4\,\text{km}$ and $\widehat{1/\alpha} = 46\,\text{km}$, respectively. One interpretation of this difference is that the composition of species has a higher rate of spatial turnover for "continuous" compared with "fragmented" habitats. A consequence of splitting up the data set into two smaller ones, is that the uncertainty, particularly in the population dynamic parameters, increases considerably. However, it serves to showcase how communities with different dynamics can generate similar variance compositions.

DISCUSSION

In this paper we showed how the dynamic species abundance distribution can be fitted using the GLMM framework. The main benefit is that we can use standard statistical tools to analyze the community dynamics more directly compared with previous approaches (Engen et al., 2002; Grøtan & Engen, 2008). With the accessibility and speed of analyzing dynamic species abundance distributions as GLMMs, population dynamics has the potential to affect the ability of ecologists to understand and evaluate different sources of variation with much greater flexibility than before.

Comparing the results from simulations of discrete population dynamic models to estimates from the GLMM illustrates how the modeling approach, i.e., the approximation of dynamics by a continuous time process, is more accurate for weaker density regulation than strong density regulation. This is due to the more volatile fluctuations that strong density regulation generates (Figure 1k), which is not accurately approximated by the continuous time population model (Equation 1). But if the dynamics are slower than a mean return time of two time units, meaning that $\gamma < 0.5$, the estimation results are more reliable (Figure 1i,j). Therefore, if the estimated strength of density regulation is higher than, say 0.5, parameter estimates should be interpreted with caution. One solution is to model the relative log abundance over shorter time intervals, for example monthly instead of yearly, if the data needed to do this are available. Among-species variation, expressed as variation in carrying capacity, can be accurately estimated even when the strength of density regulation is high (Figure 2). The general environmental effects, describing dynamics in mean log abundance, have fewer levels than the species-specific effects, describing dynamics of relative abundance, and are therefore harder to estimate with high accuracy (Figures 4c,d and 5a,b, Bolker et al., 2009; Harrison et al., 2018; Silk et al., 2020). It is, however, important to account for the variation in mean

log abundance among locations and time points by either fixed or random effects, otherwise it will bias the estimates of variation in relative log abundance.

The modeling approach in this paper makes it possible for ecologists to estimate variance components of dynamic species abundance distributions that have direct connections to population dynamics. For instance, two communities can have the same total variance, but different proportions attributed to environmental variance and species heterogeneity, which in turn will give potentially very different conclusions regarding, for example, the probability of extinction for individual species, which increases with environmental stochasticity (Lande et al., 2003). Similarly, two communities can have the same magnitude and composition of variance, but different mean return times to equilibrium, which affect for instance how species respond to changes in carrying capacity due to, for example, changes in habitat or climate. In the bat community, where the variance composition was almost identical for "continuous" and "fragmented" habitats, we estimated a much shorter temporal scaling for "fragmented" communities compared with "continuous" habitats. This can be interpreted as higher pressure through density regulation due to smaller habitats in fragmented areas, but the uncertainty makes it hard to conclude whether this is a true pattern.

For the fish community case study, the main cause of variation in community composition was species heterogeneity, expressed as variance in log carrying capacity, with 64% of the total variance, whereas 36% is attributable to environmental fluctuations. The results are consistent with other studies, although analyzed over different time ranges. Rebelo (1992) studied monthly variation from August 1987 to July 1988 in species abundance at 10 locations in the area and found that eight out of 55 fish species recorded accounted for 80% of the total abundance and six species were recorded at all time points. The findings of Rebelo (1992) suggest that the monthly dynamics of the fish community is generated by ecological heterogeneity, as some species are continuously abundant over the study period, whereas other less abundant species generate the species turnover in the community.

The correlation structure of the fitted dynamical species abundance model can be used to understand community turnover, and used as a measure of temporal β-diversity (Anderson et al., 2011; Magurran et al., 2019). If the correlation in relative log abundance rapidly approaches a limiting value as the interval between observations increases, it could indicate a fast turnover rate in the community. If this is the case, the strength of density regulation is an estimate of the rate of change in turnover along a temporal gradient (Anderson et al., 2011). Similarly, assuming the same exponential decay for spatial correlation, the parameter α expresses the turnover along a spatial gradient. For the fish

community, the synchrony due to species-specific responses to environmental fluctuations decreases by a fraction e^{-1} after roughly 8.6 km or 5.8 years, respectively. This indicates little spatial synchrony due to environmental fluctuations, as locations are mostly further away from each other, supporting previous reports of considerable spatial variation, whereas temporal diversity was more stable (Pombo et al., 2005, 2007).

Defined as a GLMM, the dynamic species abundance distribution can also be formulated as a joint species distribution model (JSDM; Warton et al., 2015). The main difference to other JSDMs is that dynamic species abundance distributions are used to model community dynamics, such as temporal correlations, whereas JSDMs are often used to analyze correlation between species to understand the complexity of species co-occurrence patterns (Ovaskainen et al., 2016; Pollock et al., 2014; Tobler et al., 2019). For JSDMs of species interactions in species rich communities, the solution to reduce the number of parameters is to apply dimension reduction techniques to the residual covariance matrix, known as generalized linear latent variable modeling (Hui et al., 2015; Niku et al., 2019; Warton et al., 2015). However, at the community level, the impact of species interactions is small compared with environmental fluctuations (Mutshinda et al., 2009, 2011). There are more complex models that incorporate density dependent dynamics, interactions and spatio-temporal variation to analyze joint dynamic species distribution models (Thorson et al., 2016), but the accessibility of GLMM analysis makes the connections presented here a valuable addition to multispecies ecological modeling.

The analysis of ecological communities with dynamic species abundance distributions draws upon themes that have been of interest to ecologists for over a century (McGill, 2011; Rosenzweig, 1995). By using the approach described in this paper, more information can be extracted by partitioning the residual variance into components so that each are associated with population processes and variation in space and time. Obtaining the population dynamic parameters that describe how the populations will behave over time is key to understanding how communities will react to environmental changes (Shoemaker et al., 2020). This is particularly important given the large environmental changes occurring globally, which both affect the spatial heterogeneity through habitat fragmentation and deterioration, and temporal heterogeneity through a more unpredictable climate. Knowledge on how communities respond to such changes is therefore pivotal to understand and act to reduce the ongoing biodiversity crisis (Johnson et al., 2017). With the implementation of dynamics species abundance distributions such as GLMMs, there are more opportunities to analyze a wide array of different data sets within this framework, which can provide greater understanding of community dynamics in a changing world.

ORCID
Erik Blystad Solbu https://orcid.org/0000-0002-6023-3100
Bert van der Veen https://orcid.org/0000-0003-2263-3880
Ivar Herfindal https://orcid.org/0000-0002-5860-9252
Knut Anders Hovstad https://orcid.org/0000-0002-7108-0787

REFERENCES
Anderson, M. J., T. O. Crist, J. M. Chase, M. Vellend, B. D. Inouye, A. L. Freestone, N. J. Sanders, et al. 2011. "Navigating the Multiple Meanings of β Diversity: A Roadmap for the Practicing Ecologist." *Ecology Letters* 14(1): 19–28.

Bellier, E., V. Grøtan, S. Engen, A. K. Schartau, I. Herfindal, and A. G. Finstad. 2014. "Distance Decay of Similarity, Effects of Environmental Noise and Ecological Heterogeneity among Species in the Spatio-Temporal Dynamics of a Dispersal-Limited Community." *Ecography* 37(2): 172–82.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. Henry, H. Stevens, and J.-S. S. White. 2009. "Generalized Linear Mixed Models: A Practical Guide for Ecology and Evolution." *Trends in Ecology and Evolution* 24(3): 127–35.

Brondizio, E. S., J. Settele, S. Díaz, and H. T. Ngo. 2019. "Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services." Bonn, Germany.

Brooks, M. E., K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Maechler, and B. M. Bolker. 2017. "glmmTMB Balances Speed and Flexibility among Packages for Zero-Inflated Generalized Linear Mixed Modeling." *The R Journal* 9(2): 378–400 https://journal.r-project.org/archive/2017/RJ-2017-066/index.html.

Browne, W. J., S. V. Subramanian, K. Jones, and H. Goldstein. 2005. "Variance Partitioning in Multilevel Logistic Models that Exhibit Overdispersion." *Journal of the Royal Statistical Society. Series A: Statistics in Society* 168(3): 599–613.

Czuppon, P., and A. Traulsen. 2021. "Understanding Evolutionary and Ecological Dynamics Using a Continuum Limit." *Ecology and Evolution* 11(11): 5857–73.

Davison, C. W., C. Rahbek, and N. Morueta-Holme. 2021. "Land-Use Change and Biodiversity: Challenges for Assembling Evidence on the Greatest Threat to Nature." *Global Change Biology* 27(21): 5414–29.

Diserud, O. H., and S. Engen. 2000. "A General and Dynamic Species Abundance Model, Embracing the Lognormal and the Gamma Models." *The American Naturalist* 155(4): 497–511.

Dornelas, M., L. H. Antão, F. Moyes, A. E. Bates, A. E. Magurran, D. Adam, A. A. Akhmetzhanova, et al. 2018. "BioTIME: A Database of Biodiversity Time Series for the Anthropocene." *Global Ecology and Biogeography* 27(7): 760–86.

Efron, B., and R. J. Tibshirani. 1994. *An Introduction to the Bootstrap*. London: Chapman / Hall/CRC.

Engen, S., and R. Lande. 1996a. "Population Dynamic Models Generating Species Abundance Distributions of the Gamma Type." *Journal of Theoretical Biology* 178(3): 325–31.

Engen, S., and R. Lande. 1996b. "Population Dynamic Models Generating the Lognormal Species Abundance Distribution." *Mathematical Biosciences* 132(2): 169–83.

Engen, S., R. Lande, T. Walla, and P. J. DeVries. 2002. "Analyzing Spatial Structure of Communities Using the Two-Dimensional Poisson Lognormal Species Abundance Model." *The American Naturalist* 160(1): 60–73.

Grøtan, V., and S. Engen. 2008. "poilog: Poisson Lognormal and Bivariate Poisson Lognormal Distribution." R Package Version 0.4. https://CRAN.R-project.org/package=poilog.

Harrison, X. A., L. Donaldson, M. E. Correa-Cano, J. Evans, D. N. Fisher, C. E. D. Goodwin, B. S. Robinson, D. J. Hodgson, and R. Inger. 2018. "A Brief Introduction to Mixed Effects Modelling and Multi-Model Inference in Ecology." *PeerJ* 2018(5): 1–32.

Hartig, F., 2021. "DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models." R Package Version 0.4.4. https://CRAN.R-project.org/package=DHARMa.

Henle, K., S. Sarre, and K. Wiegand. 2004. "The Role of Density Regulation in Extinction Processes and Population Viability Analysis." *Biodiversity and Conservation* 13(1): 9–52.

Hui, F. K. C., S. Taskinen, S. Pledger, S. D. Foster, and D. I. Warton. 2015. "Model-Based Approaches to Unconstrained Ordination." *Methods in Ecology and Evolution* 6(4): 399–411.

Johnson, C. N., A. Balmford, B. W. Brook, J. C. Buettel, M. Galetti, L. Guangchun, and J. M. Wilmshurst. 2017. "Biodiversity Losses and Conservation Responses in the Anthropocene." *Science* 356(6335): 270–5.

Lande, R., S. Engen, and B.-E. Sæther. 2003. *Stochastic Population Dynamics in Ecology and Conservation*. Oxford: Oxford University Press.

Magurran, A. E., M. Dornelas, F. Moyes, and P. A. Henderson. 2019. "Temporal β Diversity—A Macroecological Perspective." *Global Ecology and Biogeography* 28(12): 1949–60.

McGill, B. J. 2011. "Species Abundance Distributions." In *Biological Diversity: Frontiers in Measurement and Assessment*, edited by A. E. Magurran and B. J. McGill, 105–22. Oxford: Oxford University Press.

Mutshinda, C. M., R. B. O'Hara, and I. P. Woiwod. 2009. "What Drives Community Dynamics?" *Proceedings of the Royal Society B* 276(1669): 2923–9.

Mutshinda, C. M., R. B. O'Hara, and I. P. Woiwod. 2011. "A Multispecies Perspective on Ecological Impacts of Climatic Forcing." *Journal of Animal Ecology* 80(1): 101–7.

Naeem, S., J. Emmett Duffy, and E. Zavaleta. 2012. "The Functions of Biological Diversity in an Age of Extinction." *Science* 336(6087): 1401–6.

Niku, J., W. Brooks, R. Herliansyah, F. K. C. Hui, S. Taskinen, and D. I. Warton. 2019. "Efficient Estimation of Generalized Linear Latent Variable Models." *PLoS One* 14(5): e0216129.

Ovaskainen, O., N. Abrego, P. Halme, and D. Dunson. 2016. "Using Latent Variable Models to Identify Large Networks of Species-to-Species Associations at Different Spatial Scales." *Methods in Ecology and Evolution* 7(5): 549–55.

Pollock, L. J., R. Tingley, W. K. Morris, N. Golding, R. B. O'Hara, K. M. Parris, P. A. Vesk, and M. A. Mccarthy. 2014. "Understanding Co-Occurrence by Modelling Species Simultaneously with a Joint Species Distribution Model (JSDM)." *Methods in Ecology and Evolution* 5(5): 397–406.

Pombo, L., M. Elliott, and J. E. Rebelo. 2005. "Environmental Influences on Fish Assemblage Distribution of an Estuarine Coastal Lagoon, ria de Aveiro (Portugal)." *Scientia Marina* 69(1): 143–59.

Pombo, L., J. E. Rebelo, and M. Elliott. 2007. "The Structure, Diversity and Somatic Production of the Fish Community in an Estuarine Coastal Lagoon, Ria de Aveiro (Portugal)." *Hydrobiologia* 587(1): 253–68.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing https://www.R-project.org/.

Rebelo, J. E. 1992. "The Ichthyofauna and Abiotic Hydrological Environment of the ria de Aveiro, Portugal." *Estuaries* 15(3): 403–13.

Rosenzweig, M. L. 1995. *Species Diversity in Space and Time*. Cambridge: Cambridge University Press.

Shoemaker, L. G., L. L. Sullivan, I. Donohue, J. S. Cabral, R. J. Williams, M. M. Mayfield, J. M. Chase, et al. 2020. "Integrating the Underlying Structure of Stochasticity into Community Ecology." *Ecology* 101(2): 1–17.

Silk, M. J., X. A. Harrison, & D. J. Hodgson. 2020. "Perils and Pitfalls of Mixed-effects Regression Models in Biology." *PeerJ* 8: e9522.

Solbu, E. B., B. van der Veen, I. Herfindal, and K. A. Hovstad. 2022. "Code for 'Analyzing Dynamic Species Abundance Distributions using Generalized Linear Mixed Models'." Zenodo, Dataset. https://doi.org/10.5281/zenodo.6401148.

Sutherland, W. J., A. S. Pullin, P. M. Dolman, and T. M. Knight. 2004. "The Need for Evidence-Based Conservation." *Trends in Ecology & Evolution* 19(6): 305–8.

Thorson, J. T., J. N. Ianelli, E. A. Larsen, L. Ries, M. D. Scheuerell, C. Szuwalski, and E. F. Zipkin. 2016. "Joint Dynamic Species Distribution Models: A Tool for Community Ordination and Spatio-Temporal Monitoring." *Global Ecology and Biogeography* 25(9): 1144–58.

Tobler, M. W., M. Kéry, F. K. C. Hui, G. Guillera-Arroita, P. Knaus, and T. Sattler. 2019. "Joint Species Distribution Models with Species Correlations and Imperfect Detection." *Ecology* 100(8): 1–14.

Warton, D. I., F. Guillaume Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. C. Hui. 2015. "So Many Variables: Joint Modeling in Community Ecology." *Trends in Ecology and Evolution* 30(12): 766–79.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Solbu, Erik Blystad, Bert van der Veen, Ivar Herfindal, and Knut Anders Hovstad. 2022. "Analyzing Dynamic Species Abundance Distributions Using Generalized Linear Mixed Models." *Ecology* 103(9): e3742. https://doi.org/10.1002/ecy.3742