

Quantifying spatial variation in the size and structure of ecologically stratified communities

Simone Tenan^{*,†,1}, Mattia Brambilla^{1,2}, Paolo Pedrini¹ and Chris Sutherland^{†,3}

¹MUSE – Science Museum, Vertebrate Zoology Section, Corso del Lavoro e della Scienza 3, 38122 Trento, Italy; ²Department of Biodiversity and Protected Areas, Lombardy Foundation for the Environment, Largo 10 luglio 1976 1, 20822 Seveso (MB), Italy; and ³Department of Environmental Conservation, University of Massachusetts Amherst, Amherst, MA 01003, USA

Summary

1. Including ecological specialization (e.g. functional guild) in analyses performed across regions can help to study how size and structure of communities vary across environmental gradients. Multi-species occupancy models, and their extension to a multi-region framework, represent useful tools for such gradient analysis based on functional traits. However, in these models species richness is only a derived parameter and therefore explicit relationships cannot be inferred.

2. We provide a novel hierarchical multi-region community model that allows for direct modelling of trait-based patterns of species richness along environmental gradients by partitioning communities into ecologically relevant strata (e.g. guilds). We illustrate the flexibility of the model by simulations, where we estimate guild richness and community composition using both categorical and continuous species traits. In addition, we apply our model to data from eight avian communities to explore how guild richness varies across elevational gradients.

3. Simulations indicate that the model is able to estimate relationships between guild richness and environmental gradients even when the number of regions is low. Applying the model to the bird community data set, we show how the number of species in four feeding guilds varies in different ways across an elevational gradient.

4. Understanding large scale variation in species richness and how it relates to environmental gradients requires an explicit treatment of community structure. Our modelling approach allows the testing of hypotheses directly and simultaneously on strata- and community-specific species richness while accounting for the issues of rarity and detectability, and can be used to investigate the mechanisms which shape large scale spatial variation in species richness.

Key-words: community composition, ecological guilds, imperfect detection, multi-region community models, multi-species occupancy models, species richness

Introduction

A major goal of community ecology is to understand how and why communities vary across space (McGill *et al.* 2006). An intuitive approach for doing so is to investigate how particular features of a community change across quantifiable gradients. Typically, the number of species in a community, or species richness, has been the community state variable upon which much of our understanding of community dynamics is founded (Gotelli & Colwell 2001). However, it is becoming increasingly apparent that, in addition to the size of a community, it is also important to consider how community composition/structure may be affected by variation in environmental conditions (McGill *et al.* 2006; Tobler *et al.* 2015). A focus on functional traits and gradients can help the identification of general patterns (and hence prediction) in community ecology

(e.g. Grönroos & Heino 2012; Kissling, Sekercioglu & Jetz 2012) and allows investigation of the distribution of traits within and between communities and the way environmental gradients affect community structuring (McGill *et al.* 2006; Kissling, Rahbek & Böhning-Gaese 2007).

In the last decade, multi-species occupancy models (MSOM, Dorazio & Royle 2005) have proven to be useful tools for estimating community-level metrics and dynamics, thus providing estimates of species richness free from potential biases induced by issues of rarity and detectability, the two main reasons why species go undetected (Bailey, MacKenzie & Nichols 2014). Multi-species occupancy and abundance models have also been extended to potentially account for community composition (Tobler *et al.* 2015) and accommodate functional group structure (Yamaura *et al.* 2011; Chandler *et al.* 2013), where a functional group or guild is a group of species having similar ecological characteristics. Characterizing guild specific responses to environmental variation should greatly increase our understanding of how and why biodiversity changes across

*Correspondence author. E-mail: simone.tenan@muse.it

†These authors equally contributed to this work.

space and, importantly, allow for better predictive capabilities of how specific communities might respond to projected changes in environmental conditions in both space and time.

Although MSOMs produce accurate estimates of species richness and diversity, those measures are typically obtained for a single community of interest which is representatively sampled at different sites within that region. Differences in diversity measures using the MSOM approach can therefore only be assessed among sites or group of sites within a region and thus a single community, which is arguably of less interest than comparisons made across multiple independent communities. Alternatively, gradient analysis employed on multiple regions can be powerful to identify dynamics that are general across different regions (Fukami & Wardle 2005) although typically, such approaches do not account for imperfect detection. The classical single-region MSOM framework has been recently extended to multiple regions, which allows to test hypotheses on the geographic variation in species richness (Sutherland *et al.* 2016). A limitation of MSOMs and their multi-region extension is that species richness N is not directly modelled but rather derived from a zero-inflation (or data augmentation) parameter Ω . Recognizing that expected species richness $E(N) = M\Omega$, with fixed $M \gg N$, data augmentation converts the problem of estimating N to that of estimating Ω (Dorazio & Royle 2005; Dorazio *et al.* 2006; Royle, Dorazio & Link 2007) despite the latter has no intuitive biological interpretation (Schofield & Barker 2011). In short, this data augmentation formulation of the community model provides no direct information about the explicit nature of the relationship between N and a set of covariates, which in many cases would be of greatest interest.

In this paper, we develop a hierarchical multi-region community model where species richness, N , is modelled directly allowing for explicit inferences to be made about variation in community size. In our model, each community is assumed to be composed of groups of species which can be defined following different ecological or biological criteria, which we refer to as ecological stratification, and, rather than estimate total community size we model the size of each sub-community, which in turn provides estimates of community size and structure (Royle, Converse & Link 2012). Finally, having developed a framework for modelling the sub-structure of multiple communities, the multi-region feature of our stratified community model can be used to investigate how community size and, more interestingly, community structure varies across space in response to measurable region-specific covariates. The structure of the paper is as follows: first, we describe the stratified multi-region community modelling framework and, using simulated data across a range of scenarios, we evaluate the performance of the model. Then, using multi-species bird point count data collected from eight different regions, we demonstrate how our approach can be used to explore how community structure, specifically the number of species in each guild, varies across an elevation gradient.

Materials and methods

MODELLING REGIONAL VARIATION IN COMMUNITY STRUCTURE

Royle, Converse & Link (2012), and Royle & Converse (2014), described a model for stratified populations where individuals are grouped into sub-populations or strata. We apply the same concept to species within communities, where species (not individuals) are members of groups which are defined by some ecological stratification (e.g. guilds) within a community (not population). Thus, rather than estimate the number of individuals within each stratum of the population, the inference objective is to estimate the number of species within each stratum of the community. For example, suppose we want to compare community composition of $r = 1, 2, \dots, R$ communities based on $g = 1, 2, \dots, G$ functional groups. Let N_{gr} be the guild-specific species richness for each community and let $N_r^T = \sum_g N_{gr}$ be the total species richness for community r . Assuming that each species in a community can be assigned to only one functional group, each community can thus be naturally viewed as the union of all functional groups. Let variable g_{ir} indicate the functional group to which the observed species $i = 1, 2, \dots, n_r$ belongs, with the sum of (functional) group-specific sample sizes representing the total number of observed species in region r ($n_r = \sum_g n_{gr}$).

Variation in the number of species among functional groups and communities can be modelled as $N_{gr} \sim \text{Poisson}(\lambda_{gr})$, with the expected guild species richness that can be expressed as a function of p region-specific covariates (c_r), i.e. $\log(\lambda_{gr}) = \beta_{0g} + \sum_{j=1}^p \beta_{gj} c_{jr}$ (see below). The model for group-specific species richness is a Poisson GLM and standard GLM methods can be applied including, for example, an additive random effect to account for overdispersion.

With the Poisson model, by conditioning on the total number of species in the r th community N_r^T , across all G functional groups, the vector N_r of guild- and region-specific community sizes N_{gr} has a multinomial distribution:

$$N_r \sim \text{MN}(N_r^T, \pi_r) \quad \text{eqn 1}$$

with probabilities $\pi_r = (\pi_{1r}, \dots, \pi_{Gr})$ and $\pi_{gr} = \lambda_{gr} / \sum_g \lambda_{gr}$. Quantity π_{gr} represents the probability that species i , member of community in region r , belongs to group g .

Data augmentation is used to address the fact that each N_r^T is latent and thus unknown, removing it as explicit parameter by marginalizing over a $\text{Binomial}(M, \Omega_r)$ prior distribution for N_r^T , with fixed $M \gg \max(N_1^T, \dots, N_R^T)$ (note that, for convenience, M is kept equal across regions) and unknown parameters Ω_r that can be estimated.

The augmented data, with $M - n_r$ 'all-zero' encounter histories, will be represented by 'real' species, which occur with probability Ω_r , and species which are not members of the community of size N_r^T and which occur with probability $1 - \Omega_r$ (structural zeros). The latent binary variable $\omega_{ir} \sim \text{Bern}(\Omega_r)$ denotes whether species i was present in the r th community ($\omega_{ir} = 1$) or whether it is a structural zero ($\omega_{ir} = 0$). For species that were observed in a region, $\omega_{ir} = 1$ is fixed. Note that Ω_r and N_r^T are equivalent parameters, since $E[N_r^T] = \Omega_r M$. By conditioning on the total community size, Ω_r is thus confounded with $\exp(\beta_0)$, and one of the two parameters has to be fixed (Royle, Converse & Link 2012). We thus specified $\Omega_r = (\sum_g \lambda_{gr}) / M$. In addition, we assumed the functional group membership variables $g_{ir} \sim \text{Cat}(\pi_r)$, with $\pi_{gr} = \lambda_{gr} / \sum_g \lambda_{gr}$.

The matrix of true occupancy states \mathbf{Z} is a 3-dimensional array containing the elements z_{ijr} , the species-by-site-by-region occupancy states,

with site $j = 1, \dots, J_r$. Whenever $\omega_{jr} = 0$ it implies $z_{ijr} = 0$. Site- and region-specific occupancy states are described as $z_{ijr} \sim \text{Bern}(\psi_{ijr}\omega_{jr})$, where ψ_{ijr} is the species-specific occurrence probability for each site in each region and ω_{jr} is the inclusion parameter described above.

Encounter frequency data are also summarized in a 3-dimensional array Y with elements y_{ijr} . The standard binomial observation model relates the true underlying occupancy states, z_{ijr} , to the observed detection frequency data, $y_{ijr} \sim \text{Bin}(K_{jr}, p_{ijr}z_{ijr})$, with detection probability p_{ijr} , and number of sampling occasions $k = 1, \dots, K$. The observation model can be generalized to account for occasion-specific covariates, $y_{ijk} \sim \text{Bern}(p_{ijk}z_{ijr})$. Indexes of ψ_{ijr} and p_{ijk} suggest it is possible to model heterogeneity in these parameters as species-, site- or region-level fixed or random effects, or using species-, site- and region-specific covariates, in addition to time-varying effects for detectability.

This represents a novel extension of Sutherland *et al.* (2016) and can accommodate all the same variation such as differences in detectability and occurrence as a function of site-, species- or occasion-specific covariates. In addition, community- and guild-specific diversity metrics can be easily computed. For example, using occupancy probabilities, ψ_{ijr} , it is possible to directly estimate guild- and region-specific Hill numbers ($^q\Delta$; Bross, Hooten & Fitzpatrick 2014) that summarize three types of biodiversity metrics commonly used, (i) species richness ($q = 0$), (ii) Shannon diversity ($q = 1$, the Shannon entropy exponentiated), and (iii) Simpson diversity ($q = 2$, inverse of the complement of the Gini–Simpson index; Jost 2006):

$$^q\Delta_{gr} = \left(\sum_{i=1}^{N_{gr}} \left(\frac{\psi_{ijr}}{\sum_{s=1}^{N_{gr}} \psi_{sjr}} \right)^q \right)^{\frac{1}{1-q}} \quad q \neq 1, q \geq 0 \quad \text{eqn 2}$$

and

$$^1\Delta_{gr} = \exp \left(- \sum_{i=1}^{N_{gr}} \frac{\psi_{ijr}}{\sum_{s=1}^{N_{gr}} \psi_{sjr}} \log \frac{\psi_{ijr}}{\sum_{s=1}^{N_{gr}} \psi_{sjr}} \right) \quad q = 1. \quad \text{eqn 3}$$

We demonstrate the derivation of these using our case study below, and provide the script for their derivation in the supplemental model code.

SIMULATION STUDY

We conducted a simulation study to demonstrate how the multi-region model for stratified communities performs under two levels of complexity. In the simplest case (Scenario 1), we considered communities with equal expected number of species belonging to $G = 4$ groups, that can represent the levels of a categorical trait. We simulated the number of species for each functional group as $N_{gr} \sim \text{Pois}(\lambda_{gr})$, with $\log(\lambda_{gr}) = \beta_{0g}$ and with some variation in functional group size but no variation between regions, i.e. $\beta_0 = \{2.0, 2.0, 2.5, 3.0\}$, where β_0 is the vector of group-specific β_{0g} .

For the more complex scenario (Scenario 2), we simulated data where, in addition to differences in the intercept (e.g. Scenario 1), the expected number of species in each functional group also varied across regions as a quadratic function of a continuous region-specific environmental covariate, c_{1r} (e.g. elevation or latitude): $\log(\lambda_{gr}) = \beta_{0g} + \beta_{1g}c_{1r} + \beta_{2g}c_{1r}^2$. The covariate c_{1r} was simulated from a Uniform(−1.5, 1.5) distribution and we used the following values for the β parameters: $\beta_0 = \{2.4, 3.0, 4.0, 2.4\}$, $\beta_1 = \beta_2 = \{0, -0.2, -0.4, 0\}$.

In addition to modelling variation in functional group size in relation to region specific covariates, and to increase the realism of the model for assessment purposes, we also allowed for variability in detectability

by species as a function of a continuous species-specific trait (x_{ir} ; e.g. body size) generated from a normal distribution:

$$x_{ir} \sim N(\mu_{x,r}, \sigma_{x,r}^2) \quad \text{eqn 4}$$

where $\mu_{x,r}$ and $\sigma_{x,r}^2$ are, respectively, the region-specific mean ($\mu_{x,r} = 0$) and variance ($\sigma_{x,r}^2 = 1$) for the trait of species i . We assumed an effect of the species-specific covariate on detectability through the coefficient $\delta_1 = 0.3$ (eqn 5). Such an effect can be expected, for instance, in camera-trap studies where smaller-bodied animals are less likely to be sensed by devices (Marcus Rowcliffe *et al.* 2011). Species traits (e.g. size, form, presence of flowers, rarity or local population size) can also affect detectability of plant species (Garrard *et al.* 2013). We also allowed for unstructured variation in species-specific occurrence probabilities as a simple species level random effect (eqn 6). Heterogeneity in detectability and occupancy among species and regions were simulated and modelled as follows:

$$\text{logit}(p_{ijr}) = \delta_{0i} + \delta_1 x_{ir}, \quad \text{with } \delta_{0i} \sim N(\mu_{\delta_0}, \sigma_{\delta_0}^2) \quad \text{eqn 5}$$

and

$$\text{logit}(\psi_{ijr}) = \gamma_{0i}, \quad \text{with } \gamma_{0i} \sim N(\mu_{\gamma_0}, \sigma_{\gamma_0}^2). \quad \text{eqn 6}$$

We simulated species occupancy and detectability using a bivariate normal model with occupancy hyperparameters $\mu_{\gamma_0} = \text{logit}(0.2)$ and $\sigma_{\gamma_0}^2 = 1.5$, detectability hyperparameters $\mu_{\delta_0} = \text{logit}(0.15)$ and $\sigma_{\delta_0}^2 = 1.0$, and correlation coefficient $\rho = 0.75$ (Appendix S1, Supporting Information). Simulating data using this correlation structure between occurrence and detection probabilities produces species ensembles according to the biological reality that, usually, rare species are more difficult to detect (Dorazio & Royle 2005).

Note that we can easily extend models in eqns 5 and 6 by allowing for effects of site- and region-specific covariates, as well as by specifying region-specific hyperparameters. In the same way, gradient analysis could be performed by expressing hyperparameters in eqn 4 as functions of region-specific environmental covariates.

In all cases, data were generated for a ‘low’ ($R = 6$) and ‘high’ ($R = 12$) number of regions, $J = 50$ sites within each region, and $K = 5$ sampling occasions of each site within each region.

Two sets of priors were considered for the model parameters α_Ω and β_Ω (set 1: Normal(0, 100) distributions; set 2: Uniform(−5, 5) distributions). For details about prior specification on the other parameters and code for data simulation see Data S1. We used 300 simulations where models were fitted, using the Markov chain Monte Carlo framework. Summaries of the posterior distribution generated from a single Markov chain initialized with random starting values, run for 10 000 iterations after a 20 000 burn-in, and thinned every 10 draws. The mean across the 300 replicates of the Bayesian point estimate of the posterior mean for each parameter was calculated, along with the root mean squared error and the proportion of times the 95% credible interval for the estimate encompassed the real value (coverage). Models were implemented in program JAGS (Plummer 2003), that we executed from R (R Core Team 2012).

APPLICATION TO BIRD POINT COUNT DATA

We illustrate our approach by analyzing data collected at $R = 8$ spatially independent bird communities in Northern Italy and investigating the effect of an elevation gradient on guild specific species richness. Elevational gradients in species diversity are suitable for studying the factors that influence spatial patterns of biodiversity (e.g. Nogués-Bravo *et al.* 2008), and birds are popular study species since they are well known and easily surveyed.

Replicated detection non-detection data were obtained by sampling the eight communities in regions with an average altitude ranging from 290 to 1412 m asl. In each region, incidence records of all species observed within a 100-m radius around point counts were recorded. The number of point count locations in each region ranged from a minimum of $J = 11$ to a maximum of $J = 103$ (median: 31.5), and the number of sampling occasions (10-min point counts) ranged from $K = 2$ to $K = 10$ (median: 3). Point counts within regions were spaced 200–600 m apart, and average distance between regions was 95 km (minimum 23, maximum 219 km).

We modelled spatial variation in functional group size as a function of elevation for $G = 4$ feeding guilds: (i) granivores (i.e. predominantly feeding on seeds), (ii) insectivores (i.e. predominantly feeding on insects, both aerial and terrestrial/arboreal), (iii) generalists (neither definable as predominantly granivores nor insectivores), (iv) carnivores (i.e. predominantly feeding on vertebrates) (Kissling, Sekercioglu & Jetz 2012). We expected differences in the elevational distribution of feeding guilds, as a consequence of the variation in the environmental conditions along the gradient (e.g. Jankowski *et al.* 2013; Ferger *et al.* 2014).

We explored the relationship between the expected guild richness and (standardized) elevation through a quadratic effect: $\log(\lambda_{gr}) = \beta_{0g} + \beta_{1g}\text{elev}_r + \beta_{2g}\text{elev}_r^2$. Results are based on 100 000 samples from the posterior distribution of the parameters (3 chains, thinning of 50, and burn-in of 100 000). We augmented the data for all regions such that $M = 200$, i.e., by 200 – n_r species for each region ($n = \{75, 63, 29, 37, 47, 52, 55, 41\}$ for regions 1 to 8 respectively). We arbitrarily chose M so that it was likely much larger than the total number of species in the largest community (Sutherland *et al.* 2016). We allowed for heterogeneity among species in detectability and occupancy by treating them as random effects having community-level hyperparameters and a correlation structure as in Appendix S1: $\text{logit}(p_{ijr}) = \delta_{0i}$ with $\delta_{0i} \sim N(\mu_{\delta_0}, \sigma_{\delta_0}^2)$, and $\text{logit}(\psi_{ijr}) = \gamma_{0i}$ with $\gamma_{0i} \sim N(\mu_{\gamma_0}, \sigma_{\gamma_0}^2)$. In addition to guild richness, we derived two alternative guild- and region-specific diversity metrics, the Shannon and the Simpson diversity, through Hill numbers.

Results

SIMULATION STUDY

Posterior distributions for all parameters were similar for both of the prior specifications used (prior set 1 and 2) and we thus present results obtained under the Normal(0, 100) priors. Results from the models using the Uniform(–5, 5) priors are provided in Appendix S2 (Tables S3 and S4).

In the first and simplest scenario (Scenario 1), where group-specific species richness varied within region but not between regions, the 95% Bayesian credible interval (BCI) of the posterior estimates covered the true value in 93–95% of the 300 simulation replicates for $R = 6$ regions (Table S1 in Appendix S2), and in 91–97% of the times with 15 regions (Table S2 in Appendix S2).

Increasing model complexity (Scenario 2) did not entail a reduction in model performance, and parameter estimates were close to the true data generating values even when community group composition was modelled as a function of an environmental covariate (Fig. 1) and a continuous species trait was used to described community structure (Fig. S1 in Appendix S2). Coverage for all parameters ranged from 91%

to 94% ($R = 6$; Table S1 in Appendix S2) and from 91% to 96% ($R = 15$; Table S2 in Appendix S2).

APPLICATION

The minimum and maximum number of observed species belonging to the four guilds varied as follows: granivores, 5–13; insectivores, 10–31; generalists, 10–18; carnivores, 0–14. Observed and estimated patterns of elevational gradient showed differences among guilds, from ‘hump-shaped’ in granivores and generalists where there was a suggestion that species richness was highest at medium elevation and lowest at high and low elevation, to linearly decreasing in insectivores and monotonically decreasing in carnivores (Fig. 2).

While the 95% BCIs for the quadratic effect coefficient (β_{2g}) contained 0 for all functional groups and did not provide clear support for the quadratic effect (Table S5 in Appendix S2), the probability of a quadratic effect was $\Pr(\beta_{2g} < 0) = 0.78$ and $\Pr(\beta_{2g} < 0) = 0.76$ for granivores and generalists respectively, and $\Pr(\beta_{2g} > 0) = 0.90$ for carnivores. Guild-specific mean estimates of the Shannon and the Simpson diversity showed elevational patterns similar to those estimated for guild richness (Fig. 3). For other community-level parameter estimates see Table S5 in Appendix S2.

Discussion

We provide a hierarchical model to study community structure in relation to categorical and/or continuous species traits. By partitioning communities into ecologically relevant strata, such as functional guilds, our framework can be used to investigate variation in the size and structure of communities along ecological gradients. By accounting for the ecological stratification of communities we conceptually link the nature of the model, where communities are ‘stratified’, to the ecological question about guild-specific features. We do that in a multi-region and multi-species framework where gradient analysis can be applied directly to community-level parameters. At the same time and in the same fashion as classical multi-species occupancy models (MSOMs), we can test hypotheses on species-level parameters and account for imperfect detectability, and thus for the fact that not every species in each ecological group is detected. Most importantly, we model patterns of guild richness among different communities in a formulation which allows to test hypotheses directly on strata- and community-specific species richness, overcoming the inherent limitation of MSOMs (Dorazio *et al.* 2006) and of their multi-region extension (Sutherland *et al.* 2016) where the estimation of community size is indirect. MSOMs, applied either to single or multiple regions, can be extended to include species-level functional group attributes and used to estimate guild richness. Single-region case studies are available for multi-species abundance models (Yamaura *et al.* 2011; Chandler *et al.* 2013). However, both multi-species occupancy and abundance models do not allow for formal

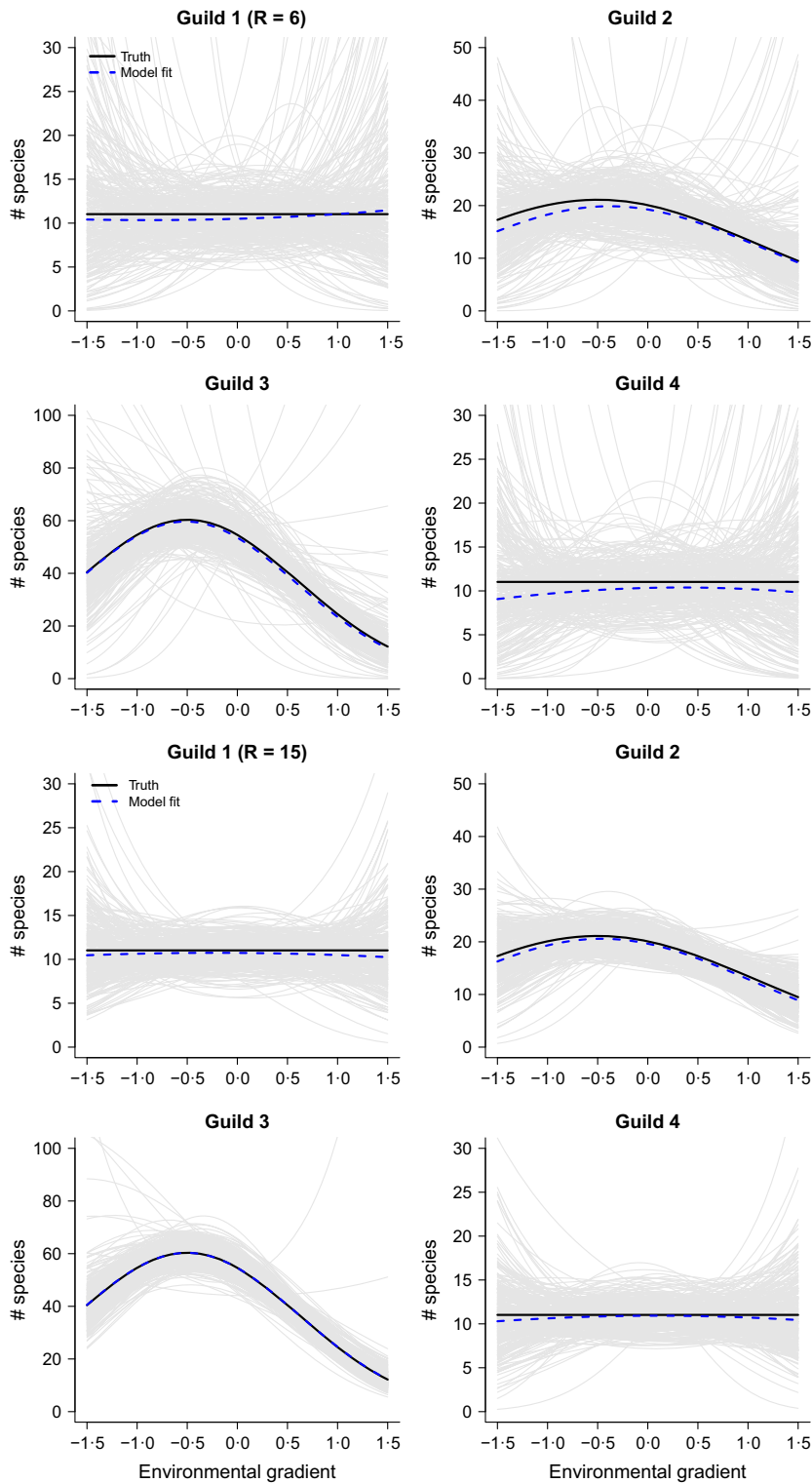


Fig. 1. Simulation results for the estimated variation in guild richness, for $G = 4$ groups, along a simulated environmental gradient. The black line represents the data-generating relationship (i.e. 'truth'), the blue dashed lines indicate the fitted relationship based on the estimated mean values (across 300 replicates) of the parameters β_{0g} , β_{1g} , and β_{2g} , for $R = 6$ and $R = 15$ regions respectively. Grey lines are the fitted relationships from each of the 300 simulations.

hypothesis testing directly on guild richness or total community size. We demonstrated that our model performed well, especially in terms of ability to estimate guild richness gradients, even when the number of regions was low ($R = 6$). Although an in-depth analysis of model performance in relation to data quality was beyond the scope of this study, we explored two main scenarios in relation to a

different number of regions R in a range of R thought to be most common among existing datasets on multiple regions and suitable in relation to our example application with $R = 8$ regions.

Gradient analysis, i.e. the comparison of community characteristics along environmental gradients, is relevant for understanding community dynamics, where useful gradients are

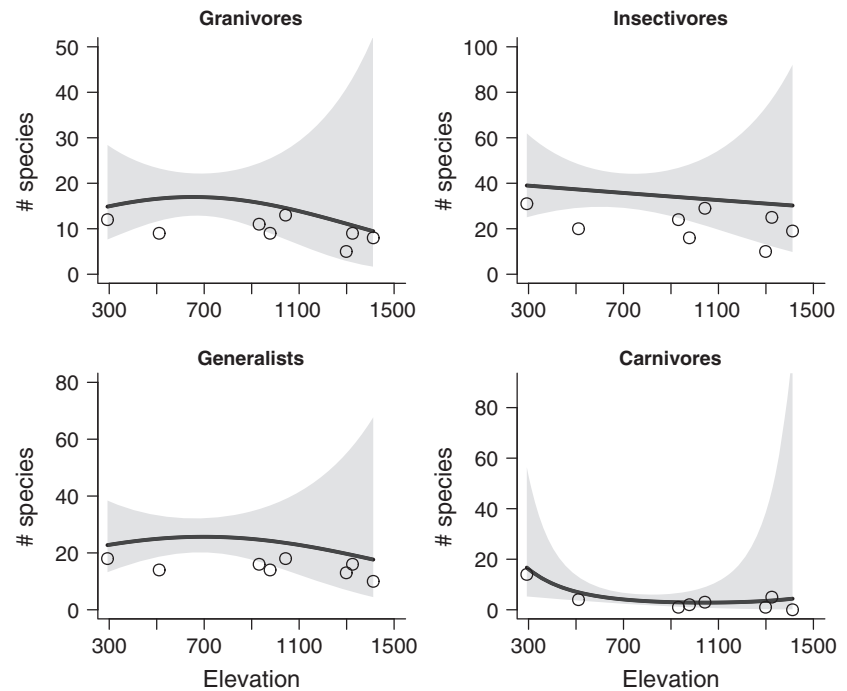


Fig. 2. Elevational gradient of guild richness for the avian communities of eight regions. Grey areas indicate 95% Bayesian credible intervals. Open circles indicate observed guild richness. Elevation is given in meters above sea level.

usually anthropogenic or natural (Fukami & Wardle 2005; McGill *et al.* 2006). Ecologists have devoted considerable effort towards characterizing variation in species richness in response to environmental gradients in search of universal patterns, and the relationship between altitude and community size is used as template for large scale gradient studies (Nogués-Bravo *et al.* 2008).

We used real data on eight spatially-separated avian communities to examine the elevational variation in the species richness of different dietary guilds. As expected, the relationship varied among functional groups. The hump-shaped patterns of granivores and generalists (Fig. 2) can be explained by vegetation structure and availability of food resources (Jankowski *et al.* 2013; Ferger *et al.* 2014). Medium elevation habitats are in fact more heterogeneous and characterized by a patchwork of herbaceous field buffers and mixed conifer and broadleaved forests. The decreasing pattern of insectivores and carnivores may also be affected by prey availability (Ferber *et al.* 2014). We also showed how to derive commonly used community metrics through Hill numbers (Broms, Hooten & Fitzpatrick 2014), stratifying them by functional groups. Deriving biodiversity metrics while accounting for imperfect detectability is possible only in a multi-species occupancy framework, and cannot be accommodated in traditional approaches like rarefaction and extrapolation procedures (Chao *et al.* 2014).

A number of mechanisms have been identified as responsible for explaining elevational diversity gradients and all the associated hypotheses are generally supported by empirical studies, suggesting that environment–richness relationships cannot be explained by a single factor (Sanders & Rahbek 2012; Ferger *et al.* 2014). Our modelling approach can be used to further investigate the myriad of mechanisms which shape spatial

variation in species richness through a GLM framework while accounting for imperfect detectability and rarity, widely known issues usually neglected in gradient analyses applied to local-scale datasets (e.g. Jankowski *et al.* 2013; Ferger *et al.* 2014). Moreover, our community model can be employed in more applied fields of ecology, such as testing guild-level responses to the configuration of conservation practices (Riffell *et al.* 2015).

To address such questions, we need to take into account that the model requires two levels of spatial replication, among sites within regions and among regions, but the model knows no scale and result interpretation should explicitly consider the spatial scale at which inference is made within and between regions. This is also related to the type and spatial resolution of the covariates used in the model.

It is an exciting time for community ecology, which has recently seen the advent of joint models of communities, namely joint species distribution models (JSDMs; Warton *et al.* 2015), a framework for the joint modelling of abundance across many taxa to quantify the effect of environmental predictors and interspecific interaction on abundance or distribution. Species associations are thus explicitly estimated by JSDMs, although they currently do not account for imperfect detection, whereas the multi-species occupancy framework, upon which our model is built, assumes independence between species (i.e. between the z_{ijr} elements of the r th incidence matrix) but explicitly accounts for imperfect detection. The two frameworks are, in fact, completely compatible (Beissinger *et al.* 2016), and integrating these approaches would provide a model that can be used to jointly investigate geographic variation in stratified communities (e.g. guild structure) and within-community species interactions.

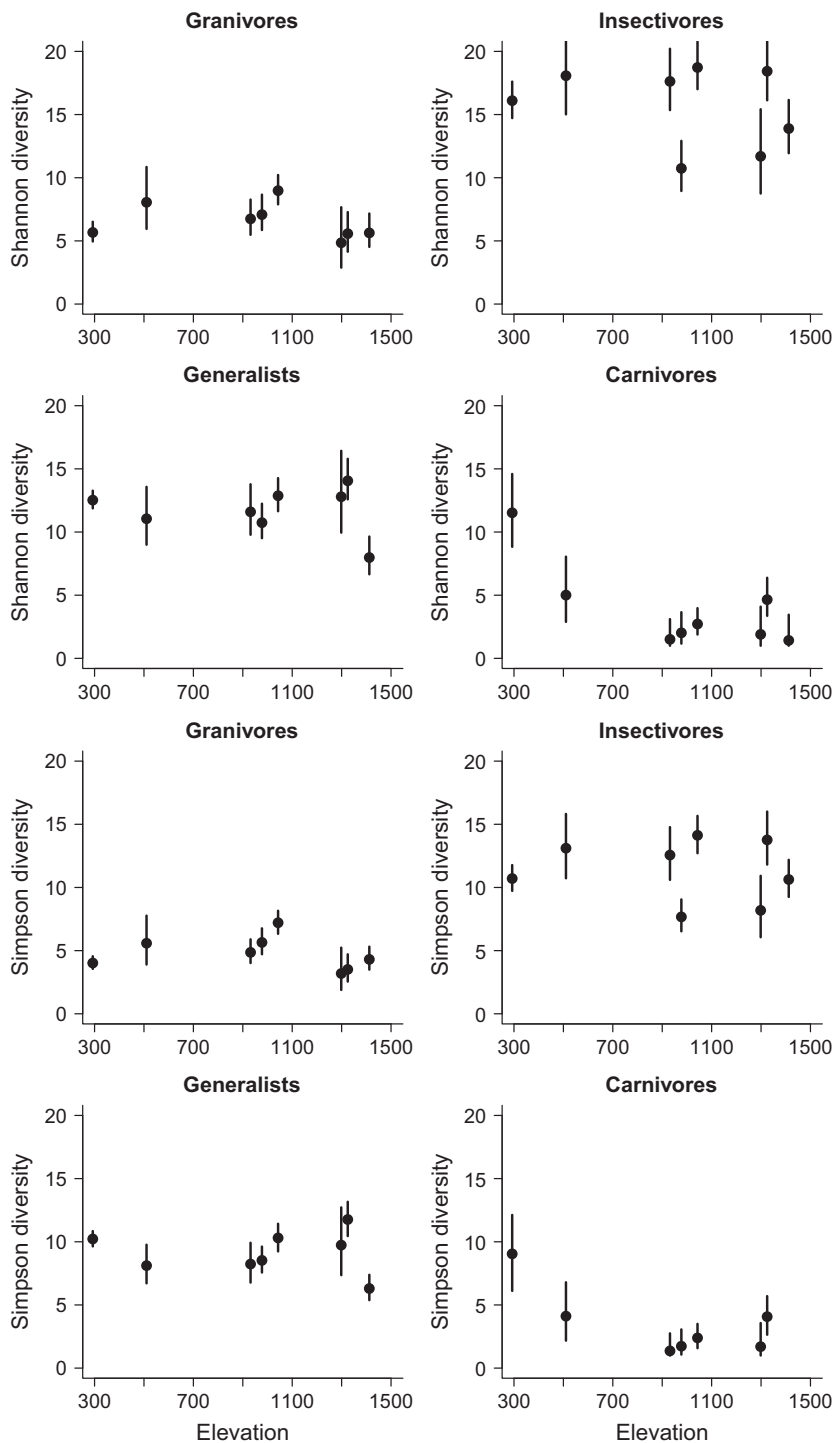


Fig. 3. Elevational gradient of guild-specific Shannon and Simpson diversity (mean and 95% Bayesian credible interval) for the avian communities of eight regions. Elevation is given in meters above sea level. Note that the Simpson diversity estimate could not be derived for carnivores in the region at the highest elevation (bottom-right panel) where no carnivore species were detected and estimated occupancies were very close to zero.

In a perspective of community ecology based on functional traits, McGill *et al.* (2006) outlined a series of questions to delineate and direct like ‘how much variation in traits and why’ or ‘how do environmental gradients affect community structuring’. Our modelling framework can be used to address these questions by examining quantitatively the trait-based community structure along with its variation across distinct regions. In this fashion, spatial variation in species traits can be tested through their hyperparameters,

such as the mean and variance of a continuous trait (e.g. body mass as in Ahumada *et al.* 2011) or the (functional) group membership probability (π_g), i.e. the proportion of species belonging to a specific guild (Bregman, Sekercioglu & Tobias 2014).

Additionally, in a view of metacommunity as a collection of sample locations considered to be ‘local’ communities (Dorazio *et al.* 2010), gradient analyses can be compared across multiple regions in a single framework that deals explicitly with

issues of imperfect detection, heterogeneity in detectability and heterogeneity in occurrence probabilities. Consistent trends in multiple regions would strongly support the idea of general processes that are robust to variation in biotic and abiotic site history (Fukami & Wardle 2005).

The multi-region model here presented, similarly to the community model in Sutherland *et al.* (2016), retains the regional level species-by-site data structure of classical (single-region) MSOMs, and thus all of the benefits and recent developments of the latter (reviewed in Iknayan *et al.* 2014) apply to the current framework. Extending the static model to a dynamic, multi-season formulation (e.g. Dorazio *et al.* 2010) would allow the study of spatio-temporal variation in trait-based colonization-extinction dynamics within or between communities, along with variation of species traits and community structure in relation to environmental gradients and time.

Authors' contributions

S.T. and C.S. conceived the ideas and designed methodology; M.B. and P.P. collected the data; S.T. and C.S. analysed the data and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Acknowledgements

We thank Marc Kéry and another anonymous referee for constructive and helpful comments that improved the manuscript. We also wish to thank Marco De Rosa, Silvano Paoli, Giuseppe Jurman, and Michele Filosi of the Fondazione Bruno Kessler for allowing and facilitating the use of the HPC Kore system. Finally, we are grateful to A. Franzoi, F. Rizzolli, F. Rossi, for the help with fieldwork, and to Aaron lemma for IT assistance. Field activity was partially funded through the project LIFE+ project 'LIFE11/IT/187 T.E.N. (Trentino Ecological Network): a focal point for a Pan-Alpine Ecological Network', by Servizio Agricoltura PAT and by 'Accordo di Programma per la Ricerca PAT-MUSE (2013-2014)'.

Data accessibility

R and BUGS script for running the model: included as Data S1. Input data for the BUGS model: Figshare <https://doi.org/10.6084/m9.figshare.4308941.v1>.

References

Ahumada, J.A., Silva, C.E., Gajapersad, K. *et al.* (2011) Community structure and diversity of tropical forest mammals: data from a global camera trap network. *Philosophical Transactions of the Royal Society B*, **366**, 2703–2711.

Bailey, L.L., MacKenzie, D.I. & Nichols, J.D. (2014) Advances and applications of occupancy models. *Methods in Ecology and Evolution*, **5**, 1269–1279.

Beissinger, S.R., Iknayan, K.J., Guillera-Aroita, G., Zipkin, E.F., Dorazio, R.M., Royle, J.A. & Kéry, M. (2016) Incorporating imperfect detection into joint models of communities: a response to Warton *et al.* *Trends in Ecology & Evolution*, **31**, 736–737.

Bregman, T.P., Sekercioglu, C.H. & Tobias, J.A. (2014) Global patterns and predictors of bird species responses to forest fragmentation: implications for ecosystem function and conservation. *Biological Conservation*, **169**, 372–383.

Broms, K.M., Hooten, M.B. & Fitzpatrick, R.M. (2014) Accounting for imperfect detection in hill numbers for biodiversity studies. *Methods in Ecology and Evolution*, **6**, 99–108.

Chandler, R.B., King, D.I., Raudales, R., Trubey, R., Chandler, C. & Arce Chávez, V.J. (2013) A small-scale land-sparing approach to conserving biological diversity in tropical agricultural landscapes. *Conservation Biology*, **27**, 785–795.

Chao, A., Gotelli, N.J., Hsieh, T., Sander, E.L., Ma, K., Colwell, R.K. & Ellison, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, **84**, 45–67.

Dorazio, R.M. & Royle, J.A. (2005) Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*, **100**, 389–398.

Dorazio, R.M., Royle, J.A., Söderström, B. & Glimskär, A. (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, **87**, 842–854.

Dorazio, R.M., Kéry, M., Royle, J.A. & Plattner, M. (2010) Models for inference in dynamic metacommunity systems. *Ecology*, **91**, 2466–2475.

Ferger, S.W., Schleuning, M., Hemp, A., Howell, K.M. & Böhning-Gaese, K. (2014) Food resources and vegetation structure mediate climatic effects on species richness of birds. *Global Ecology and Biogeography*, **23**, 541–549.

Fukami, T. & Wardle, D.A. (2005) Long-term ecological dynamics: reciprocal insights from natural and anthropogenic gradients. *Proceedings of the Royal Society B*, **272**, 2105–2115.

Garrard, G.E., McCarthy, M.A., Williams, N.S., Bekessy, S.A. & Wintle, B.A. (2013) A general model of detectability using species traits. *Methods in Ecology and Evolution*, **4**, 45–52.

Gotelli, N. & Colwell, R. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Grönroos, M. & Heino, J. (2012) Species richness at the guild level: effects of species pool and local environmental conditions on stream macroinvertebrate communities. *Journal of Animal Ecology*, **81**, 679–691.

Iknayan, K.J., Tingley, M.W., Furnas, B.J. & Beissinger, S.R. (2014) Detecting diversity: emerging methods to estimate species diversity. *Trends in Ecology and Evolution*, **29**, 1–10.

Jankowski, J.E., Merkord, C.L., Rios, W.F., Cabrera, K.G., Revilla, N.S. & Silman, M.R. (2013) The relationship of tropical bird communities to tree species composition and vegetation structure along an andean elevational gradient. *Journal of Biogeography*, **40**, 950–962.

Jost, L. (2006) Entropy and diversity. *Oikos*, **113**, 363–375.

Kissling, W.D., Rahbek, C. & Böhning-Gaese, K. (2007) Food plant diversity as broad-scale determinant of avian frugivore richness. *Proceedings of the Royal Society B*, **274**, 799–808.

Kissling, W.D., Sekercioglu, C.H. & Jetz, W. (2012) Bird dietary guild richness across latitudes, environments and biogeographic regions. *Global Ecology and Biogeography*, **21**, 328–340.

Marcus Rowcliffe, J., Carbone, C., Jansen, P.A., Kays, R. & Kranstauber, B. (2011) Quantifying the sensitivity of camera traps: an adapted distance sampling approach. *Methods in Ecology and Evolution*, **2**, 464–476.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006) Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution*, **21**, 178–185.

Nogués-Bravo, D., Araújo, M., Romdal, T. & Rahbek, C. (2008) Scale effects and human impact on the elevational species richness gradients. *Nature*, **453**, 216–219.

Plummer, M. (2003) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Technische Universität Wien, Vienna, Austria.

R Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Riffell, S.K., Monroe, A.P., Martin, J.A., Evans, K.O., Burger, L.W. & Smith, M.D. (2015) Response of non-grassland avian guilds to adjacent herbaceous field buffers: testing the configuration of targeted conservation practices in agricultural landscapes. *Journal of Applied Ecology*, **52**, 300–309.

Royle, J.A. & Converse, S.J. (2014) Hierarchical spatial capture-recapture models: modelling population density in stratified populations. *Methods in Ecology and Evolution*, **5**, 37–43.

Royle, J.A., Dorazio, R.M. & Link, W.A. (2007) Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, **16**, 67–85.

Royle, J.A., Converse, S.J. & Link, W.A. (2012) Data augmentation for hierarchical capture-recapture models. arXiv preprint arXiv:1211.5706.

Sanders, N.J. & Rahbek, C. (2012) The patterns and causes of elevational diversity gradients. *Ecography*, **35**, 1–3.

Schofield, M.R. & Barker, R.J. (2011) Full open population capture-recapture models with individual covariates. *Journal of Agricultural, Biological and Environmental Statistics*, **16**, 253–268.

Sutherland, C., Brambilla, M., Pedrini, P. & Tenan, S. (2016) A multiregion community model for inference about geographic variation in species richness. *Methods in Ecology and Evolution*, **7**, 783–791.

- Tobler, M.W., Zúñiga Hartley, A., Carrillo-Percegue, S.E. & Powell, G.V. (2015) Spatiotemporal hierarchical modelling of species richness and occupancy using camera trap data. *Journal of Applied Ecology*, **52**, 413–421.
- Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K. (2015) So many variables: joint modeling in community ecology. *Trends in Ecology & Evolution*, **30**, 766–779.
- Yamaura, Y., Andrew Royle, J., Kuboi, K., Tada, T., Ikeno, S. & Makino, S. (2011) Modelling community dynamics based on species-level abundance models from detection/nondetection data. *Journal of Applied Ecology*, **48**, 67–75.

Received 8 June 2016; accepted 6 December 2016
Handling Editor: Robert Freckleton

Supporting Information

Details of electronic Supporting Information are provided below.

Appendix S1. Modelling details.

Appendix S2. Summary of posterior parameter estimates from the simulation study and the application.

Data S1. R and JAGS code for data simulation and for the model fitted to bird point count data.