Non-stationarity in the co-occurrence patterns of species across environmental gradients

Avi Bar-Massada*,[1] and Jonathan Belmaker[2]

[1]Department of Biology and Environment, University of Haifa at Oranim, Kiryat Tivon 36006, Israel; and [2]Department of Zoology and The Steinhardt Museum of Natural History, Tel Aviv University, Tel Aviv 69978, Israel

Summary

1. Quantifying the role of biotic interactions in driving community assembly often relies on analysing species co-occurrence patterns, where segregated patterns are taken to indicate antagonistic interactions such as competition. It is unknown, however, if co-occurrence patterns are stationary across environmental gradients, as it is possible that the strength of biotic interactions that drive these patterns also depends on the environment. In this study, we aim to understand how patterns of co-occurrence change when species move from their environmental range centre towards their range periphery while isolating the potential signal of biotic interactions from confounding factors.

2. We used two separate statistical approaches (null models and joint species distribution models) to quantify pairwise co-occurrence patterns for tree species sampled in 9382 plots distributed across the conterminous US. We also analysed co-occurrence patterns that emerged from a simple meta-community model. We then assessed how patterns of species segregation and aggregation change in relation to habitat suitability while accounting for multiple factors known to confound co-occurrence analyses.

3. We found strong non-stationarity in co-occurrences, with patterns shifting from segregated at the environmental range centre towards aggregated at range margins for the majority of tree species. Patterns were in full agreement between model simulations and both empirical analyses. Model simulations suggest that this pattern is at least partly driven by variation in the relative abundances of non-focal species even when no direct biological interactions are present.

4. *Synthesis.* Patterns of tree species co-occurrence vary across environmental gradients, with increased segregation when environmental conditions are optimal and increased aggregation when the environment is less suitable. This pattern may originate from a trade-off between the abundance of the focal species pair, which decreases towards the environmental range margin, and the increasing abundance of non-focal species to which the environment is more suitable. The strong dependence of co-occurrence patterns on environmental conditions might limit the predictive ability of joint species distribution models, which couple species co-occurrences and their environmental responses, because co-occurrence patterns and environmental responses are confounded.

Key-words: biotic interactions, communities, competition, null model, species ranges, trees

Introduction

The importance of biotic interactions for understanding broad-scale distributional patterns is a major ecological question (Gotelli, Graves & Rahbek 2010; McGill 2010; Pigot & Tobias 2013; Afkhami, McIntyre & Strauss 2014; Araújo & Rozenfeld 2014; Belmaker *et al.* 2015). Apart from the theoretical interest in this question, understanding the role of biotic interactions has applied implications, as predicting the response of species to change (climate, land use) is increasingly reliant on species distribution models (SDM). These models account for environmental variability but typically ignore biotic interactions. If biotic interactions are indeed important, SDM may be missing an essential component (Gilman *et al.* 2010; Kissling *et al.* 2012; Zarnetske, Skelly & Urban 2012; Wisz *et al.* 2013; Godsoe *et al.* 2015).

Even when biotic interactions can be quantified accurately, their strength and even directionality may vary in space and time (Thompson 1988; Chamberlain, Bronstein & Rudgers 2014). Such non-stationarity in biotic interactions may substantially hamper the ability to use information on biotic interactions to improve predictions of distributional changes. Thus, identifying general patterns in the variation of the strength of biotic interactions is important for understanding to what extent, and where geographically, SDM may fail.

*Correspondence author. E-mail: barmassada@gmail.com

If the competitive ability of a species depends to some degree on the suitability of the environment it occupies (Barrio *et al.* 2013; Milazzo *et al.* 2013), we may expect high spatial segregation between competing species at the centre of their environmental (not necessarily geographical) range, and variable or aggregated patterns at their environmental range periphery. Alternatively, species may actually compete more severely when in unfavourable environments, where conditions are extreme and resources are limited. This is because species may be more sensitive to competition, and hence susceptible to local exclusion, if they are close to their environmental tolerance threshold. As an example, alpine plants were shown to be more severely impacted by competition from invasives under increased temperatures, which represent conditions closer to their environmental limits (Alexander, Diez & Levine 2015). Under these conditions, spatial segregation between competing species will be most common at their environmental range periphery.

Studies aimed at identifying variation in biotic interactions across space have typically inferred biotic interactions from spatial patterns of co-occurrence, usually using observational studies over many species. While experimental studies found evidence for directional variation in the strength of biotic interactions along environmental gradients (e.g., Callaway *et al.* 2002), patterns using co-occurrences are more equivocal (Dullinger *et al.* 2007; Maestre *et al.* 2009; López *et al.* 2013). Thus, we still do not have a general understanding of the extent to which co-occurrences patterns vary across environmental gradients.

Analyses of species co-occurrence are commonly based on null models that attempt to identify when species occurrences are aggregated, segregated or random (Gotelli 2000; Lyons *et al.* 2016). However, these analyses are confounded by several factors that can lead to non-random co-occurrence patterns even in the absence of interspecific interactions. These factors include similarities between species' environmental preferences (Bar-Massada 2015b; Royan *et al.* 2015), environmental heterogeneity (Heino & Grönroos 2013; Bar-Massada 2015b) and differences in geographic ranges (Connor, Collins & Simberloff 2013; Pollock *et al.* 2014). For example, increasing levels of niche overlap between tree species leads to more aggregated co-occurrence patterns, while increased environmental heterogeneity leads to more segregated patterns (Bar-Massada 2015b). Co-occurrence analyses should account for these confounding factors if one attempts to gain insight about relationships between co-occurrence patterns and biotic interactions. Relatively new statistical methods that address some of these issues are based on joint species distribution models (JSDM), which analyse species co-occurrences while accounting for their similarities in their environmental responses (Ovaskainen, Hottola & Siitonen 2010; Kissling *et al.* 2012; Pollock *et al.* 2014; Warton *et al.* 2015). However, even with these methods, it remains unclear if co-occurrences patterns remain stationary across environmental gradients.

In this study, we first analysed a large data base of North American trees to understand how the direction and strength of segregation and aggregation between species changes as species move from their environmental range centre towards their range periphery. By using field data composed of fine-grained local samples, we prevented the dilution of the potential signal of biotic interactions by spatial scale (i.e., the size of our forest samples was small enough to capture species interactions). In addition, we specifically accounted for the effects of potentially confounding variables on co-occurrence analyses (such as environmental heterogeneity and niche overlap between species) in an attempt to further distil the signal of biotic interactions. We then corroborated our empirical findings using a meta-community model. This allowed us to gain a deeper understanding of the mechanisms underlying both the empirical and modelled patterns. Together, the empirical and modelled analyses provide a general understanding of the expected patterns of change in pairwise co-occurrences across environmental gradients.

Materials and methods

OVERVIEW

We quantified pairwise species co-occurrence for tree species in the conterminous US, and then analysed the effects of habitat suitability (as a measure of location along an environmental gradient) on co-occurrence patterns, while accounting for multiple potential confounding variables (see Fig. 1 for a flow chart that depicts the main component of our analysis). Our current analysis builds upon a previous analysis of the same data (Bar-Massada 2015b) which focused solely on the effects of species' niches and environmental heterogeneity on co-occurrence patterns. To quantify co-occurrence patterns, we used the standard null model approach (Gotelli 2000; Lyons *et al.* 2016), which has been the most commonly utilized in recent decades. As an additional test of the relationship between co-occurrence and environmental conditions, which explicitly accounts for species-specific environmental preferences, we also analysed our data using JSDM (Pollock *et al.* 2014); however, due to lack of space, we describe that part of our study in the supplementary materials. In addition to the empirical analysis, we analysed the results of a simple meta-community model (Bar-Massada 2015a) to assess whether our empirical results reflect a more general relationship between habitat suitability and species co-occurrence patterns.

TREE DATA

We used data on forest communities obtained from the U.S. Forest Inventory and Analysis Program (FIA). Data were collected in 9382 sample plots, located in the conterminous US (plots were randomly selected from the much larger full FIA data set). Each FIA sample plot consists of four circular subplots of 7·32 m radius, in which the identity of each tree species with a diameter at breast height larger than 12·7 cm is recorded. Due to privacy issues, raw FIA data are supplied without plot locations and species identities. However, the GIS Spatial Data Services unit of the U.S. Forest Service supplied a distance matrix containing the geographic distances between each pair of plots, information on species presence in plots (denoted by unique IDs) and data on 12 environmental variables in each plot which were derived from ancillary data sets. These variables were: (i) *Climate variables*: mean annual temperatures and precipitation, and seasonality of temperature and precipitation (bio1, bio4, bio12 and bio15
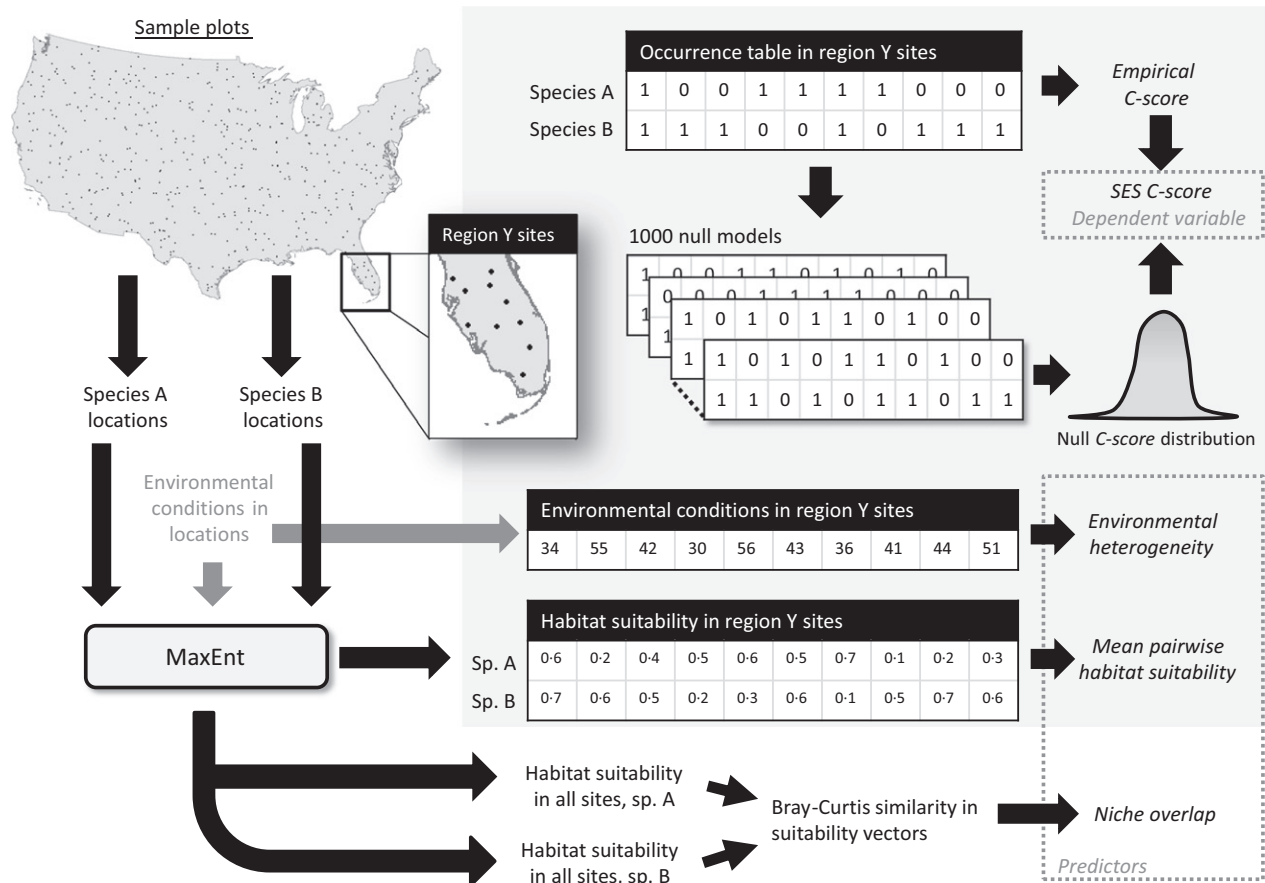
Fig. 1. A flow chart of the various stages of the empirical analysis. The shaded area of the figure marks analyses that were conducted at the regional scale; all other analyses were conducted at the continental scale. Blocks outlined by dashes highlight variables used in the statistical analysis. Due to insufficient space, the figure does not include the continental-scale process of estimating range similarity among species (Geo$_{mean}$). Numbers inside tables are arbitrary.

variables, obtained from the BioClim data base); (ii) *Topography variables*: elevation and elevation range within a 1-km cell size (obtained and subsequently processed from the U.S. National Atlas); (iii) *Coarse-scale soil data*: minimum/maximum values of water holding capacity, organic matter content, and soil permeability (obtained from the STATSGO data base, generated by the Soil Conservation Service, U.S. Department of Agriculture; available at http://water.usgs.gov/GIS/metadata/usgswrd/XML/ussoils.xml).

SPECIES CO-OCCURRENCE PATTERNS

We split the entire tree data set into 200 regions by running a cluster analysis on the site distance matrix using the Ward algorithm. We generated 200 regions to prevent community matrices from becoming too large, as this is known to lead to statistical problems in null model analyses (Fayle & Manica 2010; Gotelli & Ulrich 2012). The mean number of sites in each region was 46·3 (SD 18·6). To prevent biasing our analysis by species with very low levels of occurrence, we omitted all species with less than five occurrences in a given region. We then quantified pairwise species co-occurrence by calculating the normalized checkerboard score (C-score; Stone & Roberts 1990) for each species pair within each region.

To evaluate how these empirical C-scores deviated from those emerging by chance, we conducted null model analyses, in which we re-shuffled species occurrences across plots using a pre-defined randomization algorithm. We generated a null distribution of pairwise

species co-occurrences in each region, by running 1000 fixed-fixed null models using the trial-swap algorithm (Miklós & Podani 2004), with a thinning parameter of 10 000. Fixed-fixed null models shuffle species presences among sites in a constrained manner, which retains both species occupancies and species richness in sites, and therefore account for differences in habitat suitability across sites. After each null model run, we calculated the C-score for each species pair. Thus, at the end of the null model analysis, we had an entire distribution of null C-scores for each and every species pair in each region. To facilitate comparisons of co-occurrence patterns across species pairs and regions, we calculated the standardized effect size (SES) of the C-score by calculating the difference between the empirical C-score and the mean of its null distribution, and dividing the result by the standard deviation of the null distribution. Positive SES values denote segregated co-occurrence patterns, negative SES values denote aggregated co-occurrence patterns and SES values close to zero denote random co-occurrence patterns. We conducted all co-occurrence analyses in R (R Core Team 2013) using packages 'vegan' (Oksanen *et al.* 2013) and 'bipartite'(Dormann, Gruber & Fruend 2008).

ANALYSIS OF JOINT HABITAT SUITABILITY

To quantify habitat suitability for each species at each site, we generated a SDM based on species occurrences and the 12 environmental variables that are known to reflect the environmental requirements of tree species. For each species, we generated an SDM at the

continental scale, using all data points in our subset of the FIA data. For the SDM, we used Maximum Entropy (MaxEnt; Phillips, Anderson & Schapire 2006). As an output, MaxEnt generates a measure of relative habitat suitability ranging from 0 (unsuitable habitat) to 1 (highly suitable habitat) in any given site. To improve model performance and predictive capabilities, we restricted the analysis of habitat suitability to those species that appeared in 30 sites or more in the entire data set, which led to the generation of models for 155 tree species overall (of 305 species that are present in the data set). SDMs for all 155 species presented high cross-validated performance (AUC > 0·8). We then calculated, for each species in each region, the mean value of habitat suitability across sites as a measure of overall environmental suitability of a given region to a given species. We subsequently calculated the mean of this value for each pair of species within each region to yield a pairwise measure of habitat suitability at the regional scale (hereafter Hab_{mean}). A low Hab_{mean} corresponds with a pair of species that are found in a region that is further from the optimum of their environmental niche. Thus, a pair of species with similar environmental requirements will exhibit a gradient of increasing Hab_{mean} values from regions that are further from the optimum of their environmental niches to regions closer to their niche optimum.

QUANTIFYING FACTORS CONFOUNDING SPECIES CO-OCCURRENCE ANALYSES

Previous studies have shown that non-random species co-occurrence patterns might emerge simply due to the degree of similarity in species' niches, the amount of cohesion in their geographic distributions, and the environmental heterogeneity of the regions they inhabit (Connor, Collins & Simberloff 2013; Bar-Massada 2015b; Royan *et al.* 2015). We therefore quantified several potential confounding factors to test if they influenced our analysis.

We used the results of the SDM analysis to calculate the amount of similarity in habitat requirements for each species pair (that is, the level of agreement between their vectors of site suitability across all sites at the continental scale), as a proximate measure of realized niche overlap in terms of broad scale climatic, topographic and edaphic variables (the realized Grinellian niche; Soberón 2007). We followed Bar-Massada (2015b), who calculated continental-scale niche overlap, *D*, using the Bray–Curtis similarity metric.

To quantify environmental heterogeneity at the regional scale, we ran principal components analysis (PCA) on the same 12 environmental variables we used for modelling species distributions, after standardizing the variables to zero mean and unit variance. The first four principal components explained 79·61% of the variation in the environmental data (29·69%, 20·62%, 16·04% and 13·25% by the first four axes respectively). For each species pair, at any given geographic region, we calculated the volume of the convex hull encompassing the values of the first four principal components across all sites in which either or both species occurred. This serves as a species pair-specific measure of environmental heterogeneity at the regional scale, as it excludes sites in which neither species appeared. We calculated the convex hull using the R package 'geometry' (Habel *et al.* 2015).

In spatial analyses that relate species distributions to the environment, it is important to separate the true effect of the environment from patterns which may arise due to the cohesion of species ranges (e.g. Bahn & McGill 2007; Beale, Lennon & Gimona 2008). We thus performed spatial interpolation to estimate for each species an occurrence probability based on geographic proximity to other occupied sites but independent of environmental suitability. We then used the mean predicted probability of occurrence at the regional scale based on this geographic interpolation (Geo_{mean}) as a pairwise measure of species co-occurrence estimated from geographic proximity alone. A low Geo_{mean} corresponds to a pair of species that are likely to be found in different geographical regions, while a high value corresponds to species found in the same geographic region, independent of their environmental preference.

For the interpolation, we used Gaussian kernel density smoothers to calculate the density of occurrence for each site based on the geographic location of sites (Belmaker *et al.* 2015). We controlled for variation in sampling effort by dividing observed occurrence density by the density of sampling sites (Broennimann *et al.* 2012). This ensures that sparsely sampled areas do not receive exceptionally low densities. For each species, we randomly selected 80% of the sites to serve as training data, while the remaining 20% of the sites served as test data. We applied a kernel density smoother to the training data and generated predictions for the test data. We repeated this process 100 times and the mean predicted value for each site was used as interpolated occurrence probability for each species, based on geographic proximity to its other known occurrences.

MODELLING THE EFFECT OF HABITAT SUITABILITY ON CO-OCCURRENCE PATTERNS OF VIRTUAL SPECIES

To generalize our analysis beyond specific taxa and regional contexts, we used a meta-community model, based on the one presented in (Bar-Massada 2015a), to evaluate whether habitat suitability affects species co-occurrence patterns. This static and spatially implicit model reflects the process of habitat filtering and competition for space by different species in multiple sites (local communities) characterized by varying environmental conditions, denoted by a single environmental axis *E* bounded between 0 and 100. Species' site preference is denoted by a Gaussian niche along *E*, with an optimum μ (randomly assigned to each species) and a width σ. Initially, we used σ = 20 to enable species to establish across the entire range of *E* (albeit with varying levels of habitat suitability). Each local community hosts *J* individuals, and its environment is denoted by a single value within *E*, which is derived at random from a normal distribution with a mean of 50 and a standard deviation of 20. The abundance of different species in a given local community is determined by drawing *J* individuals from a multinomial distribution, with each species having a probability R_i to be drawn. R_i is denoted by:

$$R_i = \left[\frac{P_i}{\sum_{i=1}^{s} P_i} \right] \qquad \text{eqn 1}$$

where P_i is the suitability of site conditions for species *i*, obtained from the Gaussian relationship between its niche optimum μ, niche breadth σ, and the environmental conditions *E* in the site hosting the local community (Bar-Massada, Kent & Carmel 2014):

$$P(E) = \exp \left[-\frac{(\mu - E)^2}{2\sigma^2} \right] \qquad \text{eqn 2}$$

We used the model to generate 1000 meta-communities, each one comprising between 30 and 70 local communities. To analyse co-occurrence patterns, we calculated the C-score for a single pair of species selected at random from each meta-community. We generated 1000 null models similarly to the empirical analysis, and then calculated the SES for the C-score for the selected pair of species for a single meta-community. We repeated the entire analysis with one additional level of niche breadth (σ; 10) and two additional levels of local community sizes (*J*; 10 and 30) to assess whether our results

were robust to model parameters. As in the empirical analyses, we also quantified several potential confounding factors. Thus, for each species pair in every meta-community, we calculated a pairwise measure of environmental heterogeneity based on the range of E values in local communities that hosted at least one species from the corresponding pair (i.e., species A, species B or both). For each species pair, we also calculated the degree of fundamental niche overlap using the Bray–Curtis similarity metric between their corresponding vectors of habitat suitability across E, analogous to the method we used in the empirical analysis. Finally, for each species pair in every meta-community, we calculated the mean of pairwise habitat suitability (akin to Hab_{mean} in the empirical analysis) by averaging their vectors of habitat suitability P across all local communities.

STATISTICAL ANALYSIS

We analysed the relationship between environmental suitability and species co-occurrence patterns for each species pair separately as well as for all pairs combined. At the species pair level, we used linear models with the SES of C-score as the dependent variable and Hab_{mean}, Geo_{mean}, and environmental heterogeneity as independent variables. Niche overlap (D) was not included in this analyses as it does not vary within a species pair, but may explain patterns among species pairs (see below). We restricted this analysis to species pairs that had 10 or more observations (i.e., 10 regions). In each model, we tested if the effect size of Hab_{mean} was significant ($P < 0.05$), and if it was positive or negative. We then calculated the number of species pairs that had significantly positive or negative effects of Hab_{mean} on SES C-score and used a chi-square goodness-of-fit test to evaluate whether these deviated from the null expectation of equal numbers in both groups. As an additional test for the robustness of the effect size of Hab_{mean} on SES C-score, we calculated the confidence interval around its mean value across all species pairs (regardless of model significance), and checked if it did not encompass zero.

On top of the pair-level analysis, we also analysed all species pairs together. We used linear mixed-effect models to quantify the effects of Hab_{mean} on tree species co-occurrence patterns, while accounting for additional confounding variables that might affect the relationship. Additional fixed effects were niche overlap (D), environmental heterogeneity and Geo_{mean}. The model also included two independent random intercepts, region and species pair. Given the multitude of fixed effects, we undertook a model selection approach based on exhaustive search, using AIC as the selection criterion. We analysed the results twice, first using all species pairs, and then only for pairs exhibiting high levels of niche overlap. We defined high niche overlap as D values higher than the 90th percentile of the distribution of niche overlap (across all pairs), which corresponded to D values larger than 0·748. We fit all models using maximum likelihood, and assessed the significance of fixed effects using type II Wald's chi-square tests. We conducted the statistical analysis using the R packages 'car' (Fox & Weisberg 2011) and 'lme4' (Bates *et al.* 2015).

We repeated the analyses described above for the meta-community model data using linear models (as we had a single species pair per meta-community, there was no need for mixed-effect modelling). We developed linear models in which SES was the dependent variable, and mean habitat suitability, niche overlap, and environmental heterogeneity the independent variables. Using the same logic as in our empirical analysis, we developed two models, one with all levels of niche overlap, and another for overlap levels higher than 0·9 (which in the modelled data corresponded to the 90th percentile of niche overlap levels).

On top of these general linear models, we conducted an additional analysis in an attempt to gain better insight into the mechanism that

drives co-occurrence. We categorized each species pair in each modelled meta-community as close to their environmental range centre (top 25% of Hab_{mean}; hereafter 'range centre species'), close to their environmental range margin (bottom 25% of Hab_{mean}; hereafter 'range-margin species'), or other. For the range-centre and range-margin pairs in each meta-community, we then categorized each site as either an aggregated community (where both species occurred) or a segregated community (where only one species was present). For each group, we calculated mean pairwise habitat suitability which denotes habitat suitability in aggregated or segregated communities separately (in contrast to Hab_{mean}, which is a meta-community scale measure). We also calculated the mean abundance of species outside the focal pair (non-focal abundance) which quantifies the magnitude of space limitation in each community (as the number of sites available for establishment for the focal pair depends on the abundance of all other species). Unfortunately, we did not have abundance data for the FIA plots so we could not repeat this analysis in the empirical component of our study.

Results

ANALYSES OF TREE SPECIES ACROSS THE CONTERMINOUS US

Species pairs whose co-occurrence patterns were significantly affected by joint habitat suitability had significantly more positive than negative relationships between the standardized checkerboard score (SES C-score) and joint habitat suitability (Hab_{mean}), even after accounting for the confounding effects of environmental heterogeneity and geographic proximity (Geo_{mean}). Of the 495 species pairs that occurred in at least 10 regions, 39 (7·8%) exhibited a significant effect of Hab_{mean} on SES C-score. Of these, 31 effects were positive and 8 effects were negative (Fig. 2, $\chi^2(1) = 13\cdot56$, $P < 0\cdot001$ for a chi-square test comparing the number of observed negative and positive pairs to expectations based on equal numbers in both groups). This number increased to 39 positive and 6 negative effects when Hab_{mean} was used as a single predictor ($\chi^2(1) = 24\cdot2$, $P < 0\cdot001$). Furthermore, the mean effect size of Hab_{mean} on SES C-score for all species pairs that appeared in 10 regions or more (regardless of model significance) was 2·12, and the confidence interval ([1·46, 2·79]) did not encompass zero, implying a robust positive value of the mean effect size. Together, these results imply that increased levels of joint habitat suitability lead to more segregated co-occurrence patterns at the species pair level.

In the analysis of all species pairs combined, and when species pairs with all levels of niche overlap were considered, our model selection approach revealed that a model consisting of all fixed effects was best supported (Table S1, Supporting Information). In this model, joint habitat suitability (Hab_{mean}) had a significant positive effect on the SES of C-score ($\beta = 1\cdot33$ (SE 0·12), $\chi^2(1) = 109\cdot93$, $P < 0\cdot001$), implying that species are less aggregated and more segregated when facing more favourable environmental conditions. As expected, niche overlap (D) had a significant negative effect on SES ($\beta = -2\cdot09$ (SE 0·09), $\chi^2(1) = 532\cdot72$, $P < 0\cdot001$), and environmental heterogeneity had a significant positive effect on SES ($\beta = 0\cdot02$ (SE 0·004), $\chi^2(1) = 27\cdot71$,

$P < 0.001$). Range cohesion (Geo$_{mean}$) did not have a significant effect ($\chi^2(1) = 0.34$, $P = 0.55$).

In the analysis restricted to species pairs with high niche overlap, the best model contained only two fixed effects, joint
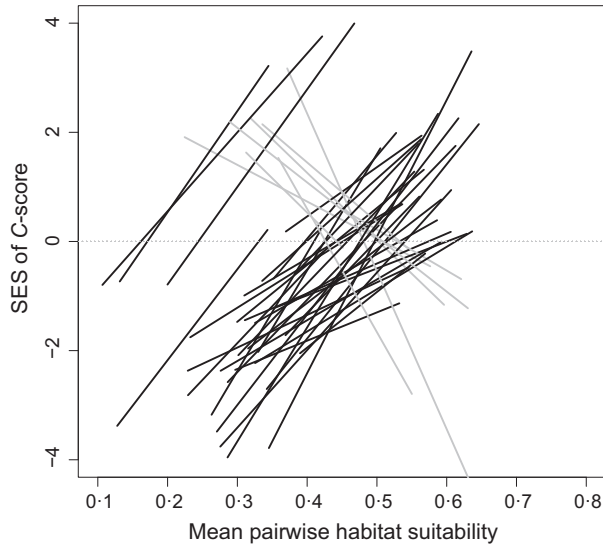
Fig. 2. Pairwise relationships between mean pairwise habitability (Hab$_{mean}$) and the standardized effect size (SES) of C-score (large negative values correspond with aggregated co-occurrence pattern and large positive values denote segregated co-occurrence patterns). Each solid line depicts the relationship for a single species pair that had a significant effect of Hab$_{mean}$ on SES in a linear model containing SES as the dependent variable, and Hab$_{mean}$, Geo$_{mean}$ and environmental heterogeneity as the independent variables. To predict the effect of Hab$_{mean}$ on SES C-score, we set Geo$_{mean}$ and environmental heterogeneity to their mean values across their observations for each species pair. Black lines highlight pairs that had significantly positive slope coefficients, and grey lines highlight pairs that had significantly negative slope coefficients.
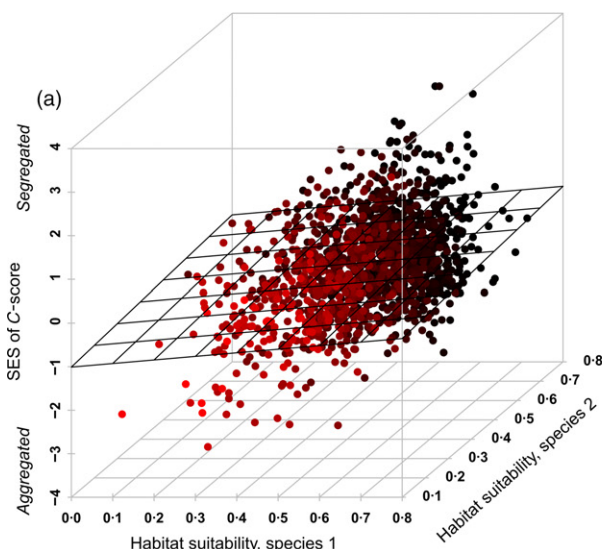
habitat suitability (Hab$_{mean}$) and range cohesion (Geo$_{mean}$). Joint habitat suitability had a significant positive effect on the SES of C-score ($\beta = 1.59$ (SE 0.42), $\chi^2(1) = 14.14$, $P < 0.001$), while range cohesion did not have a significant effect ($\chi^2(1) = 2.05$, $P = 0.15$). These two fixed effects, however, were moderately correlated (Spearman's $r = -0.79$). Just as in the models which contained all species pairs regardless of niche overlap, this analysis shows that higher habitat suitability leads to less aggregated and more segregated co-occurrence patterns.

To demonstrate that the effect of joint habitat suitability on co-occurrence patterns is not an artefact of averaging the individual suitability levels of both species, we plotted SES as a function of species' separate suitability levels (Fig. 3a). This figure reveals that indeed, suitability has a positive effect on SES for each species separately, but SES increases to its highest values (i.e., species are most segregated) when both species have high environmental suitability levels.

We found qualitatively similar results when we analysed the relationship between co-occurrence and joint habitat suitability using JSDM instead of the standard C-score and null model approach. Species occurrences tended to exhibit positive residual correlation, or *Rho*, closer to the range margin, and negative residual correlation closer to their range centre. As JSDMs inherently account for shared environmental responses between species (analogous to our measure of niche overlap), these analyses provide an independent confirmation to our main findings. The full results of the JSDM analysis are detailed in part B of the supplementary materials.

META-COMMUNITY MODEL RESULTS

Our meta-community models generated relationships between species co-occurrence patterns and mean habitat suitability that
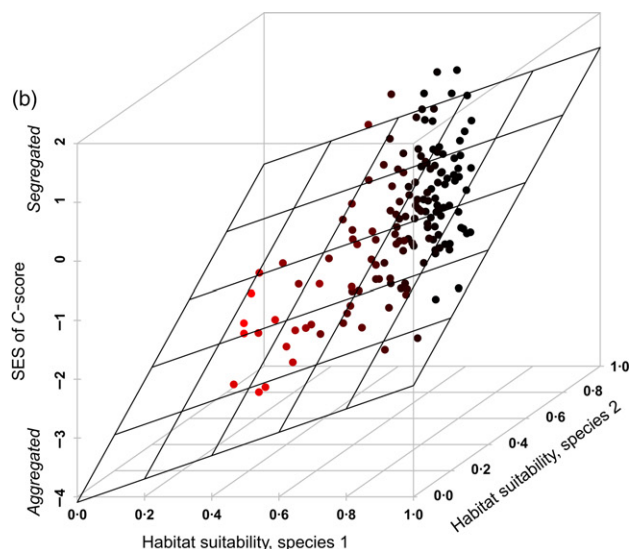
Fig. 3. Relationships between the standardized effect size (SES) of C-score and habitat suitability for species pairs with high levels of niche overlap ($D \geq 0.748$ in the empirical data, panel a; $D \geq 0.9$ in the meta-community model data, panel b). Each circle denotes a separate species pair in a given region (or meta-community in the modelled data). The suitability of each species in a given pair is represented by a separate axis. The black gridded surface, added for illustration purposes, depicts a linear model fit to the relationship between the SES and both habitat suitability levels. [Colour figure can be viewed at wileyonlinelibrary.com]

were remarkably similar to those we found in the empirical analysis. Hab_{mean} had a consistent positive effect on SES of C-score regardless of the level of niche overlap between species, and the direction of this effect was robust to changes in community size (J) and niche width (see Table S2). Furthermore, when we restricted our analysis to species pairs with very high levels of niche overlap ($D \geq 0.9$), Hab_{mean} was the only significant predictor of SES (Table S2). This result is also consistent with the empirical result. When we plotted SES as a function of species' separate suitability levels, the resulting patterns (Fig. 3b) were very similar to what we had found in the empirical analysis of forest communities (Fig. 3a).

When we analysed the results separately for range-centre and range-margin species pairs, in aggregated vs. segregated communities, we found contrasting results that might point to a statistical explanation to the relationship between co-occurrence and habitat suitability. Both range-centre and range-margin species (Fig. 4c and d, respectively) tended to be segregated when there was less space available to them (as non-focal species abundance in segregated communities was higher than in aggregated communities). In addition, range-centre species pairs had similar levels of habitat suitability in both segregated and aggregated communities (Fig. 4a). In contrast, range-margin species pairs had higher levels of habitat suitability in aggregated communities compared to segregated communities (Fig. 4b).

Discussion

We found strong and predictable non-stationarity in co-occurrences between species pairs, with patterns shifting from more segregation at the environmental range centre to more aggregation at range margins. In a separate analysis using JSDMs, we found qualitatively similar results, in which species occurrences were positively correlated at their range margin, and negatively correlated at their range centres. The agreement between the results of our model simulations and two methods of empirical data analyses suggests this is a general pattern that is independent of the particular taxa studied or the particularities of the simulations.

The strong and predictable non-stationarity of pairwise species co-occurrence patterns suggests that estimating fixed spatial or temporal associations between species from co-occurrence data may lead to inaccurate conclusions about their underlying spatial association if data were collected along an environmental gradient. For example, observations in the range centre may prove that two species are strongly segregated and hence presumably fierce competitors. However, by simply studying co-occurrence at another region in which habitat suitability for these species is lower, we may expect the pattern to change substantially towards more aggregated patterns. JSDMs and similar joint statistical models incorporate associations estimated from co-occurrences to improve species distribution modelling (Ovaskainen, Hottola & Siitonen 2010; Kissling *et al.* 2012; Pollock *et al.* 2014; Warton *et al.* 2015). Thus, patterns of co-occurrence are used to predict the current or future distributions of species taking into account both environment and presumed pairwise interspecific associations. In the light of our results, it is possible that these models may lead to inaccurate conclusions, as co-occurrence patterns can be confounded by habitat suitability, and hence such models will not lead to improved distributional predictions.
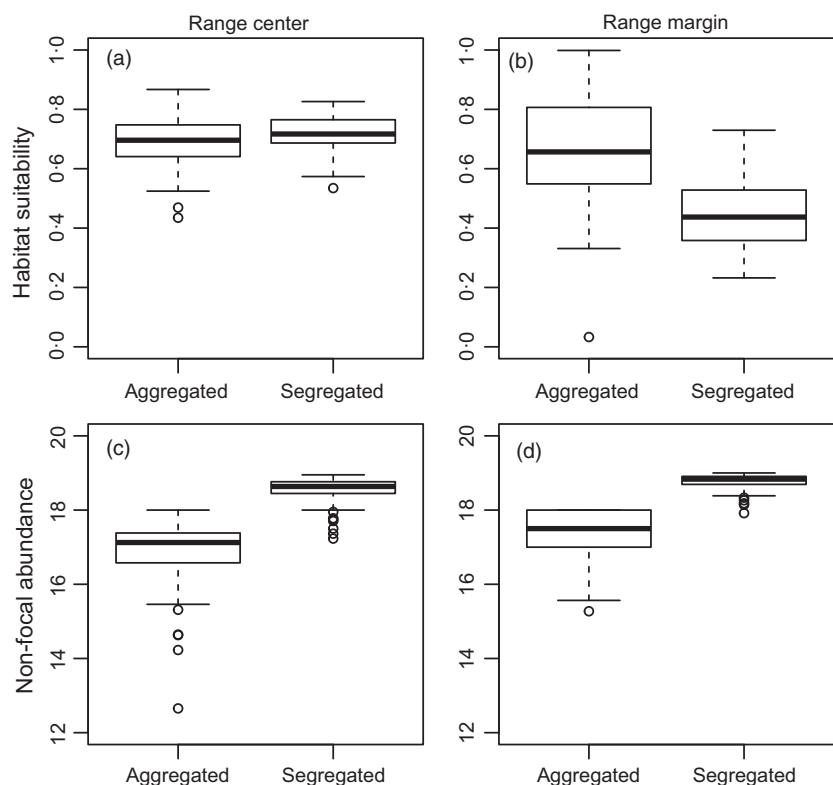
Fig. 4. Mean pairwise habitat suitability (top row) and non-focal abundance (bottom row) for species in modelled communities (niche overlap $D \geq 0.9$), in which they exhibit different co-occurrence patterns (aggregated: both species occurring; segregated: one of the two species occurring). Panels (a, c) depict species pairs at their environmental range centre (Hab_{mean} in the top 25%), whereas panels (b, d) depict species pairs at their environmental range margin (Hab_{mean} in the bottom 25%).

This study is unique as it explicitly examined several confounding factors that are often ignored (Bar-Massada 2015b), but may substantially bias the results of co-occurrence analyses. For example, species co-occurrence analyses typically conflate species interactions with environmental filtering, making separating the two difficult. Thus, the patterns observed could be formed by species interactions being weaker or by the range of environmental conditions occupied being larger at environmental range margins. By explicitly accounting for habitat heterogeneity, we were able to show that habitat suitability remains a prime predictor of the SES of C-score even after accounting for variation in environmental heterogeneity. In addition, geography is often mistaken for ecology (Bahn & McGill 2007; Warren *et al.* 2014). Indeed, we find that difference in the geographical range of species (Geo_{mean}) influences co-occurrence patterns. However, this pattern was relatively weak and the influence of environmental affinities (Hab_{mean}) remained strong even when geographical range overlap was accounted for.

If the very strong variation in the strength, and even the sign, of species co-occurrence patterns indeed bears the signal of biotic interactions, this would suggest that species compete more strongly at their environmental range centre than at their range margins. However, our simulation models did not explicitly introduce interspecific interactions. Why then did we observe patterns of segregation in environmental range centres that are in accordance with interspecific competition? We suggest that the mechanism producing this pattern is zero-sum competition for space. Given that the space available for growth is limited, both in the model and in real tree communities, an individual of a particular species being established will retard the establishment of individuals from other species. But why then does the pairwise strength of competition for space shift between environmental range centres and margins? We propose that at the range-centre species tend to be highly abundant, as well as being better suited to the local environment. In these cases, high abundance of a given species at any particular site will, by statistical necessity of the zero-sum condition, lead to low abundances of the other species within the pair. This will increase the prevalence of local extinctions, which will be manifested by more segregated co-occurrence patterns (Fig. 4c).

As species will be less abundant at their environmental range margins, the amount of area that is available for growth will depend less on the abundance of the other focal species and more on the abundance of the non-focal species. This is because at the environmental range margins, the non-focal species are more suitable and hence common in the meta-community. Hence, less suitable species will be forced to cling to the limited number of sites where their suitability is relatively high. Indeed, our analyses show that unsuitable species pairs tend to aggregate in sites that are much more suitable for them compared to the entire meta-community (Fig. 4b). In these sites, the relative abundances of superior species are low (Fig. 4d), and hence there is more room for both focal species, resulting in an aggregated pattern. As this process depends only on a constraint on total abundance along with species-specific gradients of habitat suitability, we suggest it may be a very general ecological pattern.

In the current debate about the utility of pairwise vs. community-wise approaches for understanding species interactions (Veech 2014; Soberón 2015; Arita 2016), our results suggest that pairwise co-occurrence patterns are tightly linked to community scale patterns. Pairwise analyses, therefore, should incorporate information about the abundance distribution of the entire community when these data are available. Understanding how to couple the pairwise and community-wise approaches for species co-occurrence analysis, in order to more accurately estimate biotic interactions, should become a critical focus of future research.

Our results suggest that studies on biotic interactions within communities should not focus solely on pairwise species co-occurrences, but in addition explicitly account for the distribution of habitat suitability and relative abundance within the community as a whole. As community ecologists are beginning to embrace joint models (models that simultaneously account for correlations among species and their environmental responses; Warton *et al.* 2015), we highlight the need to better understand how correlations among species change along environmental gradients. Accounting for this effect might greatly improve our ability to model and predict changes in ecological communities under environmental change.

Authors' contribution

A.B. and J.B. conceived the ideas and designed methodology; A.B. collected the data; A.B. and J.B. analysed the data and wrote the manuscript. Both authors contributed critically to the drafts and gave final approval for publication.

Acknowledgements

Data accessibility

Plot-level FIA data are available upon request from the Spatial Data Services (SDS) unit of the US Forest Service. To make a data request, visit: http://www.fia.fs.fed.us/tools-data/spatial/.

References

Afkhami, M.E., McIntyre, P.J. & Strauss, S.Y. (2014) Mutualist-mediated effects on species' range limits across large geographic scales. *Ecology Letters*, **17**, 1265–1273.

Alexander, J.M., Diez, J.M. & Levine, J.M. (2015) Novel competitors shape species' responses to climate change. *Nature*, **525**, 515–518.

Araújo, M.B. & Rozenfeld, A. (2014) The geographic scaling of biotic interactions. *Ecography*, **37**, 406–415.

Arita, H.T. (2016) Species co-occurrence analysis: pairwise versus matrix-level approaches. *Global Ecology and Biogeography*, **25**, 1397–1400.

Bahn, V. & McGill, B.J. (2007) Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, **16**, 733–742.

Bar-Massada, A. (2015a) Immigration rates and species niche characteristics affect the relationship between species richness and habitat heterogeneity in modeled meta-communities. *PeerJ*, **3**, e832.

Bar-Massada, A. (2015b) Complex relationships between species niches and environmental heterogeneity affect species co-occurrence patterns in modelled and real communities. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20150927.

Bar-Massada, A., Kent, R. & Carmel, Y. (2014) Environmental heterogeneity affects the location of modeled communities along the niche-neutrality continuum. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20133249.

Barrio, I.C., Hik, D.S., Bueno, C.G. & Cahill, J.F. (2013) Extending the stress-gradient hypothesis – Is competition among animals less common in harsh environments? *Oikos*, **122**, 516–523.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015) Fitting Linear mixed-effects models using lme4. *Journal of Statistical Software*. **67**, 1–48.

Beale, C.M., Lennon, J.J. & Gimona, A. (2008) Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences of the USA*, **105**, 14908–14912.

Belmaker, J., Zarnetske, P., Tuanmu, M.-N., Zonneveld, S., Record, S., Strecker, A. & Beaudrot, L. (2015) Empirical evidence for the scale dependence of biotic interactions. *Global Ecology and Biogeography*, **24**, 750–761.

Broennimann, O., Fitzpatrick, M.C., Pearman, P.B. *et al.* (2012) Measuring ecological niche overlap from occurrence and spatial environmental data: measuring niche overlap. *Global Ecology and Biogeography*, **21**, 481–497.

Callaway, R.M., Brooker, R.W., Choler, P. *et al.* (2002) Positive interactions among alpine plants increase with stress. *Nature*, **417**, 844–848.

Chamberlain, S.A., Bronstein, J.L. & Rudgers, J.A. (2014) How context dependent are species interactions? *Ecology Letters*, **17**, 881–890.

Connor, E.F., Collins, M.D. & Simberloff, D. (2013) The checkered history of checkerboard distributions. *Ecology*, **94**, 2403–2414.

Dormann, C.F., Gruber, B. & Fruend, J. (2008) Introducing the bipartite package: analysing ecological networks. *R News*, **8**, 8–11.

Dullinger, S., Kleinbauer, I., Pauli, H. *et al.* (2007) Weak and variable relationships between environmental severity and small-scale co-occurrence in alpine plant communities. *Journal of Ecology*, **95**, 1284–1295.

Fayle, T.M. & Manica, A. (2010) Reducing over-reporting of deterministic co-occurrence patterns in biotic communities. *Ecological Modelling*, **221**, 2237–2242.

Fox, J. & Weisberg, S. (2011) *An R Companion to Applied Regression*, 2nd edn. Sage, Thousand Oaks, CA, USA.

Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010) A framework for community interactions under climate change. *Trends in Ecology & Evolution*, **25**, 325–331.

Godsoe, W., Murray, R., Plank, M.J., Maron, A.E.J.L. & Kalisz, E.S. (2015) Information on biotic interactions improves transferability of distribution models. *The American Naturalist*, **185**, 281–290.

Gotelli, N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology*, **81**, 2606–2621.

Gotelli, N.J., Graves, G.R. & Rahbek, C. (2010) Macroecological signals of species interactions in the Danish avifauna. *Proceedings of the National Academy of Sciences of the USA*, **107**, 5030–5035.

Gotelli, N.J. & Ulrich, W. (2012) Statistical challenges in null model analysis. *Oikos*, **121**, 171–180.

Habel, K., Grasman, R., Gramacy, R.B., Stahel, A. & Sterratt, D.C. (2015) *Geometry: Mesh Generation and Surface Tesselation*. R package version 0.3-6. https://cran.r-project.org/web/packages/geometry/index.html [accessed 18 December 2016].

Heino, J. & Grönroos, M. (2013) Does environmental heterogeneity affect species co-occurrence in ecological guilds across stream macroinvertebrate metacommunities? *Ecography*, **36**, 926–936.

Kissling, W.D., Dormann, C.F., Groeneveld, J. *et al.* (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

López, R.P., Valdivia, S., Rivera, M.L. & Rios, R.S. (2013) Co-occurrence patterns along a regional aridity gradient of the subtropical andes do not support stress gradient hypotheses (ed. F. de Bello). *PLoS ONE*, **8**, e58518.

Lyons, S.K., Amatangelo, K.L., Behrensmeyer, A.K. *et al.* (2016) Holocene shifts in the assembly of plant and animal communities implicate human impacts. *Nature*, **529**, 80–83.

Maestre, F.T., Martínez, I., Escolar, C. & Escudero, A. (2009) On the relationship between abiotic stress and co-occurrence patterns: an assessment at the community level using soil lichen communities and multiple stress gradients. *Oikos*, **118**, 1015–1022.

McGill, B.J. (2010) Matters of Scale. *Science*, **328**, 575–576.

Miklós, I. & Podani, J. (2004) Randomization of presence-absence matrices: comments and new algorithms. *Ecology*, **85**, 86–92.

Milazzo, M., Mirto, S., Domenici, P. & Gristina, M. (2013) Climate change exacerbates interspecific interactions in sympatric coastal fishes (ed M Genner). *Journal of Animal Ecology*, **82**, 468–477.

Oksanen, J., Blanchet, F.G., Kindt, R. *et al.* (2013) *Vegan: Community Ecology Package*. R package version 2.3-5. https://cran.r-project.org/web/packages/vegan/index.html [accessed 18 December 2016].

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

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pigot, A.L. & Tobias, J.A. (2013) Species interactions constrain geographic range expansion over evolutionary time. *Ecology Letters*, **16**, 330–338.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. (2014) Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, **5**, 397–406.

R Core Team. (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Royan, A., Reynolds, S.J., Hannah, D.M., Prudhomme, C., Noble, D.G. & Sadler, J.P. (2015) Shared environmental responses drive co-occurrence patterns in river bird communities. *Ecography*, **38**, 1–10.

Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.

Soberón, J. (2015) Pairwise versus presence-absence approaches for analysing biodiversity patterns. *Journal of Biogeography*, **42**, 807–808.

Stone, L. & Roberts, A. (1990) The checkerboard score and species distributions. *Oecologia*, **85**, 74–79.

Thompson, J.N. (1988) Variation in interspecific interactions. *Annual Review of Ecology and Systematics*, **19**, 65–87.

Veech, J.A. (2014) The pairwise approach to analysing species co-occurrence. *Journal of Biogeography*, **41**, 1029–1035.

Warren, D.L., Cardillo, M., Rosauer, D.F. & Bolnick, D.I. (2014) Mistaking geography for biology: inferring processes from species distributions. *Trends in Ecology & Evolution*, **29**, 572–580.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K.C. (2015) So many variables: joint modeling in community ecology. *Trends in Ecology & Evolution*, **30**, 766–779.

Wisz, M.S., Pottier, J., Kissling, W.D. *et al.* (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews of the Cambridge Philosophical Society*, **88**, 15–30.

Zarnetske, P.L., Skelly, D.K. & Urban, M.C. (2012) Biotic multipliers of climate change. *Science*, **336**, 1516–1518.

Supporting Information

Details of electronic Supporting Information are provided below.

Appendix S1. Additional results and analysis of species co-occurrence using Joint Species Distribution Models (JSDM).