Complex relationships between species niches and environmental heterogeneity affect species co-occurrence patterns in modelled and real communities

Avi Bar-Massada

Department of Biology and Environment, University of Haifa, Kinyat Tfon 36006, Israel

Species co-occurrence analysis is commonly used to assess how interspecific interactions dictate community assembly. Non-random co-occurrences, however, may also emerge from niche differences as well as environmental heterogeneity. The relationships between species co-occurrence patterns, environmental heterogeneity and species niches are not fully understood, due to complex interactions among them. To analyse the relationships among these patterns and processes, I developed synthetic community models and analysed a large dataset of tree species across the conterminous United States. Niche overlap and environmental heterogeneity had significant and contrasting effects on species co-occurrence patterns, in both modelled and real communities. Niche breadth, in turn, affected the effect sizes of both variables on species co-occurrence patterns. The effect of niche breadth on the relationship between co-occurrence and niche overlap was markedly consistent between modelled and real communities, while its effect on the relationship between co-occurrence and environmental heterogeneity was mostly consistent between real and modelled data. The results of this analysis highlight the complex and interactive effects of species niche overlap, niche breadth and environmental heterogeneity on species co-occurrence patterns. Therefore, inferring ecological processes from co-occurrence patterns without accounting for these fundamental characteristics of species and environments may lead to biased conclusions.

1. Background

The question of which ecological processes drive species assembly in communities has been a major focus of ecological research in the past decades. Specifically, it has been hypothesized that competitive, facilitative or neutral interactions among species play varying roles in affecting community composition [1–3]. Analysis of patterns of species co-occurrence in real communities compared with those generated by null models has been a focal tool in the attempt to disentangle the opposite roles of interspecific interactions in determining species assembly [4–9]. While the early debate between Diamond [1] and Connor & Simberloff [10] on the utility of using such approaches has mostly subsided [11], there remain considerable methodological, statistical and theoretical challenges in the attempt to link species co-occurrence patterns to the true ecological processes behind them [11–14].

An inherent problem in co-occurrence analysis is that it ignores species resource requirements, and consequently cannot account for co-occurrence patterns (or the lack of them) that arise from niche differentiation among species. Species segregation can simply be an outcome of non-overlapping habitat requirements between species, which leads to differences in geographical distributions; therefore, significant segregation patterns can emerge with little or no interspecific interactions between species [13]. In the same manner, species can exhibit aggregation patterns simply due to similarity in habitat requirements within heterogeneous landscapes (with little or no interspecific interactions), or due to facilitative or mutualistic interactions [15–17]. Unfortunately, it is impossible to disentangle the roles of these unique processes by identifying...
Influence species co-occurrence patterns. Several recent studies attempted to devise statistical frameworks to disentangle the roles of habitat requirements from interspecific interactions in driving species co-occurrence patterns. Ovaskainen et al. [13] developed a multivariate logistic regression approach to incorporate species-specific habitat requirements into co-occurrence analysis, and tested their approach on wood-decaying fungal species in forest sites. They found that species aggregation patterns were mostly due to similarities in habitat requirements, without any evidence of interspecific interactions among species. Similarly, Pollock et al. [19] developed a joint species distribution model (SDM) based on a hierarchical multivariate probit regression model [20]. They analysed the relative roles of shared environmental responses and other ecological processes in determining species co-occurrences, according to the residual correlations in co-occurrence patterns and environmental correlations in the model, respectively. For frogs and eucalypts in Australia, they found that shared environmental responses were the main driver of species distributions, though there was some evidence of interspecific interactions. Finally, Wittman et al. [21] studied the role of thermal tolerance on ant community structure in the Siskiyou Mountains, USA. They found that species pairs with increased thermal niche overlap tended to be more spatially segregated compared with species pairs with little niche overlap.

Beyond the effect of niche overlap per se, niche breadth may also be related to species co-occurrence patterns, as it is known to affect species sorting into real-world communities, I analysed a large empirical dataset of tree species across the entire conterminous United States. I argue, however, that analysing the effect of heterogeneity on species co-occurrence patterns by facilitating spatial segregation of species within heterogeneous habitats. For example, a meta-community of stream macroinvertebrates exhibited significantly more segregated co-occurrence patterns across streams than within streams, and this pattern corresponded to differences between heterogeneity levels across these scales [26]. Another study on stream insects found that in streams with similar abiotic conditions (i.e., homogeneous patterns) species exhibited strong aggregated patterns, while strong segregated patterns were found in communities occurring across streams with different abiotic conditions [27]. Positive relationships between segregated co-occurrence patterns and resource heterogeneity were also found for weed species in agricultural fields [28] and natural grassland communities [29].

How would environmental heterogeneity, species niche overlap and species niche breadth interact to affect species co-occurrence patterns? The complex interaction among all these separate patterns and processes is difficult to envision. To quantify this relationship, I used two approaches. First, I developed a simple community model to analyse this relationship under controlled settings. Secondly, to evaluate the relationship in real-world communities, I analysed a large empirical dataset of tree species across the entire conterminous United States.

2. Material and methods

(a) Tests with synthetic species pairs

To test the effects of species’ niche similarity on their co-occurrence patterns at different levels of niche breadth and environmental heterogeneity, I developed a community model in R [30] (code available in the electronic supplementary material), based on a simplified version of a previously published model [31]. I generated, at random, communities of 10–100 species, consisting of 30–100 sites. In the initial set of simulations, each site hosts 30 individuals and has an environmental condition E, which is derived at random from a normal distribution with a mean of 50 and a standard deviation of 20. Environmental heterogeneity is defined by two variables: the range of E values and the coefficient of variation of E values across sites. Both measures of heterogeneity are based on the same environmental axis (E) that is used to quantify niche overlap and niche breadth, and thus they are not independent. I argue, however, that analysing the effect of heterogeneity on ecological communities is only informative when conducted this way [32], as the predominant ecological dogma on the importance of heterogeneity relies upon its interaction with species niches (i.e., more heterogeneous habitats comprise more niches); therefore, environmental heterogeneity should ideally be quantified in the same niche axes that determine species occurrence. Both measures of heterogeneity are calculated for each species pair separately, based only on sites that host at least one species from the corresponding pair (i.e., species A, species B, or both). Each species has a Gaussian niche with a mean (μ), chosen at random from a uniform distribution bounded by the range of environmental conditions, and a breadth (σ) between 10 and 35. The range of niche breadths
denotes a gradient from specialists to generalist species. In a
given simulation, all species have the same niche breadth, but
may have different niche optima. The suitability $P$ of a site
with environment $E$ to a given species is denoted by [33,34]:

$$P(E) = \exp \left( -\frac{(\mu - E)^2}{2\sigma^2} \right). \quad (2.1)$$

The degree of niche overlap between two species, $D$, is calculated
using the Bray–Curtis similarity metric:

$$D = 1 - \frac{\sum_{i} \min[P_1(E), P_2(E)]}{\sum_{i} \max[P_1(E), P_2(E)]}, \quad (2.2)$$

where $P_1(E)$ and $P_2(E)$ denote the suitability of environmental
condition $E$ to species 1 and species 2, respectively. $D$ ranges
from zero (no overlap) to one (complete overlap).

In each site, I used a multinomial model of species filtering to
determine the number of individuals of each species present in
the site [31]. The model is static and represents a process of filter-
ing from a regional species pool where all species have the same
abundance. The number of individuals of species $i$ that establish
in a given site is drawn from a multinomial distribution with a
probability $R_i$ and a sample size that equals the size of the site
(i.e. the number of individuals it can host, regardless of species).

The probability of colonization $R_i$ is denoted by

$$R_i = \frac{P_i}{\sum_j P_j}, \quad (2.3)$$

where $P_i$ is the suitability of site conditions for species $i$ (equation
(2.1)). Thus, the likelihood of occurrence for a species in a given site
depends only on the suitability of site conditions to the species’ niche requirement relative to their suitability to other species. However, as the number of individuals per site is finite (initially 30), species engage in interspecific competition for
space. To assess the effect of competition for space on the results,
I repeated the entire analysis four more times with different size
samples (10, 20, 40 and 50). I analysed co-occurrence patterns for
a single pair of species (hereafter the ‘focal pair’), consisting of the
third and fourth species in the community table, to avoid biasing the results by over-accounting for the environmental con-
ditions in communities with large numbers of species pairs. As
species do not interact in an explicit manner and have arbitrary
niche requirements, it does not matter which pair of species is
being analysed for co-occurrence patterns.

To analyse the effect of niche breadth on the relationship
between niche overlap, environmental heterogeneity and species
co-occurrence patterns, I conducted separate analyses for com-

tetries comprising species with different niche breadths ($\sigma$),
including each value from 10 to 35, inclusive.

Overall, I generated 1000 communities at each level of niche
breadth, which I subsequently analysed using null models to
determine whether they exhibited significant co-occurrence pat-
terns. For the focal species pair, I quantified co-occurrence
using the C-score metric of Stone & Roberts [6].

$$C_{-score} = (N_1 - N_{12})(N_2 - N_{12}), \quad (2.4)$$

where $N_1$ and $N_2$ are the number of exclusive occurrences
of species 1 and 2, respectively, and $N_{12}$ is the number of their
co-occurrences.

I generated 1000 null models for each community, using the
trial-swap algorithm [35], with a thinning parameter of 10,000.
The trial-swap algorithm generates fixed–fixed null models, which
retain both overall species abundances and species richness
levels in sites. I calculated C-score for each null model separately, and the mean and the standard deviation of the dis-
bution of null models for the focal species pair. Based on these
and on the observed C-score of the focal species pair, I calculated
the standardized effect size (SES) to quantify the number of
standard deviations by which the observed C-score differs from
the mean of the null distribution of C-scores. Positive SES
values denote segregated co-occurrence patterns, whereas nega-
tive SES values denote aggregated co-occurrence patterns, and
SES values close to zero denote random co-occurrence patterns.

Given that the SES of a truly random community has a standard
normal distribution, SES values above 1.96 (and below $-1.96$)
denote significantly segregated (and aggregated) species pairs,
respectively, which correspond with the top (and bottom) 2.5%
of the distribution around the null expectation. I conducted all
co-occurrence analyses in R [30] using packages ‘vegan’ [36] and
‘bi派p’ [37].

(b) Analysis of empirical co-occurrence patterns in
forest communities

To evaluate empirical relationships among co-occurrence patterns
of tree species, niche overlap, niche breadth and environmental
heterogeneity, I used a large database of tree species occurrence in
sample plots from the US Forest Inventory and Analysis Program (FIA).
The dataset comprised data from 10,000 permanent

clusters of plots (a randomly selected subset from the native
FIA database), each plot consisting of data aggregated from four
circular sub-plots of 7.32 m radius, in which the identity of each
tree species with a diameter at breast height larger than 5" was
noted. Sample plots were located within the entire conterminous
United States. Unfortunately, US Federal Government regulations
prevent the disclosure of actual plot locations. As an alternative,
information on species identity in plots, as well as data on the
values of multiple environmental variables in these plots were
supplied by the GIS Spatial Data Services unit of the US Forest
Service, which pre-processed the data according to the specifica-
tions requested for this study. In these data, species names and
plot locations were replaced by unique codes. For each plot, data
also included the values of 12 environmental variables that
are known to affect tree distributions, which were derived from
ancillary datasets. These variables are categorized as follows:

climate variables: mean annual temperatures and precipitation,
and seasonality of temperature and precipitation (bio1, bio4, bio12
and bio15 variables, obtained from the BioClim database; [38]);

topography variables: elevation and elevation range within a
1-km cell size (obtained and subsequently processed from the US
National Atlas); (3) coarse-scale soil data: minimum\maximum
values of water holding capacity, organic matter content, and
soil permeability (obtained from the STATSGO database, gener-
ated by the Soil Conservation Service, US Department of
Agriculture; available at http://water.usgs.gov/GIS/metadata/
usgsrwd/XML/ussoils.xml) [39].

Besides data on species occurrences and environmental condi-
tions per plot, the dataset included a distance matrix, which
denoted the geographical distances between plots. I generated
multiple subsets of the dataset (representing ad hoc tree commu-
nities), by running a cluster analysis on the distance matrix using
the Ward algorithm, which divided it into 200 unique clusters.
Each one comprising FIA sample points adjacent in geographical
space (electronic supplementary material, table S1). I chose to
generate 200 clusters to prevent community matrices from
becoming too large, as this is known to lead to statistical prob-
lems in null model analyses [11]. The average number of sites
in each cluster is therefore 46.3, with a standard deviation of
18.58. Then, for each cluster, I generated 1000 fixed–fixed null
models using the trial-swap algorithm [35]. For species that
occurred at least five times in a given community, I calculated
the C-score for each species pair within each cluster, and then
calculated the pairwise SES of the C-score. Finally, for each cluster,
I calculated the number of significant segregated (and aggregated)
species pairs after correcting for the false-detection error rate due
to multiple tests using the Bayes mean-based criterion [40].
(c) Analysis of niche overlap among species in forest communities

To quantify niche overlap among all species pairs in the FIA dataset, I employed an approach that is conceptually similar to my analysis of niche overlap in the modelled communities. First, I generated a SDM for each species in the dataset based on its occurrence sites and the 12 environmental variables detailed in §2b. For the SDM, I used Maximum Entropy (MaxEnt, [41,42]), which is a commonly used presence-only machine learning algorithm that iteratively contrasts values of environmental predictor values at occurrence locations with the values for a large background sample taken from across the entire dataset. To improve model performance and predictive capabilities, I restricted the analysis to those species that appeared in 30 sites or more in the entire dataset, which led to the generation of models for 155 tree species overall (out of 305 species that are present in the dataset). For each one of those species, I used MaxEnt to generate a measure of habitat suitability in any given site (a value between 0 and 1), which is analogous to the variable $P$ in equation (2.1). I then used equation (2.2) 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), under the assumption that it can serve as a proximate measure of realized niche overlap in terms of broad-scale climatic, topographic and edaphic variables. To quantify niche breadth for each species, I tallied the number of sites in which its predicted habitat suitability was larger than a threshold of 0.01. To evaluate the sensitivity of the analysis to the value of this threshold, I repeated the process with two additional thresholds, 0.1 and 0.001. I caution that this measure of niche breadth is only meaningful in the context of this analysis, as it depends on the distribution of environmental conditions across the available sampling sites.

To quantify environmental heterogeneity at the cluster level, I ran principal components analysis on the same 12 environmental variables I 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 geographical cluster, I calculated the averages of the range and the coefficient of variation of the first four principal components across all sites in which either or both species occurred. These serve as two species-pair-specific measures of environmental heterogeneity at the cluster scale.

(d) Statistical analysis

To quantify the effects of environmental heterogeneity and niche overlap on species co-occurrence patterns at different levels of niche breadth, I used two approaches. For the modelled communities, I developed multivariate linear models with SES as the dependent variable and niche overlap and environmental heterogeneity as the independent variables. For the empirical communities, I developed linear mixed models with the same variable designations as above, but added cluster identity as a random effect. In both model types, I compared the effects of the two heterogeneity metrics (range and coefficient of variation of environmental conditions). To account for the effect of niche breadth, I grouped species according to different levels of niche breadths and developed the models separately for each group. To group species according to niche breadth, I divided them according to their $\sigma$ value in the analysis of modelled communities; in the analysis of the empirical communities, I grouped them into bins according to their percentile in the distribution of niche breadth values (0–20%, 21–40%, 41–60%, 61–80% and 81–100%). I did not analyse the effects of niche breadth for species pairs in which each species belonged to a different niche breadth percentile.

3. Results

(a) Co-occurrence patterns of simulated species pairs

The co-occurrence patterns exhibited by species pairs in simulated communities were significantly related to both niche overlap and one measure of environmental heterogeneity, environmental range (figure 1a). Species pairs exhibited mostly segregated patterns (positive SES values) at low levels of niche overlap, and aggregated patterns (negative SES values) at high levels of niche overlap. Environmental range had a positive effect on species co-occurrence patterns, as short ranges of environmental conditions (i.e. low

Figure 1. Relationships between species co-occurrence patterns (denoted by the SES of the checkerboard score) and two predictor variables: niche overlap between pairs of species, and environmental heterogeneity (denoted by the range of the environmental axis). The gridded surface depicts the linear model fit to these data. (a) Results from modelled communities comprising species with narrow niche breadths ($\sigma = 15$, or the 20% percentile of simulated breadths). (b) Results from a large dataset of tree species in the conterminous United States, for species pairs in the 0–20% percentile bin of niche breadths. (Online version in colour.)
heterogeneity) led to aggregated co-occurrence patterns, whereas longer ranges of environmental conditions (i.e. high heterogeneity) led to more segregated co-occurrence patterns, manifested by increasingly positive SES values. These results were consistent among simulations with different parameter settings for local abundance per site (i.e. from 10 to 50). The second measure of heterogeneity, the coefficient of variation of environmental conditions, was a significant predictor of SES in only a small number of parameter combinations, regardless of model setting.

Niche breadth had a clear and profound effect on the relationship between both niche overlap and environmental heterogeneity, and species co-occurrence. The effect size of niche overlap on SES was larger (more negative) at lower levels of niche breadth, and decreased linearly (became less negative) as niche breadth grew (figure 2a). Co-occurrence patterns of specialist species, therefore, are more strongly affected by their niche overlap, compared with more generalist species. Niche breadth affected the relationship between environmental range and SES in the opposite direction, as the effect size of range decreased with increasing levels of niche breadth (figure 2b). In other words, specialist species exhibited the strongest effect of environmental heterogeneity on co-occurrence patterns, whereas generalist species were less affected by heterogeneity in terms of their co-occurrence patterns.

The effects of niche breadth on the subsequent effects of niche overlap and environmental heterogeneity on species co-occurrence depended on the parameter of local abundance in sites. Communities comprising smaller sites (or increased interspecific competition for space) were characterized by smaller (less negative) effect sizes of niche overlap on species co-occurrences patterns (figure 2a), and smaller (less positive) effect sizes of environmental heterogeneity on species co-occurrence patterns (figure 2b). Hence, the competitive pressure to establish in a site also has a clear impact on the relationships among species niche breadth, niche overlap, and environmental heterogeneity versus species co-occurrence patterns.

(b) Empirical co-occurrence patterns of forest trees in the United States

According to the fixed—fixed null models, only 8.2% of species pairs in geographical clusters generated from the FIA dataset had co-occurrence patterns that differ significantly from random ($\alpha = 0.025$). Among species pairs that exhibited significant co-occurrence patterns, about half (52.1%) were segregated, and the others (47.9%) were aggregated. The average percentage of aggregated species pairs per community was 9.6% (s.d. 3.1%), whereas the average percentage of significantly segregated species pairs was 10.6% (s.d. 3.12%).

(c) Effects of niche characteristics and environmental heterogeneity on empirical co-occurrence patterns

Similar to the modelled communities, empirical co-occurrence patterns for tree communities had significant relationships with both species niche overlap and environmental heterogeneity (figure 1b). A linear mixed model of SES with both predictors as fixed effects and cluster ID as a random effect had a significantly negative effect size for niche overlap ($\beta = -1.73$, s.e. 0.05, $t = -34.29$), and a significantly positive effect size for environmental heterogeneity, denoted by the mean range of the first four principal components of the 12 environmental variables ($\beta = 0.05$, s.e. 0.008, $t = 6.22$). The random effect had a variance of 0.015 (s.d. 0.12) and its residual had a variance of 1.07 (s.d. 1.03). When environmental heterogeneity was denoted by the mean coefficient of variation of the first four principal components of the environmental variables, it did not have a significant effect on SES.

When the analysis was conducted separately for each niche breadth bin, I found that niche breadth affected the
slopes of the relationships between niche overlap and SES in a manner mostly consistent with my findings from modelled communities. Niche overlaps between species with lower niche breadth percentiles had stronger (more negative) effect sizes on SES than niche overlaps between species with higher niche breadth percentiles; effect sizes (β) were −4.46, −3.27, −2.06, −1.12 and −0.92 for species pairs in the 0–20%, 21–40%, 41–60%, 61–80% and 81–100% percentiles of niche breadth, respectively. The trend in the effect of niche breadth was qualitatively the same when niche breadth was calculated based on different thresholds (above 0.1 and above 0.001 of MaxEnt’s predictions of habitat suitability). Environmental heterogeneity, denoted by the mean range of the first four principal component axes, exhibited consistent effects on SES only in the three out of five niche breadth percentiles. Its effect sizes decreased with increasing niche breadth percentiles (0.21, 0.12 and 0.05, for the 0–20%, 21–40% and 61–80% percentiles, respectively), in agreement with the modelled results. When different thresholds for determining niche breadth were used, the trend in the effect size of heterogeneity became less consistent (two out of five percentiles for a threshold of 0.1, and one out of five for a threshold of 0.01), but its value was always positive. When the coefficient of variation was used as the measure of environmental heterogeneity, it had a weaker overall effect on SES, which was significant only in a small number of niche breadth percentiles. Hence, it will not be discussed any further.

4. Discussion

Here, I analysed the roles of species niche characteristics and environmental heterogeneity in dictating the patterns of species co-occurrences in ecological communities. I found that species niche overlap and environmental heterogeneity interact in complex ways with niche breadth to drive species co-occurrence patterns. These results were mostly consistent between modelled communities and empirical data on tree species across broad spatial scales.

Environmental heterogeneity, which manifests in difference in environmental conditions among sites, can lead to niche differentiation [43], and consequently to segregated co-occurrence patterns [44]. A positive relationship between environmental heterogeneity and the prevalence of segregated co-occurrence patterns was reported before [27]. Another study, however, suggests that the relationship is complex, as it depends on the covariance of species responses to the environment, which depends on the similarity in niche requirements among species, or in other words, the degree of niche overlap [26]. Combining this suggestion with my modelled and empirical findings reveals that niche overlap is perhaps the most fundamental driver of species co-occurrence, but its effects on co-occurrence patterns are affected by environmental heterogeneity, as well as the breadth of the niches of individual species. Yet, the magnitude of the relationships between niche overlap and co-occurrence patterns (and to a lesser extent, the relationships between environmental range and co-occurrence patterns) is affected by species niche breadth. While the effect of niche breadth on the relationship between niche overlap and co-occurrence was consistent between both modelled and empirical analyses, its effect on the relationship between environmental heterogeneity and co-occurrence was more pronounced in modelled communities compared with real communities due to the much larger number of breadth values in the former. This difference is also present because MaxEnt models were only developed for species that had more than 30 occurrences, which makes it possible that species with truly narrow niches were not included in the empirical analyses; thus the range of niche breadth values in the real communities analysed here might be shorter compared with the modelled communities. If this is the case, the results of real communities correspond (and agree) only with the right-hand side of figure 2b, whereas the breadth levels in the left-hand size of that figure (which exhibit a slightly positive trend) were simply not reflected in the empirical data.

As suggested by previous studies [13,14], the results of the simple model I used here reveal that non-random co-occurrence patterns can emerge even when species have no pre-determined interspecific interactions. In such cases, significant segregated patterns are expected to occur simply due to spatial turnover when species have sufficiently different niche requirements (or are affected by dispersal limitations) [18]. Similarly, significant aggregated patterns are expected to occur when species with similar niche requirements inhabit heterogeneous habitats, simply because they tend to prefer similar sites that may be spatially aggregated within a matrix of unfavourable ones. These results, coupled with those of [14] reiterate the risk of inferring ecological processes from co-occurrence patterns without explicitly accounting for range differences or differences in niche requirements between species. Interestingly, the only other study that I am aware of that analysed the relationship between pairwise co-occurrence and similarity in niche requirements [21], in this case for ant species, found contrasting results to those I report here for tree species. Ant species with similar thermal tolerances tended to be more segregated than ant species pairs with increasingly different thermal tolerances. The discrepancies between study results may be an outcome of the inherent difference in the levels of biotic interactions among motile species (ants) compared with sessile species (trees), where motile species can segregate in space more readily when they encounter competitors.

The modelled results reveal that the strength of competition for space, represented by the size of sites in which species compete for establishment, affects the subsequent relationships between niche overlap, environmental heterogeneity and species co-occurrence patterns (figure 2). Moreover, the effect of competition strength varies with species niche breadth. Both niche overlap and environmental heterogeneity have stronger effects on co-occurrence when sites are large and competition for space is weak (white circles in figure 2). This highlights the role of these two variables in generating non-random co-occurrence patterns even in the absence of strong interspecific interactions. Furthermore, when niches are wide (i.e. species are generalists), both niche overlap and environmental heterogeneity have weaker effects on co-occurrence patterns. The weakening effect of niche overlap might be due to an interaction between niche breadth and niche overlap (when niches are wider on a finite niche axis, niche overlap increases). The reduced effect of environmental heterogeneity is probably due to species’ increased capability of establishment in a broader range of environmental conditions. Specialist species, in contrast, exhibit the strongest effects of both niche overlap and environmental heterogeneity on co-occurrence, especially when there is weak competition for space.
Inferring the role of interspecific interactions on the assembly of species via analyses of species co-occurrence patterns is a major challenge in contemporary community ecology. Past studies have mostly focused on perfecting the methodology, i.e. by developing (and benchmarking) more robust null-models and co-occurrence metrics [7,11], and subsequently on describing co-occurrence patterns of various taxa, such as plants [9], birds [5,14] and invertebrates [26]. In recent years, however, there has been a renewed interest in revisiting the theory behind co-occurrence analyses [14], by looking for ways to disentangle the roles of different ecological processes that result in distinct co-occurrence patterns. Recent examples are studies that incorporated species resource requirements [13,19] into statistical models of species co-occurrence. Another study [14], modified the general approach of co-occurrence analysis to account for range differences between species, but did not account for similarities in niche requirements between species. The emergence of significant co-occurrence patterns in modelled communities in this study, in which no interspecific interactions were assumed, raise yet another red flag about the potential of inferring interspecific interactions from co-occurrence patterns when species’ niche characteristics and relationships are unaccounted for.

I caution that the empirical approach used here to quantify niche characteristics has two main caveats for the purpose of this study. First, niches generated from occurrence records reflect realized environmental niches (or the ‘realized Grinnellian niche’ sensu [45]), and as such are partly dependent on the effects of interspecific interactions and dispersal limitations on species ranges [46]. Yet a previous study [47] suggests that environmental (Grinnellian) niches are mostly resilient to alterations driven by competitive exclusion in local sites as long as they are quantified based on coarse-scale climatic variables that are redundant in space. Another caveat in the empirical quantification of species niches is related to the choice of environmental variables and data availability. It is possible that the environmental data used for generating niches did not include important niche axes, and therefore the interpretation of species co-occurrence patterns in the context of niche overlap may sometimes be obscured by uncaptured differences in species’ niche requirements. Furthermore, given that environmental variables are only surrogates of actual niche parameters [25,48], the choice of environmental variables might influence the resulting levels of niche overlap between species. This problem is not unique to this study, as essentially all SDMs use environmental variables in an attempt to identify niches in space, or their corresponding habitat conditions. This, in turn, can also affect the results of similar studies such as [13,19].

When interpreting the results of the empirical analysis, one should be aware of another caveat. As both niche overlap and species co-occurrence are determined from occurrence data, it is possible that the results are biased to some extent as both quantities are confounded by patterns of species occurrence. I suggest, however, that this bias might be small because of a gap between the scales at which co-occurrence and niche overlap are quantified. Co-occurrence is determined at the site scale (i.e. four 7.32 m radius FIA plots in which one or both species occur) and its metric is calculated at the cluster scale. By contrast, species niches (and subsequently their overlap) are calculated based on coarse-resolution environmental variables (1 km cells) at the continental scale [47]. Thus, it is possible that interspecific interactions that affect the distribution of presence locations for a given species (and its co-occurrence with the species it interacts with) have little effect on its distribution in a multi-dimensional environmental space (its realized Grinnellian niche), because even though the species might be excluded by a competitor from some sites in a given set of environmental conditions, it is unlikely that it will be excluded from all of them unless these conditions are unique [45]. On a side note, the fact that the results of the empirical analysis exhibited remarkable agreement with those of the model (in which co-occurrence and niche overlap were unconfounded by design) may somewhat alleviate the concern about the bias in the results of the empirical analysis. Given the difficulty of estimating niches and niche overlap empirically [49], and especially the inherent problem in attempting to infer species niches from occurrence data [50], I suggest that if interpreted in light of these caveats, the results of this analysis may serve as a first approximation to the effects of niche overlap, niche breadth and environmental heterogeneity on species co-occurrence at broad spatial scales.

The past few years have seen a number of new promising approaches [13,14,19] that attempt to improve the methodology of inferring ecological processes from species co-occurrence patterns in order to understand species assembly in communities. It is likely that these approaches could be improved further if they combined empirical data on species co-occurrences with information on the geographical range and resource requirements of species, coupled with their availability across space, as well as the type and magnitude of interactions among species. Many of these processes are directly or indirectly related to niche characteristics, and their effects on species co-occurrence patterns are ultimately driven by the overlap in niche requirements between species. I therefore suggest that future attempts to infer ecological processes from species co-occurrence patterns should account for the effects of niche relationships among species at local and regional spatial scales, as well as the environmental heterogeneity in their habitats.

Data accessibility. FIA data used in this study can be obtained directly from the US Forest Service, via their Spatial Data Services unit (http://www.fia.fs.fed.us/tools-data/spatial/default.asp). R code for running the community model is available in the electronic supplementary material.

Competing interests. I declare I have no competing interests.

Funding. I received no funding for this study.

Acknowledgements. I thank Richard A. McCullough of the US Forest Service for pre-processing the FIA data.

References


