

Tree species co-occurrence patterns change across grains: insights from a subtropical forest

AVI BAR-MASSADA,^{1,2} QINGSONG YANG,¹ GUOCHUN SHEN,^{1,3,†} AND XIHUA WANG^{1,3}

¹Tiantong National Station for Forest Ecosystem Research, School of Ecological and Environmental Sciences,
East China Normal University, Shanghai 200241 China

²Department of Biology and Environment, University of Haifa at Oranim, Kiryat Tivon 36006 Israel

³Shanghai Institute of Pollution Control and Ecological Security, 1515 North Zhongshan Road (No. 2), Shanghai 200092 China

Citation: Bar-Massada, A., Q. Yang, G. Shen, and X. Wang. 2018. Tree species co-occurrence patterns change across grains: insights from a subtropical forest. *Ecosphere* 9(5):e02213. 10.1002/ecs2.2213

Abstract. Co-occurrence is a basic measure of spatial relationships between species. This commonly used measure has many benefits and limitations, yet a basic property that can strongly affect it has been overlooked. Co-occurrence analysis is based on discrete sampling in space, and therefore, its grain size may affect the results and their interpretation, because species interactions and their environmental responses are scale-dependent. We utilized a large dataset on tree species from a full-stem mapped forest plot in China as a template for testing the effects of grain on species co-occurrence patterns. We quantified co-occurrence patterns for large trees and saplings in nested sampling plots with increasing radii and analyzed the effect of plot size on co-occurrence. Co-occurrence patterns varied greatly across grains. More than half of the species in large trees we analyzed had significantly non-random co-occurrence patterns at some grain. In contrast, saplings exhibited much fewer non-random co-occurrences. The proportion of segregated species pairs of large trees had a unimodal relationship with grain, whereas the proportion of aggregated species was positively related to grain. These patterns disappeared in saplings, suggesting that spatial interactions among trees are more prominent among larger individuals. Therefore, co-occurrence patterns are scale-dependent, and this scale dependency reflects a mixture of ecological (interspecific interactions, environmental responses) and statistical (sampling effects) processes. Our results suggest that insights from single-grained studies cannot be generalized.

Key words: co-occurrence; C-score; forests; null models; scale.

Received 18 March 2018; accepted 26 March 2018. Corresponding Editor: Debra P. C. Peters.

Copyright: © 2018 The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** gcshen@des.ecnu.edu.cn

INTRODUCTION

Since the early debate on species interactions on islands between Diamond and Simberloff (Diamond 1975, Connor and Simberloff 1979, Diamond and Gilpin 1982, Gotelli and McCabe 2002), species co-occurrence has been one of the most widely used (and debated) measures of spatial and temporal correlations among species in ecological research. Co-occurrence patterns have been measured across many taxa from microbes through plants to animals (Wittman

et al. 2010, Götzenberger et al. 2012, Bar-Massada and Belmaker 2017), through vastly different geographical extents from the micro-biome to entire continents (Krasnov et al. 2010, Faust et al. 2012), and along temporal scales ranging from the present to hundreds of thousands of years ago (Dornelas et al. 2014, Lyons et al. 2016). Despite inherent problems in their interpretation due to the difficulty of reconstructing process from pattern, co-occurrence patterns have been used in many studies to infer on the role of species interactions in community assembly. While

their efficacy as measures of interspecific interactions is debatable (Ulrich 2004, Bar-Massada 2015), there is still merit in their usage as basic descriptors of spatial correlations among species occurrences in space and time. The question, though, remains how to properly interpret species co-occurrence, and how this pattern is affected by various ecological processes, statistical phenomena, and methodological considerations (Gotelli 2000, Ulrich et al. 2017).

One methodological consideration is the choice of spatial (or temporal) grain, which is the smallest sampling unit at which the presence of a species is recorded and subsequently analyzed. Different grain sizes are likely to generate distinct co-occurrence patterns, because many ecological processes (e.g., species interaction and environmental filtering) underlying these patterns are scale-dependent (Levin 1992). In the most basic sense, no two species can co-occur in the same point in space, while all species co-occur at the global scale (and often, at much finer scales). Therefore, the grain size of sampling and analysis becomes crucial to the ecological interpretation of co-occurrence patterns. Ideally, studies of species co-occurrence should be based on sampling grains that capture the outcome of ecological processes on site occupancy (Harms and Dinsmore 2016). While this might sound straightforward, determining the right grain size is likely to be extremely difficult, especially because many ecological processes happen across multiple scales (Belmaker et al. 2015). For example, two trees interact via competition for light (which depends on their age, height, and canopy structure relative to the slope aspect), for water (which depends on the structure of their root system, which in turn depends on soil characteristics), and for nutrients (which depends, among others, on the presence and spatial distribution of nitrogen-fixing bacteria in their root systems). Even outside the context of interspecific interactions, co-occurrence patterns can be used to infer on shared environmental responses of species (Pollock et al. 2014). Yet species' environmental responses are scale-dependent (Cushman and McGarigal 2002), while at the same time environmental conditions vary across scales (Kent et al. 2011). Meanwhile, a sample taken at a small grain size (e.g., 20 × 20 m) can reveal shared environmental responses to variables which vary

across broader scales (e.g., temperature and precipitation), but fail to show responses to environmental conditions which vary at finer spatial scales (e.g., soil nutrient contents and soil depth).

The possible grain size dependency of co-occurrence patterns makes it surprising, though, that almost no study has directly addressed this issue before. Note that we focus here on studies that quantify plot- or site-based co-occurrence at a given grain (this is in contrast to the different methodological framework of point pattern analysis (Wiegand et al. 2017), which requires data on the location of all individuals across a study area; point pattern analyses focus inherently on the spatial relationships across scales and will not be discussed here). Methodological studies on the challenges of using and interpreting species co-occurrence patterns emphasized analysis scales in terms of the overall extent of the study area, or the number of sampling sites (Gotelli and Ulrich 2012). We found only one study (McNickle et al. 2017) that tested the effect of grain size on co-occurrence patterns explicitly. That study analyzed co-occurrence patterns in tundra, grassland, and boreal and tropical forest communities and found that strongly segregated co-occurrence patterns emerged mostly at grain sizes that are much larger than plant body sizes (i.e., the grain at which the community was most segregated increased from 0.3, 1.5 m², 0.26 ha, and more than 1.4 ha in tundra, grassland, boreal, and tropical forests, respectively). Yet that study focused on community-wide co-occurrence patterns rather than pairwise patterns; this might obscure the effects of specific pairwise interactions and tends to dilute the effect of interspecific interactions in the community, as community-wide patterns are based on averaging all pairwise patterns. Furthermore, McNickle et al. (2017) quantified co-occurrence patterns regardless of plant size, and hence, their analysis cannot reveal whether and how co-occurrence patterns change along age- or size-groups. Given that species interactions can depend on individual age (or size), then their outcomes should be manifested by changes in co-occurrence patterns across age- or size-groups. At the same time, other ecological processes that affect species occurrences such as dispersal and environmental filtering operate differently across age-groups (e.g., sapling occurrence is more strongly related

to the outcomes of seed dispersal, while interspecific interactions and environmental filtering might become more prominent at large tree stages); if these processes dictate where species occur, then they might also have a direct effect on species co-occurrence patterns.

Our main objective was to quantify the scale dependency of species pairwise co-occurrence patterns within two tree size-groups in a subtropical forest, large trees and saplings. Specifically, we asked whether there are differences (or similarities) in scale dependence in co-occurrence between large trees and saplings. In general, though, in the context of our study, in which sampling grains are small compared to the grain of environmental heterogeneity, and the range of grain was short (5–25 m), we expect that aggregation would increase with grain size and that segregation will decrease (at short-to-medium distances; at broader scales, biogeographic patterns can lead to species segregations due to non-overlapping distributions). These expectations are the culmination of two main processes: (1) a sampling effect, by which increasing grain allows for the presence of more species even in the absence of interspecific interactions. This pattern might lead to increased species aggregations or random co-occurrences. (2) An interaction between species' habitat preferences and environmental heterogeneity. Increasing grain size would increase environmental variation within sites, while decreasing variation among sites. Hence at larger grain, species will be more likely to find suitable environments within each site, leading to decreased species segregation across sites, or more random co-occurrence patterns overall.

METHODS

Study area

We conducted the analysis in Tiantong National Forest Park, East China. The climate is subtropical monsoon with hot and humid summers and dry and cold winters. Precipitation mostly falls from May to August and has an annual mean of 1347 mm. Mean annual temperature is 16.2°C, and mean monthly temperatures range from 4.2° to 28.1°C in the coldest (January) and warmest (July) months, respectively. The forest is dominated by evergreen broad-leaved species, and the

common tree species are *Eurya loquaiana*, *Litsea elongata*, and *Choerospondias axillaris*.

The core area of the forest contains a 20-ha rectangular study plot (500 × 400 m), ranging in elevation from 304.26 to 602.89 m. Established in 2010, the plot was sampled for vegetation characteristics (tree species), soil conditions, and topography (mean values of elevation, slope, convexity, and aspect within 20 × 20 m blocks encompassing the entire study plot; the grain of these data was much coarser than the tree-level data, and hence, we could not use them to estimate species–environment relationships). All free-standing trees with a diameter at breast height (dbh) of 1 cm and higher were tagged, identified to the species level, measured for dbh, and the locations of their stems relative to plot boundaries were mapped (Fig. 1). Tree sampling followed standard protocols (Anderson-Teixeira et al. 2015). In 2010, there were 94,603 individual trees in the study plot, belonging to 152 species. While the study area is small relative to previous studies on tree species co-occurrence in forested ecosystems, many tree species within it exhibit significantly non-random spatial patterns (Yang et al. 2016), and as such, it is likely that co-occurrence analysis will reveal non-random associations among species. This, together with data on the exact spatial locations of all individuals, provides

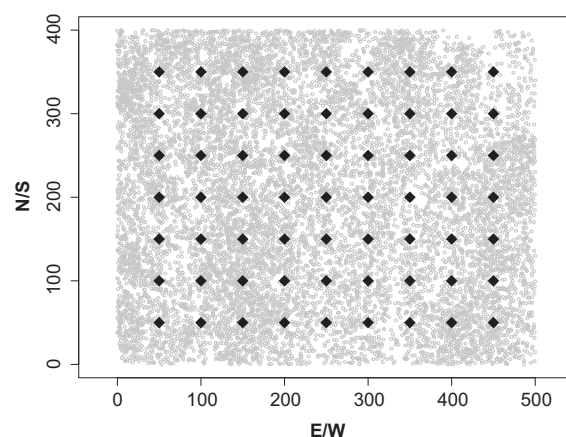


Fig. 1. The study area in Tiantong Forest Park, East China. The locations of sample plot centers are depicted by black diamonds. All individuals of the 23 tree species that are included in the analysis are depicted by gray circles.

a unique template for studying the effect of sampling grain on species co-occurrence patterns.

Data preparation

To analyze the effect of sampling grain on tree species co-occurrence in the study area, we allocated 63 plots in a systematic design on top of the stem-map of the 20-ha study area (Fig. 1). Plot centers were 50 m away from each other and from the boundary of the study area, to avoid plot overlap and to encompass the entire extent of the study area. We sampled the tree community in nested circular plots of increasing radii from 5 to 25 m, in increments of 1 m. This resulted in 21 community matrices (sites by species), each one representing a different sampling grain. In each plot, we recorded the presence/absence of all tree species at two size (age)-groups: (1) large trees, or individuals with a dbh of 10 cm or higher, and (2) saplings, with a dbh of 2 cm or less. There were 14,164 individual trees with dbh ≥ 10 cm in the study area, belonging to 107 species. There were 29,947 trees with dbh between 1 and 2 cm, belonging to 129 species. To reduce the effect of low abundance on subsequent analyses of species co-occurrence, we omitted all species that occurred in fewer than 10 sample plots. Moreover, to better reflect the effect of sampling grain on co-occurrence patterns, we restricted the analysis to species that occurred in ten plots or more in sampling radii from 5 to 10 m (since plots of different radii are nested, once a species appears in a plot at a given radius, it will occur at all larger radii; hence, the smallest number of radii a species occurred across was 16, from 10 to 25 m, and the largest number of radii a species occurred across was 21, from 5 to 25 m). This resulted in the analysis of 23 species overall (Table 1). Together, these species represented 11,971 large individuals (84.5% of the large trees in the study area) and 10,450 saplings (34.9% of the saplings). For comparison's sake, both large tree and sapling analyses were based on the same 23 species. We are well aware that these 23 species represent only a small proportion of species in the entire community, and hence, our subsequent analyses miss many interspecific interactions. Yet as our focus was mostly methodological, we suggest that our conclusions are valid for this subset of the community.

Table 1. Abundance and occurrence of the large tree species included in the analysis.

Species name	Total abundance	Occurrence in 5-m plots	Occurrence in 25-m plots
<i>Lithocarpus harlandii</i>	1462	18	61
<i>Distylium myricoides</i>	1454	16	41
<i>Litsea elongata</i>	1223	21	58
<i>Cyclobalanopsis sessilifolia</i>	1051	18	59
<i>Choerospondias axillaris</i>	897	21	60
<i>Schima superba</i>	869	13	51
<i>Cleyera japonica</i>	637	22	59
<i>Castanopsis fargesii</i>	591	11	49
<i>Machilus thunbergii</i>	465	–	58
<i>Machilus leptophylla</i>	426	–	30
<i>Castanopsis carlesii</i>	344	–	41
<i>Carpinus viminea</i>	324	–	50
<i>Cinnamomum subavenium</i>	265	–	52
<i>Acer pubinerve</i>	241	–	53
<i>Symplocos cochinchinensis</i> var. <i>laurina</i>	237	–	30
<i>Neolitsea aurata</i> var. <i>chekiangensis</i>	226	–	50
<i>Vernicia fordii</i>	220	–	33
<i>Sassafras tzumu</i>	217	–	46
<i>Ilex buergeri</i>	195	–	43
<i>Alniphyllum fortunei</i>	186	–	44
<i>Cyclobalanopsis myrsinaefolia</i>	179	–	46
<i>Liquidambar formosana</i>	137	–	39
<i>Illicium lanceolatum</i>	125	–	31

Analyses of species pairwise co-occurrence

We analyzed species co-occurrence patterns using the standard null model approach (Gotelli 2000). The approach is based on calculating a measure of co-occurrence for each pair of species based on their presences and absences in sampling sites. Since co-occurrence can also occur by random sampling processes, the next step comprises generating a null model of the community table (Gotelli and Ulrich 2012) by re-shuffling its rows and columns according to a randomization algorithm and recalculating the co-occurrence measure. The process is repeated many times to generate an entire distribution of co-occurrence measures that are based on the null model. Finally, the empirical co-occurrence measure is compared to the null distribution, to assess the degree of

deviation of the empirical co-occurrence measure from the null distribution.

Here, we used the checkerboard score, or C-score (Stone and Roberts 1990) as a measure of pairwise species co-occurrence. The C-score counts the number of checkerboard units in two vectors of species occurrences, which denote exclusive occurrence patterns (i.e., species A appears in site X but species B does not and vice versa). C-score is standardized according to the total number of occurrences of both species. As a null model, we used the species-fixed/site-equiprobable method (Jonsson 2001), which retains overall species abundances, but reassigns species occurrence in sites with a uniform probability across sites. This null model is suitable for our analysis because the study area is small, and species are able to occur throughout it. However, to evaluate whether our analysis was affected by the choice of null model, we repeated it using a species-fixed/site-fixed null model, specifically the trial-swap algorithm (Miklós and Podani 2004) which is more suitable when samples are taken from a heterogeneous region. We found that our results were qualitatively the same under both null models; hence, in the remainder of the manuscript, we will only describe the results of the fixed-equiprobable null model. In any case, we ran 1000 null models to generate the null distribution and calculated the standardized effect size (SES) of C-score as a measure of the strength and direction of co-occurrence. SES is calculated as the difference between the empirical C-score and the mean of its corresponding null distribution, divided by the standard deviation of the null distribution. Positive SES denotes segregated co-occurrence patterns, and negative SES denotes aggregated values. Assuming a standard normal distribution, SES values above 1.96 (below -1.96) represent significantly segregated (aggregated) co-occurrence patterns. We repeated the processes of quantifying pairwise species co-occurrence patterns for all species pairs across all sampling radii and age-groups. At each radius, we quantified the proportion of significantly aggregated (and segregated) species pairs. Because this analysis might be affected by inflated type I error due to multiple comparisons per species pair (as

the significance of co-occurrence for each pair was tested across multiple radii), we also calculated the proportion of species pairs with significantly non-random co-occurrences after correcting for the false discovery rate (FDR) using the Benjamini–Yekutieli (BY) method (Benjamini and Yekutieli 2001).

Surveying analysis grains in published studies

To evaluate the prevalence of single-grain studies in recent literature, we searched the Web of Science for research articles containing the topic C-score in the past 15 yr, from 2003 to 2017. We focused on C-score as it is a widely used metric in co-occurrence analysis. Obviously, there are other measures of site-based co-occurrence, but we opted to focus on C-score because it is very common and corresponds with our empirical analysis. Our search query returned all articles that included the term C-score in the title, abstract, or keywords. While the number of studies of species co-occurrence is much larger than what we found, we assumed that the sample size resulting from our search query was sufficient to represent prevailing trends in recent literature. We refined the results to include the following fields: plant sciences, ecology, marine freshwater biology, zoology, limnology, entomology, applied microbiology, fisheries, forestry, evolutionary biology, biology, and biodiversity conservation. We omitted studies that were based on published community matrices (Lehsten and Harmand 2006, Gotelli and Ulrich 2010), as it was impossible for us to obtain the sampling grains from the original studies behind them. We accessed the full text of each remaining article and reviewed its methods section to identify the study taxon and the grain of the analysis. Specifically, we noted whether studies were based on a single grain, and whether it was constant (i.e., all samples represented the same area) or variable (samples represented ecological units with varying areas). We refer to variable grains as cases where the size of the representative area of co-occurrence was not constant across samples (e.g., whole islands, habitat patches, freshwater ponds); this definition does not reject the validity of using ecological units as the grain in co-occurrence analyses.

RESULTS

Effects of sampling grain on pairwise co-occurrence patterns using SES C-score

We analyzed co-occurrence patterns for 253 species pairs overall across 16–21 radii (from 5 to 10 m sampling grain to 25 m). As we expected, we found many cases where pairwise co-occurrence varied across sampling grain; that is, SES C-score was scale-sensitive (Fig. 2). Overall, 139 species pairs (54.9% of all pairs) of large trees exhibited at least one significantly non-random co-occurrence at some sampling grain. Of these non-random co-occurrences, aggregated patterns were slightly more common than segregated patterns: 73 species pairs (28.8%) exhibited significantly aggregated co-occurrence patterns, whereas 66 species pairs (26.1%) exhibited significantly segregated co-occurrence patterns. One

species pair (*Neolitsea aurata* var. *chekiangensis*–*Cyclobalanopsis sessilifolia*) exhibited both aggregated and segregated co-occurrence patterns (Fig. 2D): It exhibited significant aggregation at a small plot radius and significant segregation at a large radius. The analysis of saplings revealed qualitatively similar results, though the overall proportion of significantly non-random co-occurrences was lower. Twenty-two pairs (8.6%) and 18 pairs (7.1%) exhibited significant aggregations and segregations at least at one grain, respectively.

Results were qualitatively the same after applying the Benjamini–Yekutieli correction for false discovery rate. Forty species pairs (15.8%) exhibited at least one significantly non-random co-occurrence patterns at some sampling grain. Of those, 29 species pairs (11.4%) exhibited significantly segregated co-occurrence patterns, whereas 11 species pairs (4.3%) exhibited

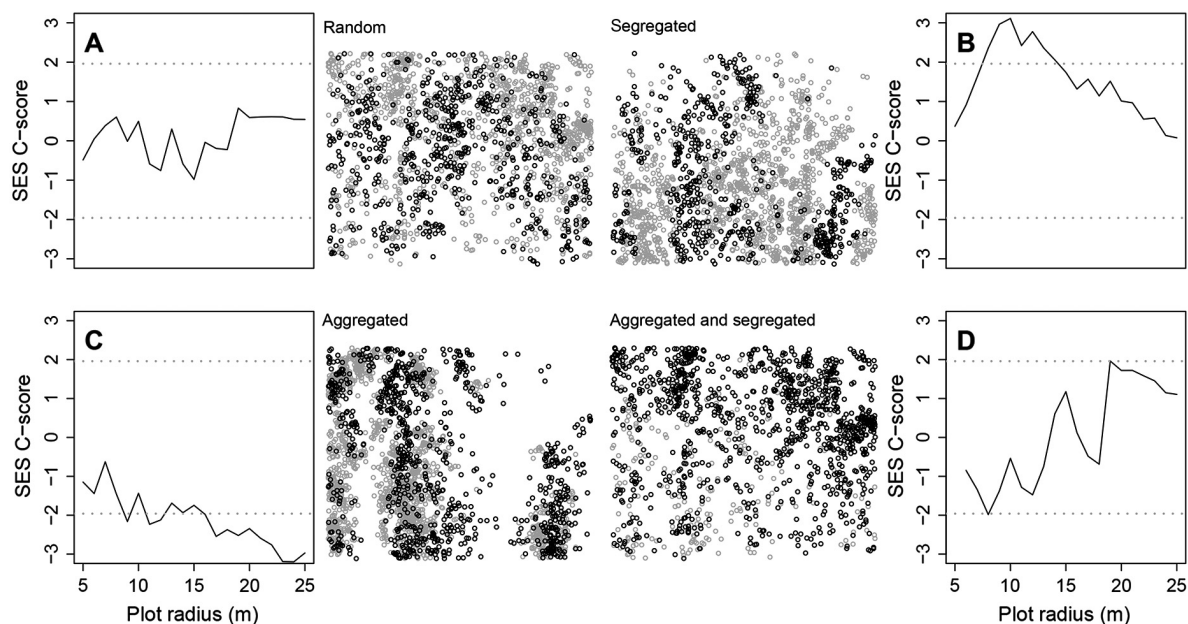


Fig. 2. Examples of four types of spatial relationships between individuals of species pairs (middle panels, the two species in a pair are depicted by black and gray circles) and their corresponding relationships between co-occurrence and sampling grain (outer panels). (A) Two species (*Cyclobalanopsis sessilifolia* and *Cleyera japonica*) that exhibit random co-occurrence patterns which are relatively consistent across sampling grains. (B) Two species (*Litsea elongata* and *Castanopsis fargesii*) that exhibit significantly segregated patterns at short-to-intermediate sampling grains. (C) Two species (*Distylium myricoides* and *Schima superba*) that exhibit significantly aggregated patterns at intermediate-to-long sampling grains. (D) Two species (*Neolitsea aurata* var. *chekiangensis* and *Cyclobalanopsis sessilifolia*) that exhibit significantly aggregated patterns at a short grain and significantly segregated patterns at a long grain. Horizontal dashed lines depict SES thresholds above 1.96 and below –1.96, which correspond with a significant SES score at $P \leq 0.025$ under a standard normal distribution.

significantly aggregated co-occurrence patterns. In the sapling analysis, these numbers dropped to five significantly segregated pairs (1.9%) and zero aggregated pairs.

In the majority of large tree species pairs, significantly non-random co-occurrence patterns (either aggregated or segregated) occurred only in a subset of sampling radii (Fig. 3). Only two out of 253 species pairs (*Castanopsis fargesii*–*Symplocos cochinchinensis* var. *laurina* and *Schima superba*–*Machilus leptophylla*) exhibited significant segregation across all sampling radii, whereas no species pair exhibited significant aggregation across all radii (though two species pairs, *Castanopsis fargesii*–*Schima superba* and *Castanopsis carlesii*–*Schima superba* exhibited consistent aggregations across 95.2% and 93.7% of radii). These results are consistent with the results of a previous study (Yang et al. 2016: Table S2) which showed that *C. fargesii* and *S. laurina* had contrasting habitat preferences, while *C. fargesii* and *S. superba*, and *C. carlesii* and *S. superba* had similar habitat preferences. Out of the 139 species pairs that had significantly non-random co-

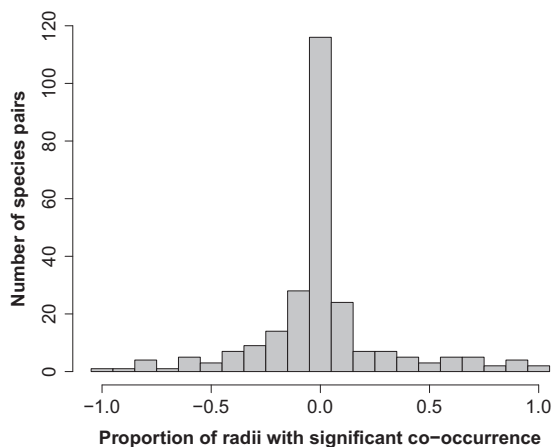


Fig. 3. Numbers of species pairs with different ratios of significantly non-random co-occurrence patterns across sampling radii. Positive x -axis values denote proportions of significant segregations, whereas negative x -axis values denote proportions of significant aggregations. For example, 1 on the x -axis corresponds with a species pair that exhibited significant segregation across all sampling radii, whereas 0 denotes species pairs that had random co-occurrence patterns regardless of sampling radius.

occurrence patterns, almost half (66 pairs, 47.5%) exhibited significant patterns (aggregated or segregated) in 20% of sampling radii or less.

When we analyzed the proportions of significantly aggregated or segregated patterns across grains (after correcting for FDR using the BY method), we found strikingly different patterns for aggregated and segregated species (Fig. 4). The proportion of species with significantly segregated patterns portrayed a significant unimodal relationship with grain ($\beta(\text{quadratic}) = -0.0003$ (SE: 2.57×10^{-5}), $P < 0.001$; $\beta(\text{linear}) = 0.0117$ (SE: 7.82×10^{-4}), $P < 0.001$; adjusted $R^2 = 0.945$), in which the largest proportion of significantly segregated pairs occurred roughly at grains from 15 to 20 m, above which this proportion declined (Fig. 4A). In contrast, the proportion of significantly aggregated pairs increased consistently with increasing grain (Fig. 4B; $\beta(\text{linear}) = 0.0017$; SE: 0.0001, $P < 0.001$; adjusted $R^2 = 0.919$). These patterns broke down when we analyzed saplings (Fig. 4 white circles; in both cases, a linear model of proportion of aggregated or segregated species vs. radius was non-significant), suggesting that spatial relationships between species are more prominent at large tree stages, whereas saplings might exhibit weaker spatial inter-relationships, as they are more strongly affected by mature trees in their surroundings.

Analysis grains in previous studies

We obtained data on analysis grain from 60 studies published between 2004 and 2017 (Appendix S1: Table S1). These studies focused on a large variety of taxonomic groups (from bacteria and fungi to mammals and trees) and differed greatly in their sampling grain (from guts of individual chironomid larvae [Lemes-Silva et al. 2014] to blocks of roughly 25 km² [von Gagen et al. 2015]). We did not find a single study besides McNickle et al. (2017) that quantified co-occurrence patterns across multiple sampling grains within a single spatial hierarchical level (e.g., a landscape), though a few studies analyzed co-occurrence patterns at more than one hierarchical level. For example, co-occurrence patterns in plant communities in Sweden were analyzed across one constant grain (quadrats) and two variable grains, patches, and entire landscapes (Reitalu et al. 2008). Another study

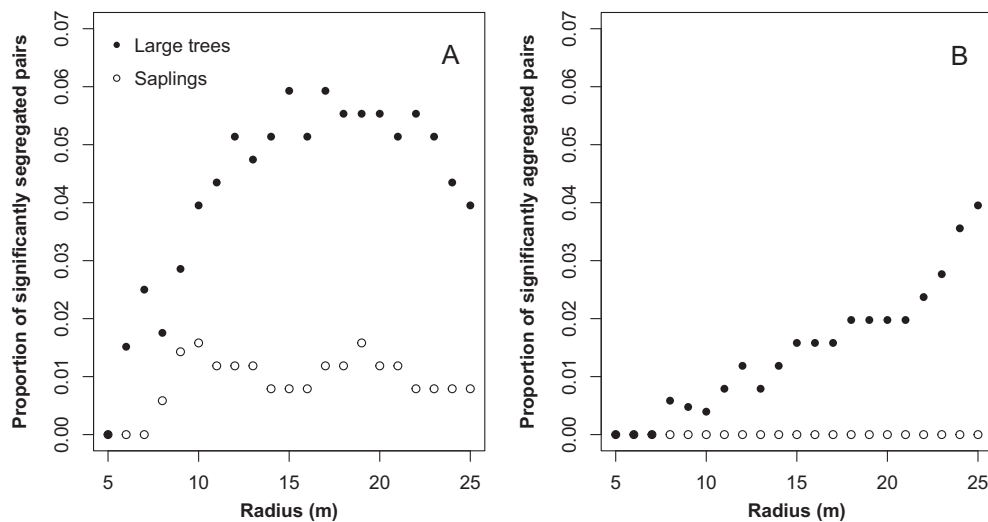


Fig. 4. Effects of sampling grain on the proportion of significantly segregated (A) and aggregated (B) species pairs. Significant pairs were identified after correcting for false detection rates using the Benjamini-Yekutieli method. Black and white circles denote large trees (dbh > 10 cm) and saplings (dbh < 2 cm), respectively.

(Boschilia et al. 2008) compared co-occurrence patterns of aquatic macrophytes in 1-m² quadrats and across entire floodplain lagoons. In general, studies were based on either constant sampling grains (60% of studies) or variable sampling grains (36.6% of studies). The two studies that operated at multiple scales had constant sampling grains at finer extents and variable sampling grains at broader extents. Studies that used variable sampling grains often had large variation in grain sizes, especially in cases where samples represented whole islands or ponds (and to a lesser degree when the grain was an individual of a different taxonomic level, for example, studying liana co-occurrence on trees [Blick and Burns 2011] or co-occurrence of ectoparasites on rodents; Krasnov et al. 2010).

DISCUSSION

We found that species co-occurrence patterns that are based on the commonly used C-score and null models are scale-dependent, with patterns differing among species, co-occurrence type (aggregation vs. segregation), and sampling grains. More than half of the tree species pairs we analyzed exhibited significant co-occurrence at some sampling grains (after correcting for FDR, the number fell to 14.8%). Furthermore, at the community level, aggregated and segregated

co-occurrence patterns of large tree species had a marked relationship with grain, where the proportion of segregated species had a unimodal relationship with grain, whereas the proportion of aggregated species had a positive relationship with grain. Saplings, in contrast, did not portray these patterns, hinting on a potential signal of both species interactions and abiotic conditions in driving co-occurrence patterns in the forest (Yang et al. 2016). From the practical context, though, these results highlight a methodological problem in studies of species co-occurrence; as the majority of pairs with non-random co-occurrence patterns exhibited non-randomness only at a small subset of sampling grains, we second the conclusion of McNickle et al. (2017) in suggesting that it seems likely that studies which are based on a single grain might result in inaccurate conclusions about spatial relationships among species.

The different scale dependencies of segregated vs. aggregated co-occurrence patterns (Fig. 4), coupled with the marked effect of age on the results, raise interesting questions about the mechanisms behind these patterns. Species co-occurrence in nature emerges due to three potential processes (Bar-Massada 2015). Species can aggregate or segregate due to facilitative or competitive interactions, respectively. However, non-random co-occurrence patterns can also emerge in the absence of interspecific interactions when

non-interacting species have similar (or different) environmental responses. Finally, even in the absence of differences in environmental responses and interspecific interactions (as in neutral communities), non-random co-occurrence patterns can emerge simply as populations aggregate or segregate due to chance dispersal events (Gotelli and McGill 2006). Can we use the information about co-occurrence scale dependency of large trees and saplings to infer on the roles of these processes in driving spatial relationships among species? While this is the pervasive problem of inferring process from pattern, we suggest that some of the patterns we found might be linked to processes by means of elimination. In our analysis, large trees of different species tended to aggregate at increasing grains. At the same time, saplings did not portray this pattern. These patterns make it likely that increased large tree aggregation is the result of the availability of more habitat conditions at larger grains. Saplings did not aggregate at larger grains probably because the spatial locations of saplings are largely the product of random dispersal across the study site; consequently, they do not yet reveal the effects of interspecific competition with individuals at the same size/age-group and/or the effects of environmental filtering on species co-occurrence patterns. In another study that focused on the spatial point pattern of trees in the same study area (Yang et al. 2016), we found that the importance of dispersal decreases with life stage, whereas the importance of environmental filtering increases with life stage. It is also possible that the lack of spatial relationships among saplings stems from the fact that at younger life stages, trees are more strongly affected by the presence of larger (adult) individuals which are superior competitors for light, water, and nutrients (Mori and Takeda 2003); yet both adult (canopy) and saplings of different tree species can exhibit a remarkable variation in responses to competition via different levels of shade tolerance and type of mycorrhizal association (Canham and Murphy 2016), which could manifest in inconsistent spatial patterns of co-occurrence across different species pairs.

Species segregations exhibited a scale dependence pattern which we did not expect beforehand. The proportion of segregated species increased initially at smaller grains but decreased

at larger grains. The decrease in segregations at larger grains is in line with our expectation, but it is still difficult to attribute it to a single ecological process. Species–area relationships suggest that just by chance larger plots will contain more species; hence, we may expect segregation numbers to be lower. The fact that saplings of the same species do not exhibit a unimodal relationship suggests that maybe this sampling effect is weak, and hence, the effect of environmental variation is indeed important, but this is impossible to prove because we did not have data on environmental conditions at a sufficiently fine grain. The remaining question is why segregation levels increase from small-to-intermediate grain sizes. Fig. 4 reveals that at the smaller grain sizes, non-random co-occurrence patterns (both aggregations and segregations) are rare. Could it be that the spatial relationships among very few individuals at the local scale are essentially random? To quote from a recent study (Chase 2014): “There are a multitude of probabilistic events (birth and death rates, dispersal, etc.) that allow each species to have a large number of sporadically distributed individuals in the habitat that it finds less favourable. As sampling scale declines to encompass fewer individuals and less habitat heterogeneity, the relative contribution of those stochastic events to the overall structure of the community increases, and we perceive this system, which is highly niche-structured at larger scales, as largely neutrally structured at smaller scale.” It becomes obvious, then, that species co-occurrence patterns at such small grain will not differ from random. Only once grain increases enough to allow for multiple individuals from different species, then patterns of segregation begin to emerge. Interestingly, a study on the effects of analysis grain on species richness and community composition in a subtropical forest in China (which is also fully mapped at the stem level) found that the total proportion of explained variation in richness and composition due to topography and spatial structure is scale invariant (Legendre et al. 2009). The authors interpreted this finding as an outcome of a scale-dependent tradeoff between the effects of topography (which becomes homogenized at coarser grains) and the pure spatial structure of the community (unobservable variables such as dispersal limitation and other

neutral processes that generate a spatial auto-correlated structure). While we caution that the smallest grain in that study was 20 m (larger than most of our grains), and the focus was on community structure (in contrast to pairwise interactions in our case), we might infer from their findings potential explanations for our results. Specifically, the possibility for increased environmental homogenization (among sites) at coarser grains can lead to more aggregation, as species will tend to appear together in more sites because larger sites will consist of more sub-habitat types, making them favorable to more species (Fig. 4B).

From a practical perspective, the scale dependency of co-occurrence analyses makes it difficult to draw conclusions about species interactions from studies based on a single grain (McNickle et al. 2017). The question is how to move forward and find ways to overcome this grain problem. The most straightforward approach would be to conduct multi-scale studies, in which species are sampled at multiple grains, and co-occurrence analyses are conducted at different scales, in a manner that can either strengthen conclusions about species relationships (i.e., when little or no scale sensitivity is found), or highlight variation in spatial relationships across scales (Boschilia et al. 2008, Reitalu et al. 2008, Laporta and Sallum 2014, Harms and Dinsmore 2016). Both outcomes are equally insightful and may increase our understanding of spatial relationships among species. In fact, the field of landscape ecology has long recognized the need to analyze species–habitat relationships across multiple scales simultaneously (Wiens et al. 1987, Wu et al. 2002), and consequently, researchers were able to better understand these fundamental relationships. On the flip side, in many cases the added effort in sampling at multiple grains may lead to practical limitations on data collection, which can limit the statistical power of co-occurrence analyses. In such cases, researchers should find ways (e.g., scaling rules; Peterson and Parker 1998) to generalize their results despite limited empirical knowledge on grain effects.

More than a third of the studies that analyzed co-occurrence patterns using C-score in the past 15 yr have not used constant grains, but rather used distinctive ecological units as their grain. These units vary in size, sometimes by an order of

magnitude. How are the findings of these studies affected by inherent variation in sampling grain is a non-trivial question. For example, if one assumes that the magnitude of co-occurrence is affected by overall species richness (either due to changes in the strength of interspecific interactions across multiple species [Levine et al. 2017] or due to statistical artifacts that arise in large community matrices; Gotelli and Ulrich 2012), then larger ecological units (e.g., islands or lakes) might exhibit different co-occurrence patterns than smaller ones. In such cases, variable co-occurrence patterns can simply be an artifact of species–area relationships (McNickle et al. 2017). Consequently, the size distribution of different ecological units might directly affect the patterns of species co-occurrence across them (an archipelago comprising many small islands may promote more segregated co-occurrence patterns compared to an archipelago with a mixture of small and large islands). For illustration, Table 1 shows the number of sampling sites each species appears in for small (5 m) sites and large (25 m) sites. For any given species, matrix fill (the number of non-empty cells) is greater at larger sampling sites. This in turn affects the number of potential unique configurations of the permuted community matrix, which can affect the statistical power of tests attempting to reveal non-random co-occurrence patterns. At the same time, larger sampling sites contain more species (and these, in turn, appear in more sites), and consequently, the number of pairwise occurrences is expected to grow with increasing grain. This phenomenon highlights the inherent risk of using an analysis grain that is too large. A possible solution to these problems is to move from presence–absence community matrices to abundance matrices. Though approaches for the analysis of abundance matrices do exist (Almeida-Neto and Ulrich 2011), most studies still use presence/absence matrices (possibly because they are much simpler to create).

To conclude, we found that species co-occurrence patterns are scale-dependent, in manners that are difficult to predict without detailed insight on the scale of interaction between species (as well as their separate and joint responses to environmental heterogeneity). Unfortunately, in the past 15 yr the overwhelming majority of studies on species co-occurrence that used the popular C-score index (including those of the lead author of this study) were based on either a

single analysis grain, or on a variable grain (which might suffer from inherent biases due to both statistical and ecological mechanisms). Hence, we caution that their results should be interpreted with care, as they mostly hold for the specific spatial grain for which they were applied to. Future studies on species co-occurrence should ideally account for the grain problem by taking a multi-scale approach, in which co-occurrence patterns are quantified across multiple grains simultaneously.

ACKNOWLEDGMENTS

We thank Heming Liu, Xiaofeng Fang, Zhanghua Wang, and many others for their help in the field. This study was supported by the National Key Research and Development Program (2016YFC0503102 to G.C.S) and the National Natural Science Foundation of China (31470487 to G.C.S).

LITERATURE CITED

- Almeida-Neto, M., and W. Ulrich. 2011. A straightforward computational approach for measuring nestedness using quantitative matrices. *Environmental Modelling & Software* 26:173–178.
- Anderson-Teixeira, K. J., et al. 2015. CTFs-ForestGEO: a worldwide network monitoring forests in an era of global change. *Global Change Biology* 21:528–549.
- Bar-Massada, A. 2015. Complex relationships between species niches and environmental heterogeneity affect species co-occurrence patterns in modelled and real communities. *Proceedings of the Royal Society of London B: Biological Sciences* 282: 20150927.
- Bar-Massada, A., and J. Belmaker. 2017. Non-stationarity in the co-occurrence patterns of species across environmental gradients. *Journal of Ecology* 105:391–399.
- Belmaker, J., P. Zarnetske, M.-N. Tuanmu, S. Zonn-evel, S. Record, A. Strecker, and L. Beaudrot. 2015. Empirical evidence for the scale dependence of biotic interactions: scaling of biotic interactions. *Global Ecology and Biogeography* 24:750–761.
- Benjamini, Y., and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 29:1165–1188.
- Blick, R. A. J., and K. C. Burns. 2011. Liana co-occurrence patterns in a temperate rainforest. *Journal of Vegetation Science* 22:868–877.
- Boschilia, S. M., E. F. Oliveira, and S. M. Thomaz. 2008. Do aquatic macrophytes co-occur randomly? An analysis of null models in a tropical floodplain. *Oecologia* 156:203–214.
- Canham, C. D., and L. Murphy. 2016. The demography of tree species response to climate: sapling and canopy tree growth. *Ecosphere* 7:e01474.
- Chase, J. M. 2014. Spatial scale resolves the niche versus neutral theory debate. *Journal of Vegetation Science* 25:319–322.
- Connor, E. F., and D. Simberloff. 1979. The assembly of species communities: Chance or competition? *Ecology* 60:1132–1140.
- Cushman, S. A., and K. McGarigal. 2002. Hierarchical, multi-scale decomposition of species-environment relationships. *Landscape Ecology* 17:637–646.
- Diamond, J. M. 1975. Assembly of species communities. *Page Ecology and Evolution of Communities*. Harvard University Press, Cambridge, Massachusetts, USA.
- Diamond, J. M., and M. E. Gilpin. 1982. Examination of the “null” model of Connor and Simberloff for species co-occurrences on Islands. *Oecologia* 52:64–74.
- Dornelas, M., N. J. Gotelli, B. McGill, H. Shimadzu, F. Moyes, C. Sievers, and A. E. Magurran. 2014. Assemblage time series reveal biodiversity change but not systematic loss. *Science* 344:296–299.
- Faust, K., J. F. Sathirapongsasuti, J. Izard, N. Segata, D. Gevers, J. Raes, and C. Huttenhower. 2012. Microbial co-occurrence relationships in the human microbiome. *PLOS Computational Biology* 8:e1002606.
- Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81:2606–2621.
- Gotelli, N. J., and D. J. McCabe. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond’s assembly rules model. *Ecology* 83:2091–2096.
- Gotelli, N. J., and B. J. McGill. 2006. Null versus neutral models: What’s the difference? *Ecography* 29:793–800.
- Gotelli, N. J., and W. Ulrich. 2010. The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia* 162:463–477.
- Gotelli, N. J., and W. Ulrich. 2012. Statistical challenges in null model analysis. *Oikos* 121:171–180.
- Götzenberger, L., et al. 2012. Ecological assembly rules in plant communities—approaches, patterns and prospects. *Biological Reviews* 87:111–127.
- Harms, T. M., and S. J. Dinsmore. 2016. Spatial scale matters when modeling avian co-occurrence. *Ecosphere* 7:e01288.
- Jonsson, B. G. 2001. A null model for randomization tests of nestedness in species assemblages. *Oecologia* 127:309–313.
- Kent, R., A. Bar-Massada, and Y. Carmel. 2011. Multiscale analyses of mammal species composition–environment relationship in the contiguous USA. *PLoS ONE* 6:e25440.

- Krasnov, B. R., S. Matthee, M. Lareschi, N. P. Koralló-Vinarskaya, and M. V. Vinarski. 2010. Co-occurrence of ectoparasites on rodent hosts: null model analyses of data from three continents. *Oikos* 119:120–128.
- Laporta, G. Z., and M. A. M. Sallum. 2014. Coexistence mechanisms at multiple scales in mosquito assemblages. *BMC Ecology* 14:30.
- Legendre, P., X. Mi, H. Ren, K. Ma, M. Yu, I.-F. Sun, and F. He. 2009. Partitioning beta diversity in a subtropical broad-leaved forest of China. *Ecology* 90:663–674.
- Lehsten, V., and P. Harmand. 2006. Null models for species co-occurrence patterns: assessing bias and minimum iteration number for the sequential swap. *Ecography* 29:786–792.
- Lemes-Silva, A. L., P. R. Pagliosa, and M. M. Petrucio. 2014. Inter- and intra-guild patterns of food resource utilization by chironomid larvae in a subtropical coastal lagoon. *Limnology* 15:1–12.
- Levin, S. A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur Award Lecture. *Ecology* 73:1943–1967.
- Levine, J. M., J. Bascompte, P. B. Adler, and S. Allesina. 2017. Beyond pairwise mechanisms of species coexistence in complex communities. *Nature* 546:56–64.
- Lyons, S. K., et al. 2016. Holocene shifts in the assembly of plant and animal communities implicate human impacts. *Nature* 529:80–83.
- McNickle, G. G., E. G. Lamb, M. Lavender, J. F. Cahill Jr., B. S. Schamp, S. D. Siciliano, R. Condit, S. P. Hubbell, and J. L. Baltzer. 2017. Checkerboard score–area relationships reveal spatial scales of plant community structure. *Oikos* 127:415–426.
- Miklós, I., and J. Podani. 2004. Randomization of presence–absence matrices: comments and new algorithms. *Ecology* 85:86–92.
- Mori, A., and H. Takeda. 2003. Light-related competitive effects of overstory trees on the understory conifer saplings in a subalpine forest. *Journal of Forest Research* 8:163–168.
- Peterson, D. L., and V. T. Parker. 1998. *Ecological Scale: Theory and Applications*. Columbia University Press, New York, NY, USA.
- Pollock, L. J., R. Tingley, W. K. Morris, N. Golding, R. B. O'Hara, K. M. Parris, P. A. Vesik, and M. A. McCarthy. 2014. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution* 5:397–406.
- Reitalu, T., H. C. Prentice, M. T. Sykes, M. Lonn, L. J. Johansson, and K. Hall. 2008. Plant species segregation on different spatial scales in semi-natural grasslands. *Journal of Vegetation Science* 19:407–416.
- Stone, L., and A. Roberts. 1990. The checkerboard score and species distributions. *Oecologia* 85:74–79.
- Ulrich, W. 2004. Species co-occurrences and neutral models: reassessing J. M. Diamond's assembly rules. *Oikos* 107:603–609.
- Ulrich, W., W. Kryszewski, P. Sewerniak, R. Puchałka, G. Strona, and N. J. Gotelli. 2017. A comprehensive framework for the study of species co-occurrences, nestedness and turnover. *Oikos* 126:1607–1616.
- von Gager, M., M. von Gager, and A. Schmitz Ornés. 2015. Problems with bins: a critical reassessment of Gotelli and Ulrich's Bayes approach using bird data. *Acta Oecologica* 69:137–145.
- Wiegand, T., M. Uriarte, N. J. B. Kraft, G. Shen, X. Wang, and F. He. 2017. Spatially explicit metrics of species diversity, functional diversity, and phylogenetic diversity: insights into plant community assembly processes. *Annual Review of Ecology, Evolution, and Systematics* 48:329–351.
- Wiens, J. A., J. T. Rotenberry, and B. V. Horne. 1987. Habitat occupancy patterns of North American shrubsteppe birds: the effects of spatial scale. *Oikos* 48:132–147.
- Wittman, S. E., N. J. Sanders, A. M. Ellison, E. S. Jules, J. S. Ratchford, and N. J. Gotelli. 2010. Species interactions and thermal constraints on ant community structure. *Oikos* 119:551–559.
- Wu, J., W. Shen, W. Sun, and P. T. Tueller. 2002. Empirical patterns of the effects of changing scale on landscape metrics. *Landscape Ecology* 17:761–782.
- Yang, Q.-S., G.-C. Shen, H.-M. Liu, Z.-H. Wang, Z.-P. Ma, X.-F. Fang, J. Zhang, and X.-H. Wang. 2016. Detangling the effects of environmental filtering and dispersal limitation on aggregated distributions of tree and shrub species: Life stage matters. *PLoS ONE* 11:e0156326.

SUPPORTING INFORMATION

Additional Supporting Information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.2213/full>