RESEARCH ARTICLE

When do we have the power to detect biological interactions in spatial point patterns?

Tuomas Rajala[1] (iD) | Sofia Charlotta Olhede[1] (iD) | David John Murrell[2] (iD)

[1]Department of Statistical Science, University College London, London, UK

[2]Centre for Biodiversity and Environment Research, University College London, London, UK

Correspondence
Tuomas Rajala
Email: t.rajala@ucl.ac.uk

Abstract

1. Uncovering the roles of biotic interactions in assembling and maintaining species-rich communities remains a major challenge in ecology. In plant communities, interactions between individuals of different species are expected to generate positive or negative spatial interspecific associations over short distances. Recent studies using individual-based point pattern datasets have concluded that (a) detectable interspecific interactions are generally rare, but (b) are most common in communities with fewer species; and (c) the most abundant species tend to have the highest frequency of interactions. However, it is unclear how the detection of spatial interactions may change with the abundances of each species, or the scale and intensity of interactions. We ask if statistical power is sufficient to explain all three key results.

2. We use a simple two-species model, assuming no habitat associations, and where the abundances, scale and intensity of interactions are controlled to simulate point pattern data. In combination with an approximation to the variance of the spatial summary statistics that we sample, we investigate the power of current spatial point pattern methods to correctly reject the null model of pairwise species independence.

3. We show the power to detect interactions is positively related to both the abundances of the species tested, and the intensity and scale of interactions, but negatively related to imbalance in abundances. Differences in detection power in combination with the abundance distributions found in natural communities are sufficient to explain all the three key empirical results, even if all pairwise interactions are identical. Critically, many hundreds of individuals of both species may be required to detect even intense interactions, implying current abundance thresholds for including species in the analyses are too low.

4. *Synthesis*. The widespread failure to reject the null model of spatial interspecific independence could be due to low power of the tests rather than any key biological process. Since we do not model habitat associations, our results represent a first step in quantifying sample sizes required to make strong statements about the role of biotic interactions in diverse plant communities. However, power

should be factored into analyses and considered when designing empirical studies.

1 | INTRODUCTION

Understanding the contribution of biological interactions to the assembly and regulation of natural communities remains a key goal in ecology. The continual development and refinement of methods to detect interactions from spatial, temporal, and spatio-temporal data have therefore been a mainstay of the literature on the subject.

A particular focus on the role of competition can be found in plant ecology, not least because plants seem to require the same few nutrients (Silvertown, 2004), but also because their sessile nature might permit an understanding of processes from the spatial pattern of individuals (Murrell, Purves, & Law, 2001), and allow for easier experimental manipulation (Goldberg & Barton, 1992). Multiple methods exist to detect interspecific interactions but in non-manipulative field conditions there are often only two choices, both of which rely upon data where the location, identity, and often size of every individual is recorded (Wiegand et al., 2017). The first option is to fit growth and/or survival models that take into account the identity and size of nearby neighbours (e.g. Comita, Muller-Landau, Aguilar, & Hubbell, 2010; Stoll, Murrell, & Newbery, 2015; Stoll & Newbery, 2005; Uriarte, Condit, Canham, & Hubbell, 2004). However, this requires repeated sampling over time in order to track the fate of every individual and very often such data are not available. Another issue is that because all interaction parameters are fitted at once, considering all pairwise interactions is very difficult due the large number of parameters. As a consequence neighbouring individuals are sometimes lumped into conspecifics and heterospecifics with the potential problem that interspecific interactions are "lost" due to cancelling out of weak and strong, and/or positive, and negative effects of different species. An extension that has been recently explored is to model survival/growth as functions of the phylogenetic or functional similarity of neighbours (Fortunel, Valencia, Wright, Garwood, & Kraft, 2016; Uriarte et al., 2010). The second option is to investigate the spatial pattern of the community to test the null hypothesis that species are independently arranged with respect to one another. Inference from a single snapshot of the community relies upon the assumption that spatial data retain the "memory" of the birth, death and growth of the individuals (Flügge, Olhede, & Murrell, 2012), and consequently the effect of interspecific interactions should show up as inter-species spatial dependence after any effect of the abiotic environment has been removed (Murrell et al., 2001). Under the assumption that all pairwise tests are independent, each pair of species can be assessed individually, and dependent interactions

are categorised as being a competitive interaction if the species are spatially segregated, and facilitative if they are positively associated in space, although confirmation via experimental manipulation is still advisable. Due to less restrictive data requirements (the test can be carried out on a single sampling of the community), the spatial snapshot option has proven to be very popular, and the test methods employ well-established spatial statistics such as Ripley's K or the pair correlation function (pcf) to test the null model of spatial independence (Wiegand et al., 2012).

The results of previous spatial analyses of multi-species communities have found only a very low frequency of interspecific spatial interactions (aggregation/segregation) over scales relevant to plant competition, implying interspecific interactions are generally rare, or weak (as discussed by Chacón-Labella, Cruz, & Escudero, 2017; Luo, Yu, Chen, Wu, & Ding, 2012; Wang et al., 2014; Wiegand et al., 2012). However, comparisons of different plant communities have also revealed a positive relationship between the frequency of spatial independence and the number of species in the community (Chacón-Labella et al., 2017; Luo et al., 2012; Perry, Miller, Enright, & Lamont, 2014; Wang et al., 2014; Wiegand et al., 2012). Spatial independence between all pairs of species is one of the key assumptions of several unifying theories for biodiversity (McGill, 2010), and the low frequency of detected interactions has been put forward in support of this assertion (Chac'on-Labella et al., 2017; Perry et al., 2014; Wiegand et al., 2012). However, classical niche theory also predicts the strength of interspecific interactions to decline as the number of coexisting species increases (equation 4 in Chesson, 2000), with the relative strength of interspecific interactions being proportional to $1/(s-1)$ for s species. Therefore, the main difference between the theories is that null models for biodiversity assume spatial independence for all communities regardless of species richness, whereas niche theory predicts spatial interactions are likely to be stronger, and therefore more frequently detected in less species-rich communities. Hence, we argue the spatial analyses appear to better support the predictions of classical niche theory.

However, both the low frequency of interspecies interactions and the relationship between species richness and species interactions could arise due to the ability of the statistical tests to detect significant interactions at the sample sizes being used. Because of the unequal treatment of the null and alternative hypothesis in classical testing, failure to reject the hypothesis of no interaction does not provide concrete proof of a lack of interactions. As pointed out by Wiegand et al. (2012), when species are rare the rate at which two species might co-occur in space is also

very low and the statistical tests used might not be able to detect any interaction, even if it were very strong. If, as is often the case, species-rich communities have few common and many rare species, then we would expect to detect few significant interactions. Indeed, several investigations have found the frequency of significant spatial associations between species to be positively related to the abundance of both species being considered (Luo et al., 2012; Wang et al., 2014; Wiegand et al., 2012), raising the possibility that interactions can only be detected amongst the most abundant species.

For all tests, a lower limit on the abundances of species to be included in the analyses must normally be set, and this acknowledges there is a limit to our ability to detect even strong interactions in small sample sizes. Previous investigations have used a range of lower abundance thresholds including 100 (Flügge, Olhede, & Murrell, 2014), 70 (Wiegand et al., 2012), 30 (Perry, Miller, Lamont, & Enright, 2017), and even 18 (Chaćon-Labella et al., 2017) individuals. However, how and why is the lower threshold of individuals selected? What are the limits of our analyses to detect significant interspecific interactions? We are unaware of any study that investigates the statistical power of the tests for spatial independence between pairs of species that are commonly used and consequently there are no guidelines for the lower abundance threshold. As such care is required when interpreting failures to reject the null hypothesis, and we argue it is hard to make strong statements about the relative roles of stochastic- and niche-based processes across different communities until we gain a better understanding of the power of the methods to detect departures from spatial independence. In other words, is spatial independence a good first approximation in species rich plant communities because of diffuse competition leading to weak interactions, or is it because the statistical methods lack the power to detect the interactions for the given sample sizes typically available?

Here, we will elaborate on the statistical power of commonly used methods to detect significant interactions from spatial point pattern data. We shall study this problem by constructing a simple model for generating bivariate patterns where we can directly control the strength of interaction, and by utilising an approximation to the variance of the spatial summary statistic. We will show how the power to detect significant interactions is very much a function of the species' abundances, the strength of the interaction (normally the variable we are trying to infer, and therefore unknown), and the spatial scale over which the test is performed. Unfortunately, it is not possible to provide definitive sample size criteria since the power also changes with the summary statistic and test method being used. In particular, for simplicity we ignore habitat associations, and as we will discuss, it is hard to tell how much our results will change when using inhomogeneous tests that take this into account. Despite this, our results are a useful first guide to understanding the sample sizes required to detect pairwise interactions. With this caveat in mind, our analyses will suggest previous abundance thresholds for species inclusion are likely too low to detect even very strong interactions in the most species-rich communities being tested, thus questioning

the previously derived conclusion of a lack of dependence between species. Since power can be estimated from Monte Carlo simulations, we hope our results will motivate ecologists to think more about the issue of sample size in future studies and therefore help to resolve the debate over the relative importance of biotic interactions in species-rich communities.

2 | MATERIALS AND METHODS

2.1 | Summary statistics for bivariate interaction

Consider data for two species labelled 1 and 2 given as two sets of locations of individuals $\mathbf{x}_1 = \{x_{11}, \ldots x_{1n_1}\}$ and $\mathbf{x}_2 = \{x_{21}, \ldots x_{2n_2}\}$ respectively, where the locations are observed in a well-defined area. We will call the combined set of points $(\mathbf{x}_1, \mathbf{x}_2)$, a bivariate point pattern, and refer to the individuals' locations simply as points. Technical details are left to Supporting Information Appendix A, but in brief we assume that the data generating mechanisms can be described by some processes X_1 and X_2, and the goal of statistical analysis is to draw conclusions about the processes using the observed set $(\mathbf{x}_1, \mathbf{x}_2)$. We start by assuming that the processes are second-order stationary, which means there is no underlying heterogeneity in the abiotic environment (e.g. elevation, soil chemistry) that also affects the distributions of the species, and implies that the statistics calculated from the data do not depend on any particular location in the observation window (see Section 4 for extensions). Although ecological communities are rarely well approximated by stationary models, we motivate studying the stationary case as this must be explored first, before any more complex scenarios can be understood.

We will focus our attention on the second-order statistic commonly known as Ripley's K (Ripley, 1979) and its derivative, the pcf; our rationale being that these two summaries are amongst the most popular when characterising joint dependence (Law et al., 2009; Perry, Miller, & Enright, 2006; Velázquez, Martínez, Getzin, Moloney, & Wiegand, 2016). First (as is standard) we define the intensity of a point process $\lambda > 0$ as the expected number of points per unit area. The cross-K or partial-K, denoted here by $K_{12}(r)$, is a function defined as the expected number of points of species 2 inside a circle of radius r placed on a random individual of species 1, scaled with intensity λ_2 to remove dimension and facilitate comparisons. Due to symmetry, it follows that $K_{12}(r) = K_{21}(r)$. The parameter r controls for spatial scale and allows for multi-scale analysis.

The derivative of K_{12} in r is denoted by $g_{12}(r)$, and is called the cross- or partial-pcf. The pcf describes the aggregation/segregation of cross-species point locations at distance r where the probability of having a species 1 individual in some small region and a species 2 individual in some small region distance r away is relative to $g_{12}(r)\lambda_1\lambda_2$. The quantities are scaled so that for independent processes the expectation is $K_{12}(r) = \pi r^2$ and $g_{12}(r) = 1$. The different statistics are used to ask subtly different questions, with $K_{12}(r)$ testing for species independence *up-to* distance r, and $g_{12}(r)$ testing for independence *at* distance r.

2.2 | Model generated data for illustration

For better understanding of the power of bivariate point pattern statistics, we develop a simple two-species model for which the level of cross-species aggregation/segregation can be controlled directly and explicitly by two parameters that determine the spatial scale and the strength of the interaction. Using this model, we can provide power estimates for different sample sizes and interaction scales and strengths using simulations. The details of the model are provided in Supporting Information Appendix B. Briefly, we assume species 1 is insensitive to the presence of species 2, but that the spatial distribution of species 2 is dependent on the spatial distribution of species 1. Asymmetric interactions are a reasonable starting point given they are thought to be quite common in plant communities (Freckleton & Watkinson, 2002) and theory suggests competitive asymmetry may help maintain diversity in competitive communities (Nattrass, Baigent, & Murrell, 2012). The locations of all n_1 individuals are given by a Poisson process, so species 1 exhibits no intraspecific spatial structure. The n_2 individuals are placed with distribution that depends on the locations of species 1. Importantly the model has

$$g_{12}(r) = 1 + bh(r),$$

where $h(r) = \exp[-r^2/(2\tau^2)]$ is a decreasing function whose exponential decay is controlled by the parameter $\tau > 0$, and has a range (h is non-zero) of approximately 2τ.

This function is analogous to the interaction or competition kernels used in spatially explicit birth-death models (Murrell, 2010; Murrell & Law, 2003). The strength of interspecies' interaction, as summarised by $g_{12}(r)$, is controlled by the parameter $b \geq -1$. If $-1 < b < 0$ the two species exhibit segregation ($g_{12} < 1$), if $b > 0$ the two species exhibit aggregation or clustering ($g_{12} > 1$), and when $b = 0$ the two species are independent. The reader should note that this model is simply a pattern generating process for illustration, rather than a mechanistic model, and we simulate patterns conditional on fixed n_1 and n_2 as we want full control over them (for the unconditional model the abundances are random, like in the birth and death processes, see e.g. Murrell, 2010). Example point patterns showing inter-species aggregation and segregation can be found in Supporting Information Figure S6 in Appendix B.

2.3 | Testing bivariate independence

We now turn our attention to the main problem of determining if the processes X_1 and X_2, as observed through the bivariate point pattern $(\mathbf{x}_1, \mathbf{x}_2)$, are statistically independent. If the processes were independent, then the observed pattern would be a *random super-position* of the two processes. We will take this as our *independence* or *null hypothesis* which now needs to be tested using the observed data.

To test if the independence hypothesis is compatible with the data, observed values of a chosen test statistic are compared to the distribution of the test statistic under the independence model. We can test

either (a) at some specific distance, which we call *pointwise tests* or (b) simultaneously over multiple distances. For both types of tests, the idea is to compute some test statistic $T \in \mathbb{R}$ from the data, and compare it to the values of T (its distribution) as if the null hypothesis were true. If the data value is sufficiently extreme, we have reason to reject the null hypothesis.

The true distribution of the test statistic under independence is rarely known in point pattern applications, and needs to be approximated by an empirical distribution derived from simulations under the independence model. This approach is known as Monte Carlo testing (Myllymäki, Mrkvička, Grabarnik, Seijo, & Hahn, 2017). We consider the observation area to be rectangular, in which case the independence simulation consists of randomly shifting pattern 1 (or 2 or both) with a toroidal wrap (Lotwick & Silverman, 1982). This keeps the intra-species statistics of the patterns intact while "breaking" any interspecies dependencies, and can also be used for inhomogeneous patterns (Cronie & van Lieshout, 2015).

For the purposes of this discussion, we will consider only the simple pointwise testing scenario, for which we can employ an analytical approach using a Gaussian approximation to the distribution corresponding to the random shift simulations. As we will show, the approximation is very useful since it is not only computationally very efficient relative to the MC simulations, but also allows some analytical insight into what affects the power of the tests. The pointwise tests we will study are comparable to simultaneous tests when the best distance to test at is known (see Supporting Information Table S1 in Appendix C). As detailed in Supporting Information Appendix A, we can choose an unbiased estimator $\widehat{K}_{12}$ for which approximately it holds:

$$\widehat{K}_{12} \sim N\left(K_{12}, \sigma^2\right) \Leftrightarrow T := \frac{\widehat{K}_{12} - K_{12}}{\sigma} \sim N(0,1), \tag{1}$$

where K_{12} is the value under the correct model. Conditional on the observed point counts n_1, n_2, the variance of $\widehat{K}_{12}(r)$ can be approximated by

$$\sigma^2 \approx c_1 (n_1 n_2)^{-1} \left[(n_1 + n_2) c_2 + c_3 \right], \tag{2}$$

where c_1, c_2, c_3 are constants depending on the distance r and the geometry of the observation area (see Supporting Information Appendix A.3). At a short distance, the constants reflect mainly the stochasticity of each point's neighbour count, and when the distance increases the "censoring" of the neighbourhood at the edge of the finite observation window contributes additional uncertainty. The form given in Equation 2 is exact when X_1 and X_2 are distributed according to a homogeneous Poisson process, but as we will see later on in Section 3.1, the approximation works quite well also for weakly internally aggregated/segregated patterns. Under strong internal aggregation, the true variance of K_{12} will be higher than the approximation given by (2), but under strong internal segregation the true variance of the cross-K will be lower than given by (2); we refer to Supporting Information Appendix A.3 for further details. Although we focus on K_{12}, the approach to

approximating the distribution is nearly identical for g_{12}, only the constants are different.

2.4 | Power of a statistical test

Denoting the null hypothesis of bivariate independence by H_0, the test statistic by T, and a confidence level of the test by $(1 - \alpha)$ where $\alpha \in (0,1)$. Recall that α is the researcher's fixed accepted margin of making a false positive decision, also known as *type I error*, defined mathematically as

$$\alpha = P\left(T > q_{1-\alpha} | H_0 \text{ true}\right),$$

where P is the distribution of T, $q_{1-\alpha}$ is the corresponding threshold value for T so that if $T > q_{1-\alpha}$ is under H_0, then we reject the null hypothesis H_0. The condition refers to T being tested. On the other hand, the *power* of a test is the probability of a true positive judgment, that is, the probability of rejection when the hypothesis H_0 does not hold. Consider first the margin of making a false negative judgment,

$$\beta = P\left(T \leq q_{1-\alpha} | H_0 \text{ not true}\right),$$

also known as *type II error*. Then the power of the test is defined as

$$\text{power} = \text{power}\left(H_0, T, \alpha\right) := 1 - \beta.$$

Therefore, a test is powerful if it can correctly reject the wrong null model with a high probability.

Consider the idealised situation of testing the cross-species independence using the pointwise summary $K_{12} = K_{12}(\tilde{r})$ for some fixed spatial scale $\tilde{r}$ only. For the test statistic K_{12}, the null hypothesis H_0: "random superposition" implies $K_{12} = k_0 = \pi \tilde{r}^2$. Let us now consider the situation that in truth $K_{12} = k_{12} \neq k_0$. Then, if we accept the approximate Gaussianity of the test statistic as shown in the previous section, it follows by elementary manipulations that

$$\text{power} \approx 1 - \Phi\left(q_{1-\alpha/2} - \frac{|k_{12} - k_0|}{\sigma}\right), \qquad (3)$$

where Φ is the cumulative distribution function of the standard normal distribution, with a-quantiles q_a (approx. 1.96 for $\alpha = 0.05$ in the two-sided test). Notice that the sign of interaction does not matter, meaning that due to symmetry of the Gaussian distribution aggregation is as easy or hard to detect as segregation of similar strength. Also notice how the power is dependent on the variance (σ^2) of the test statistic used. The smaller the variance, the higher the power, which explains why different unbiased estimators of K_{12} have been developed (see e.g. Illian, Penttinen, Stoyan, & Stoyan, 2008) and, while all being correct in the sense of bias, they can lead to different rates of detecting interactions because of different variances.

We can now use the power formula and our approximation for the variance (Equation 2) to illustrate how to

- compute the power of the test given the point counts n_1, n_2, expected true signal k_{12}, and the type I error tolerance α;
- compute the required point counts given the expected true signal k_{12}, the type I error tolerance α, and the type II error tolerance β or power.
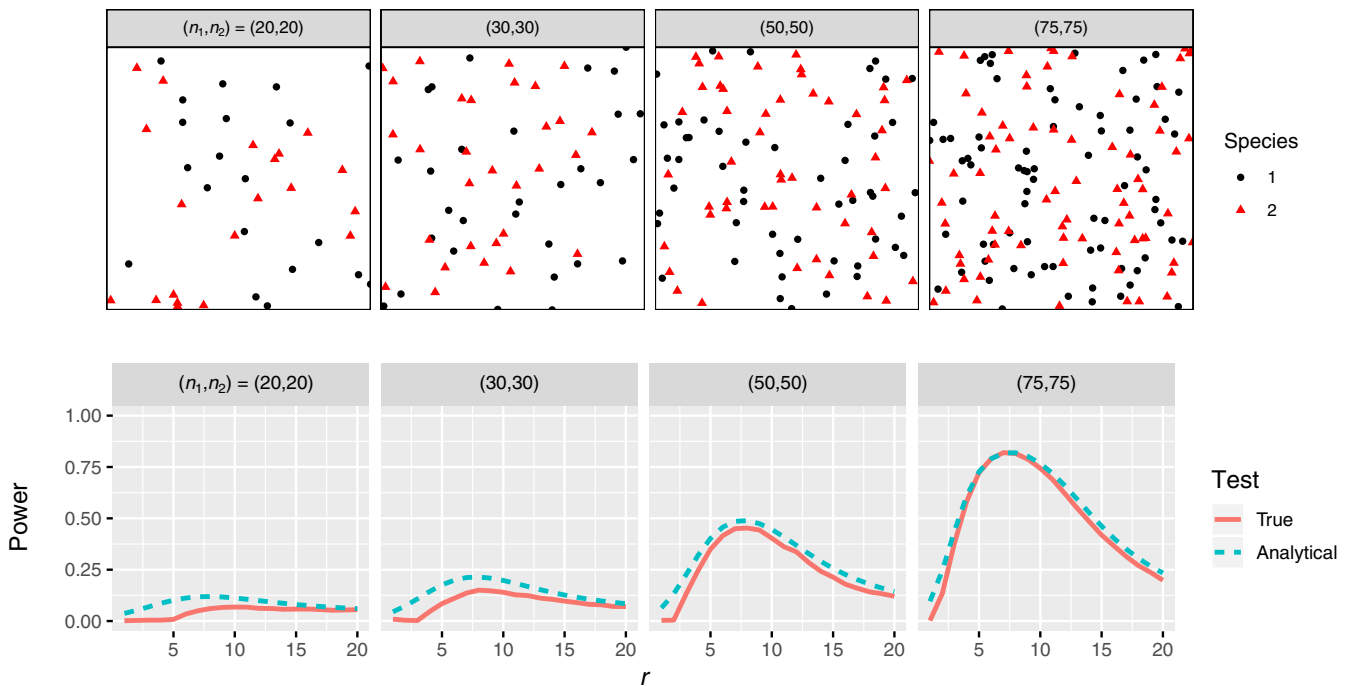
FIGURE 1   *Top*: Examples of the segregated bivariate point patterns, $b = -0.5$ and $2\tau = 10$, 100×100 window. *Bottom*: The power of K_{12}-based pointwise cross-species independence tests when species are segregated like in the example patterns. The true power is estimated using 5,000 repeated tests with 199 random shifts each

3 | RESULTS

The power formula (Equation 3) is a good approximation to the power of the toroidal shift Monte Carlo test (Figure 1). There is very little difference between the test's true power and the approximative power given by the analytical formula, with the analytical approximation slightly overestimating the power (at most 10%) due to the clustering of species 2 and subsequent underestimation of the variance via formula (2). The acceptable quality implies that we can discuss the power and its effect on ecological interpretations using the convenient analytical formula, acknowledging the small optimistic bias and the simplifying assumptions of stationarity and weak intra-species structuring.

As indicated by Equation 2, the variance of the estimator for the K_{12}-function is increased when either or both of n_1 and n_2 are small. This means that both the imbalance in population abundances as well as the total number of individuals affect our ability to detect bivariate interactions. We shall investigate each of these in turn, as well as the spatial distance of testing.

3.1 | Power in balanced scenarios and the importance of the spatial scale of testing

Figure 1 depicts the pointwise powers for different balanced ($n_1 = n_2$) low-abundance scenarios when data are segregated (aggregated results are nearly identical). Visual inspection of the example point patterns (Figure 1, top row) already gives some indication that departures from spatial independence might be hard to detect for the lowest abundances. More formal analysis of the power quantifies the increase in ability to detect interactions with increasing abundances (n_1, n_2) of the species being investigated and how this is affected by the spatial scale at which the hypothesis is tested (Figure 1, bottom row). In all cases, the power to detect the interaction at small spatial scales ($r < 2$) is low because, although the interaction is at its strongest here, the variance of K_{12} is relatively high and overwhelms the ecological signal. The trade-off between signal and noise leads to a

unimodal relationship between power and the neighbourhood radius r, with the peak being approximately at $r = 7$ for the interaction range $2\tau = 10$ for all abundance sizes considered (Figure 1). We will refer to this peak in power with r as the optimal distance for testing, and will focus on this best case scenario for the results presented below. The unimodal relationship highlights the point that having some prior knowledge about the likely ranges of biotic interactions is going to be important for detecting interactions.

Previous results based on in situ data analysis suggest detectable interactions between trees typically occur over 10–20 m (Uriarte et al., 2004). Scaling our analyses accordingly, we can use the power formula to estimate the population sizes we require in order to reliably detect an interaction of a given strength and range (Figure 2). If for example, we wish to be 75% sure a true positive is not to be missed when the interaction strength is weak ($b = -0.1$), then we require species with populations of approximately 400 individuals for the 10 unit interaction neighbourhood ($2\tau = 10$) and 250 individuals for 20 unit neighbourhood ($2\tau = 20$). This value is surprisingly large compared to what data we commonly have available to us.

In contrast, for the maximum possible negative interaction strength ($b = -1$), a similar level of power is reached with only around 35 individuals for $2\tau = 10$ unit and 18 individuals for $2\tau = 20$. Conversely, if we have a pair of species with $n_1 = n_2 = 50$, and we wish to be 75% sure a true positive is not missed, we must hope that the true interaction $|b|$ when coupled with short interaction range ($2\tau = 10$) is at least 0.7–0.75, and if coupled with long interaction range ($2\tau = 20$) is at least 0.3–0.4. It therefore seems likely that only the very strongest interactions are detectable with the number of individuals that are typically found in the species-rich datasets.

3.2 | Imbalances in species abundance

Since most communities exhibit a "hollow curve" distribution of population abundances (McGill et al., 2007), an imbalance in population sizes is very common. From the variance formula (2), it is clear that imbalance has a strong effect on the power because the term

$r = 7$ for $2\tau = 10$, $r = 14$ for $2\tau = 20$
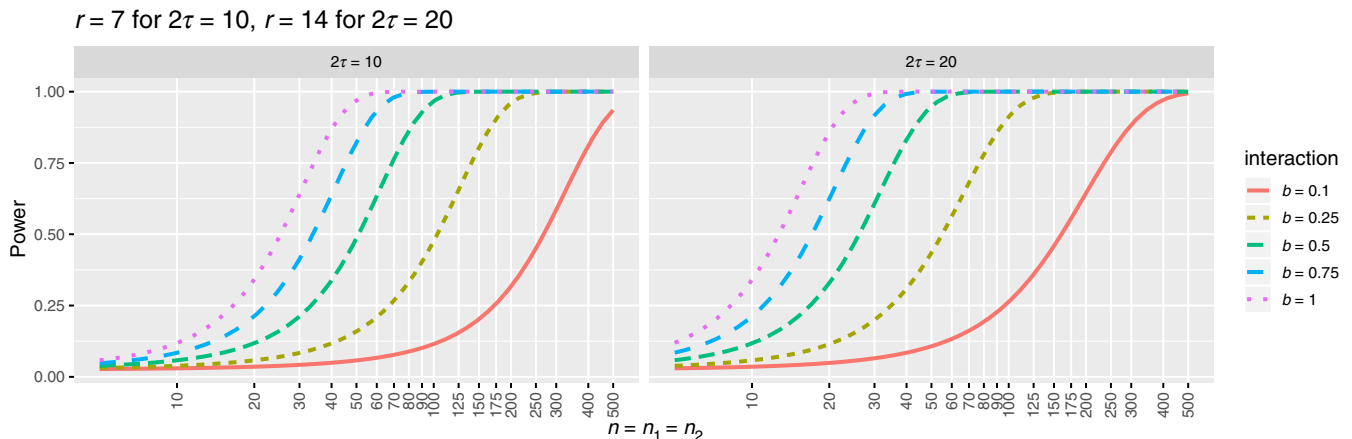
FIGURE 2 Power of K_{12}-based pointwise cross-species independence tests when abundances are balanced and testing is done on the best possible distance. Test level $\alpha = 5\%$
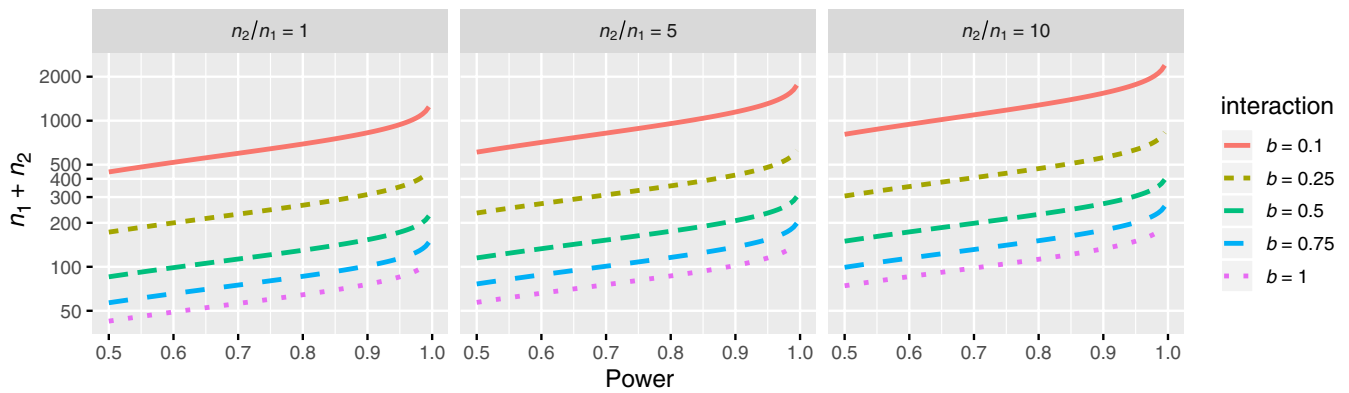
FIGURE 3 Sample size $n_1 + n_2$ requirements if testing for independence at level α = 5% with a K_{12}-based pointwise cross-species independence test in the example scenario. Interaction range 2τ = 10

$(n_1 n_2)^{-1}$, and hence the variance, increases with imbalance. This relationship is confirmed when we use the power formula to quantify the effect of population imbalance for different interaction strengths and combined population sizes (Figure 3). So, for example, for an interaction strength of $|b|$ = 0.1 and a desired power of 80%, a combined individual count of about 750 is required when the populations are perfectly balanced, but 1,000 are required when one species is five times more abundant than the second species, and a surprisingly large 1,500 required when one species is ten times more abundant than the other. Alternatively, consider that we require 90% power, and that the interactions are assumed to be $|b|$ = 0.5 and of short range, 2τ = 10. Then, to be on the safe side, we should attain samples of sizes at least (100, 100), (40, 200), or (30, 300), depending on the imbalance.

3.3 | Power at rainforest sample sizes

We now consider how our understanding of the power to detect interactions might affect results for abundance distributions typical of observed plant communities. For simplicity, let us assume interactions are of the type given by our model and that every species is interacting with every other species in an identical manner (so b and range 2τ are the same for all pairs of species). Since the power is the probability of detecting interactions, given that they exist, we can get a rough estimate of the number of detected cross-species interactions by assuming the tests are independent, and summing up the powers. This then allows a coarse comparison of recently reported frequencies of detected interactions in tropical forests (Chacón-Labella et al., 2017; Lan et al., 2016; Perry et al., 2014; Wang et al., 2014; Wiegand et al., 2012) with the expected frequency of detected interactions as a function of power.

Figure 4 shows the expected number of cross-species interactions detected as a function of abundance for various hypothetical interaction strengths and ranges. The abundances are taken from the Barro Colorado Island 1995 census (https://ctfs.si.edu/webatlas/datasets/bci/abundance) of woody plants with diameter at breast height at least 1 cm (Condit, 1998), and these are used in conjunction with our variance approximation and bivariate interaction model (so we are not using the spatial point pattern associated with the 1995 census). The abundances are highly skewed, with a large proportion of low-abundance species, and we show the power in two cases, when the pool of species consists of those with abundance at least 30 and 100. Reducing
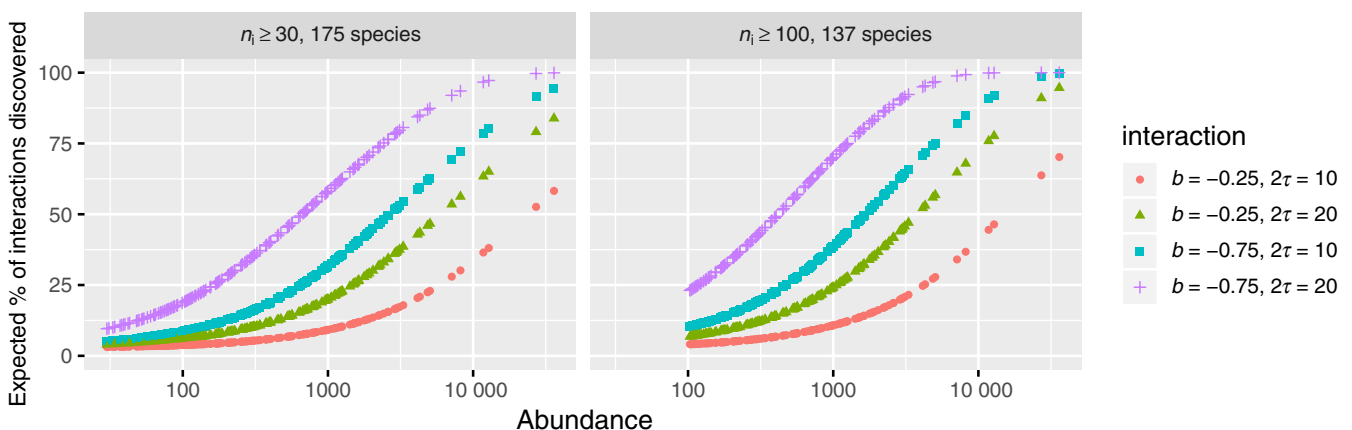
FIGURE 4 Expected fraction of interaction detected per abundance, if all pairs of species were to interact, and we tested independence with K_{12} at optimal distance

the species pool by increasing the abundance threshold naturally increases the proportions of detection, and highlights the importance of using similar thresholds when comparing different communities. It is striking how little power is to be expected for most of the species even when assuming strong interaction ($b = -0.75$). Only when the abundance of a species reaches thousands, can we be expected to detect even 50% of the interactions present. This is a very thought-provoking result, as the lack of detection might be explained simply by a lack of power in the majority of species-pairs. However, we remind the reader that our tests do not take habitat associations into account whereas the previous analyses (Chacón-Labella et al., 2017; Lan et al., 2016; Perry et al., 2014; Wang et al., 2014; Wiegand et al., 2012) approximately factor this out, and being different tests the power to detect interactions will be different. We discuss this point in more detail below.

4 | DISCUSSION

Understanding the relative strength of interspecific interactions is one of the key goals of community ecology, and the null model approach has been popular for characterising spatial point patterns of (predominantly) diverse plant communities (e.g. Chacón-Labella et al., 2017; Lan et al., 2016; Martinez, Wiegand, Gonzalez-Taboada, & Obeso, 2010; Perry et al., 2014; Velázquez, Paine, May, & Wiegand, 2015; Wang et al., 2014; Wiegand et al., 2012; Wiegand et al., 2017). However, there has been little guidance on when a given test is likely to be able to detect species associations that are present. Here, we have made a first step in closing this important gap in our understanding. Our results clarify the quantitative relationships between the strength of the underlying biological interaction, sample size (number of individuals of both species under investigation), and the spatial scale over which the test is being performed. We have also shown that statistical power may explain both the low detection rate of biological interactions in plant communities, and the negative relationship between species-richness and frequency of detected interspecific interactions in comparative studies.

Ecologists have had to rely largely upon their intuition for deciding the minimum population size to include in their analyses with the result that a range of criteria up to 100 individuals (Flügge et al., 2014) have been used. For species-rich communities, where many interspecific interactions may necessarily be weak (Chesson, 2000), abundances of both species may need to be in the hundreds of individuals before any interaction is detected (Figure 3), and this implies previous abundance thresholds are likely too low to detect many interactions. As several authors have acknowledged, the failure to reject the null hypothesis of spatial independence in so many species-pairs does not necessarily mean interspecific interactions are not occurring, or present (Chacón-Labella et al., 2017; Perry et al., 2014; Wiegand et al., 2012), and we hope our study highlights how the power of the tests can be assessed and should be factored into the interpretation of the results. The power formula can also be used in estimating the area of observation necessary to increase the power to a desirable level (Supporting Information Appendix C.4),

so can also be used to aid study design. Nonetheless, we do stress that there is still much to be learned about the power of the statistical tests used in earlier studies, given the assumptions we had to make, and that the reader should take our contribution as a guide to sample sizes that are required to make strong statements about the frequency and strength of interspecific interactions.

Although our model is clearly mis-specified as we use tests assuming that intensity is not dependent on abiotic features of the environment, the general applicability of our results will carry-over into the inhomogeneous setting. The quantitative relationships between sample sizes and statistical power will of course change for different tests, including the inhomogeneous Poisson process, but the qualitative relationships we uncover are likely to remain. In particular, we would still expect a positive relationship between population size and frequency of interactions to emerge simply due to an increase in power at larger sample sizes. Such a positive relationship has already been reported in a number of empirical studies that take habitat associations into consideration (Chacón-Labella et al., 2017; Luo et al., 2012; Wang et al., 2014; Wiegand et al., 2012). It is possible that common species are better competitors and are somehow suppressing the abundance of the weaker competitors, but without experimental manipulation, or perhaps different analyses using repeated sampling over time (Damgaard & Weiner, 2017), it is hard to distinguish whether this pattern is a result of biological processes or the ability of the statistical methods to detect interactions at different population sizes. Given the potential for low power to detect interactions for abundances typical of species-rich communities, we suggest that future tests should consider the neighbourhood from the perspective of traits or phylogenetic relatedness (e.g. Wang et al., 2016), with interactions potentially being some function of relatedness or functional similarity. In so doing, tests would consider the impact of multiple species on a focal species and would reduce some of the low power issues we highlight here. However, such an approach requires reliable phylogenies, and adequate sampling of traits that are relevant to growth, survival and fecundity of individuals, and these are still challenging issues in ecology.

The spatial scale over which tests are performed is important for the ability to detect spatial dependencies (Figure 1), and our results are similar to empirical studies that often find few negative interactions at the shortest distances, even though this is where the interactions are likely to be strongest (Chacón-Labella et al., 2017; Wang et al., 2014; Wiegand et al., 2012). Short scales suffer from having high variability due to the relatively small number of neighbours possible in a small area, but at longer distances, the effect of neighbours is weaker. Hence, there is a sweet spot where this trade-off is maximised, and the location of this is likely dependent on several factors, not least of which is the scale over which interactions are occurring (see e.g. figure 2 in Chacón-Labella et al., 2017 for an empirical example). For woody plants, there have been several studies that have fitted neighbourhood growth or survival models to individual-based data that track the fate of trees over time (e.g. Stoll & Newbery, 2005; Uriarte et al., 2004), and most results seem to point to interactions being confined to 10–30 m radius around an individual. However,

little is known about how the spatial scales of interspecific interactions change with life history stage, environmental conditions, or even species identity even though the latter has been shown to be very important for determining coexistence (Murrell & Law, 2003). Any changes to the scales of interactions will have consequences for the hypothesis testing methods discussed here, but until more is understood about the spatial scales of interactions between species, it seems sensible to test over ranges reported in earlier studies.

Our discussion up to this point has been in the context of stationary, most notably homogeneous, data. Most recent analyses have tried to factor out the effects of spatial heterogeneity in the abiotic environment by using inhomogeneous Poisson processes as the null model (Chacón-Labella et al., 2017; Punchi-Manage et al., 2015; Wiegand et al., 2012). Currently, it is hard to predict whether the power of an inhomogeneous analogue of our scenario would be lower or higher. On the one hand, we could expect higher power to detect interactions because the model better captures the underlying processes that generate the spatial distributions of the species within the community. However, we also expect variance to be increased, since extra parameters need to be estimated leaving fewer degrees of freedom per parameter. For example, tests using the inhomogeneous Poisson process method use a smoothing kernel to approximately remove the effects of large scale structure assumed to be caused by habitat associations (see e.g. Wiegand et al., 2012). Typically, the same smoothing parameter is used for all species, which is a sensible assumption when little is known about the spatial scale of habitat associations, but there is no reason to suspect a single smoothing parameter is appropriate for all species. An open challenge is to better understand how mis-specification of the smoothing parameter will bias the detection of interactions. Again, we feel that using a biologically motivated model to simulate data is a useful approach for exploring such issues.

The understanding of statistical significance versus the practical importance of any effect has been discussed in other application domains (Button et al., 2013; Hojat & Xu, 2004; Sawyer & Ball, 1981). The difficulties inherent in studies determining significance in settings with relatively small sample sizes, has for example been noted by Ioannidis (2005) and Open Science Collaboration et al. (2015). Our results underscore the difficulties of statistical testing for smaller sample sizes, especially given the unequal weighting between the null and alternate hypotheses. We also remind the reader that the spatial statistics used in the null model approach do not say anything directly about the processes that may have created the patterns, and different processes could generate the same summary statistic. Since the data are often a single snapshot of the community it is also hard to infer the importance of the results for the population dynamics of the species under scrutiny. For example, other processes such as temporal variation in the environment (Chisholm et al., 2014), and within species interactions (perhaps mediated via specialist natural enemies, LaManna et al., 2017) could both contribute more to population dynamics than any detected pairwise interaction. Similarly, interactions that are undetected due to low sample sizes could be expected to contribute little to population dynamics, especially if both species are

quite rare, although the contribution of many weak interactions might still be significant factors affecting growth, survival, and/or fecundity. As a result, we believe the spatial tests for independence should act as exploratory studies to highlight potentially significant species interactions, but understanding their biological importance requires different methods. As an alternative, model-based approaches, either in the form we use here (which include the familiar Thomas Cluster models) or birth-death models (May, Huth, & Wiegand, 2015; Rajala, Murrell, & Olhede, 2018) could also be applied to the inference of biological interactions from point pattern data (Wiegand et al., 2017). Model fitting will normally lead to estimation of parameters that can also be estimated in the field (e.g. dispersal kernels, interaction kernels), we therefore feel that their continued development will help to improve the understanding of the processes underpinning the results returned (Wiegand et al., 2017).

In conclusion, we hope our main contribution is to encourage more users to consider explicitly the ability of the spatial point pattern tests to detect significant associations between species. We have shown that the data requirements to detect even strong interactions may be quite high, mirroring results for null model tests of species co-occurrences in community matrix data (Freilich, Wieters, Broitman, Marquet, & Navarrete, 2018; Gotelli, 2000). On this basis, we suggest it is desirable to only interpret the frequency of interactions across large numbers of species once the effect of different powers to detect interactions for pairs of species of given population sizes has been (even approximately) factored out. This seems especially important in comparative analyses across different communities where the spatial scales, strengths of interactions, and the species abundance distributions may differ and affect the power to detect interactions.

AUTHORS' CONTRIBUTIONS

T.R. conceived the idea during discussions with D.J.M. and S.O.; T.R. derived the model and formulas, designed and executed computations, and contributed extensively to the manuscript; D.J.M. contributed to the rainforest experiment and extensively to the manuscript; S.O. contributed to the manuscript. All the authors contributed to the intellectual core of the manuscript.

DATA ACCESSIBILITY

The abundances in Section 3.3 for the Barro Colorado Island 1995 census are available at http://ctfs.si.edu/webatlas/datasets/bci/abundance/.

ORCID

Tuomas Rajala (iD) http://orcid.org/0000-0002-1343-8058

Sofia Charlotta Olhede (iD) http://orcid.org/0000-0003-0061-227X

David John Murrell (iD) http://orcid.org/0000-0002-4830-8966

REFERENCES

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafó, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365.

Chacón-Labella, J., de la Cruz, M., & Escudero, A. (2017). Evidence for a stochastic geometry of biodiversity: The effects of species abundance, richness and intraspecific clustering. *Journal of Ecology*, 105(2), 382–390.

Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, 31, 343–366. https://doi.org/10.1146/annurev.ecolsys.31.1.343

Chisholm, R. A., Condit, R., Rahman, K. A., Baker, P. J., Bunyavejchewin, S., Chen, Y., … Yap, S. (2014). Temporal variability of forest communities: Empirical estimates of population change in 4000 tree species. *Ecology Letters*, 17(7), 855–865.

Comita, L. S., Muller-Landau, H. C., Aguilar, S., & Hubbell, S. P. (2010). Asymmetric density dependence shapes species abundances in a tropical tree community. *Science*, 329(5989), 330–332. https://doi.org/10.1126/science.1190772

Condit, R. (1998). *Tropical forest cencus plots*. Berlin, Germany, and Georgetown, TX: Springer-Verlag and R. G. Landes Company.

Cronie, O., & van Lieshout, M. N. M. (2015). Summary statistics for inhomogeneous marked point processes. *Annals of the Institute of Statistical Mathematics*, 68, 905–928.

Damgaard, C., & Weiner, J. (2017). It's about time: A critique of macroecological inferences concerning plant competition. *Trends in Ecology & Evolution*, 32(2), 86–87.

Flügge, A. J., Olhede, S. C., & Murrell, D. J. (2012). The memory of spatial patterns-changes in local abundance and aggregation in a tropical forest. *Ecology*, 93(7), 1540–1549.

Flügge, A. J., Olhede, S. C., & Murrell, D. J. (2014). A method to detect subcommunities from multivariate spatial associations. *Methods in Ecology and Evolution*, 5(11), 1214–1224.

Fortunel, C., Valencia, R., Wright, S. J., Garwood, N. C., & Kraft, N. J. B. (2016). Functional trait differences influence neighbourhood interactions in a hyperdiverse amazonian forest. *Ecology Letters*, 19(9), 1062–1070. https://doi.org/10.1111/ele.12642

Freckleton, R. P., & Watkinson, A. R. (2002). Large-scale spatial dynamics of plants: Metapopulations, regional ensembles and patchy populations. *Journal of Ecology*, 90(3), 419–434.

Freilich, M. A., Wieters, E., Broitman, B. R., Marquet, P. A., & Navarrete, S. A. (2018). Species cooccurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities? *Ecology*, 99(3), 690–699. https://doi.org/10.1002/ecy.2142

Goldberg, D. E., & Barton, A. M. (1992). Patterns and consequences of interspecific competition in natural communities: A review of field experiments with plants. *The American Naturalist*, 139(4), 771–801.

Gotelli, N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, 81(9), 2606–2621. https://doi.org/10.1890/0012-9658(2000)081[2606:NMAOSC]2.0.CO;2

Hojat, M., & Xu, G. (2004). A visitor's guide to effect sizes–statistical significance versus practical (clinical) importance of research findings. *Advances in Health Sciences Education*, 9(3), 241–249. https://doi.org/10.1023/B:AHSE.0000038173.00909.f6

Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Chichester, UK: Wiley & Sons.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124

LaManna, J. A., Mangan, S. A., Alonso, A., Bourg, N. A., Brockelman, W. Y., Bunyavejchewin, S., … Myers, J. A. (2017). Plant diversity increases with the strength of negative density dependence at the global scale. *Science*, 356(6345), 1389–1392.

Lan, G., Zhang, Y., He, F., Hu, Y., Zhu, H., & Cao, M. (2016). Species associations of congeneric species in a tropical seasonal rain forest of China. *Journal of Tropical Ecology*, 32(2016), 201–212. https://doi.org/10.1017/S0266467416000171

Law, R., Illian, J., Burslem, D. F. R. P., Gratzer, G., Gunatilleke, C. V. S., & Gunatilleke, I. A. U. N. (2009). Ecological information from spatial patterns of plants: Insights from point process theory. *Journal of Ecology*, 97(4), 616–628.

Lotwick, H. W., & Silverman, B. W. (1982). Methods for analysing spatial process of several types of points. *Journal of the Royal Statistical Society*, 44(3), 406–413.

Luo, Z. R., Yu, M. J., Chen, D. L., Wu, Y. G., & Ding, B. Y. (2012). Spatial associations of tree species in a subtropical evergreen broad-leaved forest. *Journal of Plant Ecology*, 5(3), 346–355. https://doi.org/10.1093/jpe/rtr048

Martinez, I., Wiegand, T., Gonzalez-Taboada, F., & Obeso, J. R. (2010). Spatial associations among tree species in a temperate forest community in North-western Spain. *Forest Ecology and Management*, 260(4), 456–465.

May, F., Huth, A., & Wiegand, T. (2015). Moving beyond abundance distributions: Neutral theory and spatial patterns in a tropical forest. *Royal Society Proceedings B*, 282, 20141657–20141657.

McGill, B. J. (2010). Towards a unification of unified theories of biodiversity. *Ecology Letters*, 13(5), 627–642. https://doi.org/10.1111/j.1461-0248.2010.01449.x

McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., … White, E. P. (2007). Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, 10(10), 995–1015.

Murrell, D. J. (2010). When does local spatial structure hinder competitive coexistence and reverse competitive hierarchies? *Ecology*, 91(6), 1605–1616. https://doi.org/10.1890/09-0832.1

Murrell, D. J., & Law, R. (2003). Heteromyopia and the spatial coexistence of similar competitors. *Ecology Letters*, 6(1), 48–59. https://doi.org/10.1046/j.1461-0248.2003.00397.x

Murrell, D. J., Purves, D. W., & Law, R. (2001). Uniting pattern and process in plant ecology. *Trends in Ecology & Evolution*, 16(10), 529–530. https://doi.org/10.1016/S0169-5347(01)02292-3

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., & Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 79(2), 381–404.

Nattrass, S., Baigent, S., & Murrell, D. J. (2012). Quantifying the likelihood of co-existence for communities with asymmetric competition. *Bulletin of Mathematical Biology*, 74(10), 2315–2338. https://doi.org/10.1007/s11538-012-9755-8

Open Science Collaboration, et al. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716.

Perry, G. L. W., Miller, B. P., & Enright, N. J. (2006). A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. *Plant Ecology*, 187(1), 59–82.

Perry, G. L. W., Miller, B. P., Enright, N. J., & Lamont, B. B. (2014). Stochastic geometry best explains spatial associations among species pairs and plant functional types in species-rich shrublands. *Oikos*, 123(1), 99–110. https://doi.org/10.1111/j.1600-0706.2013.00400.x

Perry, G. L. W., Miller, B. P., Lamont, B. B., & Enright, N. J. (2017). Community-level spatial structure supports a model of stochastic geometry in species-rich shrublands. *Oikos*, *126*(6), 833–842.

Punchi-Manage, R., Wiegand, T., Wiegand, K., Getzin, S., Huth, A., Gunatilleke, C. V. S., … Metcalfe, D. B. (2015). Neighborhood diversity of large trees shows independent species patterns in a mixed dipterocarp forest in Sri Lanka. *Ecology*, *96*(7), 1823–1834. https://doi.org/10.1890/14-1477.1

Rajala, T., Murrell, D. J., & Olhede, S. (2018). Detecting multivariate interactions in spatial point patterns with Gibbs models and variable selection. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *67*, 1237–1273. https://doi.org/10.1111/rssc.12281

Ripley, B. D. (1979). Tests of 'randomness' for spatial point patterns. *Journal of the Royal Statistical Society: Series B*, *41*(3), 368–374.

Sawyer, A. G., & Ball, A. D. (1981). Statistical power and effect size in marketing research. *Journal of Marketing Research*, *18*, 275–290. https://doi.org/10.2307/3150969

Silvertown, J. (2004). Plant coexistence and the niche. *Trends in Ecology & Evolution*, *19*(11), 605–611. https://doi.org/10.1016/j.tree.2004.09.003

Stoll, P., Murrell, D. J., & Newbery, D. M. (2015). Effect sizes and standardization in neighbourhood models of forest stands: Potential biases and misinterpretations. *Methods in Ecology and Evolution*, *6*(10), 1117–1125.

Stoll, P., & Newbery, D. M. (2005). Evidence of species-specific neighborhood effects in the dipteracarpaceae of a Bornean rain forest. *Ecology*, *86*(11), 3048–3062.

Uriarte, M., Condit, R., Canham, C. D., & Hubbell, S. P. (2004). A spatially explicit model of sapling growth in a tropical forest: Does the identity of neighbours matter? *Journal of Ecology*, *92*(2), 348–360. https://doi.org/10.1111/j.0022-0477.2004.00867.x

Uriarte, M., Swenson, N. G., Chazdon, R. L., Comita, L. S., Kress, W. J., Erickson, D., … Thompson, J. (2010). Trait similarity, shared ancestry and the structure of neighbourhood interactions in a subtropical wet forest: Implications for community assembly. *Ecology Letters*, *13*(12), 1503–1514.

Velázquez, E., Martínez, I., Getzin, S., Moloney, K. A., & Wiegand, T. (2016). An evaluation of the state of spatial point pattern analysis in ecology. *Ecography*, *39*(11), 1042–1055.

Velázquez, E., Paine, C. E. T., May, F., & Wiegand, T. (2015). Linking trait similarity to interspecific spatial associations in a moist tropical forest. *Journal of Vegetation Science*, *26*(6), 1068–1079.

Wang, Q., Bao, D., Guo, Y., Lu, J., Lu, Z., Xu, Y., … Huang, H. (2014). Species associations in a species-rich subtropical forest were not well-explained by stochastic geometry of biodiversity. *PLoS ONE*, *9*(5), e97300.

Wang, X., Wiegand, T., Kraft, N. J. B., Swenson, N. G., Davies, S. J., Hao, Z., … Wolf, A. (2016). Stochastic dilution effects weaken deterministic effects of niche-based processes in species rich forests. *Ecology*, *97*(2), 347–360. https://doi.org/10.1890/14-2357.1

Wiegand, T., Huth, A., Getzin, S., Wang, X., Hao, Z., Gunatilleke, C. V. S., & Gunatilleke, I. A. U. N. (2012). Testing the independent species' arrangement assertion made by theories of stochastic geometry of biodiversity. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1741), 3312–3320. https://doi.org/10.1098/rspb.2012.0376

Wiegand, T., Uriarte, M., Kraft, N. J. B., Shen, G., Wang, X., & He, F. (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*(1), 329–351.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.