

Methods

Using codispersion analysis to quantify and understand spatial patterns in species–environment relationships

Hannah L. Buckley^{1,2}, Bradley S. Case^{2,3}, Jess K. Zimmerman⁴, Jill Thompson^{4,5}, Jonathan A. Myers⁶ and Aaron M. Ellison²

¹Department of Ecology, Lincoln University, Lincoln 7647, New Zealand; ²Harvard Forest, Harvard University, 324 North Main Street, Petersham, MA 01366, USA; ³Department of Informatics and Enabling Technologies, Lincoln University, Lincoln 7647, New Zealand; ⁴Department of Environmental Sciences, University of Puerto Rico, San Juan 00936, Puerto Rico;

⁵Centre for Ecology & Hydrology, Bush Estate, Penicuik, Midlothian, EH26 0QB, UK; ⁶Department of Biology, Washington University in St Louis, St Louis, MO 63130, USA

Summary

Author for correspondence:

Hannah L. Buckley

Tel: +64 3 423 0736

Email: Hannah.Buckley@lincoln.ac.nz

Received: 14 September 2015

Accepted: 1 February 2016

New Phytologist (2016)

doi: 10.1111/nph.13934

Key words: anisotropy, bivariate, environmental gradient, forest dynamics plot, spatial analysis, species–environment, variogram.

- The analysis of spatial patterns in species–environment relationships can provide new insights into the niche requirements and potential co-occurrence of species, but species abundance and environmental data are routinely collected at different spatial scales. Here, we investigate the use of codispersion analysis to measure and assess the scale, directionality and significance of complex relationships between plants and their environment in large forest plots.
- We applied codispersion analysis to both simulated and field data on spatially located tree species basal area and environmental variables. The significance of the observed bivariate spatial associations between the basal area of key species and underlying environmental variables was tested using three null models.
- Codispersion analysis reliably detected directionality (anisotropy) in bivariate species–environment relationships and identified relevant scales of effects. Null model-based significance tests applied to codispersion analyses of forest plot data enabled us to infer the extent to which environmental conditions, tree sizes and/or tree spatial positions underpinned the observed basal area–environment relationships, or whether relationships were a result of other unmeasured factors.
- Codispersion analysis, combined with appropriate null models, can be used to infer hypothesized ecological processes from spatial patterns, allowing us to start disentangling the possible drivers of plant species–environment relationships.

Introduction

Environmental variability is a key driver of variation in biological diversity (Chesson, 2000). The analysis of the spatial patterns in species–environment relationships can reveal clues about the niche requirements of individual species and their potential for co-occurrence with other species (Silvertown, 2004). The quantification of spatial patterns of the distribution and abundance of species can illuminate scales of variation. These patterns often suggest experimentally testable hypotheses about multiple interacting processes that may drive species distribution and abundance patterns (Hubbell, 1979; Wiegand *et al.*, 2012).

The usual approach to relating spatial patterns of environmental gradients and populations of sessile organisms (e.g. plants, ant nests, barnacles) starts with the recording of the positions of individuals or, in the case of composite, plot-based measures, such as species richness or cover values, the positions of plots. This

enumeration yields a spatial point pattern (Dale, 1999). Environmental variables are then sampled, but they often are not measured at the same spatial grain as the point pattern. Examples include soil samples collected on a regularly spaced grid (John *et al.*, 2007; Turner & Engelbrecht, 2011), elevation and slope measurements derived from a digital elevation model (Franklin, 1995) or climate variables derived from a spatial database, such as ‘WorldClim’ (Hijmans *et al.*, 2005). Relationships between point patterns and environmental data can be analyzed using nonspatial methods that emphasize causal relationships (e.g. canonical correspondence analysis, Lepš & Šmilauer, 2003; species distribution models, Elith & Leathwick, 2009; or regression models, Shen *et al.*, 2009), or by spatial methods that deal with the visualization of pattern and the quantification of scales of variability in correlations; our focus here is on the latter.

The majority of the standard spatial descriptors used by ecologists, such as semivariograms, assume that the spatial processes

underlying the distribution of organisms (spatial point pattern), the associated environmental gradient and their covariation are stationary (spatial processes are invariant under translation) and isotropic (nondirectional) within the sampling extent (Cressie & Wikle, 2011; see Table 1 for the spatial terminology used in this paper). However, although these assumptions are convenient mathematically, they are typically unrealistic for most real-world examples.

First, the strong form of spatial stationarity (invariance under translation) is unlikely to be met in any real-world case. As a result, most spatial processes are assumed to have only

second-order stationarity: only the mean, variance and covariance need to be stationary (Vieira *et al.*, 2010). However, even second-order stationarity is unlikely in many ecological cases, and we assume only the 'intrinsic hypothesis' – that the mean and the semivariance of the distribution are dependent on interpoint distances, not specific locations (Vieira *et al.*, 2010). Second, in many ecologically realistic cases, environmental gradients create anisotropic patterns in the distributions or abundances of species, where changes in the distributions or abundances of species reflect changes in the magnitude of the environmental variable(s).

Table 1 Definitions of spatial terminology used in this paper

Term	Description	References
Anisotropy	When the spatial correlation is dependent on direction (opposite to isotropy, where the correlation is the same in all directions). For example, species across a stress gradient are anisotropic when associations vary between aggregated and segregated with decreasing stress (Bertness & Callaway, 1994)	Dale (1999)
Kernel bandwidth	The bandwidth is the set of parameters used in the kernel function of the codispersion analysis that is applied across all possible raster cell-to-cell distances for each spatial lag, resulting in a spatial variation surface. In the case of $20 \times 20\text{-m}^2$ grids, we apply a 20-m bandwidth because that is the smallest scale (spatial grain) of the data	Cuevas <i>et al.</i> (2013); Buckley <i>et al.</i> (2016); this work
Codispersion	A measure of the covariation of two variables in space. For example, covariation in the basal area of two tree species measured in $20 \times 20\text{-m}^2$ grid cells in a large forest plot	Cuevas <i>et al.</i> (2013); Buckley <i>et al.</i> (2016); this work
Marks	Attributes associated with each point in a spatial point pattern. For example, diameters or diseased/healthy status of trees in a forest plot	Wiegand & Moloney (2014)
Semivariogram	A function, usually plotted as a two-dimensional graph, revealing spatial correlation among measurements from a set of samples. It has three key parameters: nugget, sill and range. The semivariogram shows at what spatial lags spatial variability occurs in a spatial dataset, that is, the scale of variation in the data	Dale (1999)
Spatial autocorrelation	Dependence of observations on spatial proximity. For example, tree sizes may be spatially autocorrelated if growth is positively influenced by a patchily distributed environmental resource; high-resource patches will contain large trees and low-resource patches will contain small trees	Wiens (1989)
Spatial lag	The distance over which a process is measured. For example, when visualizing codispersion of a species and an environmental variable, we plot the codispersion for a range of spatial lags (and directions), that is, we ask, what is their covariation at distances (lags) of 20 m, 40 m, 60 m, ...?	Cuevas <i>et al.</i> (2013); Buckley <i>et al.</i> (2016); this work
Spatial point pattern	A set of locations in X, Y space. Spatial point patterns may be simply locations (unmarked pattern) or locations with attributes (marked pattern). For example, the X, Y coordinate locations of trees in a forest plot	Dale (1999); Wiegand & Moloney (2014)
Spatial processes	A process whose action causes changes in a spatial pattern	Wiens (1989)
Stationarity	The 'strong' form of spatial stationarity is the situation in which the joint distribution of the data is invariant when the pattern of either one is moved (translated) through space. A weaker form of spatial stationarity, 'second-order stationarity', assumes that only the mean, variance and covariance must be stationary. A still weaker form of stationarity – the 'intrinsic hypothesis' – is a lack of spatial trend, such that the mean and semivariance of the distribution are dependent only on the distance between points, not their locations. Either second-order stationarity or the intrinsic hypothesis is an assumption of most spatial statistical inference methods	Dale (1999); Vieira <i>et al.</i> (2010)

A familiar example of an anisotropic relationship between environmental gradients and species distribution arises from the ‘stress gradient hypothesis’ (Bertness & Callaway, 1994). This hypothesis posits that, as the environment becomes less stressful for species (e.g. salt spray decreases with distance from the high tide line), intra- or interspecific interactions switch from predominantly facilitative to predominantly competitive. As a result, the pattern of species distributions may shift from aggregated to regular (e.g. Malkinson *et al.*, 2003; Lingua *et al.*, 2008) or even hyperdispersed. Additional processes that may influence the clumping of species across environmental gradients include dispersal limitation, habitat filtering and density-dependent interactions with natural enemies (Condit *et al.*, 2000; Morlon *et al.*, 2008; McGill, 2010). Accurate identification of the underlying causes of such complex spatial patterns requires analytical methods that are sensitive not only to the spatial grain of the pattern, but also to nonstationarity and anisotropic changes over space.

Here, we illustrate how to use codispersion analysis (Cuevas *et al.*, 2013; Buckley *et al.*, 2016) to detect and display both isotropic and anisotropic spatial relationships between a spatial point pattern of the locations and attributes of species, and associated environmental variables measured at larger spatial grain. The analysis is based on the codispersion coefficient between the ecological characteristics of a plant species (e.g. the relative abundance, biomass, size or other functional trait) and an environmental variable in a given direction and within a given distance across a particular spatial extent, such as a plot. Codispersion analysis has been applied previously only to a few data types in ecology, including the relationship between tree size and an underlying environmental gradient (topography) at a landscape-level spatial extent (Cuevas *et al.*, 2013), multivariate spectral data (Vallejos *et al.*, 2015) and species co-occurrences (Buckley *et al.*, 2016). In this study, we apply codispersion analysis first to simulated data, and then to tree location and size (diameter) data from two large forest plots, one tropical and one temperate. Our results illustrate how codispersion analysis can be used to detect spatial patterns in tree size across environmental gradients. In addition, we demonstrate a framework for the use of different null models to test the significance of these spatial patterns (i.e. the departure of the observed patterns from random expectation), and how differences in significance among null model tests can be used to generate hypotheses about, and guide the structuring of, models of underlying spatial processes. Specifically, we ask, at a $20 \times 20\text{-m}^2$ grain size, what is the direction, magnitude and spatial pattern in covariation between selected tree species and environmental variables across these two large forest plots? For the purposes of illustrating this method, we selected common species that covaried with the environmental variables in a variety of ways to reflect some of the different underlying processes that can drive species–environment relationships. For example, we can explore whether covariation is higher between the basal area of a tree species and an environmental variable within 50 m in a northerly direction than would be expected if the species was randomly distributed.

Materials and Methods

An overview of codispersion analysis

Codispersion analysis quantifies the spatial covariation of two or more spatially explicit datasets. The result is a two-dimensional codispersion graph that allows us to assess how the two datasets covary across a range of spatial lags (distances between points) and directions (Table 1; Fig. 1; Vallejos *et al.*, 2015). Codispersion analysis can be applied to datasets organized as spatial point patterns, irregular plots or rasters. Spatial point patterns depict the locations of individuals (e.g. trees) and possible attributes (‘marks’) of these individuals (e.g. diameters or other functional traits) measured at these same locations. Rasters are often used to depict measurements of continuously varying soil or topographic properties as regular grids of cells of a particular size (resolution) from interpolations of variables that have been measured within the same vicinity as, but not precisely at the locations of, the point patterns. Spatial point patterns may also be converted (up-scaled) into rasters before codispersion analysis, such as by the quantification of tree abundances (stem density) or basal areas within raster cells of a given size.

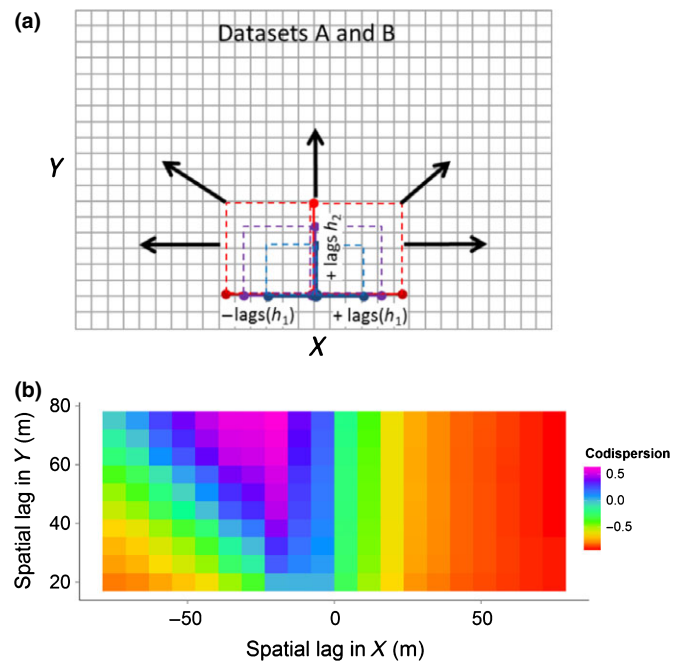


Fig. 1 (a) An illustration of the creation of directional spatial lags for ecological data organized as rasterized surfaces (both variables are represented by the large grid). The dashed lines represent different spatial lags h over which codispersion is calculated in different directions. (b) The codispersion graph. The color of each cell is the value of the codispersion coefficient of two variables for each given spatial lag h and direction in X, Y space. In this example, the graph shows negative covariation between the two variables when looking in the east direction, but positive covariation when looking in the northwest direction, indicating anisotropy in the way in which the two variables covary. The color pattern on the graph also indicates that the two variables are most negatively correlated at spatial lags > 20 m in the positive X direction, and most positively correlated at scales of $c. 20\text{--}30$ m in the negative X direction and at $c. 50\text{--}80$ m in the Y direction. Figures taken from Buckley *et al.* (2016).

In-depth statistical details of the mechanics of codispersion are given in Ruhkin & Vallejos (2008), Cuevas *et al.* (2013) and Buckley *et al.* (2016); in the latter, we consider species co-occurrences. Annotated R code (R v.3.1.2; R Core Team, 2014) for the performance of codispersion analysis, including its application to examples from this study, is provided in Supporting Information Notes S1.

In brief, codispersion analysis for two spatial datasets involves five steps.

The first involves the determination of the set of spatial lags $\mathbf{h} = \{h_1, h_2\}$: $\mathbf{h} \leq 0.25 \times$ maximum distance of the shortest side of the sample plot. The two components of \mathbf{h} are vectors representing the range of spatial lags to be analyzed for each input dataset A (e.g. tree basal area) and B (e.g. elevation above sea level). h_1 is oriented parallel to the x -axis, and ranges from $-h_{\max}$ to $+h_{\max}$ (Fig. 1a). h_2 is oriented parallel to the y -axis and ranges from 0 to h_{\max} (Fig. 1a). We note that two opposite directions are incorporated into the analysis along the x -axis (positive and negative), and so any anisotropy in the data will be more apparent along this axis. We therefore recommend that the dataset be oriented in such a way that the directionality of patterns of particular interest is along the x -axis direction, or that the data be rotated and analyzed in both directions.

Second, an Epanechnikov kernel function (Cuevas *et al.*, 2013) is applied across all possible raster cell-to-cell distances for each \mathbf{h} , resulting in a smooth spatial variation surface for each individual dataset and their intersection. The 'smoothness' of the kernel surfaces is controlled by a set of kernel bandwidth parameters $k = \{k_A, k_B, k_{AB}\}$ (Cuevas *et al.*, 2013). As rasterization of a spatial point process implies a uniform smoothing at the scale of the raster cell (Buckley *et al.*, 2016), when analyzing rasterized data, we recommend setting each element of k equal to the dimension of the raster cell to avoid unintentional repeated smoothing of the data.

Third, semivariograms for A and B and the semi-cross-variogram of the intersection of A and B are computed for the kernel-smoothed surfaces (Cuevas *et al.*, 2013).

Fourth, the empirical codispersion coefficient is computed for each lag \mathbf{h} as the semi-cross-variogram divided by the square root of the product of the semivariograms for each of the two variables. The value of the codispersion coefficient ranges from -1.0 (strong negative codispersion) to $+1.0$ (strong positive codispersion).

Finally, the codispersion values are plotted for each lag \mathbf{h} (Fig. 1b). The magnitude of the codispersion values on the graph, and the way in which codispersion values change across the graph, provide information on the strength and direction of covariation between the two datasets at different spatial grains (Fig. 1b).

Here, we first apply codispersion analysis to simulated data and use three null models to assess the significance of the observed patterns in both simulated and field data. We then apply codispersion analysis to explore the spatial relationships between tree basal areas and underlying environmental variables measured within multihectare forest plots. The results provide new insights into the potential processes underlying the observed

patterns, and can provide guidance for the development of flexible, mechanistic process-based models for the data.

Simulations

To illustrate how to apply and interpret codispersion analysis for species–environment relationships, we first generated and analyzed a range of species patterns on environmental gradients (examples in Fig. 2; the complete set of simulated patterns is given in Notes S2; R code to generate them is given in Notes S1, see later). We simulated marked point patterns in a $300 \times 300\text{-m}^2$ 'plot' by generating 1500 point locations (representing individual trees) that either were completely spatially random (CSR) or were generated by a Thomas process (using the `rThomas` function in the `spatstat` package of R; Baddeley & Turner, 2005). A Thomas process generates a clumped spatial distribution of points using parameters that describe the spatial intensity of the pattern (in this case, $\kappa = 20$ was used), the degree of variation within clumps ($\text{scale} = 0.05$) and the average number of points per cluster ($\mu = 10$). A simulated diameter (i.e. a 'mark') was assigned to each simulated 'tree'. Diameters were generated using a truncated lognormal distribution with minimum = 1, maximum = 80, mean = 40 and $\text{SD} = \log_e(80/15)$ cm. These marks were distributed across the 1500 trees either randomly, increasing or decreasing to the left side, right side, left or right top corners, or increasing as a large clump in the center of the plot (Fig. 2). We calculated the basal area of the simulated trees within each of 225 contiguous $20 \times 20\text{-m}^2$ cells within the simulated $300 \times 300\text{-m}^2$ plot; $20 \times 20\text{-m}^2$ cells were used because this is the size of typical forest inventory plots used to characterize stand structure. We then generated values for environmental variables within each raster cell. The values of the environmental variables were generated at 3600 points within the plot ($5 \times 5\text{-m}^2$ cells) and were distributed randomly among the cells or increasing or decreasing to the left side, right side, left or right top corners, or increasing towards a maximum in the center of the plot; these examples include gradient patterns at a range of angles and rotations. The environmental raster gridded into $5 \times 5\text{-m}^2$ cells was upscaled by taking the average value in $20 \times 20\text{-m}^2$ cells, so that the values were at the same locations and scale as the basal area data. For the codispersion analyses of these simulated data, we set the bandwidth $k = \{20\text{ m}, 20\text{ m}, 20\text{ m}\}$.

Forest plot data

We analyzed species–environment relationships between tree size (basal area) and environmental characteristics at two sites. The two datasets include environmental data that were collected in different ways: direct measurements in each raster cell, and spatial interpolation (downscaling) of sparser data to individual raster cells using kriging (John *et al.*, 2007).

The first dataset is from the third (2000–2002) complete census of the 16-ha Luquillo Forest Dynamics Plot (LFDP) at the Luquillo Long-Term Ecological Research Site, Puerto Rico (Thompson *et al.*, 2002). The four species selected were *Casearia arborea* (L. C. Rich.) Urban (Salicaceae), *Cecropia*

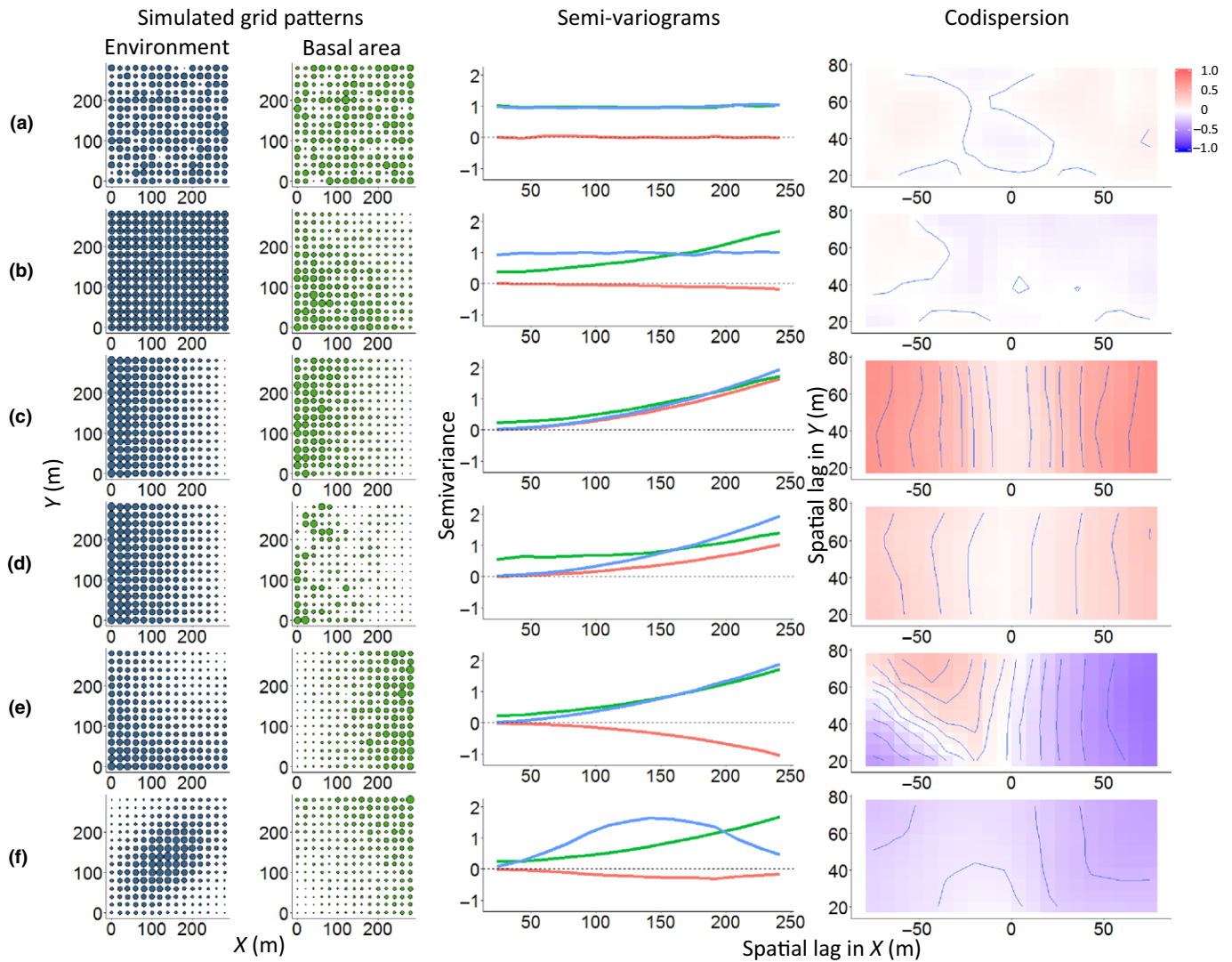


Fig. 2 Simulated species–environment patterns on $20 \times 20\text{-m}^2$ grids in $300 \times 300\text{-m}^2$ plots, their variograms and cross-variograms, and codispersion graphs. In the variograms, the blue line is the environment variogram, the green line is the species variogram and the pink line is the cross-variogram. The colors of the codispersion graphs are scaled from -1 (purple) to $+1$ (orange). The underlying pattern (environment, basal area) and mean (standard deviation) codispersion values for each analysis were: (a) CSR, CSR: 0.03 (0.04); (b) uniform, decreasing X and Y: -0.02 (0.03); (c) decreasing X, decreasing X: 0.46 (0.19); (d) decreasing X, decreasing X (underlying Thomas distribution): 0.25 (0.15); (e) decreasing X and Y, increasing X: -0.16 (0.29); and (f) bivariate normal, increasing X and Y: -0.23 (0.11).

schreberiana Miq. (Urticaceae), *Dacryodes excelsa* Vahl. (Burseraceae) and *Prestoea acuminata* var. *montana* (Willd.) H.E. Moore (Arecaceae). These are four of the most common species (out of 152 in total) in the third census of LFDP; together, they account for 44% of the total basal area of the plot (Table 2a). For each species, the basal area (m^2) of the main stem of each tree was calculated from its measured diameter; basal areas of all trees of a given species in each raster cell were summed to give the total basal area of the species for that cell. Elevation (range 333–428 m above sea level) was measured (1990–1992) and mean elevation was calculated for each cell as the mean of the elevations at the four corners of each $20 \times 20\text{-m}^2$ cell (Thompson *et al.*, 2002). Slope (range -0.7 to 65%) was calculated from the corner elevations of each $20 \times 20\text{-m}^2$ cell (Thompson *et al.*, 2002).

The basal area of *Casearia* and *Prestoea* decreases, but the basal area of *D. excelsa* increases, with elevation in LFDP as a result of the pattern of land-use history in the plot (Thompson *et al.*, 2002). The northern (lower elevation) two-thirds of the plot were logged before 1934 and used for subsistence agriculture. Logging and agriculture ceased when the area was purchased in 1934, and the regenerating forest is dominated by *Casearia*, but *Prestoea* also has its highest basal area there. *Prestoea* is often associated with slopes and ravines and disturbed areas (Weaver, 2010; Harris *et al.*, 2012). At the highest elevations and the southern third of the plot, human disturbance to the forest was limited to selective logging; *Dacryodes* dominates these areas of the plot (Thompson *et al.*, 2002). The dominance of *Cecropia* in the northern portion of the plot recorded in the third census is thought to have resulted from interactions between land-use history and hurricane

Table 2 Abundances, mean diameters (diameter at breast height, dbh) in centimeters (SD), and the means and ranges in codispersion for basal area–environment relationships for the analyzed species in the (a) Luquillo Forest Dynamics Plot and (b) Tyson Research Center Forest Plot

(a) Luquillo Forest Dynamics Plot (2000–2002 census data)							
Species	Number of stems	Mean dbh (SD)	Total basal area (m ² h ⁻¹)	Mean (SD) codispersion with elevation	Range in codispersion with elevation (min, max)	Mean (SD) codispersion with slope	Range in codispersion with slope (min, max)
<i>Dacryodes excelsa</i>	1544	21.18 (15.71)	84.28	0.00 (0.08)	−0.17, 0.14	0.03 (0.02)	−0.03, 0.10
<i>Cecropia schreberiana</i>	2902	10.02 (6.65)	32.95	0.14 (0.04)	0.06, 0.22	0.11 (0.06)	−0.05, 0.25
<i>Casearia arborea</i>	3861	5.63 (5.38)	18.39	0.05 (0.09)	−0.12, 0.21	−0.13 (0.06)	−0.24, 0.02
<i>Prestoea acuminata</i>	7707	14.29 (2.96)	128.82	−0.10 (0.07)	−0.24, 0.02	0.10 (0.03)	0.02, 0.17
(b) Tyson Research Center Plot (2013 census data)							
Species	Number of stems	Mean dbh (SD)	Total basal area (m ² h ⁻¹)	Mean (SD) codispersion with soil PC1	Range in codispersion with soil PC1 (min, max)	Mean (SD) codispersion with soil PC2	Range in codispersion with soil PC2 (min, max)
<i>Frangula caroliniana</i>	8715	2.04 (0.85)	3.34	0.41 (0.12)	0.17, 0.62	0.03 (0.10)	−0.16, 0.21
<i>Lindera benzoin</i>	4922	1.84 (0.66)	1.48	0.28 (0.14)	0.06, 0.56	0.06 (0.13)	−0.11, 0.36
<i>Quercus alba</i>	2066	29.57 (16.24)	184.66	−0.04 (0.04)	−0.14, 0.07	0.13 (0.05)	0.03, 0.24
<i>Quercus rubra</i>	1551	30.03 (17.63)	147.73	−0.39 (0.12)	−0.56, −0.15	0.03 (0.05)	−0.06, 0.13
<i>Quercus velutina</i>	691	33.46 (13.92)	71.27	−0.09 (0.09)	−0.28, 0.08	−0.09 (0.05)	−0.19, 0.03

Codispersion was estimated in the 20 × 20-m² raster cells in which environmental variables were measured.

disturbance. *Cecropia* recruited in huge numbers following Hurricane Hugo in September 1989 (Zimmerman *et al.*, 2010), such that > 95% of *Cecropia* individuals of this species recruited following this one disturbance event. Zimmerman *et al.* (1994) noted that *Casearia* was especially susceptible to uprooting during Hurricane Hugo, which opened the forest canopy. Walker (2000) found that *Cecropia* frequently recruited in soil pits caused by uprooted trees and survived longer in this area of the plot because of the persistence of canopy light gaps. Thus, the prevalence of *Cecropia* in the lowermost elevation and flatter northern portion of the plot may be the result of hurricane damage caused to *Casearia* and other species in this portion of the plot.

The second dataset is from the Tyson Research Center Plot (TRCP), a 25-ha forest dynamics plot located at Washington University in the St Louis Tyson Research Center, MO, USA (Spasojevic *et al.*, 2014). We analyzed species–environment relationships for five woody species in the central 20-ha of the plot: *Frangula caroliniana* (Walter) A. Gray (Rhamnaceae), *Lindera benzoin* L. Blume (Lauraceae), *Quercus alba* L., *Q. rubra* L. and *Q. velutina* Lam. (Fagaceae). The three *Quercus* species were some of the most widespread species in the plot, whereas *Frangula* and *Lindera* were selected because they were the two most abundant species in the plot and had interesting, highly clumped spatial patterns. Together, these five species comprised 78% of the total basal area of TRCP in the 2013 census (Table 2b). Principal components (PC) analysis (see Notes S3) was used to summarize, in two composite principal axes, the variation in 17 physicochemical soil properties that were measured at points across TRCP in 2013 and kriged to 20 × 20-m² raster cells (Spasojevic *et al.*, 2014). Maps of individual environmental variables are available on the TRCP website (<http://www.ctfs.si.edu/site/Tyson+Research+Center%2C+Missouri>) and the data used in this paper are provided in Tables S1, S2.

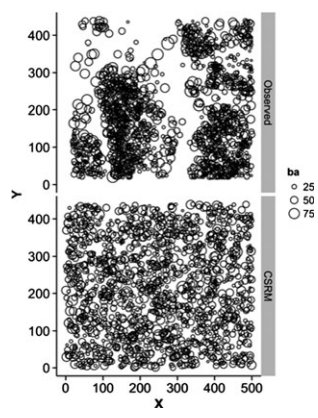
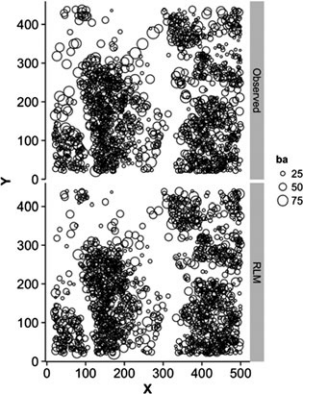
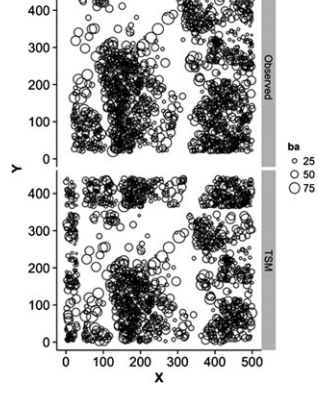
Null model analyses

To assess the significance of the observed codispersion patterns, we used three different null models to randomize aspects of the spatial point processes and their marks (diameters) (Table 3). In each, only the species location data, rather than both species and environment data, were randomized, because this was sufficient to break any spatial association of the species data with the environmental variable and allowed us to test the significance of their covariation. The three null models were a CSR model (CSR), a random labeling model (RLM) and a toroidal shift model (TSM) (see Wiegand & Moloney (2014) for detailed descriptions of these null models and other examples of their use).

The CSR generated new spatial locations for trees; the observed tree diameters were then assigned randomly (without replacement) to each tree at its new location. Comparison of the observed codispersion patterns with those generated by this null model tested whether there was any nonrandom spatial pattern in the covariation of the observed tree population (basal area within 20 × 20-m² grid cells) and the environmental variable (Table 3). One difficulty with CSR is that where species distributions are clumped, this may result in a Type I error rate that is higher than 0.05. Thus, a significant departure from the expectation of this null model may reflect the presence of clumping in the distribution of species (Table 3) and the interpretation of a significant result must be made with caution. For example, we can use a CSR to ask whether a species increases in basal area at lower elevations in the plot, but, if the spatial distribution of the species is clumped, we could obtain a ‘significant’ result even if there was no relationship between basal area and elevation. Overall, however, this significance test can be used as an initial test for spatial nonrandomness in the dataset.

The RLM permuted the observed diameters of the trees whilst retaining the observed spatial position of each tree. This null

Table 3 The three null models, an example realization of each, how they were applied in this paper and their associated null process models: for each example (which was randomized by each null model), the hypothesized ecological process is that basal area (BA) is conditional on one or more of the spatial point patterns of trees (ppp), their diameters (marks) and the spatial distribution of the environmental variable (env): BA | (ppp, marks, env)

Null model	Example	Null process	Test
Completely spatially random model (CSR)		<p>BA; (ppp, marks) env</p> <p>The spatial distribution and diameters of individual trees, from which basal area is computed, are random and therefore independent of the environment</p>	This model tests for nonrandom spatial covariation between BA and the environmental variable
Random labeling model (RLM)		<p>BA; marks (ppp, env)</p> <p>Where individual trees grow is fixed (as a result of another process, such as competition), but how they grow (size) is independent of the environment</p>	This model tests whether the environmental variable is associated with growth differences among individual trees, whose diameters are aggregated to compute BA in each raster cell
Toroidal shift model (TSM)		<p>BA; env (ppp, marks)</p> <p>Where trees grow relative to one another and the spatial distribution of their relative sizes are driven by an unknown (unmeasured) process, but where and how they grow (e.g. size) is independent of the environment</p>	This model tests for nonrandom spatial covariation between BA and the environmental variable, given the underlying marked spatial point pattern of the species

Each null model breaks apart this conditional process in a different way, as is indicated by the conditional statement (in bold type) and its associated explanation in the 'Null process' column.

model tested whether, given the underlying spatial distribution of trees (a particular autocorrelation structure), their sizes were important in determining any covariation with the environmental variable (Table 3). For example, under this null model, we can test whether covariation between basal area and soil fertility is a result of differences in the growth rates of species along a soil

fertility gradient, rather than changes in stem density. Mechanistically, in this example, the tree distributions may be driven by clumped dispersal processes that are uniform across the plot area, but the growth rates of species may vary with soil fertility.

The TSM retained the autocorrelation structure of the tree populations by retaining their relative spatial positions and

diameters, but breaking their spatial association with the environmental variable by moving the entire species pattern in a random distance and direction as though the plot was a torus. This model tested whether the observed pattern in covariation between the species and environmental variable was the same in all parts of the plot, that is, whether the pattern in covariation is stationary (Table 3). TSM is similar to CSR in that it completely breaks any association between the two variables, but it fixes the distribution pattern of the species. Thus, it distinguishes the case in which a nonrandom codispersion pattern may simply be driven by relative tree positions from a process-based link between the environment and the species. For example, under this null model, we ask whether tree basal area varies with soil fertility and whether the nature of that covariation is the same throughout the plot. When combined with the results of CSR, we can determine whether nonrandomness identified using CSR is a result of a species–environment relationship (significant TSM) or of clumping in the species distribution (nonsignificant TSM) (Table 3).

For each species, each of the three null models was used to generate 199 new datasets. For each species–environment combination, empirical tail probabilities were obtained by comparing the observed codispersion values at each spatial lag with the vector of codispersion values at the same spatial lags and directions determined from each null model. If the observed value was greater than or equal to the 195th null value or less than or equal to the fifth null value, we deemed it to be significantly different from expected (i.e. a two-tailed test; $P < 0.05$). Thus, the significance tests were made for each lag and direction for which we obtained a codispersion value.

Finally, we determined the Type I error rate for each of the three null models by comparing the observed codispersion between two CSR simulated patterns (see Notes S4) with values generated by CSR, RLM and TSM. It should be noted that the Type I error rate, our ability to identify nonsignificant codispersion values, is invariant to rotation, and the error rate tests of the null models do not address the Type II error rate (statistical power), which remains an issue of ongoing research. R code for the null model analysis is provided in Notes S1.

Results

Species–environment associations of simulated forest plot data

Codispersion plots clearly illustrated the relationships between simulated species and their environment, and detected anisotropic, positive and negative covariation between the two variables (Fig. 2). When the simulated environmental pattern was generated using a CSR process, the cross-variogram and the codispersion were both approximately zero (little or no spatial covariation), whether or not the spatial pattern in basal area was also CSR (Fig. 2a; extended results in Notes S2). When the environmental variable was generated using a uniform process across the plot, but the basal area of the species decreased from the bottom left to the top right of the plot (i.e. southwest to northeast),

the codispersion was weakly negative and weakly anisotropic. This result reflected the changing pattern of covariation in the two variables in the X - and Y -directions. By contrast, the cross-variogram was approximately zero (Fig. 2b). Sequential pattern rotations of 15° showed that codispersion analysis can also distinguish smaller changes in pattern orientation (Notes S2).

When basal area tightly covaried with the environmental variable, the cross-variogram steeply increased and the codispersion was very high, only weakening at smaller scales that approached the spatial grain of the pattern (Fig. 2c). This pattern, and indeed all pattern combinations, had lower codispersion values when the underlying point pattern of the species was clumped (Thomas process) rather than CSR (Fig. 2d; extended results in Notes S2). A difference in pattern between the left- (west) and right-hand (east) sides of the codispersion graph indicated anisotropy. For example, where the environmental variable decreased from bottom left (southwest) to top right (northeast), and the basal area increased from west to east, codispersion measured negative covariation in the west-to-east direction, but showed some positive covariation at larger scales when looking to the northeast and negative covariation at larger scales when looking to the east (Fig. 2e). This pattern was also reflected somewhat in the cross-variogram, which was flat at small lags, but negative at larger lags (Fig. 2e). Similarly, where there was some covariation in a given direction (Fig. 2f), in this case from bottom left (southwest) to top right (northeast), the codispersion map illustrated the anisotropy (the right-hand side of the plot was more negative than the left-hand side), showing a relationship that was more negative at larger scales. In this case, the cross-variogram was most negative at similar scales (100–150 m), but did not reflect the anisotropy (Fig. 2f).

For all analysis combinations of the three null models and the two underlying tree distributions (CSR and Thomas process), none of the observed codispersion values from the two CSR patterns was significantly different from that expected under either model at the 5% level. In our simulations, the CSR model resulted in only one significant cell (out of 200 cells) in the codispersion graph (see Notes S4). These results are indicative of a Type I error rate of $\leq 5\%$.

Species–environment associations of observed forest plot data

In LFDP, the basal area of *Casearia*, *Cecropia* and *Prestoea* generally decreased with increasing elevation, whereas the basal area of *Dacryodes* increased with increasing elevation (Fig. 3; Table 2a), reflecting the interaction of elevation and land-use history in the plot (Thompson *et al.*, 2002). For *Casearia*, this pattern was reflected in a weak, anisotropic codispersion pattern, where west-to-east codispersion was more positive than east-to-west codispersion, which became more negative in the northeast direction (Fig. 4a). The codispersion was weakly negative and anisotropic for the basal area of *Cecropia* (Fig. 4b), and similar, but positive, for that of *Dacryodes* (Fig. 4c). The basal area of *Prestoea* negatively covaried with elevation at the larger scales, reflecting its lower basal area at the highest elevations (Fig. 4d). The basal area

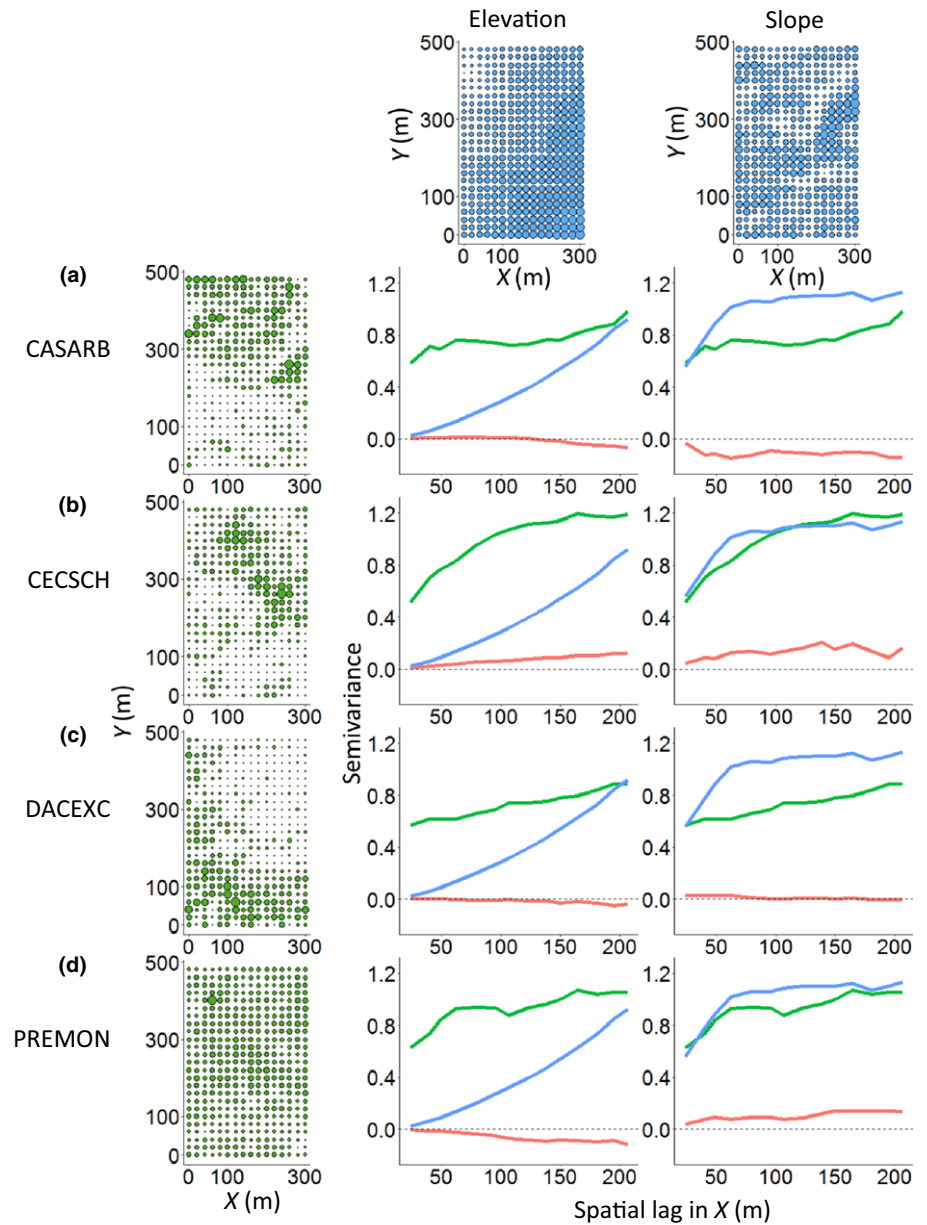


Fig. 3 Observed patterns on $20 \times 20\text{-m}^2$ grids in the 16-ha Luquillo Forest Dynamics Plot of elevation (top left), slope (top right) and basal area ($\text{m}^2 \text{ha}^{-1}$) of (a) *Casearia arborea* (CASARB), (b) *Cecropia schreberiana* (CECSCH), (c) *Dacryodes excelsa* (DACEXC) and (d) *Prestoea acuminata* (PREMON). The variogram for the environmental variable (blue line), variogram for the species (green line) and their cross-variogram (pink line) are shown for each species–environment combination; variables were centered and standardized before analysis. In each bubble plot, the dots are positioned at the center of each grid cell point and the sizes of the symbols are scaled to the variable displayed.

of *Casearia* negatively covaried with slope, whereas the basal area of *Cecropia* and *Dacryodes* positively covaried with slope. By contrast, the basal area of *Prestoea* was not strongly related to slope.

The comparison of the observed patterns with the codispersion values from CSR randomizations revealed that the observed codispersion for all of the species with both elevation and slope was different from random expectation at some, but not all, scales and directions (Fig. 4, columns 2 and 3). The only exception was for the relationship between *Prestoea* and slope, which was not significant (Fig. 4d). For all four species, the comparisons with RLM showed that the number of significant observed codispersion values was lower than expected using CSR for about one-half of the relationships tested, was higher for some and stayed the same for a few (Fig. 4, columns 4 and 5). The comparisons with TSM showed that the observed codispersion values were

significant at a few scales and directions for most species–environment combinations (Fig. 4, columns 6 and 7).

In TRCP, the first two components from the PC analysis of the soil chemistry data explained 65% of the variation in measured soil chemistry (plots and PC loadings are given in Notes S2). Variables loading strongly on PC1 were associated with soil fertility and cations (i.e. pH, base saturation, calcium, magnesium, potassium, aluminum and iron), whereas variables loading strongly on PC2 were associated with soil nitrogen availability (i.e. total nitrogen, NH_4 and nitrogen mineralization rate). These two PCs were used in the codispersion analysis of species–environment relationships for the five focal species.

The basal area of the five focal species in the $20 \times 20\text{-m}^2$ raster cells at TRCP showed a range of strong, weak, positive and negative relationships with both soil pH and cations (PC1) and soil

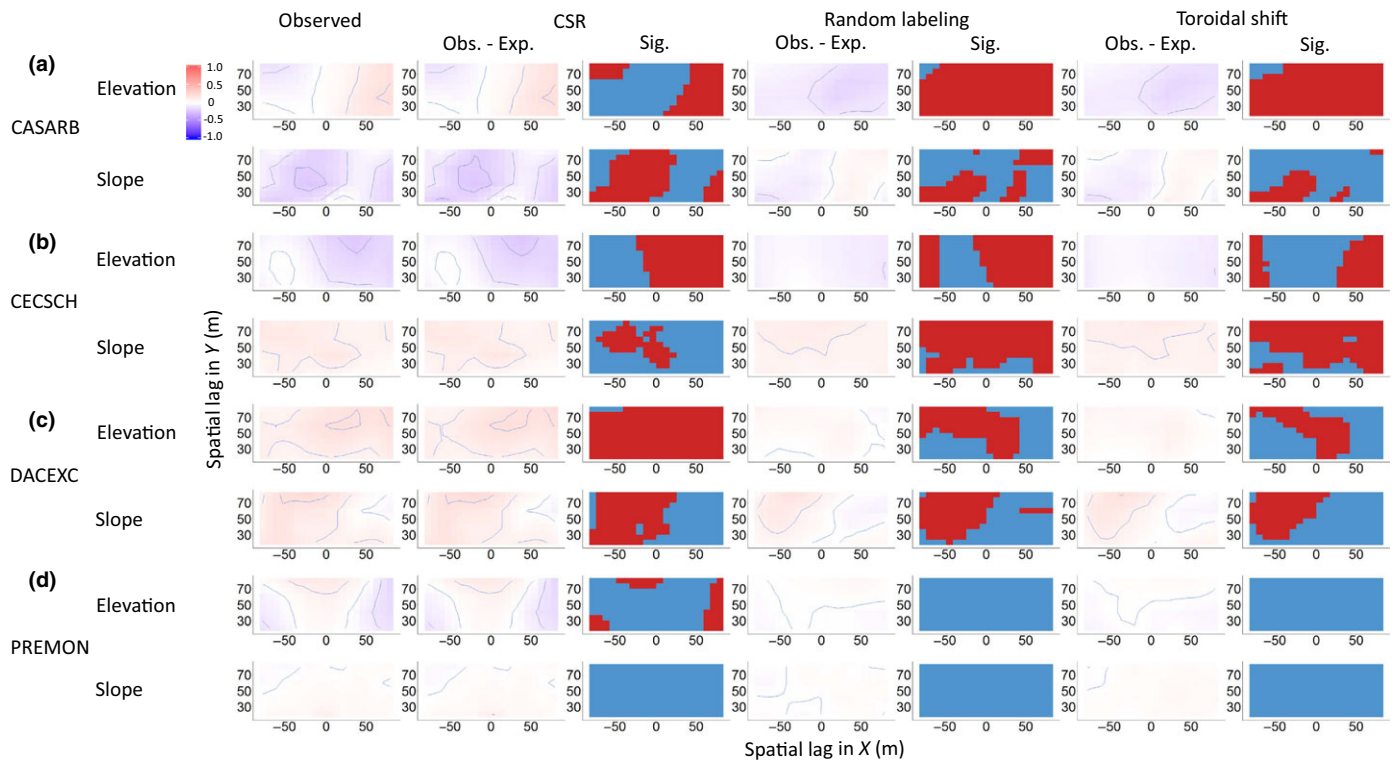


Fig. 4 Observed codispersion values, observed minus expected values and significance (red) or not (blue) at the $P < 0.05$ level relative to null expectation from three null models for bivariate species–environment combinations for four species (abbreviations as in Fig. 3) in the 16-ha Luquillo Forest Dynamics Plot. The colors on the codispersion and observed–expected graphs are scaled from -1 (purple) to $+1$ (orange); contour lines are at intervals of 0.1 . The means and ranges of the observed codispersion values are given in Table 2(a). CSR, completely spatially random.

nitrogen (PC2) (Table 2b; Fig. 5). Although abundant, *Frangula* and *Lindera* were less widespread and their populations were concentrated largely in one or a few patches that corresponded to high values on PC1, generating positive covariation (Fig. 5a,b). The three *Quercus* species (Fig. 5c–e) were more widespread within the plot; *Q. alba* was weakly and *Q. rubra* and *Q. velutina* were more strongly negatively related to more fertile soils (high values on PC1). *Quercus alba* positively covaried with nitrogen (PC2), whereas *Q. rubra* and *Q. velutina* had little or negative covariation with nitrogen (Fig. 5c–e).

Codispersion plots revealed both spatial gradients in covariation between basal area and environment and the spatial scales at which covariation was the strongest (Fig. 6, column 1). For example, anisotropic species–environment associations for *Frangula* and *Lindera* were illustrated by positive codispersion with PC2 to the east within the plot, but negative codispersion when looking to the west (Fig. 6a,b). In addition, the spatial scales of covariation differed among species. For instance, the positive covariation between *Q. alba* and PC2 was highest at large lags (> 50 m) in the east–west direction, whereas *Q. velutina* negatively covaried with PC1 at larger lags (> 60 m) in the north direction, but at smaller lags in the east–west direction (up to 50 m).

The observed patterns of species–environment associations at TRCP often differed from null expectations, but the magnitude of the effect sizes varied among the different null models. The

comparison of the observed codispersion patterns with those from the null models revealed that the weaker observed codispersion patterns with both soil fertility and cations (PC1) and soil nitrogen variables (PC2) tended not to be significant when compared with expectation when trees were distributed CSR within the plot (Fig. 6, columns 2 and 3). By contrast, comparisons with RLM (Fig. 6, columns 4 and 5) showed that the observed codispersion values were mostly higher than expected. The exceptions to this were, for some scales and directions, for *Frangula* and *Q. velutina* with PC2, and for *Q. rubra* with PC1, each of which had significantly more negative codispersion at some scales when looking to the west in the plot. The comparisons with the expected values from TSM largely mirrored those of the CSR comparisons, but with fewer significant values in most cases, such as for *Frangula* and PC2, which was nonsignificant at all lags.

Discussion

Codispersion analysis is a useful method for exploring species–environment relationships in a spatially explicit context. Simulations showed that the method correctly detected anisotropy and other spatial regularities in the covariation of the two variables, and correctly measured the scale of these effects (Fig. 2). Codispersion values in these simulations were influenced by the underlying spatial pattern of both the species and the environmental variable; more clumping in the tree distribution patterns reduced

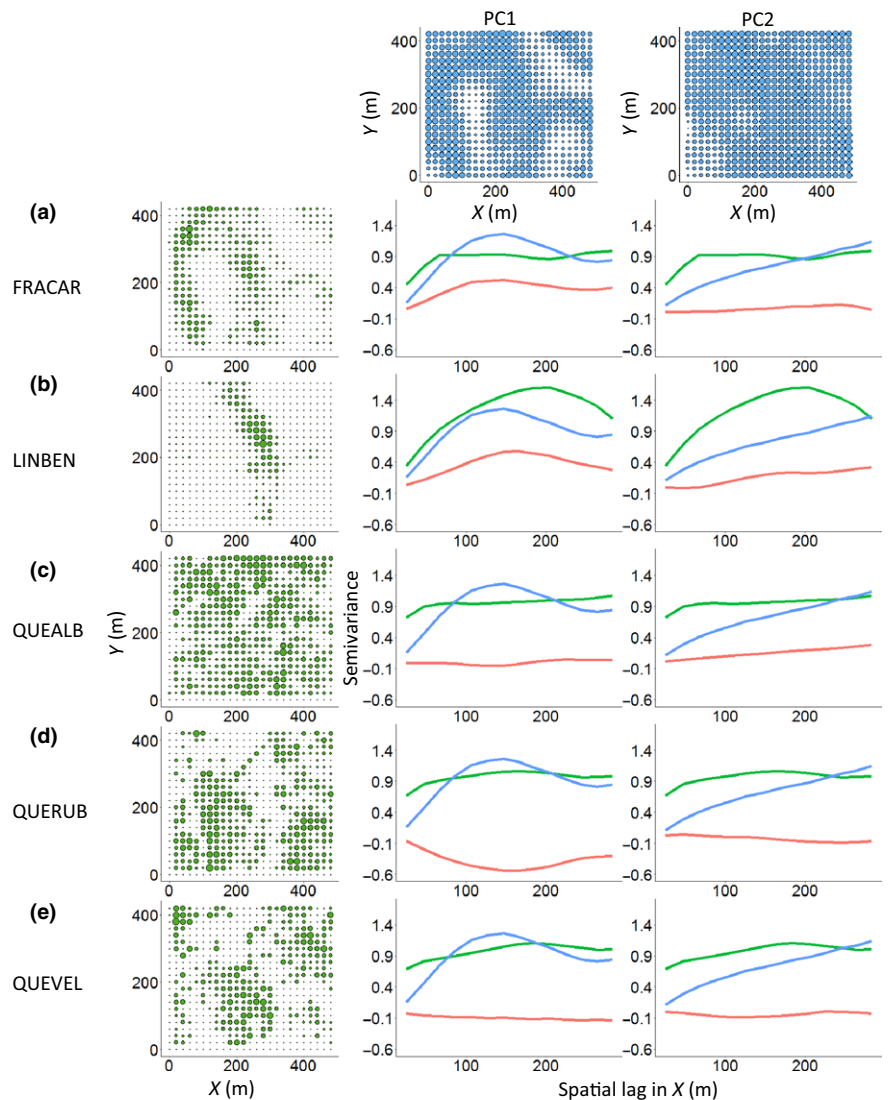


Fig. 5 Observed patterns on $20 \times 20\text{-m}^2$ grids in a 20-ha area of the Tyson Research Center Plot of soil variables represented by two principal components, PC1 (top left) and PC2 (top right), and basal area ($\text{m}^2 \text{ha}^{-1}$) of five species: (a) *Frangula caroliniana* (FRACAR), (b) *Lindera benzoin* (LINBEN), (c) *Quercus alba* (QUEALB), (d) *Quercus rubra* (QUERUB) and (e) *Quercus velutina* (QUEVEL). The variogram for the environmental variable (blue line), variogram for the species (green line) and their cross-variogram (pink line) are shown for each species–environment combination; variables were centered and standardized before analysis. In each bubble plot, the dots are positioned at the center of each grid cell point and the sizes of the symbols are scaled to the variable displayed.

the magnitude of the codispersion values, even with the same basal area and environmental gradients (Fig. 2; Notes S2). Similarly, a uniform distribution of the environmental variable led to a higher magnitude of codispersion values than resulted from a CSR environmental variable (Fig. 2; Notes S2). When observed patterns in field data were combined with null model analysis, codispersion analysis detected the scales and directions of statistically significant codispersion in basal area–environment relationships, and suggested the possible drivers of these relationships (Table 2).

The selection of appropriate null models for the analysis of spatial point patterns is especially important when the results are used to generate testable hypotheses about processes underlying the observed point patterns (Wiegand & Moloney, 2014). We suggest that comparisons of the results of the three null models used here to explore the significance of codispersion in species–environment relationship can help to tease apart possible influences on observed codispersion patterns (Table 4). In particular, whether observed patterns are

found to be significantly different from expectations for one, two or all three of the null models leads to different hypotheses about possible processes and ecological mechanisms determining the observed patterns (Table 4).

The first possibility is that the observed pattern is not significantly different from expectation of all three null models. We obtained this result when examining the codispersion of *P. acuminata* and slope at LFDP (Fig. 4d). We interpret this result as evidence that any observed spatial pattern of the basal area distribution of this species must be caused by factors that we did not measure. For example, *Prestoea* is dominant in the northern two-thirds of LFDP, which was disturbed by the land-use history, greater damage from Hurricane Hugo and is flatter than the southern third of the plot. The high abundance in the northern part of the plot as a result of the land-use history reduces the relative strength of the association with slope in this analysis. A second possibility is that the pattern is significantly different under CSR, but nonsignificantly different under TSM. This probably reflects the situation in which clumping in the species

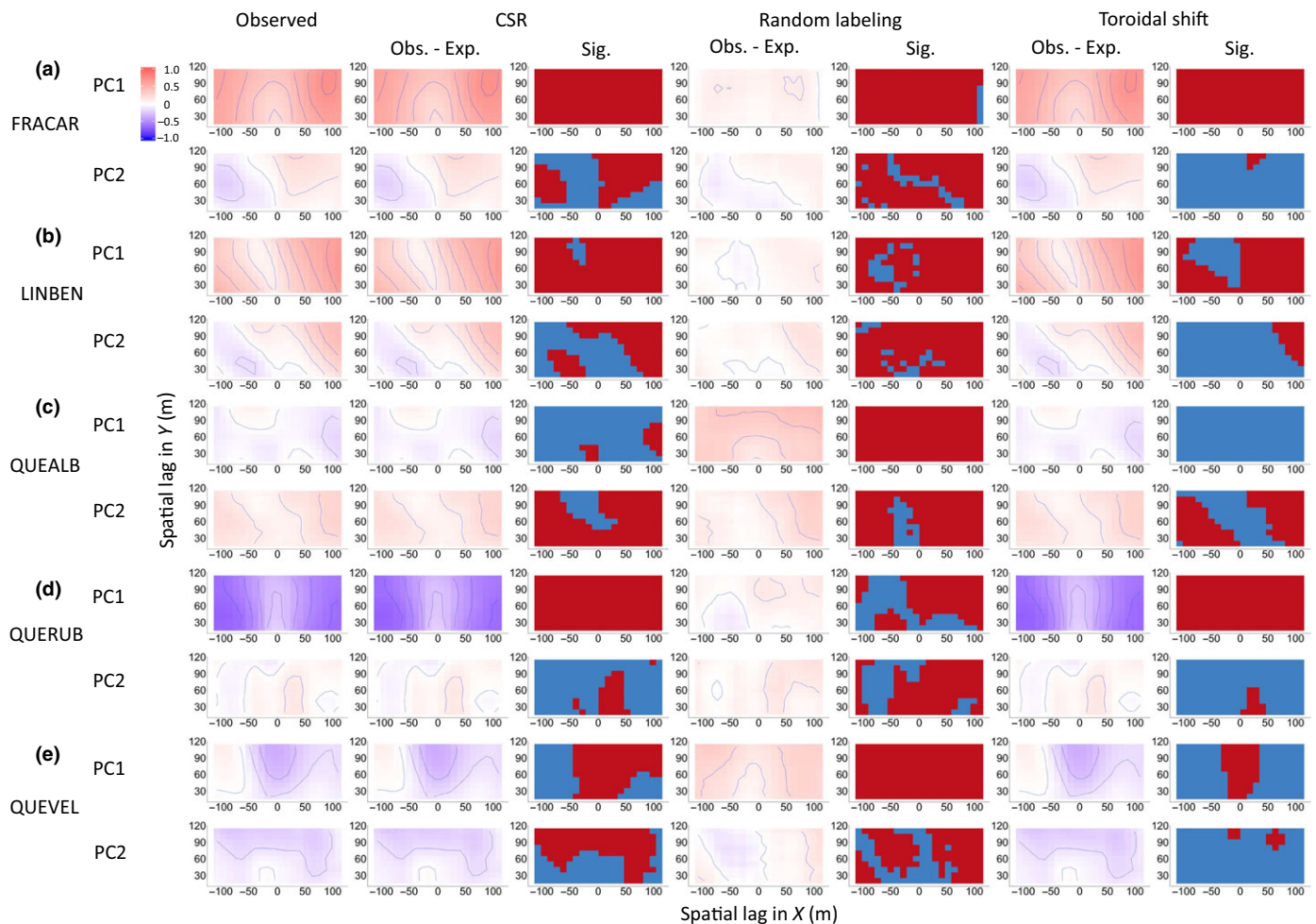


Fig. 6 Observed codispersion values, observed–expected values and significance (red) or not (blue) at the $P < 0.05$ level relative to null expectation from three null models for bivariate species–environment combinations for five species (abbreviations as in Fig. 5) in the 22-ha area of the Tyson Research Center Plot. The colors of the codispersion and observed–expected graphs are scaled from -1 (purple) to $+1$ (orange); contour lines are at intervals of 0.1 . The means and ranges of observed codispersion values are given in Table 2(b). CSR, completely spatially random.

distribution has resulted in a correlation with environment at some lags and directions, but this is not consistent across the plot, and therefore unlikely to reflect a causal dependence of species on environment. Such a result can be used to identify and understand spatial pattern in the species data.

Alternatively, the observed pattern could be significantly different from expectation for only two of the three null models. For example, at TRCP, *Q. rubra* was strongly and negatively associated with soil pH and cations at all spatial lags when assessed with CSR and TSM (Fig. 6d). However, spatial covariation was nonsignificant for a number of lags under RLM and, where it was significant, the observed codispersion was higher than expected. This suggests that, although *Q. rubra* basal area was negatively related to the soil environment, the pattern of this relationship, at least at some spatial lags and directions, was not dependent on tree size, but rather on their relative spatial positions (autocorrelation structure). Thus, the observed codispersion pattern is likely to be caused by processes that drive intraspecific clumping, such as unmeasured variation in other environmental variables or land-use history (Thompson *et al.*,

2002), interspecific interactions or dispersal limitation (e.g. Plotkin *et al.*, 2002).

Further, significant difference from expectation under TSM reveals nonstationarity in the data, which should be taken into account in subsequently developed inferential statistical models. For example, variograms for TRCP show nonstationarity in PC2 (a large-scale trend such that the variogram does not level off and therefore has no sill). The observed codispersion of PC2 (soil nitrogen variables) and *Q. alba* was significantly different from expectation at large scales, suggesting that there was nonstationarity in this pattern. If, in a subsequent model, we were interested in regressing this covariation against other variables, such as slope or elevation, we would need to account for the nonstationarity by applying a method, such as generalized least squares, in which the correlation in the errors is modeled and then specified in the regression model (Beale *et al.*, 2010).

These results, and others summarized in Table 4, demonstrate how the application of different null models to codispersion analysis can reveal subtle differences in potential causes of observed bivariate spatial relationships. Other null models that could be

Table 4 Interpretation of the null model results with examples from the two forest plot datasets

Null model results			Interpretation	Species–environment examples
CSRSM	RLM	TSM		
ns	ns	ns	Basal area is independent of the environment	<i>Prestoea acuminata</i> vs slope (Fig. 4d)
Sig.	ns	ns	Basal area is independent of the environment but aggregated in space; this pattern depends on tree spatial distributions, not tree sizes, that is, the spatial pattern of basal area is not different from expected if diameters are randomly assigned to trees	<i>Casearia arborea</i> vs elevation (Fig. 4a)
ns	Sig.	ns	Basal area is not strongly related to the environment because tree positions are independent of the environmental variable; however, the environment causes nonrandom differences in tree growth	<i>Quercus alba</i> vs PC1 (Fig. 6c)
Sig.	ns	Sig.	Basal area is nonrandomly related to the environment; this pattern depends on the relative spatial positions of trees, not their sizes	<i>Quercus rubra</i> vs PC1 (Fig. 6d)
Sig.	Sig.	ns	Tree sizes, but not necessarily their positions, depend on the environment (the environment causes differences in tree growth; tree distributions are aggregated within the plot)	<i>Cecropia schreberiana</i> vs elevation (Fig. 4b)
Sig.	Sig.	Sig.	Basal area is nonrandomly related to the environment and this depends on both tree spatial distributions and their sizes. The environment influences both where trees grow and their sizes	<i>Frangula caroliniana</i> vs PC1 (Fig. 6a)

The completely spatially random model (CSRSM) resulted in CSR tree spatial positions within the plot. The random labeling model (RLM) shuffled the marks (here, diameters) associated with each tree. The toroidal shift model (TSM) fixed the relative tree positions and their observed diameters, but moved the entire set of tree point locations in a random distance and direction as though the plot was a torus. ns, not significant; Sig., significant.

explored fruitfully in further research include pattern reconstruction methods (Wiegand & Moloney, 2014, p. 368) and spectral methods using raster data (Deblauwe *et al.*, 2015; Wagner & Dray, 2015). However, we must first understand what biological processes are being manipulated in each case to interpret observed departures from null expectations. Further, simultaneous comparisons across multiple lag distances can suffer from higher than desired Type I error rates (Loosmore & Ford, 2006; Baddeley *et al.*, 2014). Future research should address the development of a global significance test for codispersion where understanding scales of variation is important.

Finally, we note that there are three important considerations to keep in mind when applying codispersion analysis to species–environment data: the selection of values for the maximum spatial lag distance, the kernel bandwidth and the orientation of the pattern in the analysis. We recommend a maximum lag distance of no more than one-quarter of the smallest plot dimension. If the maximum lag is too large, edge effects will influence the largest scales considered. Setting the maximum lag to 25% of the smaller plot dimension ensures an adequate sample size to detect the spatial pattern and minimizes edge effects.

The selection of an appropriate kernel bandwidth is comparatively straightforward if data on a regular grid (raster) are used, as we have illustrated here. Because we rasterized the data to 20-m grid cells, the scale at which the environmental data were obtained, setting each of the three bandwidth values ($k = \{k_A, k_B, k_{AB}\}$) equal to 20 m makes sense, as 20 m is the smallest scale at which any pattern could be detected. However, if codispersion is used to analyze bivariate marked point patterns (e.g. two measurements, such as diameter and height, which are recorded for a

single point location), the values used for the bandwidth parameters will determine the scales at which the codispersion analysis can detect patterns of spatial covariation. If the scales of the two variables differ markedly, then their bandwidth parameters, and that of their cross-variogram, should be different. One possibility is to set the values of k_A , k_B and k_{AB} to the values of the nuggets of their respective variograms (for k_A , k_B) or cross-variogram (for k_{AB}). Alternatively, Cuevas *et al.* (2013) suggest an optimization method for the identification of appropriate values for k .

The X , Y orientation of the observed biological spatial pattern matters for the pattern of codispersion values displayed in the codispersion graph (but not the significance tests) because we have greater resolution of pattern in the x -axis than in the y -axis. Thus, users should think about directionality in the processes driving the spatial patterns being tested. If little is known, rotating the pattern around the midpoint and analyzing it in both directions may aid in the identification of any directionality in the spatial pattern. It should be noted that this consideration does not affect the data collection unless the plot size or shape precludes the species–environment pattern under study from being adequately sampled within the study extent; therefore, we encourage researchers to consider their hypotheses of pattern during sampling design.

Codispersion analysis is useful because it results in a graph that clearly identifies the magnitude, scale and directionality of the observed patterns. It can identify the presence and scale of anisotropy in the spatial pattern. When combined with null models, it can be used to suggest testable hypotheses of ecological process. Moreover, it can identify nonstationarity in the spatial pattern of covariation, which influences subsequent inferential

modeling choices. It can be used to address a wide range of ecological questions when we are interested in the scale and nature of spatial covariation in variables derived from point-based or grid-based sampling schemes. Such variables may be associated with any attribute of organisms or their locations. The fact that fundamentally different processes can generate similar observed patterns of clumping reinforces the need for spatial methods, combined with appropriate null models, which allow ecologists to discern the relative importance of different processes. Importantly, codispersion can be used for composite measures, such as plant community richness or biomass, and extended to more than two variables (Vallejos *et al.*, 2015), which may be a fruitful path for further ecological applications. Although this method is computationally intensive, the code provided here (Notes S1) is readily adapted for use in a parallel computing framework. Future applications of this approach across a broad range of organisms and biogeographic regions will provide new insights into the ecological causes and consequences of species–environment associations.

Acknowledgements

The Luquillo Experimental Forest Long-Term Ecological Research Program, supported by the US National Science Foundation (NSF), the University of Puerto Rico and the International Institute of Tropical Forestry supported the collection of the Luquillo Forest Dynamics Plot data. We sincerely thank the many volunteers who collected the tree census data. The Tyson Research Center, the International Center for Advanced Renewable Energy and Sustainability (I-CARES) at Washington University in St Louis, and the CTFS-ForestGEO supported the collection of the Tyson Research Center Plot (TRCP) data. We thank the Tyson Research Center staff for providing logistical support, and the more than 60 high school students, undergraduate students and researchers who contributed to the TRCP data. We thank Jim Dalling for assistance with soil sampling methods; Ben Turner, Tania Romero and the staff at the Smithsonian Tropical Research Institute Soils Laboratory for the analysis of soil samples from TRCP; and Claire Baldeck for kriging the TRCP soil variables. We thank Michael Lavine, Ronny Vallejos, Nick Gotelli and the Harvard Forest Laboratory Group for valuable discussions of these ideas, and Thorsten Weigand, Adrian Baddeley, Ege Rubak, Matt Lau and Samuel Case for help with coding and computation. H.L.B. and B.S.C. were supported by Bullard Fellowships at Harvard Forest. This work is a contribution of the Harvard Forest Long Term Ecological Research program, supported most recently by NSF grant 12-37491. We thank David Ackerly and three anonymous reviewers for comments that greatly improved this paper.

Author contributions

H.L.B., B.S.C. and A.M.E. planned and designed the research. J.T., J.K.Z. and J.A.M. collected the data. H.L.B. and B.S.C. analyzed the data. All authors contributed to writing of the manuscript.

References

- Baddeley A, Diggle PJ, Hardegen A, Lawrence T, Milne RK, Nair G. 2014. On tests of spatial pattern based on simulation envelopes. *Ecological Monographs* 84: 477–489.
- Baddeley A, Turner R. 2005. spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 12: 1–42.
- Beale CM, Lennon JJ, Yearsley JM, Brewer MJ, Elston DA. 2010. Regression analysis of spatial data. *Trends in Ecology and Evolution* 13: 246–264.
- Bertness MD, Callaway R. 1994. Positive interactions in communities. *Trends in Ecology & Evolution* 9: 191–193.
- Buckley HL, Case BS, Ellison AM. 2016. Using codispersion analysis to characterize spatial patterns in species co-occurrences. *Ecology* 97: 32–39.
- Chesson P. 2000. Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics* 31: 343–366.
- Condit R, Ashton PS, Baker P, Bunyavechewin S, Gunatilleke S, Gunatilleke N, Hubbell SP, Foster RB, Itoh A, LaFrankie JV *et al.* 2000. Spatial patterns in the distribution of tropical tree species. *Science* 288: 1414–1418.
- Cressie N, Wikle CK. 2011. *Statistics for spatiotemporal data*. Hoboken, NJ, USA: John Wiley & Sons.
- Cuevas F, Porcu E, Vallejos R. 2013. Study of spatial relationships between two sets of variables: a nonparametric approach. *Journal of Nonparametric Statistics* 25: 695–714.
- Dale MRT. 1999. *Spatial pattern analysis in plant ecology*. Cambridge, UK: Cambridge University Press.
- Deblauwe V, Kennel P, Couteron P. 2015. Testing pairwise association between spatial autocorrelated variables: a new approach using surrogate lattice data. *PLoS ONE* 7: e48766.
- Elith J, Leathwick J. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review in Ecology and Systematics* 40: 677–697.
- Franklin J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19: 474–499.
- Harris NL, Lugo AE, Brown S, Heartsill Scalley T eds. 2012. *Luquillo experimental forest: research history and opportunities*. EFR-1. Washington, DC, USA: US Department of Agriculture.
- Hijmans RJ, Cameron S, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.
- Hubbell SP. 1979. Tree dispersion, abundance, and diversity in a tropical dry forest. *Science* 203: 1299–1309.
- John R, Dalling JW, Harms KE, Yavitt JB, Stallard RF, Mirabello M, Hubbell SP, Valencia R, Navarrete H, Vallejo M *et al.* 2007. Soil nutrients influence spatial distributions of tropical tree species. *Proceedings of the National Academy of Sciences, USA* 104: 864–869.
- Lepš J, Šmilauer P. 2003. *Multivariate analysis of ecological data using CANOCO*. Cambridge, UK: Cambridge University Press.
- Lingua E, Cherubini P, Motta R, Nola P. 2008. Spatial structure along an altitudinal gradient in the Italian central Alps suggests competition and facilitation among coniferous species. *Journal of Vegetation Science* 19: 425–436.
- Loosmore BN, Ford ED. 2006. Statistical inference using the G or K point pattern spatial statistics. *Ecology* 87: 1925–1931.
- Malkinson D, Kadmon R, Cohen D. 2003. Pattern analysis in successional communities – an approach for studying shifts in ecological interactions. *Journal of Vegetation Science* 14: 213–222.
- McGill BJ. 2010. Towards a unification of unified theories of biodiversity. *Ecology Letters* 13: 627–642.
- Morlon H, Chuyong G, Condit R, Hubbell S, Kenfack D, Thomas D, Valencia R, Green JL. 2008. A general framework for the distance-decay of similarity in ecological communities. *Ecology Letters* 11: 904–917.
- Plotkin JB, Chave J, Ashton PS. 2002. Cluster analysis of spatial patterns in Malaysian tree species. *The American Naturalist* 160: 629–644.
- R Core Team. 2014. *R: a language and environment for statistical computing*. R v.3.1.2. Vienna, Austria: R Foundation for Statistical Computing. URL <http://www.R-project.org/>

- Ruhkin AL, Vallejos R. 2008. Codispersion coefficients for spatial and temporal series. *Statistics & Probability Letters* 78: 1290–1300.
- Shen G, Yu M, Hu X-S, Mi X, Ren H, Sun I-F, Ma K. 2009. Species–area relationships explained by the joint effects of dispersal limitation and habitat heterogeneity. *Ecology* 90: 3033–3041.
- Silvertown J. 2004. Plant coexistence and the niche. *Trends in Ecology and Evolution* 19: 605–611.
- Spasojevic MJ, Yablon EA, Oberle B, Myers J. 2014. Ontogenetic trait variation influences tree community assembly across environmental gradients. *Ecosphere* 5: 129.
- Thompson J, Brokaw N, Zimmerman JK, Waide RB, Everham EM III, Lodge J, Taylor CM, Garcia-Montiel D, Fluet M. 2002. Land use history, environment, and tree composition in a tropical forest. *Ecological Applications* 12: 1344–1363.
- Turner BL, Engelbrecht BMJ. 2011. Soil organic phosphorus in lowland tropical rain forests. *Biogeochemistry* 103: 297–315.
- Vallejos R, Mallea A, Herrera M, Ojeda S. 2015. A multivariate geostatistical approach for landscape classification from remotely sensed image data. *Stochastic Environmental Research and Risk Assessment* 29: 369–378.
- Vieira SR, Porto de Carvalho JR, Ceddia MB, González AP. 2010. Detrending non stationary data for geostatistical applications. *Bragantia* 69: 1–8.
- Wagner HH, Dray S. 2015. Generating spatially-constrained null models for irregular spaced data using Moran spectral randomization methods. *Methods in Ecology and Evolution* 6: 1169–1178.
- Walker LR. 2000. Seedling and sapling dynamics of treefall pits in Puerto Rico. *Biotropica* 32: 262–275.
- Weaver PL. 2010. Forest structure and composition in the lower montane rain forest of the Luquillo Mountains, Puerto Rico. *Interciencia* 35: 640–646.
- Wiegand T, Huth A, Getzin S, Wang X, Hao Z, Gunatilleke CVS, Gunatilleke IAU. 2012. Testing the independent species' arrangement assertion made by theories of stochastic geometry of biodiversity. *Proceedings of the Royal Society of London B: Biological Sciences* 279: 3312–3320.
- Wiegand T, Moloney KA. 2014. *A handbook of spatial point pattern analysis in ecology*. Boca Raton, FL, USA: CRC Press.
- Wiens JA. 1989. Spatial scaling in ecology. *Functional Ecology* 3: 385–397.
- Zimmerman JK, Comita LS, Thompson J, Uriarte M, Brokaw N. 2010. Patch dynamics and community metastability of a subtropical forest: compound effects of natural disturbance and human land use. *Landscape Ecology* 25: 1099–1111.
- Zimmerman JK, Everham EM III, Waide RB, Lodge DJ, Taylor CM, Brokaw NVL. 1994. Responses of tree species to hurricane winds in subtropical wet forest in Puerto Rico: implications for tropical tree life histories. *Journal of Ecology* 82: 911–922.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1 Species data for Tyson plot

Table S2 Environmental principal component (PC) axis data for Tyson plot

Notes S1 Annotated R code for all analyses and figures.

Notes S2 Full output from codispersion analysis of simulated point patterns.

Notes S3 Results of principal components analysis of Tyson soil chemistry data.

Notes S4 Type I error rates associated with null model analyses.

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.