

# Using codispersion analysis to characterize spatial patterns in species co-occurrences

HANNAH L. BUCKLEY,<sup>1,2,4</sup> BRADLEY S. CASE,<sup>2,3</sup> AND AARON M. ELLISON<sup>2</sup>

<sup>1</sup>*Department of Ecology, Lincoln University, Lincoln, Canterbury 7647, New Zealand*

<sup>2</sup>*Harvard Forest, Harvard University, 324 North Main Street, Petersham, Massachusetts 01366, USA*

<sup>3</sup>*Department of Informatics and Enabling Technologies, Lincoln University, Lincoln, Canterbury 7647, New Zealand*

**Abstract.** Visualizing and quantifying spatial patterns of co-occurrence (i.e., of two or more species, or of species and underlying environmental variables) can suggest hypotheses about processes that structure species assemblages and their relevant spatial scales. Statistical models of spatial co-occurrence generally assume that underlying spatial processes are isotropic and stationary, but many ecologically realistic spatial processes are anisotropic and non-stationary. Here, we introduce codispersion analysis to ecologists and use it to detect and quantify anisotropic and nonstationary patterns and their relevant spatial scales in bivariate co-occurrence data. Simulated data illustrated that codispersion analysis can accurately characterize complex spatial patterns. Analysis of co-occurrence of common tree species growing in a 35-ha plot revealed both positive and negative codispersion between different species; positive codispersion values reflected positive correlation in species abundance (aggregation), whereas negative codispersion values reflected negative correlation in species abundance (segregation). Comparisons of observed patterns with those simulated using two different null models showed that the codispersion of most species pairs differed significantly from random expectation. We conclude that codispersion analysis can be a useful exploratory tool to guide ecologists interested in modeling spatial processes.

**Key words:** anisotropy; codispersion; co-occurrence; forest dynamics plot; Harvard Forest; spatial analysis; stationarity; variogram.

## INTRODUCTION

A central inferential challenge for ecologists is the identification of mechanisms and operational scales of processes determining observed spatial patterns. We normally begin to address this challenge using spatial pattern analysis (e.g., Dale 1999, Cressie and Wikle 2011, Wiegand and Moloney 2014). In particular, analysis and interpretation of patterns of co-occurrence of two or more species, or of individual species and environmental variables, are used routinely to identify relevant spatial scales and generate testable hypotheses about processes determining observed co-occurrence patterns (e.g., Wiegand and Moloney 2014).

Spatial patterns of co-occurring species often reflect the temporal integration of pairwise species associations, varying environmental conditions, and intra- and interspecific interactions (Dale 1999). For example, spatial patterns of locations and sizes of trees in a forest reflect the life history differences of the component species as well as the cumulative history of the stand: stochastic dispersal, small-scale environmental conditions for successful germination and establishment,

self-thinning of once-dense groups of saplings and small trees, interspecific competition, loss of individuals to insects and disease, and ongoing environmental change. In many cases, environmental gradients dominate spatial patterns in forest stands, but the most widely used spatial pattern analyses assume that underlying spatial processes of the analyzed spatial patterns are stationary (spatial processes are invariant under translation) and isotropic (invariant under rotation) (Dale 1999, Cressie and Wikle 2011). Although these assumptions are mathematically convenient, they are rarely true. Alternatives, such as wavelets and other spectral methods can be used to identify scales of variation in isotropic or anisotropic spatial data, but some of these methods can analyze only limited types of ecological data (e.g., data collected on lattices; Deblauwe et al. 2012).

Codispersion analysis (Cuevas et al. 2013) is a new method for describing and visualizing complex spatial patterns of multiple co-occurring variables. In brief, codispersion quantifies covariation of two or more spatial patterns as a function of both spatial lag and direction (Cuevas et al. 2013). The analyzed spatial patterns may be any combination of point patterns, marked point patterns (where we are interested in the spatial pattern of the marks, such as tree diameter), irregularly spaced plots, or rasters (contiguous grids);

Manuscript received 30 March 2015; revised 2 July 2015; accepted 14 July 2015. Corresponding Editor: H. H. Wagner.

<sup>4</sup>E-mail: Hannah.Buckley@lincoln.ac.nz

in the latter two cases, each plot or cell is assigned a single value, such as tree density or basal area. To date, codispersion analysis has been applied to only a limited number of data sets: photographic image analysis (Ojeda et al. 2012), the relationship between tree size and an underlying environmental gradient (Cuevas et al. 2013), bivariate temporal data (Vallejos 2012), and multivariate spectral data (Vallejos et al. 2015).

Cuevas et al. (2013) suggested that codispersion analysis may be useful to describe the pattern of co-variation found in many different kinds of spatial ecological data, but it has not yet been used to analyze known (i.e., pre-determined) spatial patterns, nor have the results from observed patterns been evaluated against reasonable null expectations. Here, we describe how to apply codispersion analysis to species co-occurrence data. Using both simulated and real species co-occurrence data, we illustrate how the results can identify and quantify complex spatial patterns and spatial scales at which ecological processes may be operating. Results of codispersion analysis applied to simulated data illustrate the range of detectable patterns. We also show how to test for departure of observed codispersion from null models that (1) assume complete spatial randomness (CSR) in species co-occurrence or (2) fix the distances between points, but break the association between the two species by shifting one of the entire patterns randomly, treating the plot as a torus (toroidal shift null model).

## METHODS

Examining codispersion between two or more spatial data sets depends on the precise locations at which measurements were made and the sampling grain of the measurements (i.e., the “support”; sensu Dungan et al. 2002). If we are interested in point-wise codispersion between two marked point patterns, the variables to be compared must be measured at exactly the same locations (e.g., beetles feeding on trees or the diameter and height of individual trees). In contrast, if the variables to be compared are measured at different spatial locations, the data need to be rescaled in one of two ways prior to analysis.

One way to rescale the data is convert the point patterns to small, identically sized plots prior to analysis. If variables are measured at irregularly spaced locations, they can be up-scaled to a common plot size; each “plot” is then considered to be a point with  $x, y$  coordinates equal to the center of the plot and with marks equal to some aggregate measure of the individuals within the “plot” (e.g., Cuevas et al. 2013). For example, if soil pH is measured at 100 random points and tree diameters are measured in a 20-m diameter circular plot around each of these points, then the average tree size in each plot can

be used as the estimate of tree size at the point where soil pH was measured. Codispersion calculations then proceed as if the point patterns were measured at identical locations, but the “grain of inference” is the 20-m diameter plot. Alternatively, to compare the co-occurrence of two tree species in a large plot where, by definition, different individual trees cannot occupy identical point locations, we could first calculate the abundance of each species in individual contiguous subplots, i.e., one abundance raster for each species. Codispersion calculations then proceed as if the point patterns were measured at the centers of each of the subplots and the “grain of inference” is the size of each subplot (grid cell) in the rasters. In our worked examples, we use this latter, raster-based method.

### *A recipe for codispersion analysis of ecological data*

Codispersion coefficients across a range of spatial scales (which we illustrate for rasterized data) are generated as follows:

- (1). Generate three vectors of spatial lags (Fig. 1A). Two of these vectors of lags should be parallel to the  $x$  axis, one positive and one negative around zero ( $-\mathbf{h}_1$  to  $+\mathbf{h}_1$ ), up to one-fourth of the smallest plot dimension. The third vector of lags is parallel to the  $y$  axis, again increasing from the size of a raster up to one-fourth of the smallest plot dimension ( $\mathbf{h}_2$ ). For example, for a  $500 \times 700$  m plot, the smaller dimension is 500 m, so the maximum of  $|\mathbf{h}_1|$  or  $|\mathbf{h}_2|$  is  $500/4 = 125$  m. This ensures an adequate sample size for calculating codispersion at the largest lag. The smallest lag size should be the grain size of the raster.
- (2). Apply a kernel function across all possible cell-wise distances for each lag to compute a variation surface for all lag distances and directions  $\mathbf{h}$  (the two dimensions are defined by the  $\{\mathbf{h}_1, \mathbf{h}_2\}$  coordinate pairs) for each data set individually and the intersection of the two data sets. The way the kernel surfaces characterize spatial variation within and between data sets A and B is controlled by specifying appropriate kernel bandwidth parameters  $k = \{k_x, k_y, k_{xy}\}$  (Cuevas et al. 2013). If data have been rasterized, we recommend setting each element of  $k$  equal to the grid cell size of the raster.
- (3). Compute semi-variograms for each variable ( $\gamma_x, \gamma_y$ ), and the semi-cross-variogram ( $\gamma_{xy}$ ) across all kernel-smoothed lag vectors  $\mathbf{h}$  using a Nadaraya-Watson type estimator as follows:

$$\hat{\gamma}_{XY}(\mathbf{h}) = \frac{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{\mathbf{h} - (s_i - s_j)}{k}\right) (X(s_i) - X(s_j))(Y(s_i) - Y(s_j))}{2 \sum_{i=1}^n \sum_{j=1}^n K\left(\frac{\mathbf{h} - (s_i - s_j)}{k}\right)} \quad (1)$$

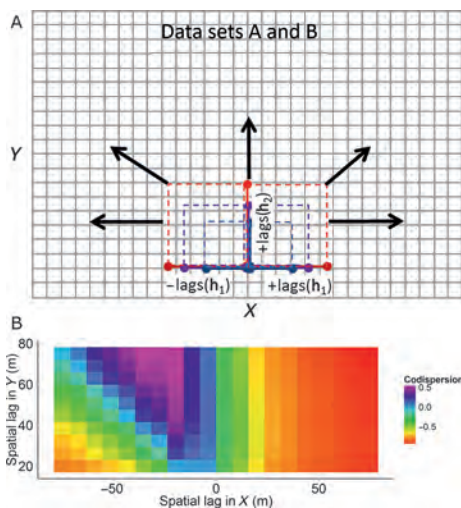


FIG. 1. (A) Illustration of the generation of spatial lags in the positive and negative  $X$  directions ( $+\text{lags}(h_x)$  and  $-\text{lags}(h_x)$ ), and the positive  $Y$  direction ( $+\text{lags}(h_y)$ ), for two rasterized surfaces (data sets A and B) used as input to codispersion analysis. The colored boxes illustrate increasingly large, two-dimensional, kernel-smoothed spatial lags over which codispersion coefficients are calculated.  $X$  and  $Y$  are actual distances for each plot data set (A and B). (B) A codispersion plot. The color of each cell is the value of the codispersion coefficient of two variables for all analyzed spatial lags and directions in  $X, Y$  space. Positive codispersion values indicate positive covariation (aggregation) and negative codispersion values indicate negative covariation (segregation).

where  $s$  is the set of spatial locations, and  $K(\cdot)$  is a symmetric and strictly positive kernel function with bandwidth  $k_{XY}$  (García-Soidán 2007, Cuevas et al. 2013).

- (4). Compute the empirical codispersion coefficient (Matheron 1965) for each lag ( $\mathbf{h}$ ) as the semi-cross-variogram of the two variables ( $\hat{\gamma}_{XY}(\mathbf{h})$ ) normalized by the square root of the product of the semi-variograms of each of the two variables:

$$\hat{\rho}_{XY}(\mathbf{h}) = \frac{\hat{\gamma}_{XY}(\mathbf{h})}{\sqrt{\hat{\gamma}_X(\mathbf{h})\hat{\gamma}_Y(\mathbf{h})}} \quad (2)$$

where the formula for the empirical semi-variogram is:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{(i,j) \in N(\mathbf{h})} (Z(s_i) - Z(s_j))^2 \quad (3)$$

In Eqs 2 and 3,  $\mathbf{h}$  is the lag distance,  $N(\mathbf{h})$  denotes the sets of pairs of observations,  $s$  is the set of spatial locations, and  $Z$  is the value of interest at a given location.

- (5). Plot the codispersion values for each spatial lag  $\mathbf{h}$  (Fig. 1B). Positive codispersion values indicate positive covariation (aggregation) and negative codispersion values indicate negative covariation (segregation) for lag  $\mathbf{h}$  with a given distance (in  $x, y$  space) and direction. Positive and negative lags on the  $x$ -axis refer to “looking right” (e.g.,

east) and “looking left” (e.g., west) within the plot, respectively. Positive lags on the  $y$ -axis refer to “looking up” (e.g., north) within the plot.

- (6). An appropriate set of null models should be selected to compare against the observed codispersion values. The choice of null model depends on the ecological question asked and the processes hypothesized to generate the observed spatial patterns. For instance, a CSR null model allows us to ask whether or not the observed pattern is spatially nonrandom, i.e., the species are distributed independently. Application of a toroidal shift null model asks whether the association between the species is nonrandom, given their univariate spatial patterns; thus, we are assessing their co-variation in space while excluding any effect of individual species' autocorrelation structures. Other, process-based null models (Wiegand and Moloney 2014) may be appropriate in certain circumstances, if sensible ideas about the processes generating spatial patterns can be formulated (e.g., Wiegand et al. 2009).

#### Illustrating codispersion analysis using simulations and real data

We first generated and analyzed a range of bivariate spatial patterns (Fig. 2; a complete set of simulated patterns is in Appendix S1; pseudocode in Appendix S4; and accompanying R code is in the Supplement). Because we were interested in comparing simulated results with observed data from a large, gridded, forest inventory plot (see next paragraph), we simulated species abundance patterns as a raster of 225 contiguous  $20 \times 20$  m grid cells arrayed in a  $300 \times 300$  m “plot.” Abundance values in grid cells were distributed either completely spatially randomly (CSR) among grid cells, increasing or decreasing to the left side, right side, left or right top corners, or in one large clump in the center of the plot. We analyzed a wide range of the possible pairs of these simulated distributions (Appendix S1).

Second, we analyzed all observed pairwise bivariate spatial patterns of the four most abundant tree species in the Harvard Forest long-term forest dynamics plot (Fig. 3A; information available online).<sup>4</sup> This fully censused 35-ha plot is part of the Smithsonian Tropical Research Institute's Center for Tropical Forest Science – Forest Global Earth Observatory (CTFS-ForestGEO) network of plots (information available online).<sup>5</sup> In this plot, a total of 116 227 woody stems  $>1$  cm diameter were mapped, tagged, and measured between June 2010 and March 2014. The four most common species, *Acer rubrum* L. (red maple, Sapindaceae), *Pinus strobus* L. (white pine, Pinaceae), *Quercus rubra* L. (red oak, Fagaceae), and *Tsuga canadensis* (L.) Carrière (eastern

<sup>4</sup> <http://harvardforest.fas.harvard.edu:8080/exist/apps/datasets/showData.html?id=hf253>

<sup>5</sup> <http://www.forestgeo.si.edu>

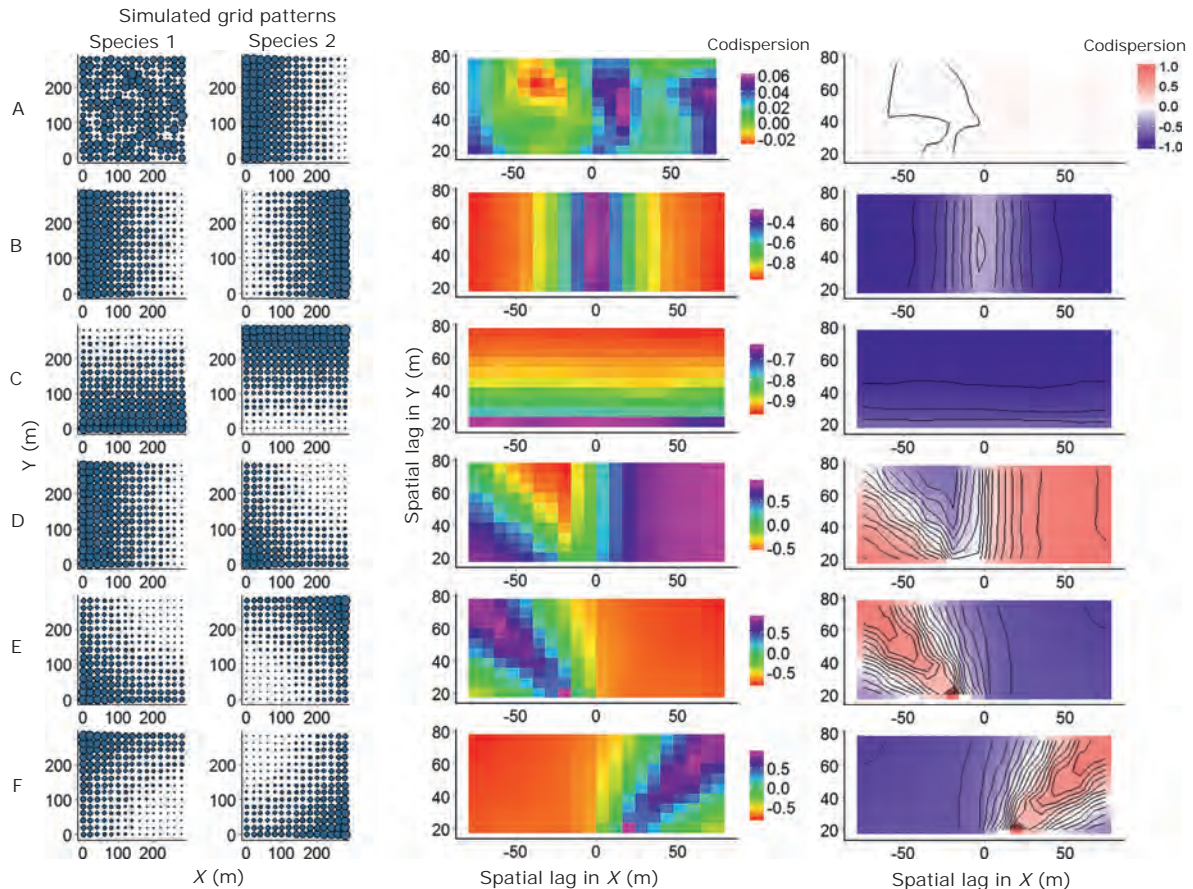


FIG. 2. Simulated species co-occurrence patterns on  $20 \times 20$ -m grids in  $300 \times 300$ -m plots and their resultant codispersion graphs. In each case, the colors on the codispersion graphs on the left are scaled to the range of values for that plot and those on the right are scaled from  $-1$  to  $+1$ . The mean ( $\pm$ SD) codispersion values for each analysis were (A) 0.02 (0.02), (B)  $-0.77$  (0.21), (C)  $-0.87$  (0.09), (D) 0.38 (0.48), (E)  $-0.28$  (0.56) and (F)  $-0.28$  (0.56).

hemlock, Pinaceae), together comprise  $>90\%$  of the total basal area in the plot. Data are available from the Harvard Forest data archive (Orwig et al. 2015).

We calculated the number of individuals of each of these four species within  $20 \times 20$ -m contiguous grid cells covering the  $500 \times 700$ -m plot (a total of 875 grid cells) and used these cell-level abundance data for all spatial analyses. We aggregated these data into  $20 \times 20$ -m cells for the following reasons: This is the approximate canopy diameter of the dominant tree species in the Harvard Forest plot;  $20 \times 20$  m (0.04 ha) is a common plot size used by foresters and ecologists to collect and analyze forest stand data (Kangas 2006); and a  $20 \times 20$ -m grid is the standard of collection and aggregation for ForestGeo data (Condit 1998).

We computed the codispersion of each pair of species at spatial lags ranging from 20 to 120 m. The maximum spatial lag equaled just under one-fourth of the length of the shortest side of the plot and was used to ensure adequate sample sizes for the largest spatial lag. To assess the significance of the observed codispersion patterns, we compared the observed codispersion values

for each species pair calculated for each spatial lag and direction to values generated using two null models. The first was a CSR model, where one species distribution was fixed and the point locations of the other species were distributed completely spatially randomly across the plot. The second was a toroidal shift model, where the positions of trees were fixed, thus maintaining their autocorrelation structure, but the entire plot was shifted in a random direction and distance around a torus (Wiegand and Moloney 2014).

For each comparison, the null models were used to generate 199 new data sets for one of the species of each pair; 199 null simulations was a large enough number to confidently determine significant differences between observed and expected, and small enough to generate expected values on a desktop computer within a few days. Only one of the species pair needed to be randomized because this was enough to break their spatial association, allowing us to test the significance of their co-variation. The observed codispersion values at each spatial lag were then compared to the vector of codispersion values at the same spatial lags and directions

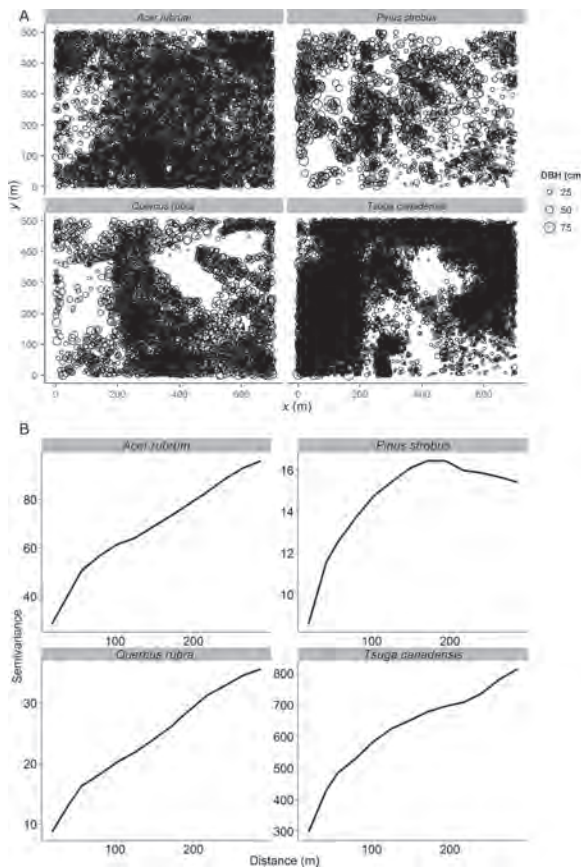


FIG. 3. Observed (A) point patterns and (B) semi-variograms for the four most abundant species in the Harvard Forest 35-ha forest dynamics plot, calculated using species' abundances in  $20 \times 20$  m rasters. In (A), the sizes of the symbols are scaled to each tree's dbh: diameter measured at breast height (1.3 m above ground).

under each null model to estimate tail probabilities; if the observed value was  $\geq 195$ th value or  $\leq 5$ th value, we deemed it to be significantly different from expected (i.e., a two-tailed test;  $P < 0.05$ ). Finally, we calculated the Type I error rate of the CSR and toroidal shift null models by comparing the observed codispersion between two CSR simulated patterns (Appendix S2) to values generated under the CSR and toroidal shift models.

## RESULTS

### Simulations

Codispersion analysis accurately detected both positive and negative covariation in abundance in simulated bivariate spatial patterns (Fig. 2). For cases in which we simulated no strong covariation between two species (i.e., at least one "species" was CSR), codispersion values at all lags were around zero (Fig. 2A; Appendix S1). When the two species were strongly segregated

or aggregated (i.e., negative or positive covariance, respectively, between them; Fig. 2B,C), the codispersion values were similarly highly negative or positive. When the patterns of abundance of the two species were strongly anisotropic in the east–west direction (the  $x$  [horizontal] dimension of the plot), such that in some areas of the plot the species were aggregated and in other areas they were segregated (Fig. 2D–F), the analysis illustrated the anisotropy by having different patterns for positive and negative lags on the  $x$ -axis of the codispersion graph. Positive codispersion values reflected lags and directions over which species were both either increasing or decreasing in abundance (Fig. 2D–F). In contrast, negative values of codispersion represented lags and directions for which the abundances of the two species were negatively correlated, e.g., one species was high in abundance when the other was low. Rotating the species' patterns illustrated that the method was sensitive to the orientation of the plot (compare Fig. 2B,C). Analysis of the mirrored reflection of the patterns illustrated that identical results were obtained for positive and negative lags in the  $y$  direction (compare Fig. 2E,F). For both the CSR and toroidal shift null models, none of the observed codispersion values from the two CSR patterns were significantly different from that expected under either model at the 5% level, indicating a Type I error rate  $\leq 5\%$  for both null models (Appendix S2).

### Real data

Empirical semi-variograms illustrated that spatial autocorrelation of *P. strobus* in the 35-ha forest dynamics plot was apparent up to lags of  $\sim 180$  m, but that abundances of the other three species were autocorrelated at scales of at least 300 m (variograms showed a linear trend, with no sill; Fig. 3B). *Tsuga canadensis*, the dominant species, negatively co-varied with (i.e., was spatially segregated from) the three other species (Fig. 4). This species showed weak anisotropy in its covariation with *Q. rubra* and *A. rubrum*, as indicated by different values of the codispersion coefficient on the right- and left-hand sides of the codispersion graph, but not with *P. strobus*. In contrast, the three sub-dominant species all positively co-varied (were aggregated) at most spatial lags; the positive codispersion was strongest between *Q. rubra* and *A. rubrum* (Fig. 4). The observed codispersion patterns largely were significantly different from those expected under the two null models, except for *P. strobus* and *Q. rubra*, which showed only weak positive covariation (Fig. 4). Because the toroidal shift null model maintained the autocorrelation structure of the individual species' patterns while breaking their bivariate spatial association, observed codispersion values were significantly different from null expectation at a smaller number of lags than we observed with the CSR model (Fig. 4).

## DISCUSSION

Codispersion analysis is an effective method for quantifying and visualizing the pairwise covariation of two or more variables in space (Vallejos et al. 2015). One of the key benefits of this method is that it gives a two-dimensional graph illustrating the sign (positive or negative), magnitude, scale, and direction of covariation between two species. This information is especially useful for choosing appropriate models for subsequent inference about underlying spatial processes.

Most methods used to model spatial data assume that the data are stationary (but see Wiegand et al. 2007, Getzin et al. 2008). The “strong” form of spatial stationarity is the situation in which the joint probability distribution of the data is invariant when the pattern is moved (translated) through space. For ecological data, this assumption is rarely, if ever, true. A weaker form of spatial stationarity, “second-order stationarity,” assumes that only the mean, variance, and covariance must be stationary (Vieira et al. 2010). Even this assumption is rarely satisfied. However, most spatial statistical methods can be used if the data meet the assumption of the “intrinsic hypothesis”: That the mean and the semi-variance of the distribution are dependent only on distance between points, not on their location; i.e., there is no underlying large-scale spatial “trend” in the data (Vieira et al. 2010).

A common way to determine if there is spatial trend in the data is to compare the semi-variograms between raw and adequately detrended data: Raw data with a spatial trend will have a semi-variogram that lacks a sill, whereas the semi-variogram of data without a spatial trend will have an obvious and stable sill (Vieira et al. 2010). However, simply detrending data and comparing semi-variograms does not identify directionality in the data. In contrast, codispersion plots illustrate distances and directions at which significant spatial covariance occur.

In our examples, codispersion plots correctly detected isotropic and anisotropic positive and negative spatial covariation in simulated abundances of two species on a grid for a variety of ecologically interpretable patterns (Fig. 1; Appendix S1). Analysis of pairwise co-occurrences of the abundances of forest tree species in a 35-ha plot (Fig. 3) also showed that codispersion analysis could detect subtle variation in spatial co-occurrences among species. Comparisons of observed values with those obtained from repeated realizations of null models also could be used to evaluate the statistical significance of observed patterns of anisotropic spatial covariance.

We emphasize that codispersion analysis can only detect and illustrate covariation in species distribution patterns; like a semi-variogram, a codispersion graph it is not explicitly designed to reveal the processes that gave rise to the observed patterns. For instance, there are at least two possible reasons that species’ distributions can co-vary in space: (1) interactions that

lead to aggregation or segregation; or (2) similarity or differences in species’ habitat preferences or other underlying (and unmeasured) variables, such as soil nutrients. As shown by the simulations (e.g., Fig. 1D), patterns that are caused by different spatial processes will show high codispersion if they co-vary spatially. However, comparison of observed codispersion to that expected under different null models can help determine the nature of the observed spatial pattern and how to develop predictive process models. For example, *P. strobus* and *Q. rubra* showed no significant codispersion at all but the largest spatial scales under either the CSR or toroidal shift null model. This result suggests that these data meet the assumption of second-order stationarity and that process models to describe their covariance could proceed without detrending the data. In contrast, all other pairs of species showed some significant codispersion relative to the toroidal shift model, suggesting that at best, process models of their covariance could lend support to the intrinsic hypothesis, not second-order stationarity.

The choice of null models also is critical in any description of spatial pattern. The CSR null model, as we applied it, did not account for tree size in randomly rearranging their positions, so it is possible that trees in the null patterns were closer than is realistic, given their size. Further, the toroidal shift model can lead to artificial significance if large-scale clustering occurs at the plot edges, creating edge effects in the null realizations (Wiegand and Moloney 2014: 365). In the case of the data from the Harvard Forest’s 35-ha forest dynamics plot, variable land-use history within the plot area precluded using a model based on a simple spatial process, such as a Thomas cluster process, which would assume spatial homogeneity in the plot. Future research should pursue more alternative null models; recent research suggests that spectral randomization methods (e.g., Deblauwe et al. 2012, Wagner and Dray 2015) and pattern reconstruction (Wiegand and Moloney 2014: 368) may be useful approaches to this problem. The key consideration when applying a null model to a spatial pattern is that we understand the process being tested by the model; this may not be trivial for more complex null models.

When using codispersion analysis with any null model, there are three technical issues that are important to keep in mind. First, a maximum lag distance should be selected that is not more than one-quarter of the smallest dimension of the plot. This ensures that an adequate number of pairs at all combinations of lags (especially the maximum lag) and directions for calculating the codispersion coefficient are available.

Second, the choice of bandwidth for the kernel function is critical. By default, our code (based on Cuevas et al. 2013) uses the same kernel for both variables and their intersection (the cross-variogram). However, if the spatial variation in the two variables differs substantially, it may be appropriate to select different bandwidths for the different variables. A

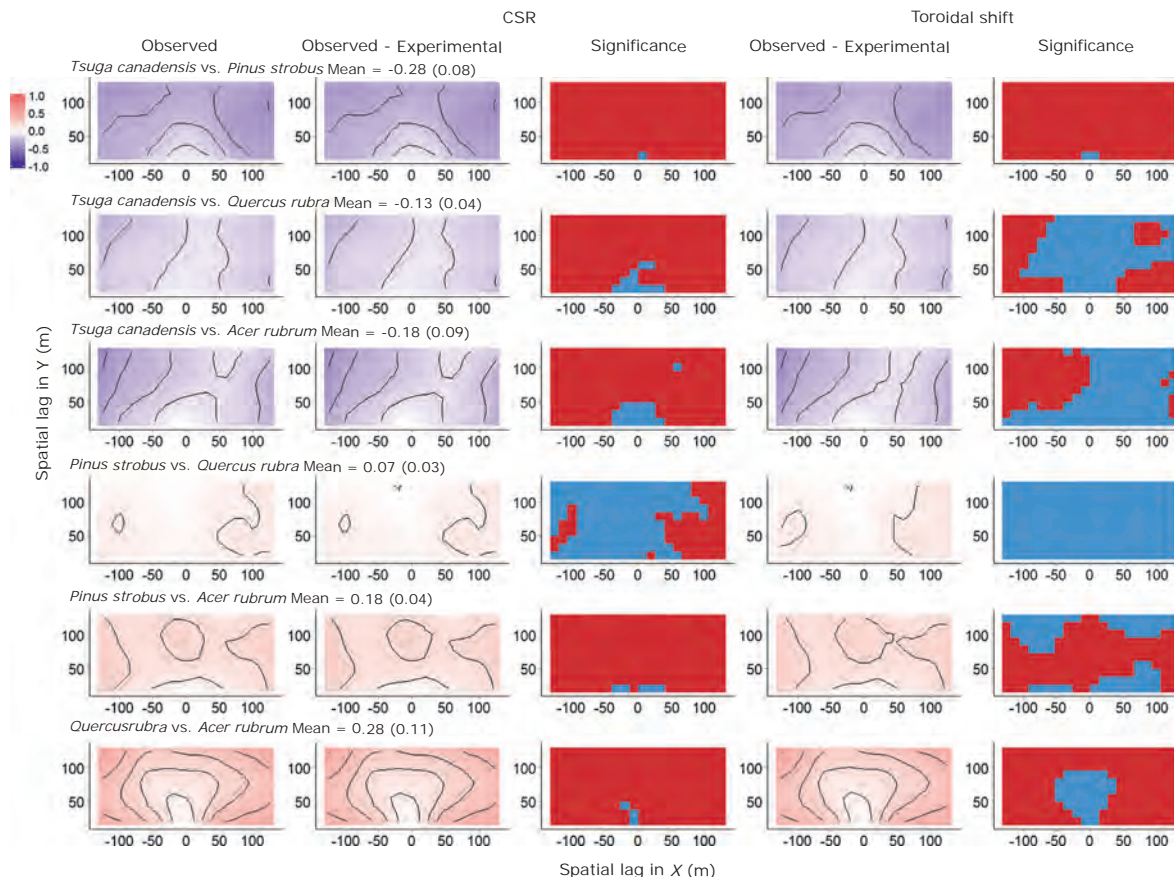


FIG. 4. Observed codispersion values and their significance (red) or not (blue) when compared to null expectation for bivariate co-occurrence data from all species pairs of the four most common tree species in the Harvard Forest 35-ha forest dynamics plot, calculated using species' abundances in  $20 \times 20$  m grid cells. Scaled ( $-1$  to  $+1$ ) codispersion graphs are shown with 0.1-unit contours. On the left, the mean codispersion values for each codispersion graph (standard deviation) are given for each species pair. Significance of codispersion values in each grid cell were calculated by comparing the observed value with 199 codispersion values from a complete spatial randomness (CSR) null model and a toroidal shift null model; if the observed value was  $\geq 195$ th value or  $\leq 5$ th value, the cell was labeled significant.

sensible value for the kernel bandwidth should be selected that is no smaller than the grain size (support) of the data and not so large that it smooths across lags (which results in no differentiation across the codispersion graph). Cuevas et al. (2013) recommend using an optimization method to select appropriate bandwidth parameters for the kernel function. More easily, the range of the univariate variograms and bivariate cross-variogram (if they exist) might be used to select an appropriate bandwidth for each variable and their covariance. We note, however, that detrending the data to obtain stable variograms with sills (as suggested by Vieira et al. 2010) prior to running codispersion analysis can be expected to eliminate the pattern that codispersion aims to detect (Appendix S3).

Clearly, the kernel bandwidth will differ if raw point-pattern data are used or if the data are rasterized. If the observations are of individual locations in space (i.e., a point pattern), the initial selection of the scale

at which the data could be rasterized (or not) should be determined based on biological considerations. If all variables are measured at identical points, no rasterization is necessary. However, in many ecological data sets, individuals of two or more different species do not co-occur at identical locations (due to physical constraints or the scale of sampling), so rasterizing species co-occurrences makes sense. Note, however, that rasterizing a point pattern so that grid cells include  $\geq 1$  observation is equivalent to applying a uniform bivariate kernel with bandwidth equal to the width of a grid cell. Thus, when calculating codispersion values as illustrated in Fig. 3, we set the bandwidth parameter to 20 m so as to not re-smooth the data any further than they had already been rasterized.

Finally, codispersion analysis is particularly useful for examining anisotropic patterns and processes. As a result, the orientation of the data matters. The  $x$  and  $y$  dimensions of the values in the codispersion

graphs (Figs 1 and 3) describe lags in the “left,” “right,” and “up” directions, and anisotropy is illustrated in most detail across the  $x$  dimension (see also Appendix S1). Therefore, we recommend that the data be oriented in a manner that reflects the directionality of patterns of particular interest, or, that the pattern is rotated and analyzed in both directions. In this way, interesting patterns are more likely to be identified and used to suggest new and testable ecological hypotheses. Future research also should directly compare the ability of this method to detect complex multivariate spatial patterns with that of other methods, including spectral analysis (Deblauwe et al. 2012), geographically weighted regression (Fotheringham et al. 2002), and Moran eigenvector maps (Wagner and Dray 2015).

#### ACKNOWLEDGMENTS

We thank Michael Lavine, Ronny Vallejos, Nick Gotelli, and the Harvard Forest Lab Group for valuable discussions of these ideas; three anonymous reviewers and Helene Wagner for constructive comments on the manuscript; Dave Orwig for use of the Harvard Forest ForestGeo plot data; and Thorsten Wiegand, Adrian Baddeley, Ege Rubak, Matt Lau, and Samuel Case for help with coding and computation. H. L. Buckley and B. S. Case were supported by a Charles Bullard Fellowship at Harvard Forest while conducting this research. A. M. Ellison was supported by NSF grant 12-37491. This paper is a contribution of the Harvard Forest LTER program.

#### LITERATURE CITED

- Condit, R. 1998. Tropical forest census plots: methods and results from Barro Colorado Island, Panama and a comparison with other plots. Springer-Verlag, Berlin, Germany.
- Cressie, N., and C. K. Wike. 2011. Statistics for spatiotemporal data. John Wiley and Sons, Hoboken, New Jersey, USA.
- Cuevas, F., E. Porcu, and R. Vallejos. 2013. Study of spatial relationships between two sets of variables: a nonparametric approach. *Journal of Nonparametric Statistics* 25:695–714.
- Dale, M. R. T. 1999. Spatial pattern analysis in plant ecology. Cambridge University Press, Cambridge, UK.
- Deblauwe, V., P. Kennel, and P. Couteron. 2012. Testing pairwise association between spatial autocorrelated variables: a new approach using surrogate lattice data. *PLoS One* 7:e48766.
- Dungan, J. L., J. T. Perry, M. R. T. Dale, P. Legendre, S. Citron-Pousty, and M.-J. Fortin. 2002. A balanced view of scale in spatial statistical analysis. *Ecography* 25:626–640.
- Fotheringham, S. A., C. Brunsdon, and M. Charlton. 2002. Geographically weighted regression: The analysis of spatially varying relationships. John Wiley and Sons, Chichester, UK.
- García-Soidán, P. 2007. Asymptotic normality of the Nadaraya-Watson semivariogram estimators. *Test* 16:479–503.
- Getzin, S., T. Wiegand, K. Wiegand, and F. He. 2008. Heterogeneity influences spatial patterns and demographics in forest stands. *Journal of Ecology* 96:807–820.
- Kangas, A. 2006. Mensurational aspects. Pages 53–64 in A. Kangas and M. Maltamo, editors. Forest inventory: methodology and applications. Springer, Dordrecht, The Netherlands.
- Matheron, G. 1965. Les variables régionalisées et leur estimation. Masson, Paris, France.
- Ojeda, S., R. Vallejos, and P. Lamberti. 2012. Measure of similarity between images based on the codispersion coefficient. *Journal of Electronic Imaging* 21:023019.
- Orwig, D. A., D. R. Foster, and A. M. Ellison. 2015. Harvard Forest CTFE-ForestGEO Mapped Forest Plot since 2014. Harvard Forest Data Archive: HF253. Available online: <http://harvardforest.fas.harvard.edu:8080/exist/apps/datasets/showData.html?id=hf253>.
- Vallejos, R. 2012. Testing for the absence of correlation between two spatial or temporal sequences. *Pattern Recognition Letters* 33:1741–1748.
- Vallejos, R., A. Mallea, M. Herrera, and S. Ojeda. 2015. A multivariate geostatistical approach for landscape classification from remotely sensed image data. *Stochastic Environmental Research and Risk Assessment* 29:369–378.
- Vieira, S. R., J. R. Porto de Carvalho, M. B. Ceddia, and A. P. González. 2010. Detrending non stationary data for geostatistical applications. *Bragantia* 69:1–8.
- Wagner, H. H., and S. Dray. 2015. Generating spatially-constrained null models for irregularly spaced data using Moran spectral randomization methods. *Methods in Ecology and Evolution* 6: 1169–1178.
- Wiegand, T., and K. A. Moloney. 2014. A handbook of spatial point pattern analysis in ecology. CRC Press, Boca Raton, Florida, USA.
- Wiegand, T., C. V. S. Gunatilleke, I. A. U. N. Gunatilleke, and A. Huth. 2007. How individual species structure diversity in tropical forests. *Proceedings of the National Academy of Sciences USA* 104:19029–19033.
- Wiegand, T., I. Martinez, and A. Huth. 2009. Recruitment in tropical tree species: revealing complex spatial patterns. *American Naturalist* 174:E106–E140.

#### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1890/15-0578.1/supinfo>