1   **Using codispersion analysis to quantify and understand spatial patterns in species-**

2   **environment relationships**

3   Hannah L. Buckley[1,2], Bradley S. Case[2,3], Jess K. Zimmerman[4], Jill Thompson[4,5], Jonathan A.

4   Myers[6], Aaron M. Ellison[2]

5   [1] Department of Ecology, Lincoln University, Lincoln, 7647, New Zealand

6   [2] Harvard Forest, Harvard University, 324 North Main Street, Petersham, Massachusetts, 01366,

7   USA

8   [3] Department of Informatics and Enabling Technologies, Lincoln University, Lincoln, 7647, New

9   Zealand

10   [4] Department of Environmental Sciences, University of Puerto Rico, San Juan, 00936, Puerto

11   Rico

12   [5] Centre for Ecology & Hydrology, Bush Estate, Penicuik, Midlothian, Scotland, EH26 0QB, UK

13   [6] Department of Biology, Washington University in St. Louis, St. Louis, Missouri, 63130, USA

14

15   Author for correspondence: Hannah Buckley, Hannah.Buckley@lincoln.ac.nz, phone: +64-3-

16   423-0736

17   Running title: Codispersion analysis of species-environment relationships

Total word count (excluding summary, references and legends)	6,747 (6500 limit)
Summary	197 (200 limit)
Introduction	951
Materials and Methods	2,660
Results	1,329
Discussion	1,537
Acknowledgements	238
No. of figures	6 (all in color)
No. of tables	4
No of Supporting Information files	4 (**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

18

Summary

1. Analysis of spatial patterns in species-environment relationships can provide new insights about 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.

2. We applied codispersion analysis to both simulated and field data on spatially-located tree species basal area and environmental variables. The significance of observed bivariate spatial associations between the basal area of key species and underlying environmental variables was tested using three null models.

3. 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 observed basal area-environment relationships, or whether relationships were due to other unmeasured factors.

4. 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.

KEYWORDS: Anisotropy, bivariate, environmental gradient, forest dynamics plot, spatial analysis, species-environment, variogram

Introduction

Environmental variability is a key driver of variation in biological diversity (Chesson 2000). Analysis of the spatial patterns in species-environment relationships can reveal clues about niche requirements of individual species and their potential for co-occurrence with other species (Silvertown 2004). Quantification of spatial patterns of species' distribution and abundance can illuminate scales of variation. These patterns often suggest experimentally testable hypotheses

48   about multiple interacting processes that may drive species distribution and abundance patterns

49   (Hubbell 1979; Weigand et al. 2012).

50   The usual approach to relating spatial patterns of environmental gradients and

51   populations of sessile organisms (e.g., plants, ant nests, barnacles) starts with recording the

52   positions of individuals, or in the case of composite, plot-based measures, such as species

53   richness or cover values, the positions of plots. This enumeration yields a spatial point pattern

54   (Dale 1999). Environmental variables are then sampled, but they often are not measured at the

55   same spatial grain as the point pattern. Examples include soil samples collected on a regularly-

56   spaced grid (John et al. 2007; Turner and Engelbrecht. 2011), elevation and slope measurements

57   derived from a digital elevation model (Franklin 1995) or climate variables derived from a

58   spatial database, such as 'WorldClim' (Hijmans 2005). Relationships between point patterns and

59   environmental data can be analyzed using non-spatial methods that emphasize causal

60   relationships (e.g., canonical correspondence analysis [Lepš and Šmilauer 2003], species

61   distribution models [Elith and Leathwick 2009], or regression models [Shen et al. 2009]), or by

62   spatial methods that deal with the visualization of pattern and quantification of scales of

63   variability in correlations; our focus here is on the latter.

64   The majority of the standard spatial descriptors used by ecologists, such as semi-

65   variograms, assume that the spatial processes underlying the distribution of organisms (spatial

66   point pattern), the associated environmental gradient, and their covariation are stationary (spatial

67   processes are invariant under translation) and isotropic (non-directional) within the sampling

68   extent (Cressie and Wikle 2011; see Table 1 for spatial terminology used in this paper).

69   However, whilst these assumptions are convenient mathematically, they are typically unrealistic

70   for most real-world examples.

71   First, the strong form of spatial stationarity (invariance under translation) is unlikely to be

72   met in any real-world case. As a result, most spatial processes are assumed to have only second-

73   order stationarity: only the mean, variance, and covariance need be stationary (Vieira et al.

74   2010). However, even second-order stationarity is unlikely in many ecological cases, and we

75   assume only the "intrinsic hypothesis" – that the mean and the semi-variance of the distribution

76   are dependent on inter-point distances, not specific locations (Vieira et al. 2010). Second, in

77   many ecologically realistic cases, environmental gradients create anisotropic patterns in the

78   distributions or abundances of species, where changes in species' distributions or abundances

79   reflect changes in the magnitude of the environmental variable(s).

80   A familiar example of anisotropic relationships between environmental gradients and

81   species distribution arises from the 'stress gradient hypothesis' (Bertness and Callaway 1994).

82   This hypothesis posits that as the environment becomes less stressful for species (e.g., salt spray

83   decreases with distance from the high tide line), intra- or interspecific interactions switch from

84   predominantly facilitative to predominantly competitive. As a result, the pattern of species

85   distributions may shift from aggregated to regular (e.g., Malkinson et al. 2003; Lingua et al.

86   2008) or even hyper-dispersed. Additional processes that may influence clumping of species

87   across environmental gradients include dispersal limitation, habitat filtering, and density-

88   dependent interactions with natural enemies (Condit 2000; Morlon et al. 2008; McGill 2010).

89   Accurate identification of the underlying causes of such complex spatial patterns requires

90   analytical methods that are sensitive not only to the spatial grain of the pattern, but also to non-

91   stationarity and anisotropic changes over space.

92   Here, we illustrate how to use codispersion analysis (Cuevas et al. 2013; Buckley et al.

93   2016) to detect and display both isotropic and anisotropic spatial relationships between a spatial

94   point pattern of species' locations and attributes, and associated environmental variables

95   measured at larger spatial grain. The analysis is based on the codispersion coefficient between

96   the ecological characteristics of a plant species (e.g., the relative abundance, biomass, size or

97   other functional trait) and an environmental variable in a given direction and within a given

98   distance across a particular spatial extent, such as a plot. Codispersion analysis has been applied

99   previously only to a few data types in ecology, including the relationship between tree size and

100   an underlying environmental gradient (topography) at a landscape-level spatial extent (Cuevas et

101   al. 2013), multivariate spectral data (Vallejos et al. 2015), and species co-occurrences (Buckley

102   et al. 2016). In this study, we apply codispersion analysis first to simulated data, and then to tree

103   location and size (diameter) data from two large forest plots, one tropical and one temperate. Our

104   results illustrate how codispersion analysis can be used to detect spatial patterns in tree size

105   across environmental gradients. In addition, we demonstrate a framework for using different null

106   models to test the significance of these spatial patterns (i.e., the departure of observed patterns

107   from random expectation), and how differences in significance among null model tests can be

4

108   used to generate hypotheses about, and guide the structuring of, models of underlying spatial

109   processes. Specifically, we ask, at a 20 × 20-m grain size, what is the direction, magnitude, and

110   spatial pattern in covariation between selected tree species and environmental variables across

111   these two large forest plots? For the purposes of illustrating this method, we selected common

112   species that covaried with the environmental variables in a variety of ways to reflect some of the

113   different underlying processes that can drive species-environment relationships. For example, we

114   can explore if covariation is higher between a tree species' basal area and an environmental

115   variable within 50 m in a northerly direction than would be expected if the species was randomly

116   distributed.

117

118   **Materials and Methods**

119   *An overview of codispersion analysis*

120   Codispersion analysis quantifies the spatial covariation of two or more spatially-explicit

121   datasets. The result is a two-dimensional codispersion graph that allows us to assess how the two

122   datasets co-vary across a range of spatial lags (distances between points) and directions (Table 1;

123   Fig. 1; Vallejos et al. 2014). Codispersion analysis can be applied to datasets organized as spatial

124   point patterns, irregular plots, or rasters. Spatial point patterns depict the locations of individuals

125   (e.g., trees) and possible attributes ("marks") of these individuals (e.g., diameters or other

126   functional traits) measured at these same locations. Rasters often are used to depict

127   measurements of continuously-varying soil or topographic properties as regular grids of cells of

128   a particular size (resolution), from interpolations of variables that have been measured within the

129   same vicinity as, but not precisely at the locations of the point patterns. Spatial point patterns

130   also may be converted (up-scaled) into rasters prior to codispersion analysis, such as by

131   quantifying tree abundances (stem density) or basal areas within raster cells of a given size.

132   In-depth statistical details of the mechanics of codispersion are given in Ruhkin and

133   Vallejos (2008), Cuevas et al. (2013), and Buckley et al. (2016); in the latter we consider species

134   co-occurrences. Annotated R code (R version 3.1.2, R Core Team 2014) for performing

135   codispersion analysis, including its application to examples from this study, is provided in

136   Supporting Information Notes S1.

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

138   First, determine the set of spatial lags $h = \{h_1, h_2\}$: $h \leq 0.25 \times$maximum distance of the

139   shortest side of the sample plot. The two components of h are vectors representing the range of

140   spatial lags to be analyzed for each input dataset X (e.g., tree basal area) and Y (e.g., elevation

141   above sea level). h_1 is oriented parallel to the x axis, and ranges from $-h_{max}$ to $+h_{max}$ (Fig. 1A). h_2

142   is oriented parallel to the y axis and ranges from 0 to h_{max} (Fig. 1A). We note that two opposite

143   directions are incorporated into the analysis along the x axis (positive and negative), so any

144   anisotropy in the data will be more apparent along this axis. We therefore recommend that the

145   dataset be oriented in such a way that the directionality of patterns of particular interest is along

146   the x axis direction, or, that the data be rotated and analyzed in both directions.

147   Second, an Epanechnikov kernel function (Cuevas et al. 2013) is applied across all

148   possible raster cell-to-cell distances for each h, resulting in a smooth spatial variation surface for

149   each individual dataset and their intersection. The "smoothness" of the kernel surfaces is

150   controlled by a set of kernel bandwidth parameters $k = \{k_X, k_Y, k_{XY}\}$ (Cuevas et al. 2013). As

151   rasterization of a spatial point process implies a uniform smoothing at the scale of the raster cell

152   (Buckley et al. 2016), when analyzing rasterized data, we recommend setting each element of k

153   equal to the dimension of the raster cell to avoid unintentional repeated smoothing of the data.

154   Third, semi-variograms for X and Y and the semi-cross-variogram of the intersection of X

155   and Y are computed for the kernel-smoothed surfaces (Cuevas et al. 2013).

156   Fourth, the empirical codispersion coefficient (Matheron 1965) is computed for each lag

157   h as the semi-cross-variogram divided by the square root of the product of the semi-variograms

158   for each of the two variables. The value of the codispersion coefficient ranges from -1.0 (strong

159   negative codispersion) to + 1.0 (strong positive codispersion).

160   Finally, the codispersion values are plotted for each lag h (Fig. 1B). The magnitude of the

161   codispersion values on the graph, and the way in which codispersion values change across the

162   graph, provide information regarding the strength and direction of covariation between the two

163   datasets at different spatial grains (Fig. 1B).

164   Here, we first apply codispersion analysis to simulated data and use three null models to

165   assess the significance of the observed patterns in both simulated and field data. We then apply

166   codispersion analysis to explore spatial relationships between tree basal areas and underlying

167   environmental variables measured within multi-hectare forest plots. The results provide new

168   insights into potential processes underlying observed patterns, and can provide guidance for the

169   development of flexible, mechanistic process-based models for the data.

170   *Simulations*

171   To illustrate how to apply and interpret codispersion analysis for species-environment

172   relationships, we first generated and analyzed a range of species patterns on environmental

173   gradients (examples in Fig. 2; the complete set of simulated patterns is in Supporting Information

174   Notes S2; R code to generate them is in Supporting Information Notes S4). We simulated

175   marked point patterns in a 300 × 300-m "plot" by generating 1500 point locations (representing

176   individual trees) that either were completely spatially random (CSR) or were generated by a

177   Thomas process (using the `rThomas` function in the `spatstat` package of R [Baddeley and

178   Turner 2005]). A Thomas process generates a clumped spatial distribution of points using

179   parameters that describe the spatial intensity of the pattern (in this case, *kappa* = 20 was used),

180   the degree of variation within clumps (*scale* = 0.05), and the average number of points per

181   cluster (*mu* = 10). A simulated diameter (i.e., a "mark") was assigned to each simulated "tree".

182   Diameters were generated using a truncated lognormal distribution with minimum = 1,

183   maximum = 80, mean = 40, and standard deviation = ln(80/15) cm. These marks were distributed

184   across the 1500 trees either randomly, increasing or decreasing to the left side, right side, left or

185   right top corners, or increasing as a large clump in the center of the plot (Fig. 2). We calculated

186   the basal area of the simulated trees within each of 225 contiguous 20 × 20-m cells within the

187   simulated 300 × 300-m plot; 20 × 20-m cells were used because this is the size of typical forest

188   inventory plots used to characterize stand structure. We then generated values for environmental

189   variables within each raster cell. The values of the environmental variable were generated at

190   3600 points within the plot (5 × 5 m cells) and were distributed randomly among the cells or

191   increasing or decreasing to the left side, right side, left or right top corners, or increasing towards

192   a maximum in the center of the plot; these examples include gradient patterns at a range of

193   angles and rotations. The environmental raster gridded into 5 × 5-m cells was upscaled by taking

194   the average value in 20 × 20-m cells so that the values were at the same locations and scale as the

195   basal area data. For the codispersion analyses of these simulated data, we set the bandwidth k =

196   {20 m, 20 m, 20 m}.

197   *Forest plot data*

198   We analyzed species-environment relationships between tree size (basal area) and environmental

199   characteristics at two sites. The two datasets include environmental data that were collected in

200   different ways: (1) direct measurements in each raster cell and (2) spatial interpolation

201   (downscaling) of sparser data to individual raster cells using kriging (John et al. 2007).

202   The first data set is from the third (2000-2002) complete census of the 16-ha Luquillo

203   Forest Dynamics Plot (LFDP) at the Luquillo Long-Term Ecological Research Site, Puerto Rico

204   (Thompson et al. 2002). The four species selected were *Casearia arborea* (L. C. Rich.) Urban

205   (Salicaeae), *Cecropia schreberiana* Miq. (Urticaceae), *Dacryodes excelsa* Vahl. (Burseraceae),

206   and *Prestoea acuminata* var *montana* (Willd.) H.E. Moore (Arecaceae). These are four of the

207   most common species (out of 152 total) in the third census of the LFDP; together they account

208   for 44% of the total basal area of the plot (Table 2A). For each species, the basal area (m^2) of the

209   main stem of each tree was calculated from its measured diameter; basal areas of all trees of a

210   given species in each raster cell were summed to give the species' total basal area for that cell.

211   Elevation (range 333 – 428 m a. s. l.) was measured (1990 – 1992) and mean elevation was

212   calculated for each cell as the mean of the elevations at the four corners of each 20 × 20-m cell

213   (Thompson et al. 2002). Slope (range -0.7 – 65%) was calculated from the corner elevations of

214   each 20 × 20-m cell (Thompson et al. 2002).

215   Basal area of *Casearia* and *Prestoea* decreases but basal area of *D. exclesa* increases with

216   elevation in the LFDP due to the pattern of land-use history in the plot (Thompson et al. 2002).

217   The northern (lower elevation) two-thirds of the plot were logged prior to 1934 and used for

218   subsistence agriculture. Logging and agriculture ceased when the area was purchased in 1934,

219   and the regenerating forest is dominated by *Casearia,* but *Prestoea* also has its highest basal area

220   there. *Prestoea* is often associated with slopes and ravines and disturbed areas (Weaver 2010,

221   Harris et al 2012). At the highest elevations and the southern third of the plot, human disturbance

222   to the forest was limited to selective logging; *Dacryodes* dominates these areas of the plot

223   (Thompson et al. 2002). The dominance of *Cecropia* in the northern portion of the plot recorded

224   in the third census is thought to have resulted from interactions between land-use history and

225   hurricane disturbance. *Cecropia* recruited in huge numbers following Hurricane Hugo in

226   September 1989 (Zimmerman et al. 2010), such that more than 95% of *Cecropia* individuals of

227   this species recruited following this one disturbance event. Zimmerman et al. (1994) noted that

228   *Casearia* was especially susceptible to uprooting during Hurricane Hugo, which opened the

229   forest canopy. Walker (2000) found that *Cecropia* frequently recruited in soil pits caused by

230   uprooted trees and survived longer in this area of the plot because of the persistence of canopy

231   light gaps. Thus, the prevalence of *Cecropia* in the lowermost elevation and flatter northern

232   portion of the plot may be the result of hurricane damage caused to *Caesearia* and other species

233   in this portion of the plot.

234   The second dataset is from the Tyson Research Center Plot (TRCP), a 25-ha forest

235   dynamics plot located at Washington University in St. Louis' Tyson Research Center, Missouri,

236   USA (Spasojevic et al. 2014). We analyzed species-environment relationships for five woody

237   species in the central 20-ha of the plot: *Frangula caroliniana* (Walter) A. Gray (Rhamnaceae),

238   *Lindera benzoin* L. Blume (Lauraceae), *Quercus alba* L., *Q. rubra* L., and *Q. velutina* Lam.

239   (Fagaceae). The three *Quercus* species were some of the most widespread species in the plot,

240   whilst *Frangula* and *Lindera* were selected because they were the two most abundant species in

241   the plot and had interesting, highly-clumped spatial patterns. Together these five species

242   comprised 78% of the total basal area of the TRCP in the 2013 census (Table 2B). Principal

243   components analysis (see Supporting Information Notes S3) was used to summarize, in two

244   composite principal axes, the variation in 17 physico-chemical soil properties that were

245   measured at points across the TRCP in 2013 and kriged to 20 × 20-m raster cells (Spasojevic et

246   al. 2014). Maps of individual environmental variables are available on the TRCP website:

247   http://www.ctfs.si.edu/site/Tyson+Research+Center%2C+Missouri.

248   ***Null model analyses***

249   To assess the significance of the observed codispersion patterns, we used three different null

250   models to randomize aspects of the spatial point processes and their marks (diameters) (Table 3).

251   In each, only the species location data, rather than both species and environment data, were

252   randomized because this was sufficient to break any spatial association of the species data with

253   the environmental variable and allowed us to test the significance of their covariation. The three

254   null models were a CSR model (CSRM), a random labelling model (RLM), and a toroidal shift

255   model (TSM) (see Weigand and Moloney [2014] for detailed descriptions of these null models

256   and other examples of their use).

257   The CSRM generated new spatial locations for trees; observed tree diameters then were

258   assigned randomly (without replacement) to each tree at its new location. Comparison of the

259   observed codispersion patterns with those generated by this null model tested whether there was

260   any non-random spatial pattern in the covariation of the observed tree population (basal area

261   within 20×20-m grid cells) and the environmental variable (Table 3). One difficulty with the

262   CSR is that where species distributions are clumped, this may result in a Type I error rate that is

263   higher than 0.05. Thus, a significant departure from the expectation of this null model may

264   reflect the presence of clumping in the species' distribution (Table 3) and the interpretation of a

265   significant result must be made cautiously. For example, we can use a CSRM to ask if a species

266   increases in basal area at lower elevations in the plot but if the spatial distribution of the species

267   is clumped, we could obtain a "significant" result even if there were no relationship between

268   basal area and elevation. Overall, however, this significance test can be used as an initial test for

269   spatial non-randomness in the dataset.

270   The RLM permuted the observed diameters of the trees while retaining the observed

271   spatial position of each tree. This null model tested whether, given the underlying spatial

272   distribution of trees (a particular autocorrelation structure), their sizes were important in

273   determining any covariation with the environmental variable (Table 3). For example, under this

274   null model, we can test whether covariaton between basal area and soil fertility is due to

275   differences in species' growth rates along a soil fertility gradient, rather than changes in stem

276   density. Mechanistically, in this example, the tree distributions may be driven by clumped

277   dispersal processes that are uniform across the plot area but species' growth rates may vary with

278   soil fertility.

279   The TSM retained the autocorrelation structure of the tree populations by retaining their

280   relative spatial positions and diameters but breaking their spatial association with the

281   environmental variable by moving the entire species pattern in a random distance and direction

282   as though the plot was a torus. This model tested whether the observed pattern in covariation

283   between the species and environmental variable was the same in all parts of the plot, i.e., whether

284   the pattern in covariation is stationary (Table 3). The TSM is similar to the CSRM in that it

285   completely breaks any association between the two variables, but it fixes the distribution pattern

286   of the species. Thus, it distinguishes the case in which a non-random codispersion pattern may

287   simply be driven by relative tree positions from a process-based link between the environment

288   and the species. For example, under this null model, we ask if tree basal area varies with soil

289   fertility and if the nature of that covariation is the same throughout the plot. When combined

290   with the results of the CSRM, we can determine if non-randomness identified by using the

291   CSRM is due to a species-environment relationship (significant TSM) or due to clumping in the

292   species distribution (non-significant TSM) (Table 3).

293   For each species, each of three null models was used to generate 199 new datasets. For

294   each species-environment combination, empirical tail probabilities were obtained by comparing

295   the observed codispersion values at each spatial lag with the vector of codispersion values at the

296   same spatial lags and directions determined from each null model. If the observed value was

297   greater than or equal to the 195th null value or less than or equal to the 5th null value, we deemed

298   it to be significantly different from expected (i.e., a two-tailed test; $P < 0.05$). Thus, the

299   significance tests were made for each lag and direction for which we obtained a codispersion

300   value.

301   Finally, we determined the Type I error rate for each of the three null models by

302   comparing the observed codispersion between two CSR simulated patterns (see Supporting

303   Information Notes S4) to values generated by the CSRM, RLM, and TSM. Note that the Type I

304   error rate, our ability to identify non-significant codispersion values, is invariant to rotation and

305   the error rate tests of the null models do not address the Type II error rate (statistical power),

306   which remains an issue of ongoing research. R code for the null model analysis is provided in

307   Supporting Information Notes S1.

308

309   **Results**

310   *Species-environment associations of simulated forest plot data*

311   Codispersion plots clearly illustrated the relationships between simulated species and their

312   environment, and detected anisotropic, positive, and negative covariation between the two

313   variables (Fig. 2). When the simulated environmental pattern was generated using a CSR

314   process, the cross-variogram and the codispersion both ≈ 0 (little or no spatial covariation),

11

315   whether or not the spatial pattern in basal area was also CSR (Fig. 2A; extended results in the

316   Supporting Information Notes S2). When the environmental variable was generated using a

317   uniform process across the plot, but the basal area of the species decreased from the bottom left

318   to the top right of the plot (i.e., southwest to northeast), the codispersion was weakly negative

319   and weakly anisotropic. This result reflected the changing pattern of covariation in the two

320   variables in the *x*- and *y*-directions. In contrast, the cross-variogram ≈ 0 (Fig. 2B). Sequential

321   pattern rotations of 15 degrees showed that codispersion analysis can also distinguish smaller

322   changes in pattern orientation (Supporting Information Notes S2).

323   When basal area tightly co-varied with the environmental variable, the cross-variogram

324   steeply increased and the codispersion was very high, only weakening at smaller scales that

325   approached the spatial grain of the pattern (Fig. 2C). This pattern, and in fact all pattern

326   combinations, had lower codispersion values when the underlying point pattern of the species

327   was clumped (Thomas process) rather than CSR (Fig. 2D; extended results in Supporting

328   Information Notes S2). A difference in pattern between the left- (west) and right-hand (east)

329   sides of the codispersion graph indicated anisotropy. For example, where the environmental

330   variable decreased from bottom left (SW) to top right (NE), and basal area increased from west

331   to east, codispersion measured negative covariation in the west-to-east direction, but showed

332   some positive covariation at larger scales when looking to the northeast and negative covariation

333   at larger scales when looking to the east (Fig. 2E). This pattern was also reflected somewhat in

334   the cross-variogram, which was flat at small lags but negative at larger lags (Fig. 2E). Similarly,

335   where there was some covariation in a given direction (Fig. 2F), in this case from bottom left

336   (SW) to top right (NE), the codispersion map illustrated the anisotropy (the right hand side of the

337   plot was more negative than the left hand side), showing a relationship that was more negative at

338   larger scales. In this case, the cross-variogram was most negative at similar scales (100-150 m),

339   but did not reflect the anisotropy (Fig. 2F).

340   For all analysis combinations of the three null models and the two underlying tree

341   distributions (CSR- and Thomas-process), none of the observed codispersion values from the

342   two CSR patterns were significantly different from those expected under either model at the 5%

343   level. In our simulations, the CSR model resulted in only one significant cell (out of 200 cells) in

12

344   the codispersion graph (see Supporting Information Notes S4). These results are indicative of a

345   Type I error rate of less than or equal to 5%.

Species-environment associations of observed forest plot data

347   In the LFDP, basal area of *Casearia*, *Cecropia* and *Prestoea* generally decreased with increasing

348   elevation, whilst basal area of *Dacryodes* increased with increasing elevation (Fig. 3, Table 2A),

349   reflecting the interaction of elevation and land-use history in the plot (Thompson et al. 2002).

350   For *Casearia*, this pattern was reflected in a weak, anisotropic codispersion pattern where west-

351   to-east codispersion was more positive than east-to-west codispersion, which became more

352   negative in the north-east direction (Fig. 4A). The codispersion was weakly negative and

353   anisotropic for the basal area of *Cecropia* (Fig. 4B) and similar, but positive, for that of

354   *Dacryodes* (Fig. 4C). Basal area of *Prestoea* negatively co-varied with elevation at the larger

355   scales, reflecting its lower basal area at the highest elevations (Fig. 4D). Basal area of *Casearia*

356   negatively co-varied with slope, whilst basal area of *Cecropia* and *Dacryodes* positively co-

357   varied with slope. In contrast, basal area of *Prestoea* was not strongly related to slope.

358   The comparison of the observed patterns with the codispersion values from the CSRM

359   randomizations revealed that the observed codispersion for all of the species with both elevation

360   and slope was different from random expectation at some, but not all, scales and directions (Fig.

361   4, columns 2 and 3). The only exception was for the relationship between *Prestoea* and slope,

362   which was not significant (Fig. 4D). For all four species, the comparisons with the RLM showed

363   that the number of significant observed codispersion values was lower than expected using the

364   CSRM for about half of the relationships tested, was higher for some, and stayed the same for a

365   few (Figure 4, columns 4 and 5). The comparisons with the TSM showed that the observed

366   codispersion values were significant at few scales and directions for most species-environment

367   combinations (Figure 4, columns 6 and 7).

368   In the TRCP, the first two components from the principal component analysis of the soil

369   chemistry data explained 65% of the variation in measured soil chemistry (plots and PC loadings

370   are given in Supporting Information Notes S2). Variables loading strongly on PC1 were

371   associated with soil fertility and cations (i.e., pH, base saturation, calcium, magnesium,

372   potassium, aluminum, and iron), whilst variables loading strongly on PC2 were associated with

373   soil nitrogen availability (i.e., total nitrogen, NH_4, and nitrogen mineralization rate). These two

374   principal components were used in the codispersion analysis of species-environment

375   relationships for the five focal species.

376   The basal area of the five focal species in the 20×20-m raster cells at TRCP showed a

377   range of strong, weak, positive, and negative relationships with both soil pH and cations (PC1)

378   and soil nitrogen (PC2) (Table 2B, Fig. 5). Although abundant, *Frangula* and *Lindera* were less

379   widespread and their populations were concentrated largely in one or a few patches that

380   corresponded to high values on PC1, generating positive covariation (Fig. 5A−B). The three

381   *Quercus* species (Fig. 5C−E) were more widespread within the plot; *Q. alba* was weakly and *Q.*

382   *rubra* and *Q. velutina* were more strongly negatively related to more fertile soils (high values on

383   PC1). *Quercus alba* positively co-varied with nitrogen (PC2), whilst *Q. rubra* and *Q. velutina*

384   had little or negative co-variation with nitrogen (Fig 5C−E).

385   Codispersion plots revealed both spatial gradients in covariation between basal area and

386   environment and the spatial scales at which covariation was the strongest (Fig. 6, column 1). For

387   example, anisotropic species-environment associations for *Frangula* and *Lindera* were illustrated

388   by positive codispersion with PC2 to the east within the plot, but negative codispersion when

389   looking to the west (Fig. 6A, B). In addition, the spatial scales of covariation differed among

390   species. For instance, the positive co-variation between *Quercus alba* and PC2 was highest at

391   large lags (greater than 50 m) in the east-west direction, whilst *Q. velutina* negatively co-varied

392   with PC1 at larger lags (greater than 60 m) in the north direction, but at smaller lags in the east-

393   west direction (up to 50 m).

394   Observed patterns of species-environment associations at the TRCP often differed from

395   null expectations, but the magnitude of the effect sizes varied among the different null models.

396   The comparison of observed codispersion patterns with those from the null models revealed that

397   the weaker observed codispersion patterns with both soil fertility and cations (PC1) and soil

398   nitrogen variables (PC2) tended not to be significant when compared to expectation when trees

399   were distributed CSR within the plot (Fig. 6, columns 2 and 3). In contrast, comparisons with the

400   RLM (Fig. 6, columns 4 and 5) showed that observed codispersion values were mostly higher

401   than expected. The exceptions to this were, for some scales and directions, for *Frangula* and *Q.*

402   *velutina* with PC2, and for *Q. rubra* with PC1, each of which had significantly more negative

14

403   codispersion at some scales when looking to the west in the plot. The comparisons with the

404   expected values from the TSM largely mirrored those of the CSR comparisons, but with fewer

405   significant values in most cases, such as for *Frangula* and PC2, which was non-significant at all

406   lags.

407

408   **Discussion**

409   Codispersion analysis is a useful method for exploring species-environment relationships in a

410   spatially-explicit context. Simulations showed that the method correctly detected anisotropy and

411   other spatial regularities in the co-variation of the two variables and correctly measured the scale

412   of these effects (Fig. 2). Codispersion values in these simulations were influenced by the

413   underlying spatial pattern of both the species and the environmental variable; more clumping in

414   the tree distribution patterns reduced the magnitude of the codispersion values, even with the

415   same basal area and environmental gradients (Fig. 2; Supporting Information Notes S2).

416   Similarly, a uniform distribution of the environmental variable led to higher magnitude of

417   codispersion values than resulted from a CSR environmental variable (Fig. 2; Supporting

418   Information Notes S2). When observed patterns in field data were combined with null model

419   analysis, codispersion analysis detected the scales and directions of statistically significant

420   codispersion in basal area-environment relationships, and suggested the possible drivers of those

421   relationships (Table 2).

422   The selection of appropriate null models for analyzing spatial point patterns is especially

423   important when the results are used to generate testable hypotheses about processes underlying

424   the observed point patterns (Weigand and Moloney 2014). We suggest that comparisons of the

425   results of the three null models we used to explore significance of codispersion in species-

426   environment relationship can help to tease apart possible influences on observed codispersion

427   patterns (Table 4). In particular, whether observed patterns are found to be significantly different

428   from expectations for one, two, or all three of the null models leads to different hypotheses about

429   possible processes and ecological mechanisms determining the observed patterns (Table 4).

430   The first possibility is that the observed pattern is not significantly different from

431   expectation of all three null models. We obtained this result when examining codisperson of

432   *Prestoea acuminata* and slope at the LFDP (Fig. 4D). We interpret this result as evidence that

433   any observed spatial pattern of the basal area distribution of this species must be due to factors

434   we did not measure. For example, *Prestoea* is dominant in the northern two thirds of the LFDP,

435   which was disturbed by the land use history, greater damage from Hurricane Hugo and is flatter

436   than the southern third of the plot. The high abundance in the northern part of the plot as a result

437   of the land use history reduces the relative strength of the association with slope in this analysis.

438   A second possibility is that the pattern is significantly different under the CSRM, but non

439   significantly different under the TSM. This likely reflects the situation where clumping in the

440   species distribution has resulted in correlation with environment at some lags and directions, but

441   this is not consistent across the plot and therefore, unlikely to reflect a causal dependence of

442   species on environment. Such a result can be used to identify and understand spatial pattern in

443   the species data.

444   Alternatively, the observed pattern could be significantly different from expectation for

445   only two of the three null models. For example, at TRCP, *Q. rubra* was strongly and negatively

446   associated with soil pH and cations at all spatial lags when assessed with the CSRM and TSM

447   (Fig. 6D). However, spatial co-variation was non-significant for a number of lags under the RLM

448   and where it was significant, the observed codispersion was higher than expected. This suggests

449   that although *Q. rubra* basal area was negatively related to the soil environment, the pattern of

450   this relationship, at least at some spatial lags and directions, was not dependent on tree sizes, but

451   rather on their relative spatial positions (autocorrelation structure). Thus, the observed

452   codispersion patterns is likely to be due to processes that drive intraspecific clumping such as

453   unmeasured variation in other environmental variables or land-use history (Thompson et al.

454   2002), interspecific interactions, or dispersal limitation (e.g., Plotkin et al. 2002).

455   Further, significant difference from expectation under the toroidal shift model reveals

456   non-stationarity in the data, which should be taken into account in subsequently developed

457   inferential statistical models. For example, variograms for the TRCP show non-stationarity in

458   PC2 (a large scale trend such that the variogram does not level off and therefore has no sill).

459   Observed codispersion of PC2 (soil nitrogen variables) and *Quercus alba* was significantly

460   different from expectation at large scales suggesting that there was non-stationarity in this

461   pattern. If, in a subsequent model, we were interested in regressing this covariation against other

462   variables, such as slope or elevation, we would need to account for the non-stationarity by

463   applying a method such as generalized least squares, where the correlation in the errors is

464   modelled and then specified in the regression model (Beale et al 2010).

465   These results, and others summarized in Table 4, demonstrate how the application of

466   different null models to codispersion analysis can reveal subtle differences in potential causes of

467   observed bivariate spatial relationships. Other null models that could be explored fruitfully in

468   further research include pattern reconstruction methods (Wiegand and Moloney, pp 368) and

469   spectral methods using raster data (Deblauwe et al. 2015; Wagner and Dray 2015). However, we

470   must first understand what biological processes are being manipulated in each case to interpret

471   observed departures from null expectations. Further, simultaneous comparisons across multiple

472   lag distances can suffer from higher than desired Type I error rates (Loosmore and Ford 2006;

473   Baddeley et al 2014). Future research should address developing a global significance test for

474   codispersion where understanding scales of variation is important.

475   Finally, we note that there are three important considerations to keep in mind when

476   applying codispersion analysis to species-environment data: selecting values for the maximum

477   spatial lag distance, the kernel bandwidth, and the orientation of the pattern in the analysis. We

478   recommend a maximum lag distance of no more than one-quarter of the smallest plot dimension.

479   If the maximum lag is too large, edge effects will influence the largest scales considered. Setting

480   the maximum lag to 25% of the smaller plot dimension ensures an adequate sample size to detect

481   the spatial pattern and minimizes edge effects.

482   The selection of an appropriate kernel bandwidth is comparatively straightforward if data

483   on a regular grid (raster) are used, as we have illustrated here. Because we rasterized the data to

484   20-m grid cells, the scale at which the environmental data were obtained, setting each of the

485   three bandwidth values ($k = \{k_X, k_Y, k_{XY}\}$) equal to 20 m makes sense, as 20 m is the smallest

486   scale at which any pattern could be detected. However, if codispersion is used to analyze

487   bivariate marked point patterns (e.g., two measurements, such as diameter and height, which are

488   recorded for a single point location), the values used for the bandwidth parameters will

489   determine the scales at which the codispersion analysis can detect patterns of spatial co-variation.

490   If the scales of the two variables differ markedly, then their bandwidth parameters, and that of

491   their cross-variogram, should be different. One possibility is to set the values of k_X, k_Y, and k_{XY} to

17

492   the values of the nuggets of their respective variograms (for k_X, k_Y) or cross-variogram (for k_{XY}).

493   Alternatively, Cuevas et al. (2013) suggest an optimization method for identifying appropriate

494   values for $\boldsymbol{k}$.

495   The x,y orientation of the observed biological spatial pattern matters for pattern of

496   codispersion values displayed in the codispersion graph (but not the significance tests) because

497   we have greater resolution of pattern in the x-axis than in the y-axis. Thus, users should think

498   about directionality in the processes driving the spatial patterns being tested. If little is known,

499   rotating the pattern around the midpoint and analyzing it in both directions may aid in identifying

500   directionality in the spatial pattern. Note that this consideration does not affect the data collection

501   unless the plot size or shape precludes the species-environment pattern under study from being

502   adequately sampled within the study extent; therefore, we encourage researchers to consider their

503   hypotheses of pattern during sampling design.

504   Codispersion analysis is useful because it results in a graph that clearly identifies the

505   magnitude, scale and directionality of the observed patterns; it can identify the presence and

506   scale of anisotropy in the spatial pattern; when combined with null models, it can be used to

507   suggest testable hypotheses of ecological process; and it can identify non-stationarity in the

508   spatial pattern of covariation, which influences subsequent inferential modelling choices. It can

509   be used to address a wide range of ecological questions where we are interested in the scale and

510   nature of spatial covariation in variables derived from point-based or grid-based sampling

511   schemes. Such variables may be associated with any attribute of organisms or their locations.

512   The fact that fundamentally different processes can generate similar observed pattern of

513   clumping reinforces the need for spatial methods, combined with appropriate null models, that

514   allow ecologists to discern the relative importance of different processes. Importantly,

515   codispersion can be used for composite measures, such as plant community richness or biomass,

516   and extended to more than two variables (Vallejos et al. 2015), which may be a fruitful path for

517   further ecological applications. Although this method is computationally intensive, the code

518   provided here (Supporting Information Notes S1) is readily adapted for use in a parallel

519   computing framework. Future applications of this approach across a broad range of organisms

520   and biogeographic regions will provide new insights into the ecological causes and consequences

521   of species-environment associations.

522

542 **Author contributions**

543 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

544 the data. H.L.B. and B.S.C. analyzed the data. All authors contributed to writing the manuscript.

545 **References**

546 **Baddeley A, Diggle PJ, Hardegen A, Lawrence T, Milne RK, Nair G. 2014.** On tests of

547   spatial pattern based on simulation envelopes. *Ecological Monographs* **84**: 477–489.

548 **Baddeley A, Turner R. 2005.** spatstat: An R Package for Analyzing Spatial Point Patterns.

549   *Journal of Statistical Software* **12**: 1-42.

550   **Beale CM, Lennon JJ, Yearsley JM, Brewer MJ, Elston DA. 2010.** Regression analysis of
551   spatial data. *Trends in Ecology and Evolution* **13**: 246-264.

552   **Bertness MD, Callaway R. 1994.** Positive interactions in communities. *Trends in Ecology &*
553   *Evolution* **9**: 191-193

554   **Buckley HL, Case BS, Ellison AM. 2016.** Using codispersion analysis to characterize spatial
555   patterns in species co-occurrences. *Ecology* **97**: 32–39.

556   **Chesson P. 2000.** Mechanisms of maintenance of species diversity. *Annual Review of Ecology*
557   *and Systematics* **31**: 343–366.

558   **Condit R, Ashton PS, Baker P, Bunyavejchewin S, Gunatilleke S, Gunatilleke N, Hubbell**
559   **SP, Foster RB, Itoh A, LaFrankie JV, Lee HS, Losos E, Manokaran N, Sukumar R,**
560   **Yamakura, T. 2000.** Spatial patterns in the distribution of tropical tree species. *Science*
561   **288**: 1414–1418.

562   **Cressie N, Wikle CK. 2011.** *Statistics for spatiotemporal data*. John Wiley and Sons: Hoboken,
563   New Jersey.

564   **Cuevas F, Porcu E, Vallejos R. 2013.** Study of spatial relationships between two sets of
565   variables: a nonparametric approach. *Journal of Nonparametric Statistics* **25**:695–714.

566   **Dale MRT. 1999.** *Spatial pattern analysis in plant ecology*. Cambridge University Press:
567   Cambridge, UK.

568   **Deblauwe V, Kennel P, Couteron P. 2015.** Testing pairwise association between spatial
569   autocorrelated variables: A new approach using surrogate lattice data. *PLoS ONE* **7**:
570   e48766.

571   **Elith J, Leathwick J. 2009.** Species distribution models: Ecological explanation and prediction
572   across space and time. *Annual Reviews in Ecology and Systematics* **40**:677–97.

573   **Franklin J. 1995.** Predictive vegetation mapping: geographic modelling of biospatial patterns in
574   relation to environmental gradients. *Progress in Physical Geography* **19**: 474-499.

575   **Harris NL, Lugo AE, Brown S, Heartsill Scalley T. (Eds.). 2012.** Luquillo Experimental
576   Forest: Research history and opportunities. EFR-1. Washington, DC: U.S. Department of
577   Agriculture.

578   **Hijmans RJ, Cameron S, Parra JL, Jones PG, Jarvis A. 2005.** Very high resolution
579   interpolated climate surfaces for global land areas. *International journal of climatology*
580   **25**: 1965-1978.

581   **Hubbell SP. 1979.** Tree Dispersion, Abundance, and Diversity in a Tropical Dry Forest. *Science*

582   **203**: 1299–1309.

583   **John R, Dalling JW, Harms KE, Yavitt JB, Stallard RF, Mirabello M, Hubbell SP,**

584   **Valencia R, Navarrete H, Vallejo M, Foster RB. 2007.** Soil nutrients influence spatial

585   distributions of tropical tree species. *Proceedings of the National Academy of Sciences of*

586   *the United States of America* **104**: 864–869.

587   **Lepš J, Šmilauer P. 2003.** *Multivariate analysis of ecological data using CANOCO*. Cambridge

588   University Press: Cambridge.

589   **Lingua E, Cherubini P, Motta R, Nola P. 2008.** Spatial structure along an altitudinal gradient

590   in the Italian central Alps suggests competition and facilitation among coniferous species.

591   *Journal of Vegetation Science* **19**: 425-436.

592   **Loosmore BN, Ford ED. 2006.** Statistical inference using the G or K point pattern spatial

593   statistics. *Ecology* **87**: 1925-1931.

594   **Malkinson D, Kadmon R, Cohen D. 2003.** Pattern analysis in successional communities – An

595   approach for studying shifts in ecological interactions. *Journal of Vegetation Science* **14**:

596   213–222.

597   **McGill B J. 2010.** Towards a unification of unified theories of biodiversity. *Ecology letters* **13**:

598   627–42.

599   **Morlon H, Chuyong G, Condit R, Hubbell S, Kenfack D, Thomas D, Valencia R, Green JL.**

600   **2008.** A general framework for the distance-decay of similarity in ecological communities.

601   *Ecology letters* **11**: 904–17.

602   **Plotkin JB, Chave J, Ashton PS. 2002.** Cluster analysis of spatial patterns in Malaysian tree

603   species. *The American Naturalist* **160**: 629–644.

604   **R Core Team. 2014.** *R: A language and environment for statistical computing*. R Foundation for

605   Statistical Computing: Vienna, Austria. URL http://www.R-project.org/. R version 3.1.2.

606   **Ruhkin AL, Vallejos R. 2008.** Codispersion coefficients for spatial and temporal series.

607   *Statistics & Probability Letters* **78**: 1290–1300.

608   **Shen G, Yu M, Hu X-S, Mi X, Ren H, Sun I-F, Ma K. 2009.** Species-area relationships

609   explained by the joint effects of dispersal limitation and habitat heterogeneity. *Ecology* **90**:

610   3033-3041.

611   **Silvertown J. 2004.** Plant coexistence and the niche. *Trends in Ecology and Evolution* **19**: 605–
612   611.

613   **Spasojevic MJ, Yablon EA, Oberle B, Myers J. 2014.** Ontogenetic trait variation influences
614   tree community assembly across environmental gradients. *Ecosphere* **5**: 129.

615   **Thompson J, Brokaw N, Zimmerman JK, Waide RB, Everham III, EM, Lodge J, Taylor**
616   **CM, García-Montiel D, Fluet M. 2002.** Land use history, environment, and tree
617   composition in a tropical forest. *Ecological Applications* **12**: 1344–1363.

618   **Turner BL, Engelbrecht BMJ. 2011.** Soil organic phosphorus in lowland tropical rain forests.
619   *Biogeochemistry* **103**: 297–315.

620   **Vallejos R, Mallea A, Herrera M, Ojeda S. 2015.** A multivariate geostatistical approach for
621   landscape classification from remotely sensed image data. *Stochastic Environmental*
622   *Research and Risk Assessment* **29**: 369–378.

623   **Vieira SR, Porto de Carvalho JR, Ceddia MB, González AP. 2010.** Detrending non stationary
624   data for geostatistical applications. *Bragantia* **69**: 1–8.

625   **Wagner HH, Dray S. 2015.** Generating spatially-constrained null models for irregular spaced
626   data using Moran spectral randomization methods. *Methods in Ecology and Evolution* **6**:
627   1169–1178.

628   **Walker LR. 2000.** Seedling and sapling dynamics of treefall pits in Puerto Rico. *Biotropica* **32**:
629   262–275.

630   **Weaver PL. 2010.** Forest structure and composition in the lower montane rain forest of the
631   Luquillo Mountains, Puerto Rico. *Interciencia* **35**: 640-646.

632   **Wiegand T, Huth A, Getzin S, Wang X, Hao Z, Gunatilleke CVS, Gunatilleke IAU. 2012.**
633   Testing the independent species' arrangement assertion made by theories of stochastic
634   geometry of biodiversity. *Proceedings of the Royal Society B*. **279**: 3312–3320.

635   **Wiegand T, Moloney KA. 2014.** *A handbook of spatial point pattern analysis in ecology*. CRC
636   Press: Boca Raton, Florida.

637   **Wiens JA. 1989.** Spatial scaling in ecology. *Functional Ecology* **3**: 385–397.

638   **Zimmerman JK, Comita LS, Thompson J, Uriarte M, Brokaw N. 2010.** Patch dynamics and
639   community metastability of a subtropical forest: compound effects of natural disturbance
640   and human land use. *Landscape Ecology* **25**:1099–1111.

641   **Zimmerman JK, Everham III EM, Waide RB, Lodge DJ, Taylor CM, Brokaw NVL. 1994.**

642   Responses of tree species to hurricane winds in subtropical wet forest in Puerto Rico:

643   implications for tropical tree life histories. *Journal of Ecology* **82**: 911–922.

24

644   **Supporting Information**

645   **Notes S1**: Annotated R code for all analyses and figures

646   **Notes S2**: Full output from codispersion analysis of simulated point patterns

647   **Notes S3**: Results of principal components analysis of Tyson soil chemistry data

648   **Notes S4**: Type I error rates associated with null model analyses

649   Please note: Wiley-Blackwell are not responsible for the content or functionality of any

650   supporting information supplied by the authors. Any queries (other than missing material) should

651   be directed to the *New Phytologist* Central Office.

652

FIGURE CAPTIONS

654   Figure 1. A. An illustration of the creation of directional spatial lags for ecological data

655   organized as rasterized surfaces (both variables are represented by the large grid). The dashed

656   lines represent different spatial lags *h* over which codispersion is calculated in different

657   directions. B. The codispersion graph. The color of each cell is the value of the codispersion

658   coefficient of two variables for each given spatial lag *h* and direction in *x,y* space. In this

659   example, the graph shows negative covariation between the two variables when looking in the

660   east direction, but positive covariation when looking in the northwest direction, indicating

661   anisotropy in the way the two variables covary. The colour pattern on the graph also indicates

662   that the two variables are most negatively correlated at spatial lags above 20m in the positive x

663   direction, and most positively correlated at scales of about 20-30 m in the negative x direction

664   and at about 50-80 m in the y direction. Figures taken from Buckley et al. (2016).

665

666   Figure 2. Simulated species-environment patterns on 20 × 20-m grids in 300 × 300-m plots, their

667   variograms and cross-variograms, and codispersion graphs. In the variograms, the blue line is the

668   environment variogram, the green line is the species variogram and the pink line is the cross-

669   variogram. The colours of the codispersion graphs are scaled from –1 (purple) to +1 (orange).

670   The underlying pattern (environment, basal area) and mean (standard deviation) codispersion

671   values for each analysis were: (**A**) CSR, CSR: 0.03 (0.04), (**B**) uniform, decreasing *x* and *y*: 0.13

672   (0.04) , (**C**) decreasing *x*, decreasing *x*: 0.46 (0.19), (**D**) decreasing *x*, decreasing *x* (underlying

673   Thomas distribution): 0.25 (0.15), (**E**) decreasing *x* and *y*, increasing *x*: -0.16 (0.29), and (**F**)

674   bivariate normal, increasing*x* and *y*: -0.23 (0.11).

675

676   Figure 3. Observed patterns on 20 × 20-m grids in the 16-ha Luquillo Forest Dynamics Plot of

677   elevation (top left), slope (top right), and basal area ($m^2.ha^{-1}$) of *Casearia arborea* (CASARB),

678   *Cecropia schreberiana* (CECSCH), *Dacryodes excelsa* (DACEXC) and *Prestoea acuminata*

679   (PREMON). The variogram for the environmental variable (blue line), variogram for the species

680   (green line) and their cross-variogram (pink line) are shown for each species-environment

681   combination; variables were centered and standardized prior to analysis. In each bubble plot, the

682   dots are positioned at the center of each grid cell point and the sizes of the symbols are scaled to

683   the variable displayed.

684

685   Figure 4. Observed codispersion values; observed minus expected values; and significance (red)

686   or not (blue) at the $P < 0.05$ level relative to null expectation from three null models for bivariate

687   species-environment combinations for four species (abbreviations as in Fig. 3) in the 16-ha

688   Luquillo Forest Dynamics Plot. The colours on the codispersion and observed−expected graphs

689   are scaled from −1 (purple) to +1 (orange); contour lines are at intervals of 0.1.The means and

690   ranges of observed codispersion values are given in Table 2A.

691

692   Figure 5. Observed patterns on 20 × 20-m grids in a 20-ha area of the Tyson Research Center

693   Plot of soil variables represented by two principal components PC1 (top left), PC2 (top right),

694   and basal area ($m^2.ha^{-1}$) of five species: *Frangula caroliniana* (FRACAR), *Lindera benzoin*

695   (LINBEN), *Quercus alba* (QUEALB), *Quercus rubra* (QUERUB), and *Quercus velutina*

696   (QUEVEL). The variogram for the environmental variable (blue line), variogram for the species

697   (green line) and their cross-variogram (pink line) are shown for each species-environment

698   combination; variables were centered and standardized prior to analysis. In each bubble plot, the

699   dots are positioned at the center of each grid cell point and the sizes of the symbols are scaled to

700   the variable displayed.

701

702   Figure 6. Observed codispersion values; observed minus expected values; and significance (red)

703   or not (blue) at the $P < 0.05$ level relative to null expectation from three null models for bivariate

704   species-environment combinations for five species (abbreviations as in Fig. 5) in the 22-ha area

705   of the Tyson Research Center Plot. The colours of the codispersion and observed−expected

706   graphs are scaled from −1 (purple) to +1 (orange); contour lines are at intervals of 0.1.The means

707   and ranges of observed codispersion values are given in Table 2B.

708

709   TABLES

710   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 and 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 20×20-m grids, we apply a 20-m bandwidth because that is the smallest scale (spatial grain) of the data.	Cuevas et al. 2013; Buckley et al. (2016); this paper
Codispersion	A measure of the covariation of two variables in space. For example, covariation in the basal area of two tree species measured in 20×20-m grid cells in a large forest plot.	Cuevas et al. 2013; Buckley et al. (2016); this paper
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 and Moloney 2014
Semi-variogram	A function, usually plotted as a two dimensional	Dale 1999

graph, revealing spatial correlation among measurements from a set of samples. It has three key parameters: nugget, sill and range. The semi-variogram shows at what spatial lags spatial variability occurs in a spatial dataset, i.e., the scale of variation in the data.

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), i.e., we ask, what is their covariation at distances (lags) of 20 m, 40 m, 60 m, …?	Cuevas et al. 2013; Buckley et al. (2016); this paper
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; Weigand and 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 semi-variance 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 et al. 2010 |

711

712

713   Table 2: Abundances, mean diameters (DBH) in centimeters (standard deviation), and the means and ranges in codispersion for basal

714   area-environment relationships for the analyzed species in the (A) Luquillo Forest Dynamics Plot and (B) Tyson Research Center

715   Forest Plot

A. Luquillo Forest Dynamics Plot (2000-2002 census data)

Species	Number of stems	Mean DBH (s.d.)	Total basal area (m²h⁻¹)	Mean (s.d.) codispersion with elevation	Range in codispersion with elevation (min, max)	Mean (s.d.) codispersion with slope	Range in codispersion with slope (min, max)
Dacryodes excelsa	1544	21.18 (15.71)	84.28	0.00 (0.08)	-0.17, 0.14	0.03 (0.02)	-0.03, 0.10
Cecropia schreberiana	2902	10.02 (6.65)	32.95	0.14 (0.04)	0.06, 0.22	0.11 (0.06)	-0.05, 0.25
Casearia arborea	3861	5.63 (5.38)	18.39	0.05 (0.09)	-0.12, 0.21	-0.13 (0.06)	-0.24, 0.02
Prestoea acuminata	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 (s.d.)	Total basal area (m²h⁻¹)	Mean (s.d.) codispersion with soil PC1	Range in codispersion with soil PC1 (min, max)	Mean (s.d.) codispersion with soil PC2	Range in codispersion with soil PC2 (min, max)
Frangula caroliniana	8715	2.04 (0.85)	3.34	0.41 (0.12)	0.17, 0.62	0.03 (0.10)	-0.16, 0.21
Lindera benzoin	4922	1.84 (0.66)	1.48	0.28 (0.14)	0.06, 0.56	0.06 (0.13)	-0.11, 0.36
Quercus alba	2066	29.57 (16.24)	184.66	-0.04 (0.04)	-0.14, 0.07	0.13 (0.05)	0.03, 0.24

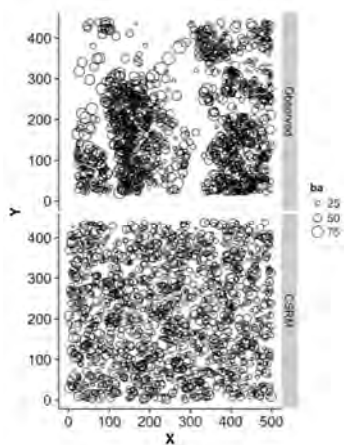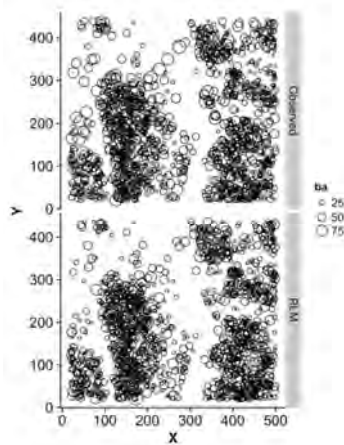|							
Quercus rubra	1551	30.03 (17.63)	147.73	-0.39 (0.12)	-0.56, -0.15	0.03 (0.05)	-0.06, 0.13
Quercus velutina	691	33.46 (13.92)	71.27	-0.09 (0.09)	-0.28, 0.08	-0.09 (0.05)	-0.19, 0.03

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

717

718   Table 3: The three null models, an example realization of each, how they were applied in this

719   paper, and their associated null process models. For each example, (which were randomized by

720   each null model), the hypothesized ecological process is that basal area (BA) is conditional on

721   one or more of the spatial point pattern of trees (ppp), their diameters (marks) and the spatial

722   distribution of the environmental variable (env): BA | (ppp, marks, env). Each null model breaks

723   apart this conditional process in a different way, as is indicated by the conditional statement (in

724   **bold** type) and its associated explanation in the "Null process" column.

Null model	Example	Null process	Test
Completely spatially random (CSRM)		**BA; (ppp, marks)	env** The spatial distribution and diameters of individual trees, from which basal area is computed, are random and therefore independent of the environment.	This model tests for non-random spatial covariation between BA and the environmental variable.
Random labelling model (RLM)		**BA; marks	(ppp, env)** Where individual trees grow is fixed (due to another process, such as competition), but how they grow (size) is independent of the environment.	This model tests if the environmental variable is associated with growth differences among individual trees, whose diameters are aggregated to compute BA in each raster cell.

32
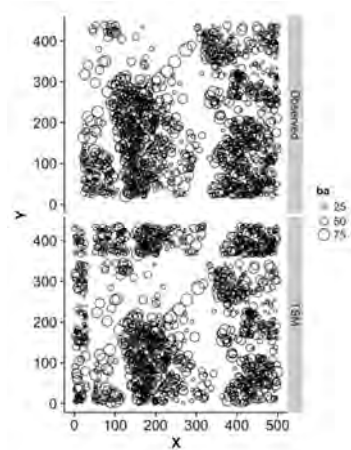
| Toroidal shift model (TSM) | | **BA; env \| (ppp, marks)** Where trees grow relative to one another and the spatial distribution of their relative sizes is driven by an unknown (unmeasured) process, but where and how they grow (e.g., size) is independent of the environment. | This model tests for non-random spatial co-variation between BA and the environmental variable, given the underlying marked spatial point pattern of the species. |

725

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

Null model results			Interpretation	Species-environment examples
CSRM	RLM	TSM		
N.S.	N.S.	N.S.	Basal area is independent of the environment.	*Prestoea acuminata* vs. slope (Fig. 4D)
Sig.	N.S.	N.S.	Basal area is independent of the environment but aggregated in space; this pattern depends on tree spatial distributions, not tree sizes, i.e., the spatial pattern of basal area is not different than expected if diameters were randomly assigned to trees.	*Casearia arborea* vs. elevation (Fig. 4A)
N.S.	**Sig**.	N.S.	Basal area is not strongly related to the environment because tree positions are independent of the environmental variable; however, the environment causes non-random differences in tree growth.	*Quercus alba* vs. PC1 (Fig. 6C)
Sig.	N.S.	**Sig**.	Basal area is non-randomly related to the environment; this pattern depends on the relative spatial positions of trees, not their sizes.	*Quercus rubra* vs. PC1 (Fig. 6D)
Sig.	**Sig.**	N.S.	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.	*Cecropia schreberiana* vs. elevation (Fig. 4B)
Sig.	**Sig.**	**Sig.**	Basal area is non-randomly related to the	*Frangula caroliniana*

environment and this depends on both tree   vs. PC1 (Fig. 6A)

spatial distributions and their sizes. The

environment influences both where trees

grow and their sizes.

---

727   The CSR model (CSRM) resulted in completely spatially random (CSR) tree spatial positions

728   within the plot. The random labelling model (RLM) shuffled the marks (here, diameters)

729   associated with each tree. The toroidal shift model (TSM) fixed the relative tree positions and

730   their observed diameters, but moved the entire set of tree point locations in a random distance

731   and direction as though the plot is a torus.

A.
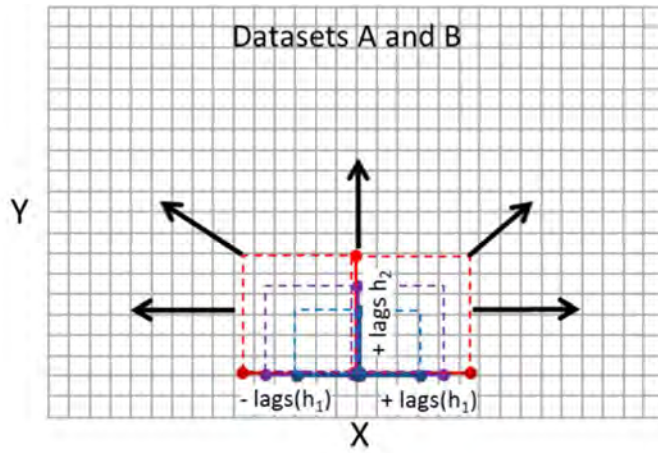
B.
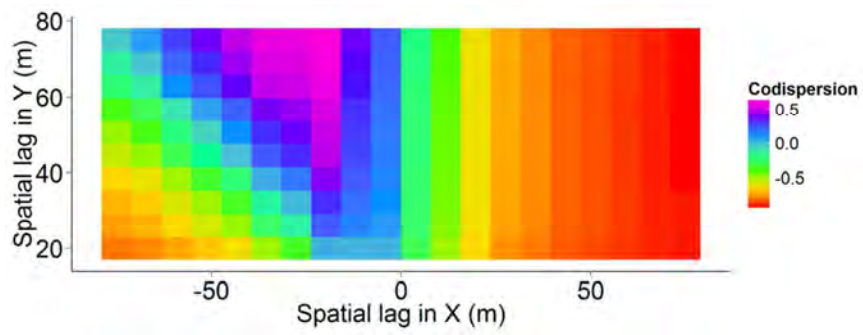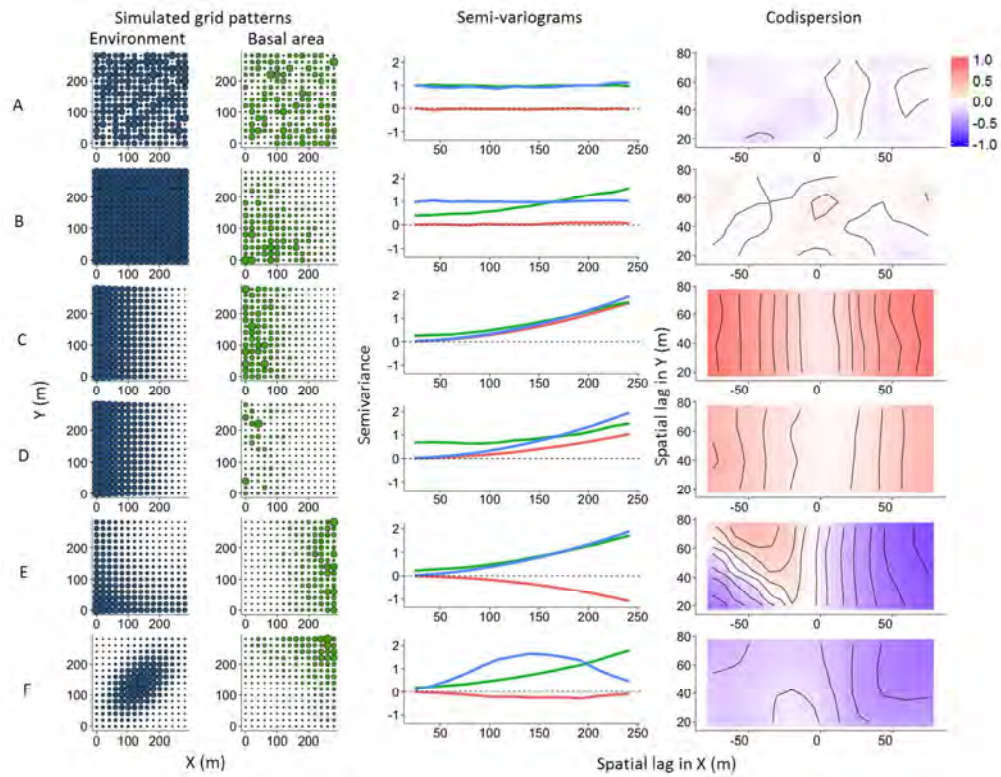
Figure 2. Simulated species-environment patterns on 20 × 20-m grids in 300 × 300-m plots, their variograms and cross-variograms, and codispersion graphs. In the variograms, the blue line is the environment variogram, the green line is the species variogram and the pink line is the cross-variogram. The colours 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: 0.13 (0.04) , (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, increasingx and y: -0.23 (0.11).
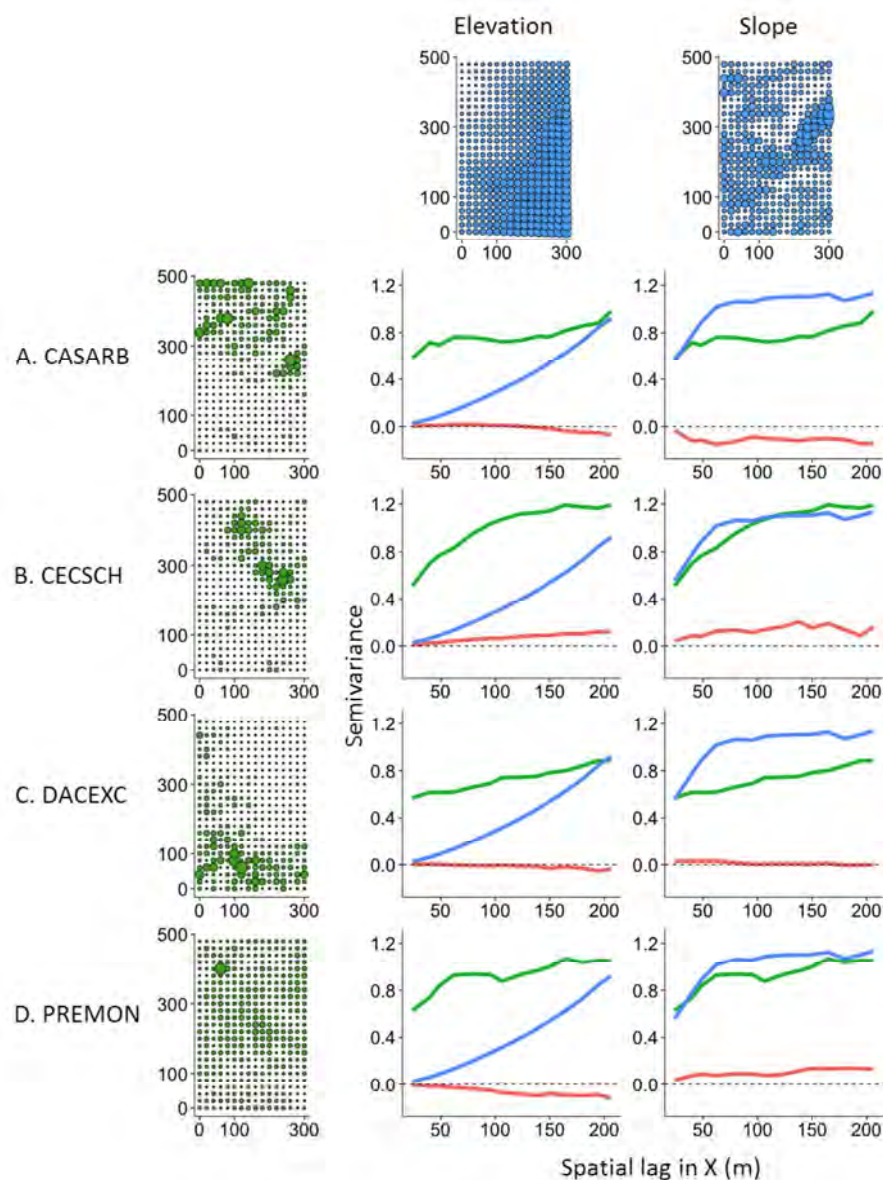348x274mm (300 x 300 DPI)

Figure 3. Observed patterns on 20 × 20-m grids in the 16-ha Luquillo Forest Dynamics Plot of elevation (top left), slope (top right), and basal area (m2.ha-1) of Casearia arborea (CASARB), Cecropia schreberiana (CECSCH), Dacryodes excelsa (DACEXC) and 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 prior to 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.
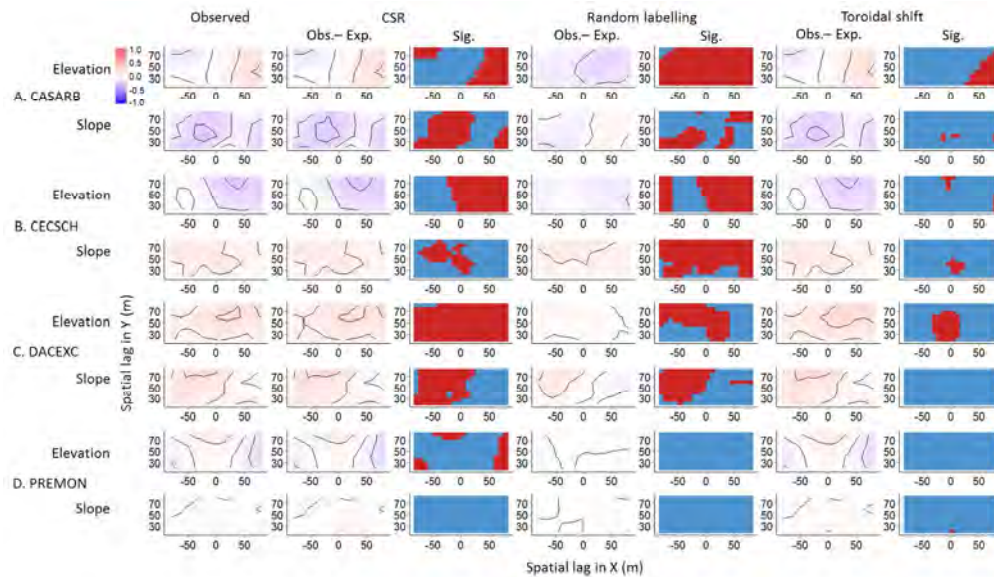370x500mm (300 x 300 DPI)

Figure 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 colours 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 observed codispersion values are given in Table 2A.
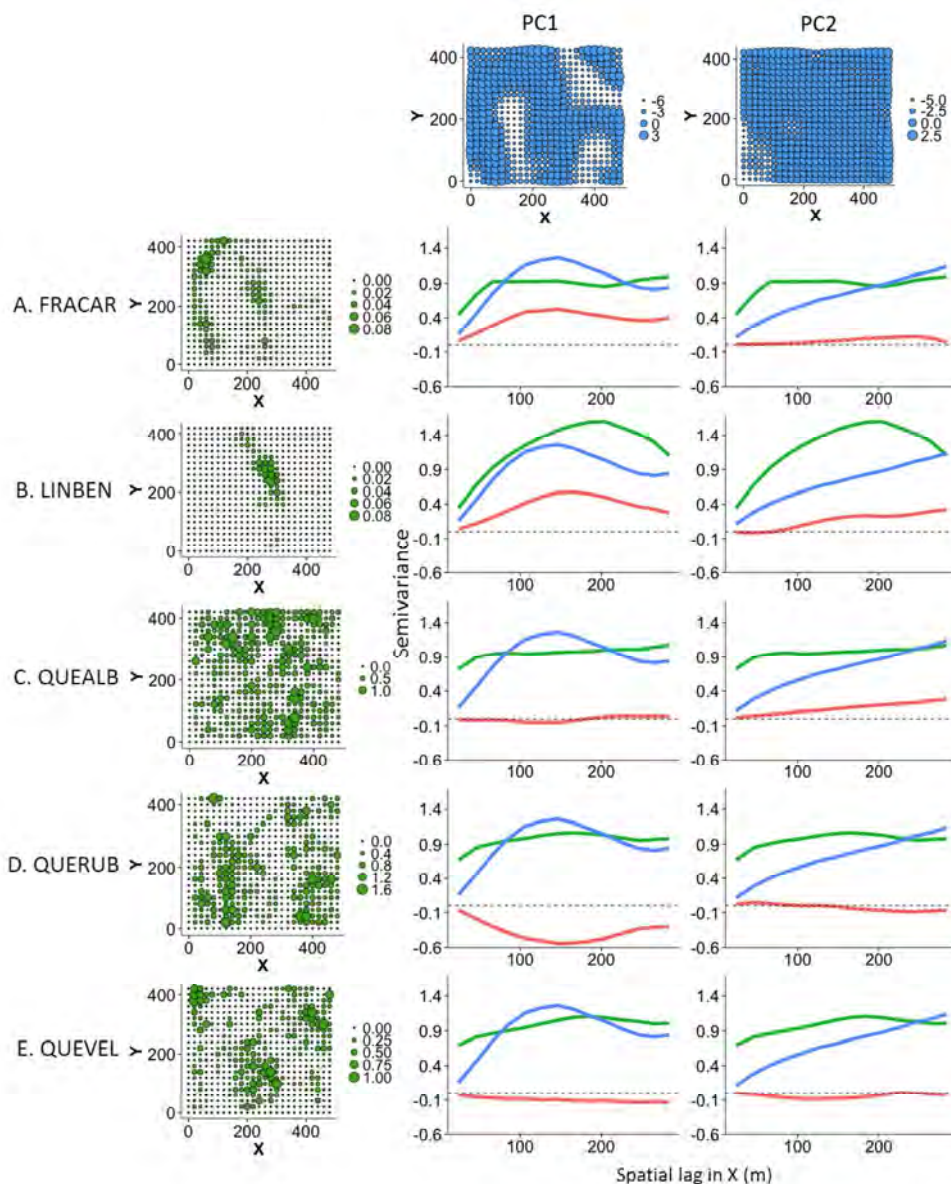308x180mm (300 x 300 DPI)

Figure 5. Observed patterns on 20 × 20-m grids in a 20-ha area of the Tyson Research Center Plot of soil variables represented by two principal components PC1 (top left), PC2 (top right), and basal area (m2.ha-1) of five species: Frangula caroliniana (FRACAR), Lindera benzoin (LINBEN), Quercus alba (QUEALB), Quercus rubra (QUERUB), and 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 prior to 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.
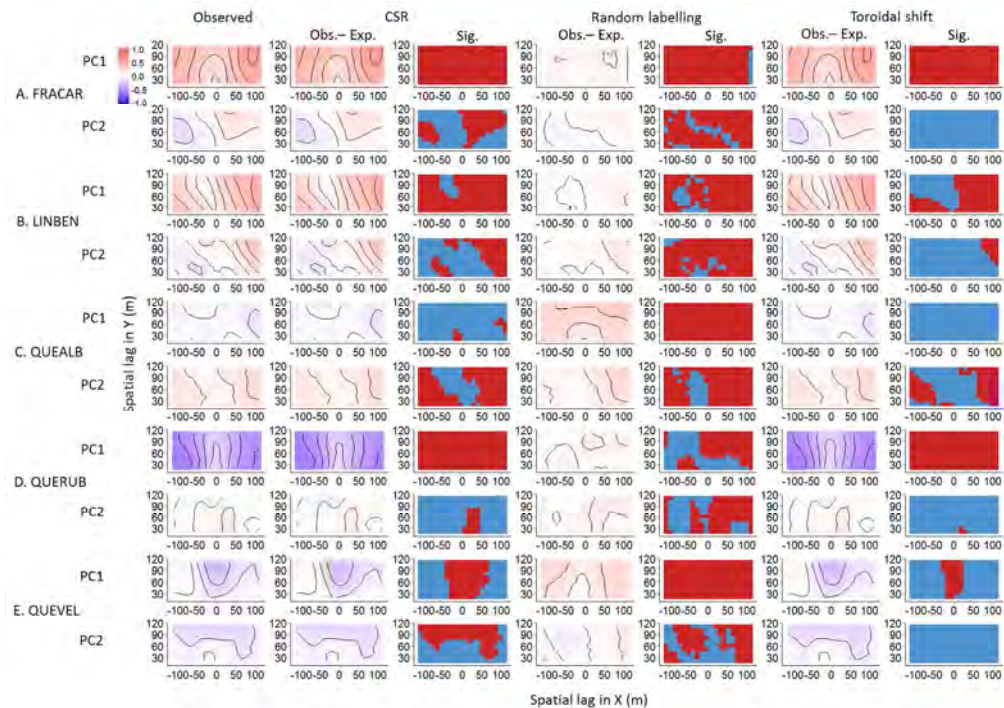407x505mm (300 x 300 DPI)

Figure 6. 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 five species (abbreviations as in Fig. 5) in the 22-ha area of the Tyson Research Center Plot. The colours 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 2B.

377x270mm (300 x 300 DPI)