Spatially disentangling within- and between-species components of community variation reveals processes driving community assembly

Daniel J. McGlinn[*] and Allen H. Hurlbert

Biology Department, University of North Carolina, Chapel Hill, NC

Email correspondence: *danmcglinn@gmail.com*

[*]Current institution: *Utah State University, Logan, UT*

Abstract

Ecological processes influence the spatial structure of ecological communities, and thus ecologists have often attempted to infer process from spatial community patterns. However, typically only one facet of community variation, intra-specific aggregation, is examined in a spatial context while inter-specific associations have been ignored. We demonstrate how a spatially explicit, joint consideration of these two fractions of community variation can be used to estimate the strength of ecological drivers of community assembly. Specifically, we estimate the relative importance of environmental filtering along a productivity gradient and dispersal limitation in shaping the spatial pattern of intra- and inter-specific aggregation in avian communities in the eastern and western regions of the United States. We accomplished this by combining community variograms, randomization-based null models, and a spatially explicit stochastic simulation model into a single cohesive inferential framework. Our analyses indicated that although the bird species in both geographic regions displayed similar patterns of fractal-like intra-specific aggregation, they differed strongly in their spatial patterns of inter-specific association. Nevertheless, our simulation model suggested that the estimated strength of environmental filtering and dispersal limitation were fairly similar in the two regions. Our analytical framework provides a powerful new method of examining the influence of key processes in community assembly on community variability.

Introduction

Ecological processes, such as competition, environmental filtering, and dispersal, influence the spatial distribution of species. As such, it is reasonable to hope that the spatial structure of individuals within and between species might help to reveal the most important processes driving community assembly. Historically, however, space has often been ignored or treated as a nuisance variable in community ecology. For example, randomization-based null models were developed as an important tool in the study of community assembly rules (Connor and Simberloff 1979, Gotelli 2000, Gotelli et al. 2010), and these have only rarely been made spatially explicit (Palmer and van der Maarel 1995, Roxburgh and Matsuki 1999, Lennon 2000). Alternatively, spatial autocorrelation is something that has been 'corrected for' when examining the relationship between two other variables in a spatial context (Legendre 1993, Dormann et al. 2007). Although ecologists have been interested in deciphering the causes of spatial variation in communities for some time, analyses which explicitly examine the spatial components of community variability have only recently been adopted into mainstream ecological analyses (Fortin and Dale 2005, Wagner and Fortin 2005).

The spatially explicit patterns that have received attention from community ecologists are the species-area relationship (SAR; Drakare et al. 2006) and the distance decay of similarity (DD; Nekola and White 1999). These two relationships are interconnected in that the shape of the SAR is intimately related to the rate of distance decay in species composition (Harte and Kinzig 1997, Morlon et al. 2008). While the SAR and DD explicitly consider the spatial structure of ecological communities and are potentially useful tools for testing models of community assembly, they primarily reflect only one facet of community variation, intra-specific aggregation, which characterizes the degree to which individuals within a species are clustered or

3

overdispersed (Table 1). However, SAR and DD patterns are insensitive to inter-specific aggregation, which reflects the degree to which species positively or negatively co-vary with each other in space (Table 1). Specifically, models in which individual species are spatially aggregated but independent of each other are sufficient to explain the majority of variance in the SAR and DD (Plotkin et al. 2000, Martín and Goldenfeld 2006, White and Gilchrist 2007, Morlon et al. 2008, Shen et al. 2009). This implies that tests of theory based upon the SAR and DD only gauge how accurately the theory can reproduce patterns of intra-specific aggregation, and that these empirical relationships are not as sensitive to the pattern of inter-specific aggregation. Strong intra-specific aggregation is typically attributed to strong dispersal limitation or specialization on patchily distributed environments or resources (Bell et al. 2006, van Teeffelen and Ovaskainen 2007). In contrast, negative inter-specific aggregation (i.e., negative species covariance) has been typically interpreted as a signature of inter-specific competition (Diamond 1975, Schluter 1984, Gotelli et al. 2010), but may reflect strong dispersal limitation and environmental filtering as well (Ulrich 2004, Bell 2005). Because many different processes can produce relatively similar patterns, it is unwise to attribute patterns of intra- and inter-specific aggregation to one particular process *a priori* (McGill 2010).

An inferential framework based upon both intra- and inter-specific spatial patterns should be more powerful in elucidating the roles of ecological processes than a framework utilizing only one set of patterns or the other. Nevertheless, there have so far been no successful attempts to link ecological processes with intra- and inter-specific spatial variation in community composition. However, Wagner (2003, 2004) developed a useful statistical tool that could provide the basis of such a framework: the community variogram. The community variogram is a powerful and elegant tool for simultaneously quantifying the spatial pattern of intra- and inter-

4

specific variation by spatially decomposing the species covariance matrix ($\mathbf{C}$, Table 1) (Wagner 2003, 2004). In a spatial context, the diagonal elements of $\mathbf{C}$ reflect the degree of intra-specific aggregation and the off-diagonal elements of $\mathbf{C}$ reflect the degree of inter-specific aggregation (Wagner 2003).

Although community variograms provide a straightforward way of quantifying these two types of spatial patterns, it is still necessary to link these patterns to multiple, interacting ecological processes. One approach to this problem is to simulate community assembly under known combinations and strengths of processes and examine the extent to which empirical community patterns reflect the signature of these processes (Gotelli et al. 2009, Colwell and Rangel 2010, Scheiner et al. 2011, Stegen and Hurlbert 2011, Hartig et al. 2011). While this sort of pattern-oriented modeling (*sensu* Grimm et al. 2005) can be a powerful approach, it still has limitations. Strong quantitative comparisons between simulated and empirical patterns can be hindered by unknown differences in numbers of individuals, species, time periods, and spatial extents. One solution to this problem is to couple a simulation model with randomization-based null models which are applied to both the simulated and empirical communities. There are two important benefits of randomization-based null models: 1) they may be designed to only 'nullify' specific aspects of empirical patterns that are of interest to the process we wish to make inferences about, and 2) they provide a means of relativizing quantitative comparisons between simulated and empirical communities. Therefore, the goal of this study is to combine community variograms, pattern-oriented simulation modeling, and randomization-based null models to develop an inferential framework that links spatial community variation and the ecological processes of environmental filtering and dispersal limitation. To accomplish this goal we: 1) demonstrate how community variograms can be used to detect non-random spatial community

variation, 2) develop a simulation model that incorporates the ecological processes we are

interested in, and 3) demonstrate the utility of this combined approach using empirical data on

North American bird assemblages.

Framework for Statistical Inference

Decomposition of community variance

A simple and cohesive framework for detecting spatial signatures of intra- and inter-specific

aggregation may be developed by first recognizing that total variance of a community matrix

(i.e., a site-by-species matrix populated either with 1's and 0's for presence/absence, or with

values of abundance, Table 1) may be expressed as the sum of the within- and between-species

components of variance. For *S* species this may be expressed as:

$$\mathrm{Var}\left(\sum_i^S \mathbf{x}_i\right) = \sum_i^S \sum_j^S \mathrm{Cov}(\mathbf{x}_i, \mathbf{x}_j)$$

$$= \underbrace{\sum_i^S \mathrm{Var}(\mathbf{x}_i)}_{} + \underbrace{2\sum_i^{S-1} \sum_{j>i}^S \mathrm{Cov}(\mathbf{x}_i, \mathbf{x}_j)}_{}$$

(1)

$$\underbrace{\phantom{\mathrm{Var}}}_{\substack{\text{Community}\\\text{Variance}}} \quad \underbrace{\phantom{\mathrm{Var}}}_{\substack{\text{Within-species}\\\text{Variance}}} \quad \underbrace{\phantom{\mathrm{Cov}}}_{\substack{\text{Between-species}\\\text{Covariance}}}$$

where $\mathbf{x}_i$ is a vector of the *i*th species across *n* sites of a given measure of occurrence (e.g.,

abundance, presence/absence data). Although within-species patterns of variance are interesting

in their own right, many authors have used equation 1 to examine whether species are

independently distributed. Specifically, the difference or ratio of total community variance to

within-species variance is a metric of inter-specific association (Pielou 1972, Schluter 1984,

Palmer 1987, Wagner 2003). This ratio is typically referred to as the variance ratio (*V* =

community variance / within-species variance), and it is based upon the assumption that the

between species-variance will be zero when species are independent (Schluter 1984). When V is greater than one, species are on average positively associated, and when V is less than one, species are negatively associated. It is important to recognize that the V-ratio expresses the direction and magnitude of total between-species covariance.

The V-ratio is an intuitive and straightforward community-level metric of inter-specific association, but there are at least three critical problems that may be encountered when interpreting this statistic. Palmer and van der Maarel (1995) highlighted two of these problems, namely that V values less than one (i.e., apparent negative species associations) can arise simply from (1) limitations on the number of individuals that have been sampled and (2) spatial dependence in the population. The solution to avoiding these Type I errors is to compare empirical values of V with a spatially-explicit, randomization-based null model that maintains the observed number of occurrences and the spatial dependence within-species but nullifies between species associations (Palmer and van der Maarel 1995, Roxburgh and Chesson 1998, Roxburgh and Matsuki 1999).

The third problem that plagues the V-ratio occurs when considering communities characterized by strong positive and negative associations. In these communities, the V-ratio may falsely suggest species are independent, a Type II error. To see this consider that when summing across all of the pairwise species covariances the strong associations of opposite signs may cancel each other and result in total between-species variance close to zero and thus an erroneous inference that species are independently distributed. This particular problem becomes more obvious if we recognize that V is directly related to the average of all possible species correlation coefficients: $V = (S-1)\bar{r} + 1$, where $\bar{r}$ is the average correlation between all S species (McCulloch 1985). Therefore, like any metric of central tendency, V is unable to distinguish

7

communities that have multi-modal or skewed distributions of association from those that are the assumed normal distribution. In Figure 1, we have illustrated four possible types of distributions of correlation coefficients which would be falsely assumed to be consistent with the null hypothesis that species are randomly distributed (dashed lines in Figure 1). Although this problem was recognized early on by Pielou (1972) no general solution has emerged.

The potential Type II errors that may arise when examining the distributions in Figure 1 suggest that a more useful analysis of the variance in community properties should instead separately consider the positive and negative components of between-species covariance. As such, total community variance may be expressed as the sum of three components:

$$\mathrm{Var}\left(\sum_i^S \mathbf{x}_i\right) = \sum_i^S \mathrm{Var}(\mathbf{x}_i) + 2\sum_{i<j}^{S-1}\sum_j^S \mathrm{Cov}_+(\mathbf{x}_i,\mathbf{x}_j) + 2\sum_{i<j}^{S-1}\sum_j^S \mathrm{Cov}_-(\mathbf{x}_i,\mathbf{x}_j) \tag{2}$$

where Cov_+ and Cov^- represent the positive and negative components of between-species covariance respectively. In a presence/absence, site-by-species matrix, Cov_+ is the number of 'aggregation units' or $\begin{smallmatrix} 1 & 1 \\ 0 & 0 \end{smallmatrix}$ submatrices present (Table 1). In contrast, Cov^- is the number of 'segregation units' (i.e., 'checkerboard units' *sensu* Stone and Roberts [1980]) or $\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}$ submatrices (Table 1). Here we are implying that Cov_+ is a metric of how aggregated different species are with respect to each other while Cov^- is a metric of how segregated species are. The joint consideration of Cov_+ and Cov^- allows each distribution in Figure 1 to be distinguished from the null distribution. Cov^- is algebraically related to the popular *C*-score, commonly used in studies attempting to detect negative associations in communities (Stone and Roberts 1990). The *C*-score is simply Cov^- standardized by the number of unique species comparisons ($S \cdot (S - 1)/2$) (Bell 2005).

8

We have argued that equation 2 provides a reasonable way to partition total community variation, but one of our primary goals is to understand how community variation changes depending upon the spatial scale considered. To accomplish this goal, we will describe how each component of equation 2 may be quantified using spatial variograms. In particular, we take advantage of the fact that community variance may be summarized by a set of distance-dependent species variance-covariance matrices, $\mathbf{C}(h)$. The particular cells of this matrix, $c_{ij}(h)$, are populated using Eq. 6 of Wagner (2003):

$$c_{ij}(h) = \frac{1}{2n_h} \sum_{a,b|h_{ab} \approx h} (x_{ia} - x_{ib})(x_{ja} - x_{jb}) \tag{3}$$

where c_{ij} is the covariance between species i and j, n_h is the number of pairs of samples separated by a distance of h, x_{ia} and x_{jb} are the abundance or presence/absence of species i and j in samples a and b respectively which are separated by distance h. Equation 3 provides estimates of within-species variance at a given spatial lag when $i = j$, and estimates of between-species covariance when $i \neq j$. Therefore we may state that:

$$\mathrm{Var}(\mathbf{x}_i, h) = \frac{1}{2n_h} \sum_{a,b|h_{ab} \approx h} (x_{ia} - x_{ib})^2 \tag{4}$$

$$\mathrm{Cov}_+(\mathbf{x}_i, \mathbf{x}_j, h \mid i \neq j) = \frac{1}{2n_h} \sum_{a,b|h_{ab} \approx h} (x_{ia} - x_{ib})(x_{ja} - x_{jb}) \text{ if } (x_{ia} - x_{ib})(x_{ja} - x_{jb}) > 0 \tag{5}$$

$$\mathrm{Cov}_-(\mathbf{x}_i, \mathbf{x}_j, h \mid i \neq j) = \frac{1}{2n_h} \sum_{a,b|h_{ab} \approx h} (x_{ia} - x_{ib})(x_{ja} - x_{jb}) \text{ if } (x_{ia} - x_{ib})(x_{ja} - x_{jb}) < 0 \tag{6}$$

Equation 4 is a metric of intra-specific spatial aggregation, and Eq. 5 and 6 are equal to 0 when their conditions are not met and are expressions of inter-specific spatial aggregation and segregation, respectively, at a given distance class. We provide additional discussion of the statistics described above in Supplementary material Appendix 1.

9

Null model comparison

Randomization-based null models are necessary to examine whether the observed pattern of species occurrences, quantified by the fractions of community variance, differ from the pattern expected under a specific null hypothesis (Gotelli and Graves 1996). Over the years, many different null models that permute species occurrences have been proposed (Gotelli 2000). Here we will consider three different null models that vary in how constrained the randomization is with respect to the observed distribution of species occupancy and spatial pattern of intra-specific aggregation (Table 1-2). The Random Assignment (RA) null model, in which the cells of the community matrix are swapped randomly, generates a null expectation for the non-spatial, intra-specific variance. The Independent Assignment (IA) null model independently shuffles species occurrences across the sites (maintaining observed levels of species occupancy), and this algorithm provides the null expectation for the spatial, intra-specific variance and non-spatial, inter-specific variance. The Spatial Assignment (SA) null model maintains the observed spatial dependence for each species (maintaining observed levels of species occupancy and intra-specific aggregation) but nullifies between-species spatial dependence (referred to as the Random Patterns null model by Roxburgh and Chesson 1998). Three null models are necessary because different fractions of community variance are sensitive to different aspects of community structure and therefore require different empirical constraints to provide meaningful comparisons. Supplementary material Appendix 1 contains a detailed description of the null models and an explanation of their appropriateness for each of the statistical metrics we compare. Additionally, we provide the source code and specific details on our application of the SA null model in the supplementary material Appendix 2.

An example

We will illustrate how the statistical machinery developed above can be used by comparing

two artificial communities A and B respectively (Figure 2). These artificial communities are

meant to illustrate 1) how fractions of community variance can change with spatial lag, and 2)

how a joint consideration of positive and negative fractions of between-species covariance can

be more informative than the total combined between-species covariance. The species in the two

communities have similar patterns of intra-specific aggregation (i.e, species in both communities

occupy similar fractions of the landscape with similar patchiness, Figure 2a,b), but they differ in

their patterns of inter-specific aggregation (Figure 2c,d). Similarity in the communities' intra-

specific spatial structure is visible in panel (b) of Figure 2 where species in both communities

show strong patterns of spatial aggregation until a distance of five pixels. In contrast, the non-

spatial analysis, which simply considers the pattern averaged over all possible spatial scales,

suggests that the within-species component of variance is no different than expected under the

Random Assignment null model (see *z*-score and *p*-value in lower right corner of each panel for

the non-spatial test statistics).

Both the spatial and non-spatial patterns of total community covariance (Figure 2c) suggest

that overall the two communities differ strongly in their inter-specific aggregation where species

in A are positively associated and species in B are negatively associated, but importantly the

strength and direction of this inference changes depending upon the spatial scale of the

comparison. For example, in community B the species appear independent until spatial lags of

greater than 10 pixels at which point the species are strongly negatively associated. The positive

and negative fractions of total covariance provide additional resolution of the inter-specific

spatial structure (Figure 2d). This can be clearly seen at a spatial lag of 3 (compare vertical

11

dashed line in panels (c) and (d) of Figure 2). In community A, at a spatial lag of 3, total

covariance suggests that the species are independent but the positive and negative fractions of

covariance indicate that the species are actually simultaneously displaying strong positive and

negative associations. Thus when considering the patterns of species association in these

communities, it appears that the simultaneous decomposition of community covariance in terms

of spatial scale and positive and negative fractions is much more informative than the classical

non-spatial analyses of total covariance.

Framework for Ecological Inference

Non-random patterns of within- and between-species variance on their own cannot be used to

infer specific ecological processes. We can begin to detect the signatures of various ecological

processes on community structure only when we couple these statistical patterns with

manipulative experiments, explanatory statistical models, or simulation studies that target

biological processes. For the purposes of demonstration here we examine the ecological insights

that may be gained by linking empirical patterns of bird community structure with simulation

models that incorporate varying levels of environmental filtering and dispersal limitation, two

ecological processes that are thought to play an important role in shaping community structure at

the spatial scales we are considering. We use a pattern matching approach to examine under

which parameter combinations the simulated communities most closely approximate the

empirical patterns (Rangel et al. 2007, Gotelli et al. 2009). Implicit in pattern matching

techniques is that similar patterns imply similar driving processes.

A key limitation of using a simulation approach is that invariably many of the parameters

cannot be empirically constrained (e.g., number of individuals in the community, local carrying

capacity)(Gotelli et al. 2009). Therefore, frequently the output of a simulation model is on a

different numerical scale than the empirical community which prevents meaningful direct comparisons. One potential solution to this problem is to only compare the relative the degree of deviation of the simulated and empirical patterns from their corresponding null models. Specifically, standardized effect size (SES) statistics provide a measure of community structure that is comparable between simulated and empirical communities. For this study we used z-scores at each spatial lag for each of the three metrics of community structure (z_{Var}, z_{Cov+}, and z_{Cov-}) as our SES metrics, calculated by subtracting the mean of the null randomizations from the observed value, and dividing that difference by the standard deviation of the null randomizations. We examined which set of simulation parameters (see below) resulted in the smallest sum of squares deviation between the z-scores of the empirical data and the z-scores of the simulated data over all spatial lags.

Empirical communities

We evaluated our inferential framework using data from the North American Breeding Bird Survey (BBS, Sauer et al. 2008). Avian community data provide a suitable test case for our approach because bird species occurrences are well known to be spatially structured due to their environmental preferences (i.e., environmental filtering, Bahn and McGill 2007, White and Hurlbert 2010) and their expression of natal philopatry which results in spatial autocorrelation in their distribution through time (i.e., dispersal limitation, Weatherhead and Forbes 1994, Paradis et al. 1998). Each survey in the BBS is conducted during the breeding season (usually June) and consists of 50 evenly spaced point counts along a 40-km roadside route. During each 3 minute point count, a single observer records all birds seen or heard within 400 m. We considered a species present along a given route if it was observed at least once during a 5-year window from 2003-2007. A temporal window helps to minimize the number of false absences simply because

a species is difficult to detect (McGill 2003, Hurlbert and White 2005). We did not consider bird groups that are not well surveyed by the BBS including water-birds, nocturnal birds, and raptors. Maps of the geographical pattern of route density and average route species richness are in Supplemental material Appendix 3, Fig. A3.1.

To apply the SA null model it is necessary to work with gridded occurrence data, and so we projected the BBS route coordinates using a Lambert-Azimuthal equal-area projection and rasterized the occurrence records to a spatial grain of 100 x 100 km. We defined two 1800 x 1800 km grids randomly but with the criteria that they did not overlap and at least 90% of the grid cells (292 of 324) contained estimates of species occupancy (Supplemental material Appendix A3.2). Species occupancy of a given cell was estimated as the ratio of the number of observations of that species by the total number of surveys that were conducted in that cell over the five year period. For quality control purposes we did not estimate occupancy if there were fewer than two survey routes in a cell or if more than 80% of the area of the cell (800 km^2) was covered by water. The SA null model was implemented such that pixels failing to meet these criteria were not considered. In the western grid a total of 309 species were observed over 3594 samples, and in the eastern grid 199 species were observed over 5485 samples.

Simulated communities

To generate simulated spatial patterns under different levels of environmental filtering and dispersal limitation, we used a modified version of the spatially explicit, individual-based simulation model conceived and initially explored by Bell (2000, 2003, 2005) and more recently operationalized by Smith and Lundholm (2010). Through time individuals are born, disperse, and die on a square n x n lattice in which grid cells vary with respect to some environmental variable. Because one obvious hypothesis is that spatial patterns are generated by patterns in the

14

underlying environmental template, we utilized the observed spatial pattern of one environmental variable, the normalized difference vegetation index (NDVI, Clark Labs 2001), which is a remotely sensed proxy for productivity (Chong et al. 1993, Paruelo et al. 1997) and which is thought to drive patterns of avian species composition (Stegen et al. in press, Hurlbert and Haskell 2003, Bonn et al. 2004). For each 100 km grid cell, we calculated the mean value of June NDVI based on the long-term average (1982-2000, excluding 1994) using the original 0.1° resolution data (Clark Labs 2001), and this was used as the environmental template for all simulations. If 80% or more of the grid cell was covered by water we did not estimate NDVI for that cell. Demographic rates in a grid cell were modeled as a function of the difference between the environment at that location [$g(x,y)$] and the species' environmental optimum (μ_i) using two symmetrical Gaussian functions:

$$B_i(x, y) = b\exp\{-1/2[g(x, y) - \mu_i]^2/L\sigma^2_{\text{Envi}}\}$$
$$D_i(x, y) = 1 - (1 - d)\exp\{-1/2[g(x, y) - \mu_i]^2/L\sigma^2_{\text{Envi}}\}$$

where B_i and D_i are the probabilities of birth and death respectively for species i at the spatial coordinates x and y, b and d are the probabilities of birth and death under optimal conditions, σ_{Envi} reflects a species' habitat breadth, and L is the gradient length over the entire lattice. Species were assigned environmental optima (μ_i) along the NDVI gradient from a random uniform distribution that spanned the internal 80% of the observed NDVI range for a given region (i.e., no species were assigned an optimum in either extreme tail of the NDVI distribution). This was done to ensure that each species had a similar effective habitat breadth.

   In addition to including the environmental influence on demographic rates, the simulation model also includes a role for dispersal limitation. In each time step an individual can produce at most a single offspring (i.e., fecundity was equal to 1), and that offspring disperses a random

15

distance and direction from the parent. The dispersal distance function was defined as a random half-Normal distribution [i.e., |Norm(0, $n\sigma_{\text{Disp}}$)|] where n is the linear dimension of the square lattice and σ_{Disp} is the dispersal breadth. According to our specification of the half-Normal distribution, the mean and standard deviation of the distance dispersed should be $n\sigma_{\text{Disp}}\sqrt{2/\pi}$ and $n\sigma_{\text{Disp}}\sqrt{1-2/\pi}$, respectively. If a dispersal distance arose that placed the offspring off the grid, then it was discarded and a new dispersal distance was resampled. The number of individuals within each cell was constrained to be less than a fixed carrying capacity ($K = 500$), which was the same for all cells of sufficient land area and with $K = 0$ elsewhere. The carrying capacity was chosen to be large enough to support coexisting species but small enough to be computationally efficient. At time $t = 0$, the individuals of every species were evenly divided amongst the cells. However, through time when the total number of individuals in a cell exceeded the carrying capacity, individuals in that cell were randomly removed (as in Smith and Lundholm 2010). In order to avoid ecological drift to a monoculture, individuals of each species were allowed to immigrate into the perimeter cells of the lattice. The probability of immigration was a small fraction of σ_{Disp} because large immigration rates will swamp out internal community dynamics (Chisholm and Lichstein 2009). The strength of environmental filtering and dispersal limitation in our simulation are inversely related to σ_{Envi} and σ_{Disp}, respectively. For example, when σ_{Envi} is small, species have low probabilities of birth and high probabilities of death in environments that differ from their optima, resulting in strong environmental filtering. When σ_{Disp} is large, then species are easily able to disperse across the lattice in a single time step, and therefore dispersal limitation is weak. In turn it is important to recognize that the influence of both the environment and dispersal will depend in part on the spatial configuration of the

environment except in purely neutral simulations (i.e., when σ_{Envi} is large). We intentionally excluded grid cells from our simulated landscape to mirror our lack of empirical data for certain regions.

We carried out our simulations in the R statistical environment using the neutral.vp package (Smith and Lundholm 2010). We modified the functions of neutral.vp to take into consideration the following differences between our simulation and that of Smith and Lundholm (2010): 1) we assigned species optima using a random uniform distribution rather than a fixed increment along the gradient, 2) we modeled species responses to the environment as a Gaussian function rather than a piece-wise linear function, 3) we modeled dispersal distance as a draw from a half-Normal distribution rather than a random walk, and 4) dispersal was repeated until the new propagule landed on the grid.

Our simulation model was parameterized using the values in Table 3, which were based in part on the values used by Smith and Lundholm (2010). Each simulation was conducted using a particular level of environmental filtering and of dispersal limitation. We conducted simulations across a parameter space that included all possible combinations of 10 levels each of σ_{Envi} and σ_{Disp}, ranging from 0.01 to 1. The computational cost of this simulation precluded an exhaustive search of all possible parameter space. Because our empirical analysis was based upon estimates of species occupancy, we converted the simulated community matrix to a binary presence/absence matrix prior to analysis, although future insights may be gained through the inclusion of abundance data. For each set of parameter values we examined the average patterns (cf. Figure 2) that emerged across 15 replicate simulations that were each run for 5000 time steps. We carried out 499 permutations of the IA and SA null models on each simulated community.

Results

East and west bird communities displayed similar patterns of intra-specific aggregation but differed strongly in their patterns of inter-specific association (Fig. 3). The IA null model suggested that intra-specific spatial patterns in both regions displayed strong patterns of aggregation until spatial distances of approximately 900-1000 km at which point the patterns appeared more overdispersed, or segregated, than random. In both regions, the within-species pattern of aggregation was approximately fractal (i.e., within-species variance increased linearly on a log-log variogram, Palmer 1988 see Supplemental material Appendix 4, Fig. A4.1), and the rate of change in variance (i.e., rate of species accumulation) was only slightly higher in the east.

The western bird community displayed significant positive total inter-specific covariance (Fig. 3a). The comparisons of the positive and negative fractions of covariance relative to the SA null model indicated that this was primarily due to the occurrence of many more aggregation units in the community rather than fewer segregation units. However, at the largest spatial lags (1000-1200 km) both components of covariance indicated significant positive associations were present in the west; that is, species exhibited more aggregation units and fewer segregation units than expected. In the east, although the non-spatial analysis of total covariance was positive (Fig. 3b middle panel), the spatial analysis indicated species were overall significantly segregated. At intermediate to large spatial scales there was a slight pattern of more aggregation units than expected (a signal of positive association); however, this was not as strong as the over-representation of segregation units in the community. The non-spatial analyses generally agreed with the results of the spatial analysis with the exception of total community covariance in the eastern region.

The combinations of dispersal breadth (σ_{Disp}) and habitat breadth (σ_{Envi}) that most closely reproduced empirical patterns of variance and covariance were similar for both eastern and western assemblages (Fig. 4). Goodness of fit was assessed by calculating the average difference between empirical and simulated values across all three SES statistics (i.e. the average residual) for each combination of σ_{Envi} and σ_{Disp}. The non-spatial residuals (Fig. 4a) indicated that there was a wider range of σ_{Envi} than σ_{Disp} that adequately fit the data. The spatial residuals (Fig. 4b) generally agreed with the non-spatial analyses but identified a narrower region in the process space as optimal and thus were more precise. According to the spatial analysis the best fitting combination of parameters were $\sigma_{Disp} = 0.215$ and $\sigma_{Envi} = 0.129$ in the west and $\sigma_{Disp} = 0.215$ and $\sigma_{Envi} = 0.028$ in the east. However, neither of these combinations reflected well-defined local minima (Fig. 4b). In both regions, the average residuals displayed a region of good fit (i.e., low residual values) which ran from a combination in which environmental filtering was very strong and dispersal limitation was weak to a combination in which environmental filtering was very weak and dispersal limitation was moderate.

The extent to which empirically observed variance components differed from the null expectations varied with spatial lag. The spatial pattern of the observed and simulated SES statistics for the best fitting parameter combinations are displayed in Fig. 5. This comparison indicated that the simulation was able to accurately recreate the observed pattern of intra-specific spatial aggregation, but was less successful at capturing strong, non-random patterns of covariance. Thus, in the west there was a poor fit for z_{Cov+}, and in the east there was a poor fit for z_{Cov-}. Although the magnitudes of z_{Cov+} (west) and z_{Cov-} (east) were not accurately fit, the observed pattern of scaling (i.e., the slope of the line) for these fractions was well approximated by the simulation. In contrast, simulated patterns for z_{Cov+} (east) and z_{Cov-} (west) were of similar

magnitude to the empirical patterns, but the pattern of spatial scaling was quite different, especially at longer lags.

The non-spatial residuals for each of the distinct variance components indicated that for both geographic regions the observed pattern of intra-specific variance was well characterized by a wide range of parameter combinations, but that the region of adequate fit for inter-specific variance was more narrowly defined (Supplemental material Appendix 4, Fig. A4.2). The spatial, intra-specific residual surfaces displayed narrower regions of adequate fit in comparison to the non-spatial analysis (Supplemental material Appendix 4, Fig. A4.3). Additionally, the spatial, intra-specific residual surfaces were either accurately fit by a very wide range of parameter combinations (when the observed pattern did not diverge strongly from species independence) or were not adequately fit by any parameter combinations (when the observed pattern displayed strong species associations). Therefore, the general overall pattern of fit observed in Fig. 5 for the best fitting parameter combination, applies to other parameter combinations as well.

Discussion

The spatial decomposition of a community's intra- and inter-specific components of variance has the potential to implicate the processes driving community assembly. However, ecological inference using these statistical patterns will ultimately depend upon a combination of clear a priori hypotheses and manipulative experiments or simulation modeling which are tailored to the system of interest. The empirical examples we examined using the BBS data illustrate several important aspects of community variance decomposition. Conceptually, the empirical examples emphasize that a joint spatial consideration of the degree of clumping and association in species occurrences is more powerful at elucidating the processes driving community structure than an individual, non-spatial examination of these community characteristics. From a biological

20

perspective our analysis suggests that bird species from both the western and eastern US display spatial dependence not only in their pattern of occurrences with conspecifics but also in their associations with heterospecifics. Furthermore although the patterns of inter-specific covariance were very different between eastern and western bird communities, our analyses, which were based upon regional spatial patterns of NDVI, suggested that these communities were potentially driven by similar magnitudes of dispersal limitation and environmental filtering.

Why is community variance decomposition relevant?

Throughout this manuscript we have emphasized that total community variance, which we defined as variance in species richness (when presence/absence data is considered) or total site abundance (when abundance data is considered), may be decomposed into different fractions which provide useful descriptions of community structure. Ultimately the goal is to provide a means of linking observed patterns of community structure with ecological and evolutionary processes. As such, our interest in spatially decomposing variance into within- and between-species components is primarily motivated by a need for sensitive and interpretable statistics that summarize the complexity of a community's spatial structure and facilitate the linking of pattern with process.

Our study builds upon the theoretical ground work that was developed for community-scale examinations of species associations (Pielou 1972, Schluter 1984). These authors proposed using within-species variance as a null expectation for total community variance to provide a measure of the strength of inter-specific associations (i.e., how different is total variance relative to that expected when between-species covariance is assumed to be zero). Although Palmer and van der Maarel (1995) demonstrated that within-species spatial dependence biased the inference of this kind of analysis, it was not until Wagner (2003) developed the multivariate community

variogram that spatial patterns in within- and between-species variance could be truly appreciated in a unified manner. Our study builds upon Wagner (2003) in three important ways: 1) we demonstrated the utility of decomposing total between-species variance further into its positive and negative components, 2) we linked the mathematical machinery she developed with a spatially explicit null model (the SA null model) when interpreting between-species variance, and 3) we developed a framework for linking the empirical patterns with process via a stochastic simulation model.

On a conceptual level, the different variance components (within- or between-species) are more linked to some processes than others. For example, neutral processes like dispersal limitation should be more closely tied to within-species patterns, and non-neutral processes, where species differ in their environmental optima, are more likely to influence between-species patterns of variance. Nevertheless, the influence of particular ecological processes is not necessarily restricted to a single variance component. For example, although dispersal limitation is typically considered a determinant of within-species spatial aggregation, it can also produce non-random patterns of species association (Ulrich 2004, Bell 2005). We confirmed this result with our own simulation models, observing how dispersal limitation on its own (i.e., when environmental filtering was very weak) can result in strong patterns of positive and negative species associations. Thus when McGill (2010) points out that most models of biodiversity start from an assumption of independent species, this does not necessarily indicate that the models predict independent species associations. The primary reason then for examining patterns of between-species covariance is not to examine the assumptions of the models discussed by McGill (2010), but rather to provide a more detailed examination of model predictions and fit to empirical data. Thus far most tests of unified theories of biodiversity have focused on examining

patterns that reflect within-species variance such as the abundance distribution, the species-area relationship, and the distance-decay relationship. By examining both within- and between-species variance we can gain greater insight into how a community is structured than could be achieved by simply examining variance in richness or total site abundance.

Spatial versus non-spatial perspectives of variance decomposition

Our study is the first to link the spatial decomposition of community variance with spatially explicit null models. This is necessary if patterns of community structure change as a function of spatial scale. Since this is generally the case, considering the spatial decomposition of variance and spatially defined null models will be relevant for the majority of empirical datasets.

Throughout this study we have provided the results for each analysis from both spatial and non-spatial perspectives. In the avian dataset, this side-by-side comparison revealed that 1) results from spatial and non-spatial analyses do not always indicate the same direction and magnitude of effect, and 2) spatial variance decomposition provides a more finely resolved impression of which ecological processes are most likely responsible for observed patterns. In the eastern bird community, total between-species covariance differed fundamentally between the spatial and non-spatial analysis. The non-spatial analysis indicated that species were positively associated, but the spatial analysis indicated that they were negatively associated over all the spatial scales we examined. Therefore, the strong positive association detected in the non-spatial analysis was an artifact due to the presence of within-species spatial aggregation, which the SA null model controls for, and not a signal of strong between-species associations (Palmer and van der Maarel 1995, Roxburgh and Chesson 1998). Additionally, the spatial and non-spatial analyses identified similar regions of process space as providing the best fits to the empirical data, but the spatial analyses were more precise in identifying a narrower region as optimal (Fig.

23

4). This result makes intuitive sense given that the spatial decomposition is more information rich than the global non-spatial statistics and that we are attempting to infer the importance of inherently spatial ecological processes.

Ecological insight into avian communities

The western and eastern bird communities displayed similar patterns of fractal-like intra-specific aggregation but differed strongly in their spatial patterns of between-species covariance. Although our finding of a fractal-like community scaling pattern in which species are spatially aggregated with conspecifics is not novel (Bahn and McGill 2007, van Teeffelen and Ovaskainen 2007, Storch et al. 2008), few other studies have examined the spatial structure of community covariance (McGill 2010). Across all the spatial scales we examined, there was strong inter-specific segregation in the eastern U.S., indicating a greater than expected number of checkerboards units. In contrast, species distributions in the west exhibited strong inter-specific aggregation indicating a greater than expected number of joint presences and joint absences. The magnitude of these effects increased in strength as a function of spatial lag (see the pattern of observed z-scores in Fig. 5). Our simulation models suggested that the observed community variation in both regions reflected the joint influence of intermediate-to-strong environmental filtering and intermediate dispersal limitation.

Although there are discrepancies between exactly which combination of processes best fit the eastern and western empirical data, the general region of best fitting parameters was similar in the two communities (blue colored regions in Fig. 4b). This result is surprising because the empirical patterns of between-species variance differed so strongly between the two regions (Fig. 3), highlighting the importance of the environmental template in interpreting spatial pattern and process. That template could hardly differ more between the arid and topographically

24

heterogeneous west and the eastern landscape of deciduous forest, agriculture, and suburbia. We

focused on the vegetation index NDVI to reflect these differences, and indeed NDVI in the east

had a higher mean and lower variance than in the west ($NDVI_{east} = 0.64 \pm 0.07$, $NDVI_{west} = 0.41$

± 0.13). These differences in NDVI were mirrored in the patterns of average grid cell richness

between the two regions as well ($\hat{S}_{east} = 53.2 \pm 6.8$, $\hat{S}_{west} = 34.8 \pm 10.8$). It seems likely that in

the west, birds appear positively associated because many different bird species occur together in

patches with high NDVI (e.g., riparian areas) and are jointly absent from the harshest desert

regions. In contrast in the east, the gradient in species richness is much weaker implying that

turnover in species composition is not the result of a nested assemblage, but instead potentially

segregates out between the different suites of species typically found in open versus wooded

habitats. Despite these differences in the environmental template, the relative importance of

environmental filtering and dispersal limitation to avian community assembly may be fairly

consistent across North America.

Strengths and weaknesses of the simulation approach

   Our ecological interpretations depend in large part on the simulation results, and thus it is

important to examine the strengths and weaknesses of this approach. The major advantage of a

stochastic simulation model over statistical modeling approaches (*sensu* Hartig et al. 2011) is

that the simulation model is able to mechanistically model the process of environmental filtering

and dispersal limitation in a spatially explicit way that would otherwise not be mathematically

tractable. In our model, community structure emerges directly from the processes we are

interested in, and we may examine how aspects of this structure vary continuously across a range

of process strengths and spatial scales in a manner that would be unfeasible to consider in an

experimental setting.

Although our simulation approach has key advantages, it is important to recognize the weakness of this approach as well. The most important and obvious weakness of the simulation approach is that it is based upon an assumption that similar patterns are produced by similar processes. This is a problem inherent to any study that attempts to make inference based upon the fit of a model to data and for which there is no simple solution. Another general criticism of many simulation studies in ecology is the large number of arbitrary parameters. We intentionally chose parameter values based upon precedents set in the literature (Bell 2005, Smith and Lundholm 2010) or that were empirically informed when appropriate (e.g., environmental gradient, size of species pool). Future applications of our framework should attempt to empirically constrain additional parameters such as rates of birth, death, and migration. Additionally it would be interesting to use empirically estimated environmental optima for each species rather than simply randomly distributing these along the gradient. In fact, non-random distributions of environmental optima may be necessary for recreating observed species richness gradients (e.g., Bonn et al. 2004). Future examinations could also examine the influence of allowing habitat specificity and dispersal ability to vary among species. Lastly, although our model mechanistically models the process of random natal dispersal we cannot rule out the possibility that variation attributed to dispersal-width (σ_{Disp}) is due to spatially structured environmental variation uncorrelated with NDVI. However, this is a problem that is common to both statistical and simulation modeling frameworks that attempt to infer the strength of dispersal limitation (Gilbert and Lechowicz 2004, Svenning et al. 2006), and is therefore not a shortcoming of the simulation framework per se.

Conclusions

Ecological communities are non-randomly spatially structured. We have developed a modeling framework in which the components of this spatial variation may be examined and attributed to ecological processes. The avian community we examined demonstrated that species have clumped, covarying distributions with one another, and that a joint consideration of both within and between-species fractions of community variance relative to a stochastic simulation model is a powerful method of uncovering what processes drive community assembly. Despite the limitations of a simulation approach, it will continue to act as a potentially powerful tool for exploring pattern and process at spatial scales over which it would be impossible to conduct an experimental study.

Acknowledgements

Literature Cited

Bacaro, G. and Ricotta, C. 2007. A spatially explicit measure of beta diversity. - Community Ecol. 8: 41-46.

Bahn, V. and McGill, B. J. 2007. Can niche-based distribution models outperform spatial interpolation? - Global Ecol. Biogeogr. 16: 733-742.

Bell, G. 2000. The distribution of abundance in neutral communities. - Am. Nat. 155: 606-617.

Bell, G. 2003. The interpretation of biological surveys. - Proceedings of the Royal Society of London Series B-Biological Sciences 270: 2531-2542.

Bell, G. 2005. The co-distribution of species in relation to the neutral theory of community ecology. - Ecology 86: 1757-1770.

Bell, G. et al. 2006. The comparative evidence relating to functional and neutral interpretations of biological communities. - Ecology 87: 1378-1386.

Bonn, A. et al. 2004. Structure of the species–energy relationship. - Proceedings of the Royal Society of London. Series B: Biological Sciences 271: 1685-1691.

Chisholm, R. A. and Lichstein, J. W. 2009. Linking dispersal, immigration and scale in the neutral theory of biodiversity. - Ecol. Lett. 12: 1385-1393.

Chong, D. L. S. et al. 1993. Relating the global vegetation index to net primary productivity and actual evapotranspiration over Africa. - International Journal of Remote Sensing 14: 1517-1546.

Clark Labs 2001. 0.1 Degree Global Monthly Vegetation Index (NDVI) 1981–2000. - Clark Labs, MA.

Colwell, R. K. and Rangel, T. F. 2010. A stochastic, evolutionary model for range shifts and richness on tropical elevational gradients under Quaternary glacial cycles. - Philosophical Transactions of the Royal Society B: Biological Sciences 365: 3695-3707.

Connor, E. F. and Simberloff, D. 1979. The Assembly of Species Communities: Chance or Competition? - Ecology 60: 1132-1140.

Diamond, J. M. 1975. Assembly of species communities. - In: Cody, M. L. and Diamond, J. M. (eds), Ecology and evolution of communities. Belknap Press of Harvard University, pp. 342-444.

Dormann, C. F. et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. - Ecography 30: 609-628.

Drakare, S. et al. 2006. The imprint of the geographical, evolutionary and ecological context on species-area relationships. - Ecol. Lett. 9: 215-227.

Fortin, M. J. and Dale, M. 2005. Spatial Analysis: A Guide for Ecologists. - Cambridge University Press.

Gilbert, B. and Lechowicz, M. J. 2004. Neutrality, niches, and dispersal in a temperate forest understory. - Proceedings of the National Academy of Sciences of the United States of America 101: 7651-7656.

Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. - Ecology 81: 2606-2621.

Gotelli, N. J. and Graves, G. R. 1996. Null Models in Ecology. - Smithsonian Institution.

Gotelli, N. J. et al. 2009. Patterns and causes of species richness: a general simulation model for macroecology. - Ecol. Lett. 12: 873-886.

Gotelli, N. J. et al. 2010. Macroecological signals of species interactions in the Danish avifauna. - Proceedings of the National Academy of Sciences 107: 5030-5035.

Grimm, V. et al. 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. - Science 310: 987-991.

Harte, J. and Kinzig, A. P. 1997. On the implications of species-area relationships for endemism, spatial turnover, and food web patterns. - Oikos 80: 417-427.

Hartig, F. et al. 2011. Statistical inference for stochastic simulation models - theory and application. - Ecol. Lett. 14: 816-827.

Hurlbert, A. H. and Haskell, J. P. 2003. The effect of energy and seasonality on avian species richness and community composition. - Am. Nat. 161: 83-97.

Hurlbert, A. H. and White, E. P. 2005. Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. - Ecol. Lett. 8: 319-327.

Koleff, P. et al. 2003. Measuring beta diversity for presence-absence data. - Journal of Animal Ecology 72: 367-382.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? - Ecology 74: 1659-1673.

Lennon, J. J. 2000. Red-shifts and red herrings in geographical ecology. - Ecography 23: 101-113.

Martín, H. G. and Goldenfeld, N. 2006. On the origin and robustness of power-law species-area relationships in ecology. - Proceedings of the National Academy of Sciences of the United States of America 103: 10310-10315.

McCulloch, C. E. 1985. Variance tests for species association. - Ecology 66: 1676-1681.

McGill, B. J. 2003. A test of the unified neutral theory of biodiversity. - Nature 422: 881-885.

McGill, B. J. 2010. Towards a unification of unified theories of biodiversity. - Ecol. Lett. 13: 627-642.

McGlinn, D. J. and Palmer, M. W. 2011. Quantifying the influence of environmental texture on the rate of species turnover: evidence from two habitats. - Plant Ecology 212: 495-506.

Morlon, H. et al. 2008. A general framework for the distance-decay of similarity in ecological communities. - Ecol. Lett. 11: 904-917.

Nekola, J. C. and White, P. S. 1999. The distance decay of similarity in biogeography and ecology. - Journal Of Biogeography 26: 867-878.

Palmer, M. W. 1987. Variability in species richness within Minnesota oldfields - a use of the variance test. - Vegetatio 70: 61-64.

Palmer, M. W. 1988. Fractal geometry: a tool for describing spatial patterns of plant communities. - Vegetatio 74: 91-102.

Palmer, M. W. and van der Maarel, E. 1995. Variance in species richness, species association, and niche limitation. - Oikos 73: 203-213.

Paradis, E. et al. 1998. Patterns of Natal and Breeding Dispersal in Birds. - Journal of Animal Ecology 67: 518-536.

Paruelo, J. M. et al. 1997. ANPP Estimates from NDVI for the Central Grassland Region of the United States. - Ecology 78: 953-958.

Pielou, E. C. 1972. 2k Contingency tables in ecology. - Journal of Theoretical Biology 34: 337-352.

Plotkin, J. B. et al. 2000. Species-area curves, spatial aggregation, and habitat specialization in tropical forests. - Journal of Theoretical Biology 207: 81-99.

Rangel, T. F. et al. 2007. Species richness and evolutionary niche dynamics: A spatial pattern-oriented simulation experiment. - Am. Nat. 170: 602-616.

Roxburgh, S. H. and Chesson, P. 1998. A new method for detecting species associations with spatially autocorrelated data. - Ecology 79: 2180-2192.

Roxburgh, S. H. and Matsuki, M. 1999. The statistical validation of null models used in spatial association analyses. - Oikos 85: 68-78.

Sauer, J. R. et al. 2008. The North American Breeding Bird Survey, Results and Analysis 1966 - 2007. Version 5.15.2008.

Scheiner, S. M. et al. 2011. The underpinnings of the relationship between space, time, and species richness. - Ecological Monographs 81: 195-213.

Schluter, D. 1984. A variance test for detecting species associations, with some example applications. - Ecology 65: 998-1005.

Shen, G. et al. 2009. Species–area relationships explained by the joint effects of dispersal limitation and habitat heterogeneity. - Ecology 90: 3033-3041.

Smith, T. W. and Lundholm, J. T. 2010. Variation partitioning as a tool to distinguish between niche and neutral processes. - Ecography 33: 648-655.

Stegen, J. C. and Hurlbert, A. H. 2011. Inferring Ecological Processes from Taxonomic, Phylogenetic and Functional Trait β-Diversity. - PLoS ONE 6: e20906.

Stegen, J. C. et al. Stochastic and deterministic drivers of spatial and temporal turnover in breeding bird communities. - Global Ecology and Biogeography in press.

Stone, L. and Roberts, A. 1990. The checkerboard score and species distributions. - Oecologia 85: 74-79.

Storch, D. et al. 2008. The quest for a null model for macroecological patterns: geometry of species distributions at multiple spatial scales. - Ecol. Lett. 11: 771-784.

Svenning, J. C. et al. 2006. The relative roles of environment, history and local dispersal in controlling the distributions of common tree and shrub species in a tropical forest landscape, Panama. - Journal of Tropical Ecology 22: 575-586.

Ulrich, W. 2004. Species co-occurrences and neutral models: reassessing J. M. Diamond's assembly rules. - Oikos 107: 603-609.

van Teeffelen, A. J. A. and Ovaskainen, O. 2007. Can the cause of aggregation be inferred from species distributions? - Oikos 116: 4-16.

Wagner, H. H. 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. - Ecology 84: 1045-1057.

Wagner, H. H. 2004. Direct multi-scale ordination with canonical correspondence analysis. - Ecology 85: 342-351.

Wagner, H. H. and Fortin, M. J. 2005. Spatial analysis of landscapes: concepts and statistics. - Ecology 86: 1975-1987.

Weatherhead, P. J. and Forbes, M. R. L. 1994. Natal philopatry in passerine birds: genetic or ecological influences? - Behavioral Ecology 5: 426-433.

White, E. P. and Gilchrist, M. A. 2007. Effects of population-level aggregation, autocorrelation, and interspecific association on the species-time relationship in two desert communities. - Evolutionary Ecology Research 9: 1329-1347.

White, E. P. and Hurlbert, A. H. 2010. The Combined Influence of the Local Environment and Regional Enrichment on Bird Species Richness. - The American Naturalist 175: E35-E43.

Tables

Table 1. A glossary of key technical terms related to spatial community patterns.

Term	Definition
community matrix ($\mathbf{X}$)	A site-by-species matrix populated by a non-negative measure of species prevalence.
covariance matrix ($\mathbf{C}$)	A symmetrical species-by-species covariance matrix. The i,j position of the matrix reflects the covariance between the ith and jth species.
spatial lag (h)	The spatial distance separating a pairwise comparison.
intra-specific aggregation	The pattern of spatial clustering in occurrences of a single species (Eq. 4).
inter-specific aggregation	The pattern of spatial clustering between two or more species, interspecific covariance in space (Eq. 5-6).
aggregation units	Units of positive covariance or the number of times in a presence-absence matrix that two species co-occur and are co-absent for a given pair of plots.
segregation units	Units of negative covariance or the number of times in a presence-absence matrix that two species are segregated (i.e. the number of checkerboard units).
random assignment (RA)	A null model in which the cells of the community matrix are swapped randomly (Table 2).
independent assignment (IA)	A null model in which the cells of each column (i.e., species) are independently shuffled across the rows (i.e., sites) (Table 2).
spatial assignment (SA)	A null model in which the cells of each column (i.e. species) are randomly swapped until the observed pattern of intraspecific aggregation is approximated but the observed pattern of interspecific aggregation is nullified (Table 2).

Table 2. The components of the community nullified by each of the three null models and the fractions of variance for which they provide an appropriate random expectation. A zero indicates that the null model does not maintain that community component, a one indicates that the null model does maintain that component, and spatial and non-spatial designations indicate whether the fraction considered is spatially defined or not.

Null Model Name	Maintained Community Component			Variance Fraction Tested	
	Species-covariance	Spatial-dependence	Species-occupancy	Intra-specific	Inter-specific
Random Assign.	0	0	0	non-spatial	---
Independent Assign.	0	0	1	spatial	non-spatial
Spatial Assign.	0	1	1	---	spatial

Table 3. The values used to parameterize the community simulation model.

Parameter	Symbol	Values
Landscape simulation:		
Landscape dimensions	n x n	18 x 18
Value of environment at location (x,y)	$g(x,y)$	NDVI(x, y)
Environmental gradient length	L	max(g) – min(g);
		West = 0.66, East = 0.35
Species dynamics:		
Size of species pool	S_p	West = 309, East = 199
Number of generations		5000
Carrying capacity at location (x, y)	$K(x,y)$	$\begin{cases} 500 \text{ if land area } (x, y) \geq 2000\,\text{km}^2 \\ 0 \text{ if land area } (x, y) < 2000\,\text{km}^2 \end{cases}$
Initial cell density of each species		$K(x,y) / S_p$
Environmental optima	μ	~ UNIF(0.1,0.9) $\cdot L$
Habitat breadth	σ_{Envi}	0.010, 0.017, 0.028, 0.046, 0.077, 0.129, 0.215, 0.359, 0.599, 1.000
Dispersal breadth	σ_{Disp}	0.010, 0.017, 0.028, 0.046, 0.077, 0.129, 0.215, 0.359, 0.599, 1.000
Immigration rate		0.005$\cdot\sigma_{\text{Disp}}$
Base birth rate	b	0.8
Base death rate	d	0.2
Fecundity		1.0
Number of replicate simulations		15

Figures
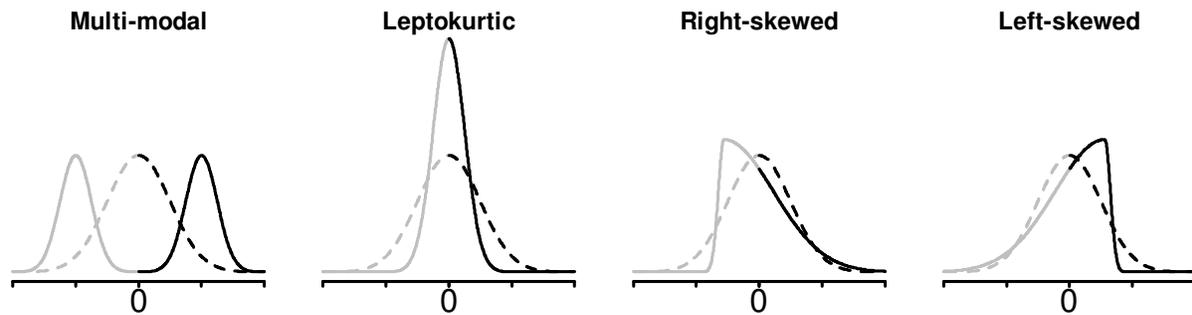
Figure 1. Four potential distributions (solid curves) of community correlation coefficients (r) in relation to the null expectation that coefficients are normally distributed (dashed curves). All of the distributions have a mean of zero (i.e., the inference based upon $\bar{r}$ or the V-ratio would be that species are independent); however, they differ strongly from normality in their patterns of negative and positive associations. The grey lines represent the distribution of negative associations and the black lines represent the distribution of positive associations.
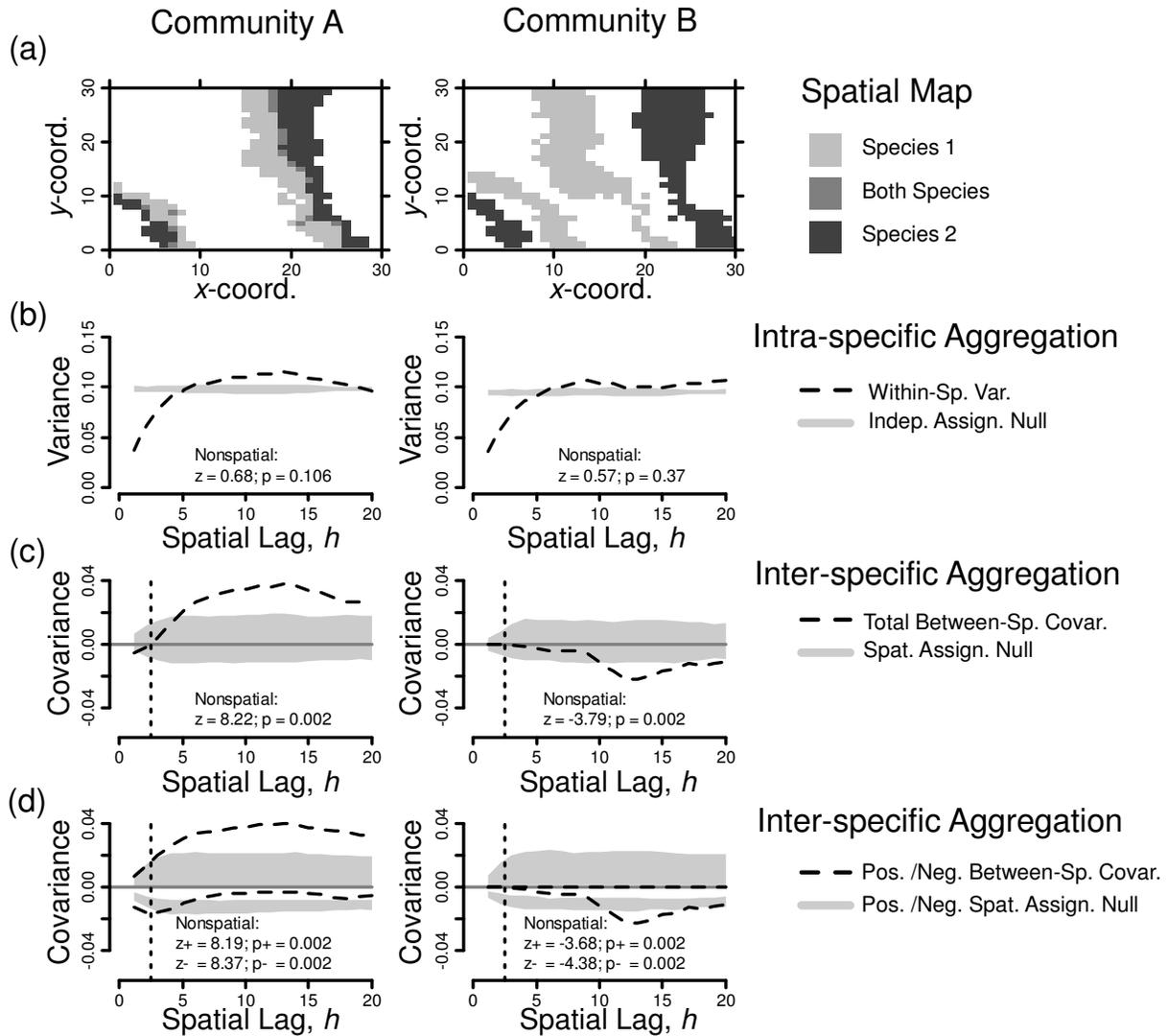
Figure 2. Conceptual figure that uses the spatial distribution of two species (a) to illustrate how

the within (b) and between (c, d) species components of spatial community variation may be

visualized. The two columns represent two communities, A and B respectively, which differ in

their inter-specific but not their intra-specific spatial structure. In panels (b-d), dashed lines

illustrate how a given variance component varies with spatial lag, while the shaded polygon

encompasses the 95% confidence envelopes from the relevant null model. In addition, the

standardized effect statistic (z-score) and associated p-value for non-spatial analyses are

displayed in the lower right corner of each graph. (b) The degree of intra-specific aggregation is

quantified by comparing within-species variance to the Independent Assignment null model of species distributions. (c) A classical metric of inter-specific association is the total community covariance, which is compared to the total covariance expected from the Spatial Assignment null model. However, this metric obscures the signature of simultaneously strong positive and negative patterns of covariance (compare panels (c) and (d) at a spatial lag of 3, the dashed vertical line). At this spatial lag, although total covariance is zero in both communities, panel (c) indicates that the species in community A are actually displaying strong inter-specific aggregation (positive covariance) and segregation (negative covariance) which is an example of a type II error. (d) It is more informative to consider positive and negative components of covariance separated relative to the Spatial Assignment null which maintains observed levels of intra-specific aggregation. Note that in panels (c) and (d) the zero line is marked as a solid grey line simply to aid visual interpretation.
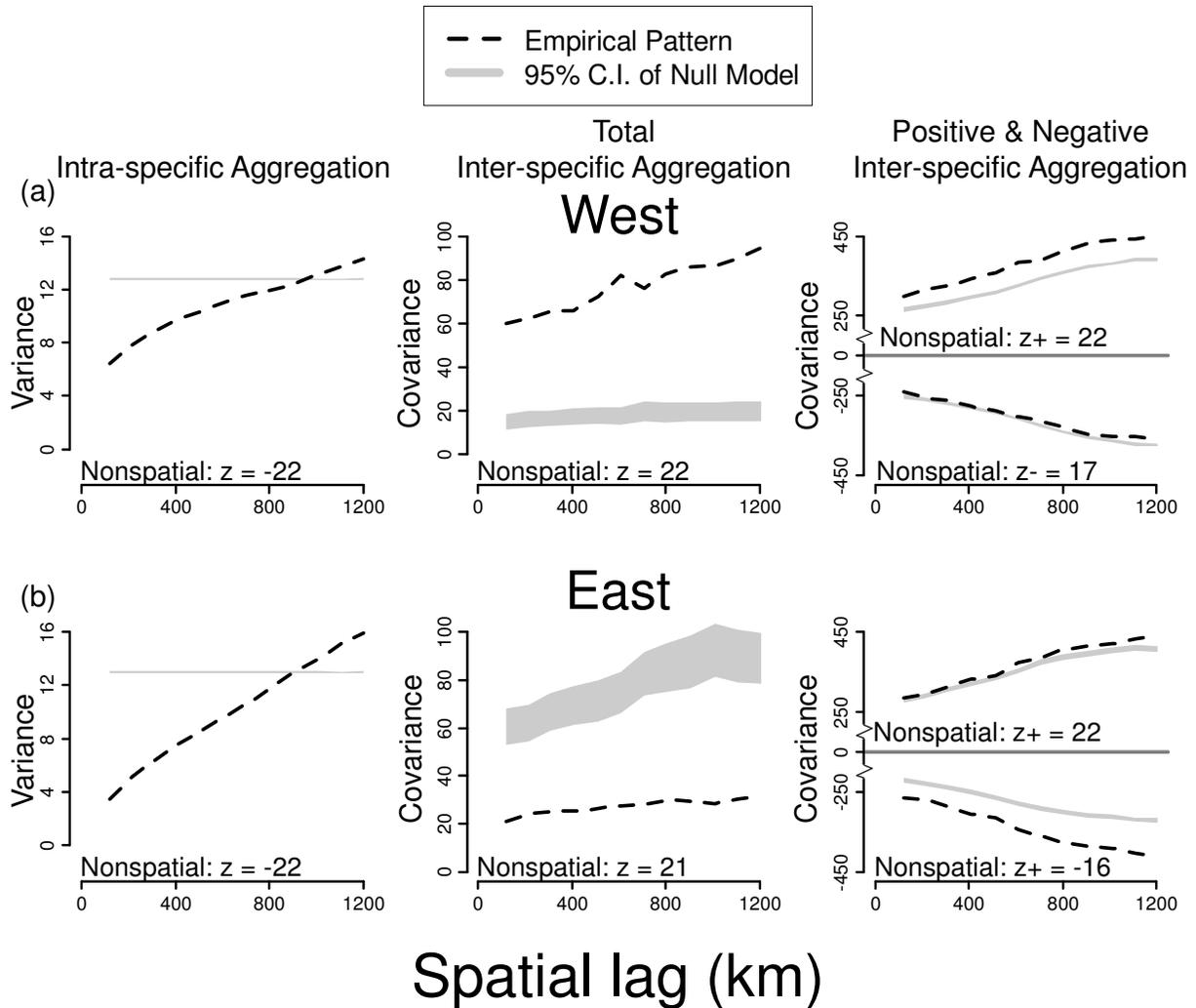
Figure 3. Empirical spatial patterns of intra- and inter-specific variance (dashed lines) and the

95% confidence intervals of the relevant null models (grey polygons which in some cases are

extremely narrow, see Table 1) for birds in the west (a) and in the east (b). The patterns of intra-

specific aggregation for the two regions (left column) are fairly similar in shape and both

transition from aggregated to overdispersed spatial patterns at approximately the same spatial

scale (900-1000 km). However, the patterns of inter-specific association (middle and right

columns) are quite different between the western and eastern regions. In the west, the bird

community shows a strong pattern of positive associations (middle column). This is due

primarily to the occurrence of many more aggregation units rather than a lack of segregation

units in the observed community relative to independent species distributions (right column). In

the east, although total covariance is positive, the null model indicates that species are

significantly negatively associated (middle column). This is due primarily to the occurrence of

many more segregation units rather than fewer aggregation units relative to expectation of

species independence (right column). The z-scores reported in each panel are the non-spatial SES

statistics ($p = 0.002$ for all analyses). Note that total inter-specific association is simply the sum

of the positive and negative fractions of inter-specific association.
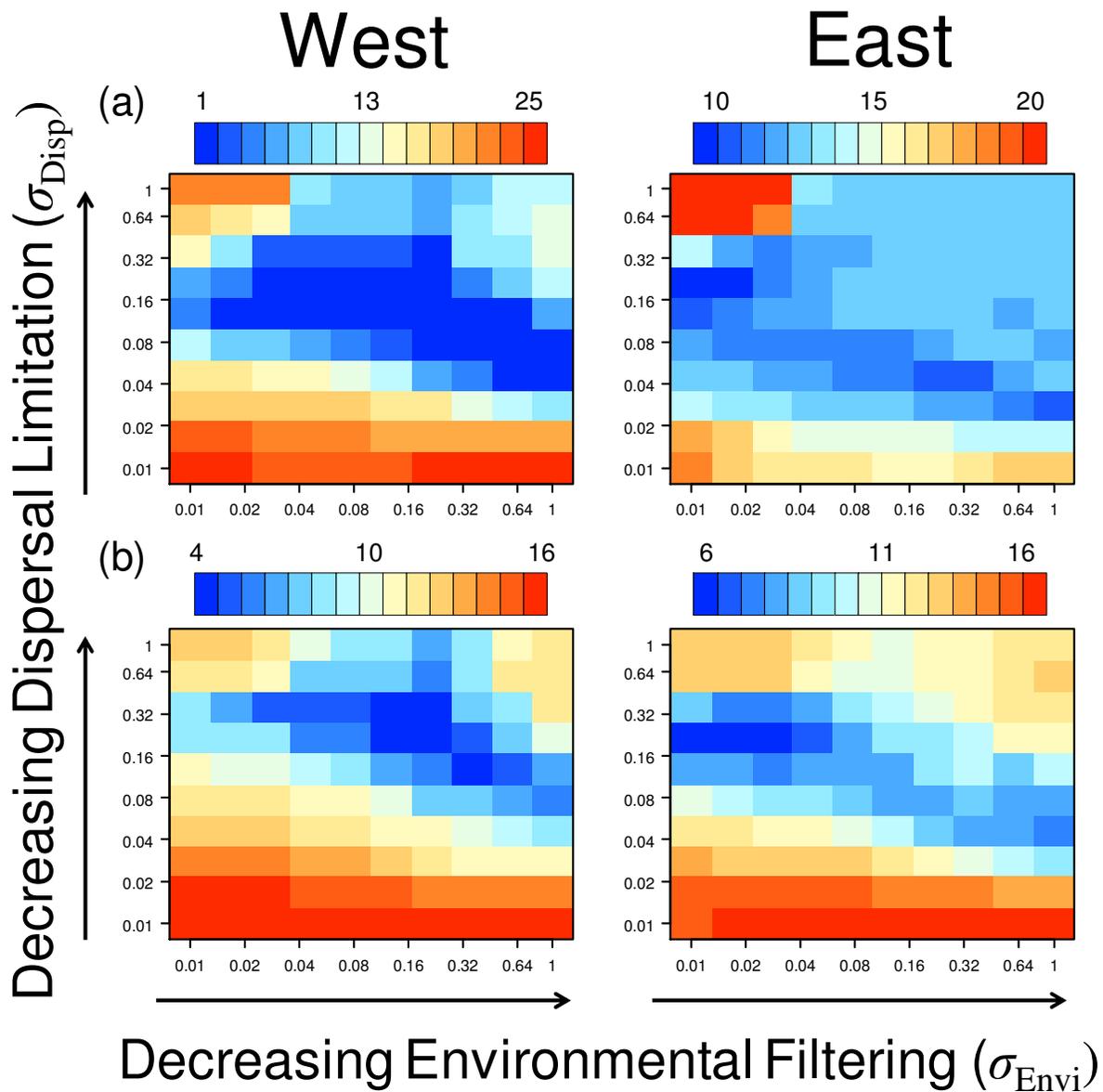
Figure 4. The absolute value of the difference (i.e. residual) between the empirical and simulated z-scores (z_{Var}, z_{Cov+}, z_{Cov-}) for each combination of dispersal breadth (σ_{Disp}) and habitat breadth (σ_{Envi}) evaluated in the simulations. Smaller values (i.e., cooler colors) indicate between fit. The results of the non-spatial decomposition of community variance are displayed in panel (a) and the spatial results are in panel (b) for the west and east regions (left and right columns respectively). For the spatial analysis the residuals represent the average over all spatial lags.

40

Note that the *x*- and *y*-axes are $\log_2$ transformed.
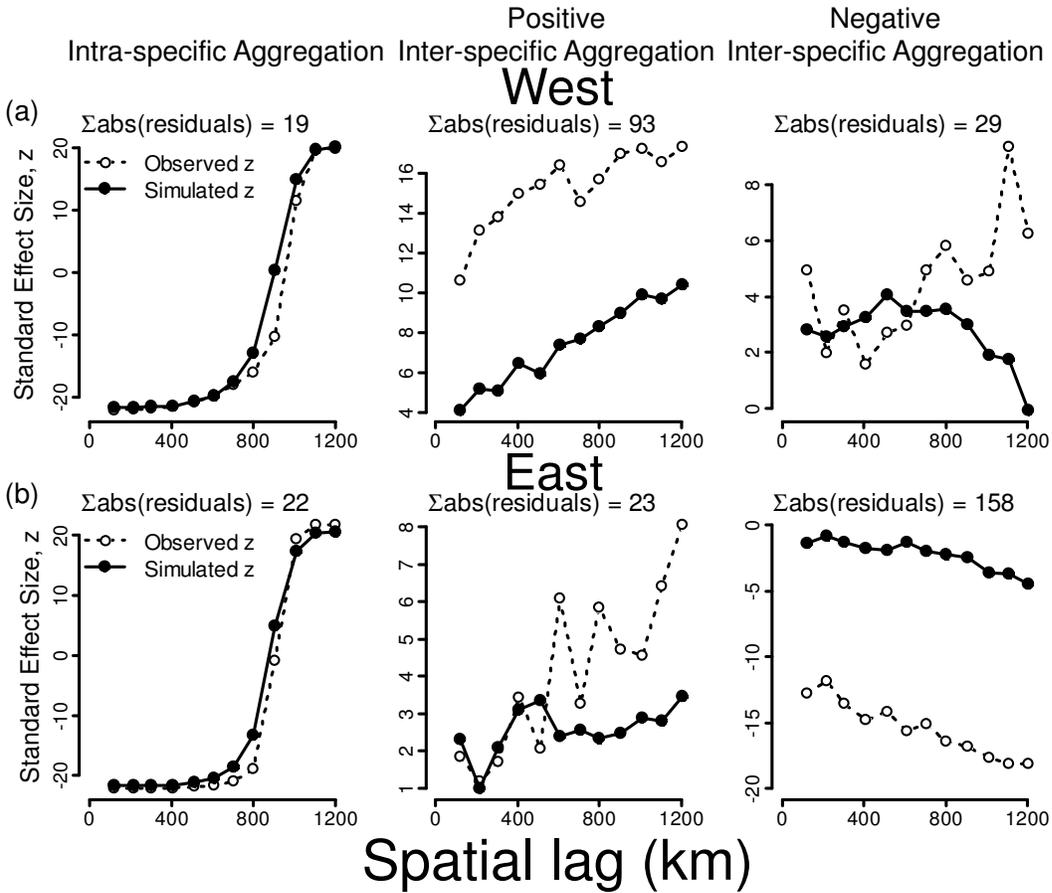
Figure 5. The empirical and simulated z-scores (z_{Var}, z_{Cov+}, z_{Cov-}) as a function of spatial lag for

the best fitting parameter combinations: west $\sigma_{Disp} = 0.215$, $\sigma_{Envi} = 0.129$ and east $\sigma_{Disp} = 0.215$,

$\sigma_{Envi} = 0.028$. The intra-specific patterns of aggregation were well approximated by the

simulation model in both the west (a) and the east (b). Although, the positive fraction of inter-

specific aggregation was not as well fit in the west as it was in the east, the observed pattern of

scaling (i.e., the slope of the line) was well approximated by the simulation in the west. A similar

pattern was observed for the negative fraction of inter-specific aggregation, although in this case

the eastern data was poorly fit in magnitude but was well approximated in slope. The sums of

the residuals reported above each panel are the summed absolute value of the difference between

the observed and simulated z-scores.

Supplementary Material

Appendix 1 - Spatial Statistics and Null Modeling Approaches

Spatial Statistics

Each of the non-spatial components of community variance/covariance given in Eq. 2 is equal to the product of its spatial counterpart in Eqs. 4-6 summed across all possible spatial lags and species combinations and multiplied by n_h. Note that Wagner (2003) referred to Eq. 4 as the 'variogram of complementarity' and when considered over presence/absence data this metric is equal to the number of unique species gained on average when comparing sites separated by a distance *h* and therefore is closely aligned with a myriad of pairwise species turnover metrics (Koleff et al. 2003, Bacaro and Ricotta 2007, McGlinn and Palmer 2011).

Null Models

The first and least constrained null model is the Random Assignment (RA) null model in which the cells of the site-by-species matrix are swapped completely at random between rows and columns. The only constraint on this null model is that it maintains the total sum of the empirical site-by-species matrix, and thus it provides a null community if occurrences are random with respect to species and sites. We used the RA null model to test whether the non-spatial, within-species variance was different than random. Note that in a non-spatial context, that when species occupancy is maintained (as in the subsequent null models, see Table 1) that the null and empirical fractions of within-species variance must be identical.

The second null model is the Independent Assignment (IA) null model in which species' occurrences are independently shuffled across the sites (i.e., rows). The total number of occurrences of each species is fixed at the observed values (i.e., species occupancy or column sums are fixed); however, the pattern of co-occurrence between species and the richness of each

site is randomized. We used the IA null model in combination with eq. 4 to test whether species were spatially aggregated (i.e., displayed nonrandom intra-specific aggregation) because it specifically nullifies spatial dependence within and between species but maintains the frequency of each species. We also used the IA null model in combination with eqs. 5-6, to examine the non-spatial, between-species variance. The IA null is appropriate to test the between-species fraction of variance because, as noted above, in a non-spatial context the null and empirical communities have identical fractions of within-species variance and thus deviations between the IA null and observed between-species variance should only reflect differences due to nonrandom covariance patterns between species.

The third null model is the Spatial Assignment (SA) null model which maintains the observed spatial dependence for each species (i.e., observed intra-specific aggregation) but nullifies between-species spatial dependence (referred to as the Random Patterns null model by Roxburgh and Chesson 1998). As with the IA null, the SA null maintains the total number of occurrences of each species (i.e., species occupancy or column sums are fixed). Roxburgh and Matsuki (1999) demonstrated that the SA null model outperformed other spatial null models at accurately maintaining the empirical pattern of intra-specific aggregation. Additionally, in this study we consider spatial grids of data that are not completely sampled, and the SA null model can be easily adjusted to handle missing data unlike spatial null models that represent independent reflections or rotations of the spatial sampling domain (Palmer and van der Maarel 1995). We used the SA null model to test whether species co-occurrences were non-random after factoring in the spatial dependence each species displayed empirically. Thus the SA null model was appropriate for testing for non-randomness in the spatial, between-species variance fraction. It is critical that the pattern of intra-specific spatial aggregation is maintained when examining

species associations because strong spatial dependence within-species can result in strong negative associations (Palmer and van der Maarel 1995).We provide the source code and specific details on our application of the SA null model in Appendix 2.

Literature Cited

Bacaro, G. and Ricotta, C. 2007. A spatially explicit measure of beta diversity. - Community Ecol. 8: 41-46.

Koleff, P. et al. 2003. Measuring beta diversity for presence-absence data. - Journal of Animal Ecology 72: 367-382.

McGlinn, D. J. and Palmer, M. W. 2011. Quantifying the influence of environmental texture on the rate of species turnover: evidence from two habitats. - Plant Ecology 212: 495-506.

Palmer, M. W. and van der Maarel, E. 1995. Variance in species richness, species association, and niche limitation. - Oikos 73: 203-213.

Roxburgh, S. H. and Chesson, P. 1998. A new method for detecting species associations with spatially autocorrelated data. - Ecology 79: 2180-2192.

Roxburgh, S. H. and Matsuki, M. 1999. The statistical validation of null models used in spatial association analyses. - Oikos 85: 68-78.

Wagner, H. H. 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. - Ecology 84: 1045-1057.

Appendix 2 – Source code

Author: Dan McGlinn
Date: 10/25/10
Purpose: to provide necessary support functions to simulate a community influenced by neutral and niche processes and then examine the spatial pattern of associations in the resulting community

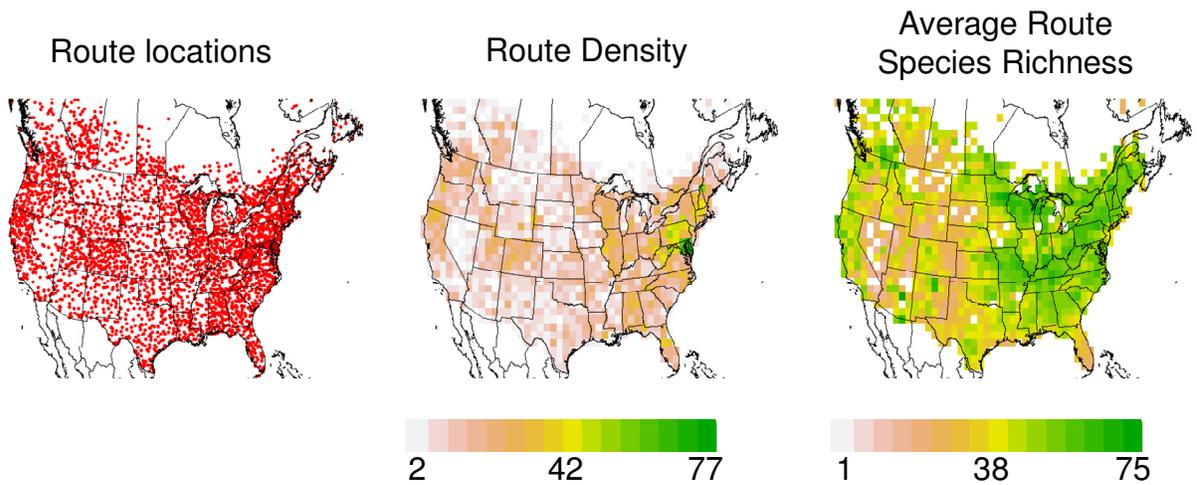*Appendix 3 – BBS sampling regime and NDVI patterns*

Figure A3.1. Geographic distribution of route locations, route density, and average species

richness

Figure A3.2. Map of sampling design with NDVI
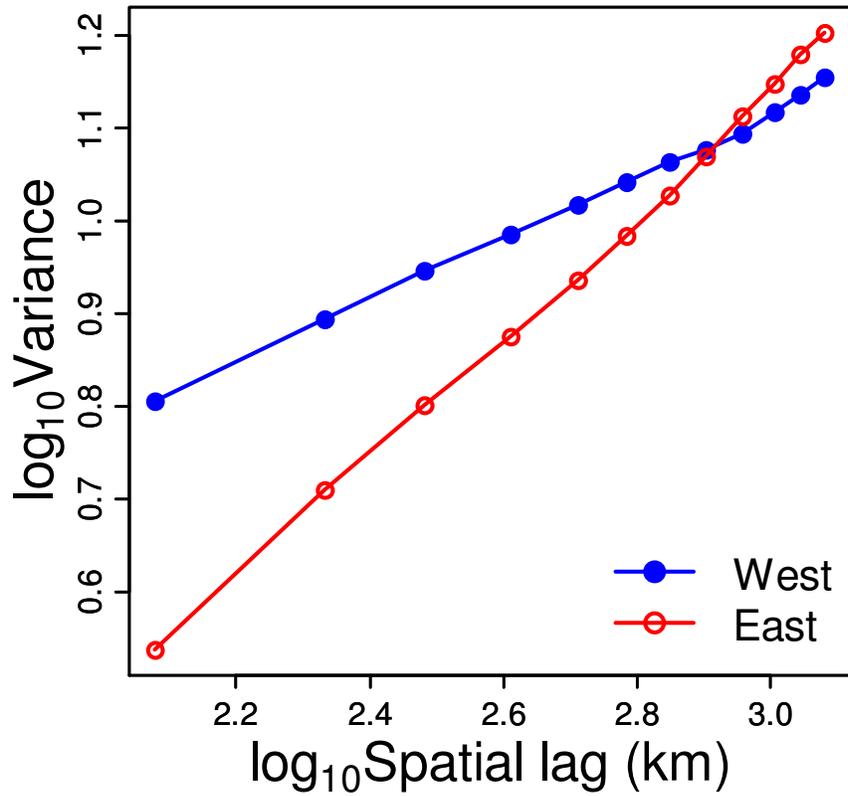
Figure A4.1. The empirical spatial patterns of intra-specific variance on log-log axes for the western and eastern bird communities. The overall linear shape of both trends suggests that clumping pattern of species in both regions was approximately fractal (Palmer 1988).
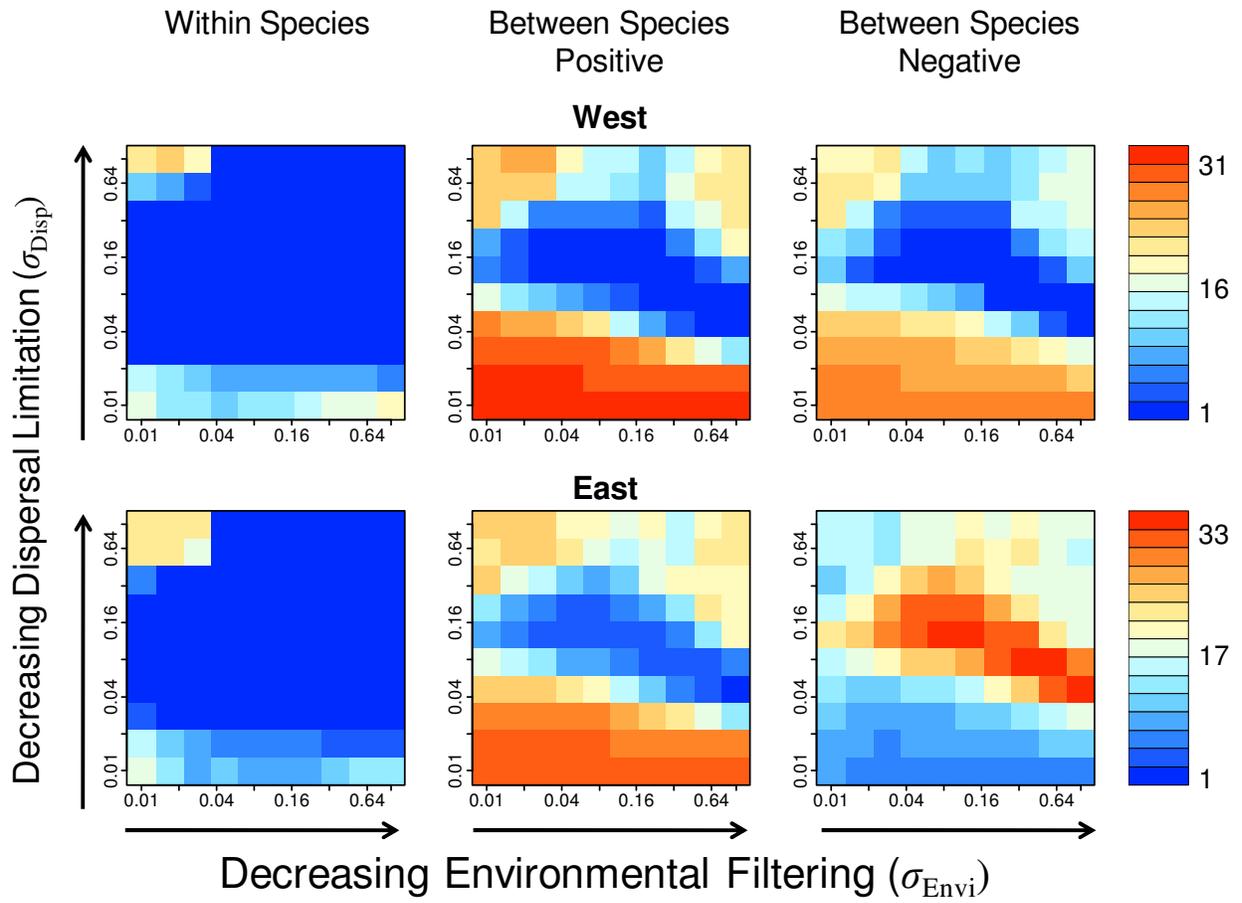
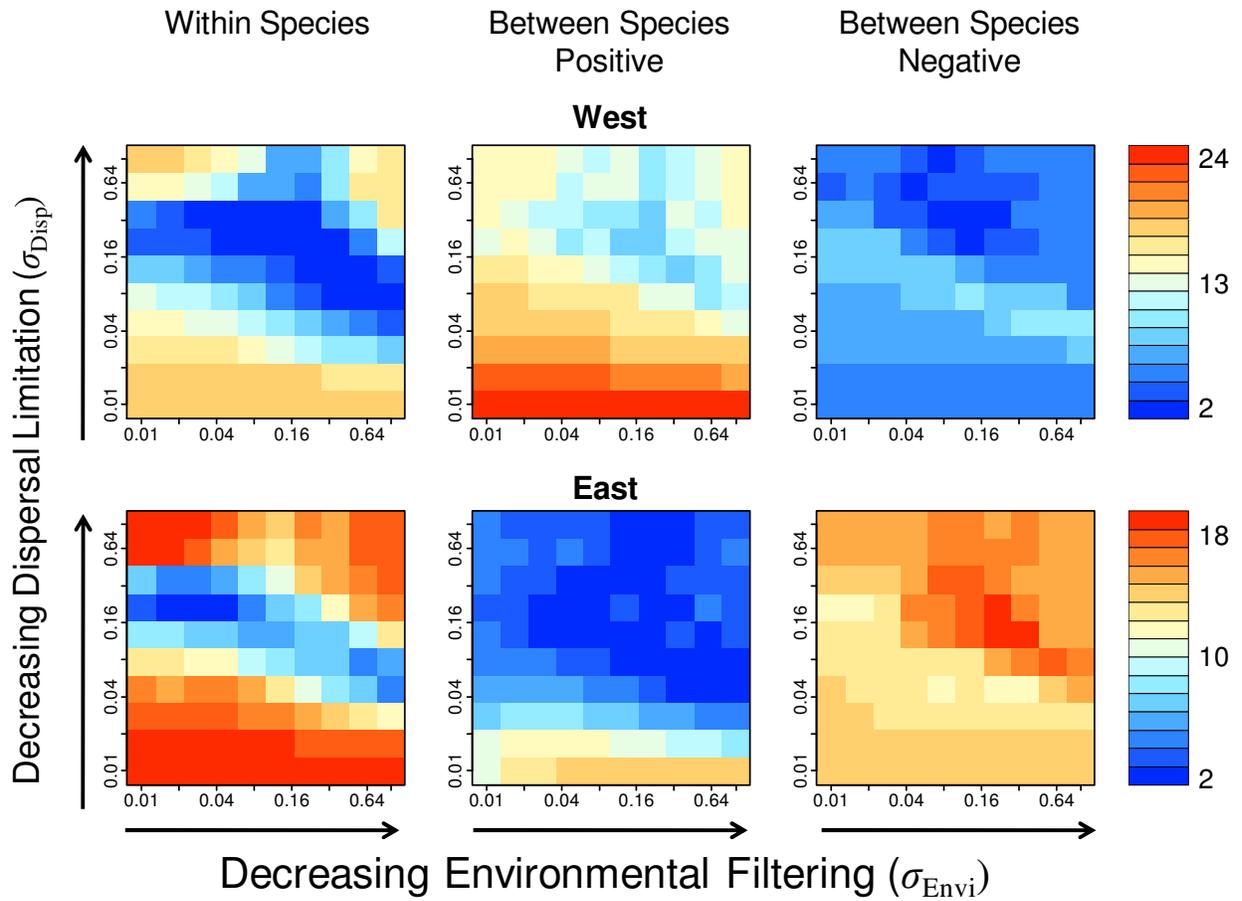Figure A4.2. Non-spatial residual surfaces for each fraction of variance

Figure A4.3. Spatial residual surfaces for each fraction of variance