

SPATIAL MODELING IN ECOLOGY: THE FLEXIBILITY OF EIGENFUNCTION SPATIAL ANALYSES

DANIEL A. GRIFFITH¹ AND PEDRO R. PERES-NETO^{2,3}

¹*School of Social Sciences, University of Texas–Dallas, P.O. Box 830688, Richardson, Texas 75083-0688 USA*

²*Department of Biology, University of Regina, Regina S4S 0A2 Canada*

Abstract. Recently, analytical approaches based on the eigenfunctions of spatial configuration matrices have been proposed in order to consider explicitly spatial predictors. The present study demonstrates the usefulness of eigenfunctions in spatial modeling applied to ecological problems and shows equivalencies of and differences between the two current implementations of this methodology. The two approaches in this category are the distance-based (DB) eigenvector maps proposed by P. Legendre and his colleagues, and spatial filtering based upon geographic connectivity matrices (i.e., topology-based; CB) developed by D. A. Griffith and his colleagues. In both cases, the goal is to create spatial predictors that can be easily incorporated into conventional regression models. One important advantage of these two approaches over any other spatial approach is that they provide a flexible tool that allows the full range of general and generalized linear modeling theory to be applied to ecological and geographical problems in the presence of nonzero spatial autocorrelation.

Key words: *ecological community; eigenvalue; eigenvector; Moran coefficient; principal coordinates of neighbor matrices; spatial autocorrelation; spatial filter; spatial model; spatial structure.*

INTRODUCTION

A major issue in contemporary ecology pertains to the identification and explanation of geographic variability affiliated with ecological communities (Cormack and Ord 1979, Smith 2002). Geographic space operates either as a factor that molds or constrains ecological structures, or as a confounding variable that introduces biases when it is overlooked during an ecological analysis (see Legendre 1993, Legendre and Legendre 1998, Keitt et al. 2002). On the one hand, spatial autocorrelation is known to influence the interpretation of statistical models by affecting tests of significance of the association between species distributions and environmental factors (Selmi and Boulinier 2001, Legendre et al. 2002, Peres-Neto 2004), as well as calculated correlations among such variables (Clifford et al. 1988, Richardson 1990, Dutilleul 1993). On the other hand, geographically contagious biotic processes (such as population growth, geographic dispersal, differential fertility or mortality, social organization, or competition dynamics) also can promote spatial autocorrelation in species distributions. In many instances, these two influences operate simultaneously, inducing spatial heterogeneity in ecological communities. The goal of

employing spatial statistical models is to explicitly account for the effects of these two sources of influence. Taken together, space can be seen as a predictor, when the goal is to exploit the mechanisms (e.g., shared common factors, geographic diffusion) that generate spatial autocorrelation in species distributions in space, and/or as a covariable when the goal is to adjust for (i.e., account for and filter out) spatial variation when testing for associations between species distributions and environmental factors.

Various modeling approaches can be utilized by ecologists to explicitly account for spatial autocorrelation in ecological phenomena (e.g., autoregression models, trend surface analysis; see Keitt et al. [2002] and Lichstein et al. [2002] for reviews). Paralleling the history of statistics, initially model specifications involved autoregressive versions of regression models based up the normality assumptions (Ripley 1988). While this approach had considerable, but limited, success, a number of ecological phenomena can be better described by the Poisson (i.e., counts of rare events) or the binomial (i.e., binary or percentage measures) probability models, given that species data typically are quantified with abundance or presence–absence values (i.e., some sites are occupied while others are not). The implementation of auto-specifications (akin to autoregressive models) of these latter two models remained elusive for years, even after generalized linear modeling (GLM) theory was reasonably well developed in the statistics literature. The purpose of this paper is threefold: (1) to describe how eigenfunction

Manuscript received 4 November 2005; revised 1 March 2006; accepted 2 March 2006; final version received 10 April 2006. Corresponding Editor: F. He.

³ Corresponding author. Present address: Département des Sciences Biologiques, Université du Québec à Montréal (UQAM), CP 8888, Succ. Centre-Ville, Montréal, Québec QC H3C 3P8 Canada. E-mail: pedro.peres-neto@uregina.ca

spatial analysis provides a flexible methodology for incorporating spatial variation in ecological models, regardless of the adopted probability model, that accounts for observed spatial autocorrelation by explicitly casting spatial variability in terms of predictors and/or covariates; (2) to unify the existing methods related to eigenfunction spatial analysis; and (3) to demonstrate how eigenfunction spatial analysis can be used to estimate the spatial component structuring communities and to filter out the effects of spatial autocorrelation when modeling species distributions according to their environment. The class of methods presented is flexible enough that it can be applied to any type of distribution under general as well as generalized linear model (e.g., logistic/binomial and Poisson regressions) procedures.

BACKGROUND TO THE PROBLEM

Ecologists are increasingly aware of the problems introduced by spatial autocorrelation when assessing relationships between distributional patterns and local environmental characteristics. The problem arises in these situations because statistical models and tests overestimate the effects of environmental factors driving species distribution if solely based on environmental predictors without accounting for spatial variation (Legendre and Legendre 1998, Dale and Fortin 2002). Probably the most common approach used for partitioning out the common effects due to spatial processes is to use trend surface polynomial regression (e.g., Wartenberg 1985, Borcard et al. 1992, Legendre and Legendre 1998, Lichstein et al. 2002 [and references therein]). Trend surface analysis, however, presents reasonable solutions only when the sample area is approximately homogeneous and the sampling design is close to regular (Norcliffe 1969, Scarlett 1972). Additionally, the spatial structure to be modeled is rather simple and global, such as a gradient, a single wave, or a saddle (Legendre and Legendre 1998:739). Moreover, the use of trend surfaces introduces an arbitrary choice for the degree of the polynomial functions, with their lack of independence possibly hindering model selection. Although orthogonalization is always possible (e.g., Gram-Schmidt, PCA), the degree of interpretability of linear combinations of polynomials can become difficult.

Lichstein et al. (2002) advocated the use of regression models based on trend surface polynomials to account for broad-scale spatial patterns, whereas autoregressive models should be used to account for fine-scale variations. The problem with this approach is that ecologists are forced to work with different classes of models that can potentially hinder interpretation since results cannot be directly combined. In addition, analysis can only be conducted at the species level since there is no analogue for the autoregressive model in the realms of multispecies analysis, such as direct gradient analysis. Even when the auto-normal probability model (i.e., a conventional autoregressive model) is employed,

accounting for spatial autocorrelation detected in the geographic distribution of some ecological phenomena can be problematic. The most severe complication here arises when autoregressive models are used to describe large georeferenced data sets, since the accompanying Jacobian term, which is necessary to estimate the autoregressive parameters, is based upon the determinant of an $n \times n$ matrix. A second complication is the need to forego classical GLMs in favor of somewhat less developed models, such as the auto-Poisson and auto-logistic regression models. These models can produce maximum-likelihood estimates only under certain limited conditions (Cressie 1991), although some computational solutions are available (see Kaiser and Cressie 1997, He et al. 2003).

Recent advances in spatial modeling that seek to avoid the complications involved in estimating autoregressive parameters, as well as to exploit established ordinary least-squares (OLS) and GLM theory, fall under the heading of spatial-filtering methodology. In the linear-regression context, spatial filtering transforms a variable containing spatial dependence into one free of spatial dependence by partitioning the original georeferenced attribute variable into two synthetic variates that can be employed with a GLM framework: a spatial-filter variate capturing latent spatial dependency that otherwise would remain in the response residuals, and a nonspatial variate that is free of spatial dependence. Three forms of spatial filtering have emerged. Getis (1990, 1995) devised a multistep procedure based upon Ripley's second-order statistic (Ripley 1977)—now called the K function—and the G_i spatial statistic developed by Getis and Ord (1992). Griffith (e.g., 2000a) proposes a transformation procedure that depends on the eigenfunctions of matrix $(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)$ —where \mathbf{I} is an $n \times n$ identity matrix, $\mathbf{1}$ is an $n \times 1$ vector of ones, T denotes matrix transpose, n is the number of sampling locations, and \mathbf{C} is a binary connectivity matrix that is a term appearing in the numerator of the Moran coefficient (MC) spatial autocorrelation index. Note that the analysis also could be based upon the Geary ratio, the other popular spatial autocorrelation index. In a somewhat similar fashion, Borcard and Legendre (2002; see also Borcard et al. 2004) proposed a transformation procedure, called principal coordinates of neighbor matrices (PCNM), that depends on the eigenfunctions of the matrix of truncated geographic distances among locations. Getis and Griffith (2002) furnished a comparison between their individual formulations. The current study outlines the conceptual framework upon which Griffith's spatial-filtering and PCNM methodology are built, demonstrating how these methods are linked to spatial autocorrelation. Comparisons between the two methodologies are illustrated with the georeferenced oribatid mite data set that already has been analyzed by Borcard et al. (2004).

EIGENFUNCTION-BASED SPATIAL-FILTERING METHODOLOGIES

Paralleling principal components analysis in multivariate statistics, the Griffith and the Borcard-Legendre methods begin by diagonalizing an $n \times n$ geographic structure matrix (e.g., the aforementioned matrix **C**). The resulting eigenvectors, themselves, then are used directly as synthetic explanatory variables in regression analysis. This modeling approach is semiparametric in nature, casting spatial autocorrelation as some unknown function (nonparametric), which must be estimated from a given data set; the function is additively coupled with a set of covariates whose coefficients need to be estimated (parametric). Linear combinations of judiciously selected subsets of the aforementioned eigenvectors constitute the unknown function.

The Griffith topology-based spatial-filtering methodology rests on the following property (Griffith 2003): The first eigenvector, say \mathbf{E}_1 , is the set of numerical values that has the largest MC achievable by any set for the spatial arrangement defined by the geographic connectivity matrix **C**. The second eigenvector is the set of values that has the largest achievable MC by any set that is uncorrelated with \mathbf{E}_1 . The third eigenvector is the third such set of values. And so on. This sequential construction of eigenvectors continues through \mathbf{E}_n , the set of values that has the largest negative MC achievable by any set that is uncorrelated with the preceding $(n - 1)$ eigenvectors. As such, Griffith (2000a) argued that these eigenvectors furnish distinct map pattern descriptions of latent spatial autocorrelation in georeferenced variables. Note that the number of positively and negatively autocorrelated eigenvectors changes depending on the neighborhood matrix, but the number of negatively autocorrelated eigenvectors is always larger than the number of positively autocorrelated eigenvectors for irregular tessellations. Moreover, MC values are not necessarily consistent across the two sets of eigenvectors (i.e., the largest positive MC does not equal the largest negative MC).

In ecological applications, a spatial filter is constructed by first restricting attention to the set of candidate eigenvectors that represent positive spatial autocorrelation, which is the one known to overestimate the contribution of environmental factors (Legendre et al. 2002). Next, judicious selection of eigenvectors can be made with a forward-regression routine (see Appendix B). Of note is that the eigenvectors in the resulting subset representing distinct map patterns are both mutually orthogonal and linearly uncorrelated in their numerical form (Griffith 2000b), a property that is corrupted by the weighting involved in computing GLM parameter estimates, but could be correlated with covariates contained in a model specification. This multicollinearity can be eliminated by extracting the eigenfunctions of matrix $[\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\mathbf{C}[\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]$, where the $n \times (p + 1)$ matrix **X** contains the p predictor variables plus a vector of ones

(Tiefelsdorf and Griffith 2006). This is the numerator of the MC for a mean-response linear combination of predictor variables contained in matrix **X**.

The PCNM spatial-filtering method (Borcard and Legendre 2002) begins with an eigenfunction decomposition of a truncated matrix of geographic distances among locations. Eigenvectors corresponding to positive eigenvalues are used as spatial descriptors in regression or canonical analysis. This method can be applied to any set of locations providing a good coverage of a given geographic landscape. In the original method described by Borcard and Legendre (2002), the truncated matrix of geographic distances is built in such a way that it considers the influence of a sampling location on itself (i.e., the geographic distance matrix has nonzero values in the main diagonal). Although this consideration could be seen as difficult to justify, there are examples of spatial models where it has been applied (Bavaud 1998). However, in order to make the method fully compatible with the MC framework, and therefore more similar to the Griffith topology-based spatial-filtering methodology, Dray et al. (2006) implemented a modification in the original Borcard-Legendre (2002) PCNM spatial-filtering method. Hereafter, we will refer to the results of this modified version as distance-based eigenvector maps, to distinguish it from Griffith's method, whose results can be referred to as topology-based eigenvector maps. The distance-based eigenvector procedure (after Dray et al. 2006) can be summarized with the following steps:

- 1) Compute a pairwise Euclidean (geographic) distance matrix **D** among sampling units ($\mathbf{D} = [d_{ij}]$).
- 2) Choose a threshold value t and construct a truncated connectivity matrix **W** (i.e., not all sites are connected) using the following rule:

$$\mathbf{W} = (w_{ij}) = \begin{cases} 0 & \text{if } i = j \\ 0 & \text{if } d_{ij} > t \\ [1 - (d_{ij}/4t)^2] & \text{if } d_{ij} \leq t \end{cases}$$

where t is chosen as the maximum distance that maintains all sampling units being connected using a minimum spanning tree algorithm (Legendre and Legendre 1998).

- 3) Compute the eigenvectors of the centered connectivity matrix, $(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)\mathbf{W}(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)$. As for the topology-based spatial-filtering methodology, this centering process guarantees that the extracted eigenvectors are orthogonal and linearly independent (Griffith 2000b).

Given the non-Euclidean nature of spatial connectivity matrices, both positive and negative eigenvalues are produced. The non-Euclidean part is introduced by the fact that only certain connections among sites, and not all, are considered in matrix **W**. As in Griffith's method, the extracted eigenvectors represent the decomposition of the MC into all mutually orthogonal and linearly

uncorrelated map patterns. Eigenvectors having associated positive eigenvalues represent positive spatial association, whereas eigenvectors having negative eigenvalues represent negative spatial association. An MC for any eigenvector \mathbf{v} can be directly calculated as follows (Dray et al. 2006):

$$\begin{aligned} \text{MC}(\mathbf{v}) &= \frac{n}{\mathbf{1}^T \mathbf{S} \mathbf{1}} \mathbf{v}^T \left(\mathbf{I} - \frac{\mathbf{1} \mathbf{1}^T}{n} \right) \mathbf{W} \left(\mathbf{I} - \frac{\mathbf{1} \mathbf{1}^T}{n} \right) \mathbf{v} \\ &= \frac{n}{\mathbf{1}^T \mathbf{S} \mathbf{1}} \mathbf{v}^T \mathbf{W} \mathbf{v}. \end{aligned}$$

Distance-based eigenvector maps with large eigenvalues represent coarse scales of variability or landscape-wide trends (e.g., global); eigenvectors with intermediate size eigenvalues represent medium scales (e.g., regional); eigenvectors with small eigenvalues represent fine scales or patchiness (e.g., local). Therefore, distance-based eigenvector maps capture a range of geographic scales encapsulated in a given data set, restricted by the landscape boundary extent of sample locations and the threshold value used to truncate distance, and by the smallest interlocation distances existing in a landscape. Again, only prominent eigenvector maps (i.e., those that significantly contribute to the explanation of a response variable) should be utilized.

To provide a clear description of the topology-based and the distance-based eigenvector maps, we provide an example involving a small number of sampling locations, where all the main analytical steps and the matrixes involved are presented (see Fig. 1). Eigenvector maps are treated just as other predictors of interest; after extraction, eigenvector maps are selected (see Appendix B) and considered jointly with environmental predictors in the model (Fig. 1). In addition, we make available a complete set of Matlab routines that calculate eigenvector predictors based on both methods (see Supplements 1–3). We also show the spatial patterns depicted by three distance-based eigenvector maps constructed for a 40×40 regular grid and a linear combination of eigenvectors (Fig. 2). In this latter case, each of 40 randomly chosen eigenvectors (out of 745 portraying prominent degrees of positive spatial autocorrelation) was multiplied by a randomly chosen coefficient from the interval $[-1, 1]$,

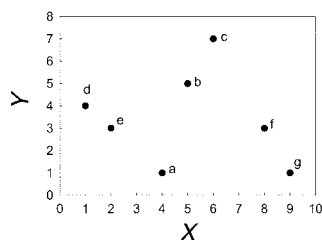
and then the weighted set was summed. This procedure mimics a possible outcome of a linear combination of eigenvector maps created with a model selection technique.

THE ORIBATID MITE DATA SET REVISITED: AN EMPIRICAL COMPARISON OF THE TWO EIGENFUNCTION METHODS

In June 1989, 70 *Sphagnum* moss cores (5 cm in diameter and 7 cm deep) were sampled from a 10×2.5 m area in a floating moss and peat blanket that extended from a forest border into a lake in St. Hippolyte, Québec (Borcard et al. 1992). The resulting ecological data set comprises the following: 35 oribatid mite species counts; spatial coordinates (u , v) of the 70 sample cores; and 11 environmental variables, i.e., bulk density of the substratum (measured in grams per liter of dry uncompressed matter), water content (measured in grams per liter of raw uncompressed material), and Helmert orthogonal contrasts coding four species of *Sphagnum* moss, ligneous litter, bare peat, the interface between two substrates, three categories of shrubs, and microtopography (blanket or hummock).

In this example, 28 topology-based eigenvectors and 22 distance-based eigenvectors having positive eigenvalues were retained as spatial descriptors. In order to compare the two sets of predictors, we use canonical analysis as a template. Canonical analyses, such as redundancy analysis (RDA, Rao 1964) and canonical correspondence analysis (CCA, ter Braak 1986), are widely used tools for modeling communities with environmental and spatial predictors (Legendre and Legendre 1998). They provide the means of conducting direct explanatory analyses in which associations among species can be studied with respect to their common and unique relationships with environmental variables and spatial predictors. Redundancy analysis can be best understood as a method for extending multiple regression, which has a single response \mathbf{Y} and multiple predictors \mathbf{X} (e.g., several spatial and environmental predictors), to multiple regression involving multiple response variables (e.g., several species) and a common matrix of predictors \mathbf{X} . Prior to analysis, the species data was Hellinger-transformed (Legendre and Gallagher 2001). Using a few fabricated examples, Legendre and

FIG. 1. A schematic representation of the steps involved in the distance-based and topology-based eigenvector map extraction procedures. Sampling units are identified by letters, and columns and rows in the matrices follow the same order. For the distance-based eigenvector maps (left panel), the following five steps are involved: (1) compute a pairwise Euclidian distance matrix $\mathbf{D} = [d_{ij}]$ for the sampling localities; (2) chose a threshold t (here a value of 3.6056 was employed, which was found by a minimum spanning tree [see *Eigenfunction-based spatial-filtering methodologies*]), and construct a truncated connectivity matrix $\mathbf{W} = [w_{ij}]$ using the following rule: $w_{ij} = 0$ if $i = j$; $w_{ij} = 0$ if $d_{ij} > t$; and $w_{ij} = [1 - (d_{ij}/4t)^2]$; and (3) center the truncated connectivity matrix (see *Eigenfunction-based spatial-filtering methodologies*) using the projection matrix $(\mathbf{I} - \mathbf{1} \mathbf{1}^T/n)$, and then extract the associated eigenvectors. Meanwhile, the following are the steps for the topology-based eigenvector maps (right panel): (1) compute a binary 0–1 connectivity matrix based upon a particular surface partitioning (landscape topology); and (2) center the connectivity matrix using the projection matrix $(\mathbf{I} - \mathbf{1} \mathbf{1}^T/n)$, and then extract the associated eigenvectors. In the present example, Thiessen polygons are employed to establish a topological surface articulation. Once eigenvector maps are extracted with both methods, a selection scheme (see *Eigenfunction-based spatial-filtering methodologies*) is used to retain only prominent eigenvectors. Finally, species distributions are modeled as functions of the retained eigenvector maps and environmental factors. As a fictional example, we show a model based on three environmental variables (X_1 , X_2 , and X_3) and two eigenvector maps (EV1 and EV3).



	X	Y
a	4	1
b	5	5
c	6	7
d	1	4
e	2	3
f	8	3
g	9	1

Geographic coordinates

Distance-based eigenvector maps

Euclidean distance matrix

0.00	4.12	6.32	4.24	2.83	4.47	5.00
4.12	0.00	2.24	4.12	3.61	3.61	5.66
6.32	2.24	0.00	5.83	5.66	4.47	6.71
4.24	4.12	5.83	0.00	1.41	7.07	8.54
2.83	3.61	5.66	1.41	0.00	6.00	7.28
4.47	3.61	4.47	7.07	6.00	0.00	2.24
5.00	5.66	6.71	8.54	7.28	2.24	0.00

Truncated connectivity matrix

0.00	0.00	0.00	0.00	0.96	0.00	0.00
0.00	0.00	0.98	0.00	0.94	0.94	0.00
0.00	0.98	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.99	0.00	0.00
0.96	0.94	0.00	0.99	0.00	0.00	0.00
0.00	0.94	0.00	0.00	0.00	0.00	0.98
0.00	0.00	0.00	0.00	0.00	0.98	0.00

Centered truncated connectivity matrix

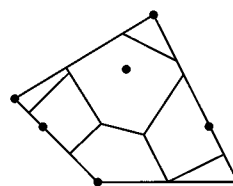
-0.04	-0.31	-0.04	-0.04	0.65	-0.17	-0.04
-0.31	-0.58	0.67	-0.31	0.35	0.49	-0.31
-0.04	0.67	-0.04	-0.05	-0.32	-0.18	-0.04
-0.04	-0.31	-0.05	-0.05	0.67	-0.18	-0.05
0.65	0.35	-0.32	0.67	-0.59	-0.45	-0.32
-0.17	0.49	-0.18	-0.18	-0.45	-0.31	0.80
-0.04	-0.31	-0.04	-0.05	-0.32	0.80	-0.04

Positive eigenvalues and eigenvectors

122.94	0.37	0.19	0.72
63.11	-0.16	-0.49	-0.01
0.02	-0.15	-0.63	0.01
	0.38	0.19	-0.70
	0.48	0.02	-0.01
	-0.50	0.21	-0.01
	-0.43	0.50	0.00

Topology-based eigenvector maps

Surface partitioning



Connectivity matrix

0	1	0	0	1	1	1
1	0	1	1	1	1	0
0	1	0	0	0	0	0
0	1	0	0	1	0	0
1	1	0	1	0	0	0
1	1	0	0	0	0	1
1	0	0	0	0	1	0

Centered connectivity matrix

-0.73	0.12	-0.31	-0.45	0.41	0.41	0.55
0.12	-1.02	0.55	0.41	0.27	0.27	-0.59
-0.31	0.55	0.12	-0.02	-0.16	-0.16	-0.02
-0.45	0.41	-0.02	-0.16	0.69	-0.31	-0.16
0.41	0.27	-0.16	0.69	-0.45	-0.45	-0.31
0.41	0.27	-0.16	-0.31	-0.45	-0.45	0.69
0.55	-0.59	-0.02	-0.16	-0.31	0.69	-0.16

Positive eigenvalues and eigenvectors

1.30	-0.28	0.25
0.42	0.24	-0.16
	0.20	-0.80
	0.47	0.26
	0.34	0.44
	-0.43	-0.06
	-0.54	0.05

Selection of eigenvector maps

Species distribution = $f(\text{Environmental Factors, selected Eigenvector maps}) + \text{error}$
 (e.g., Species distribution = $b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4EV_1 + b_5EV_3 + \text{error}$)

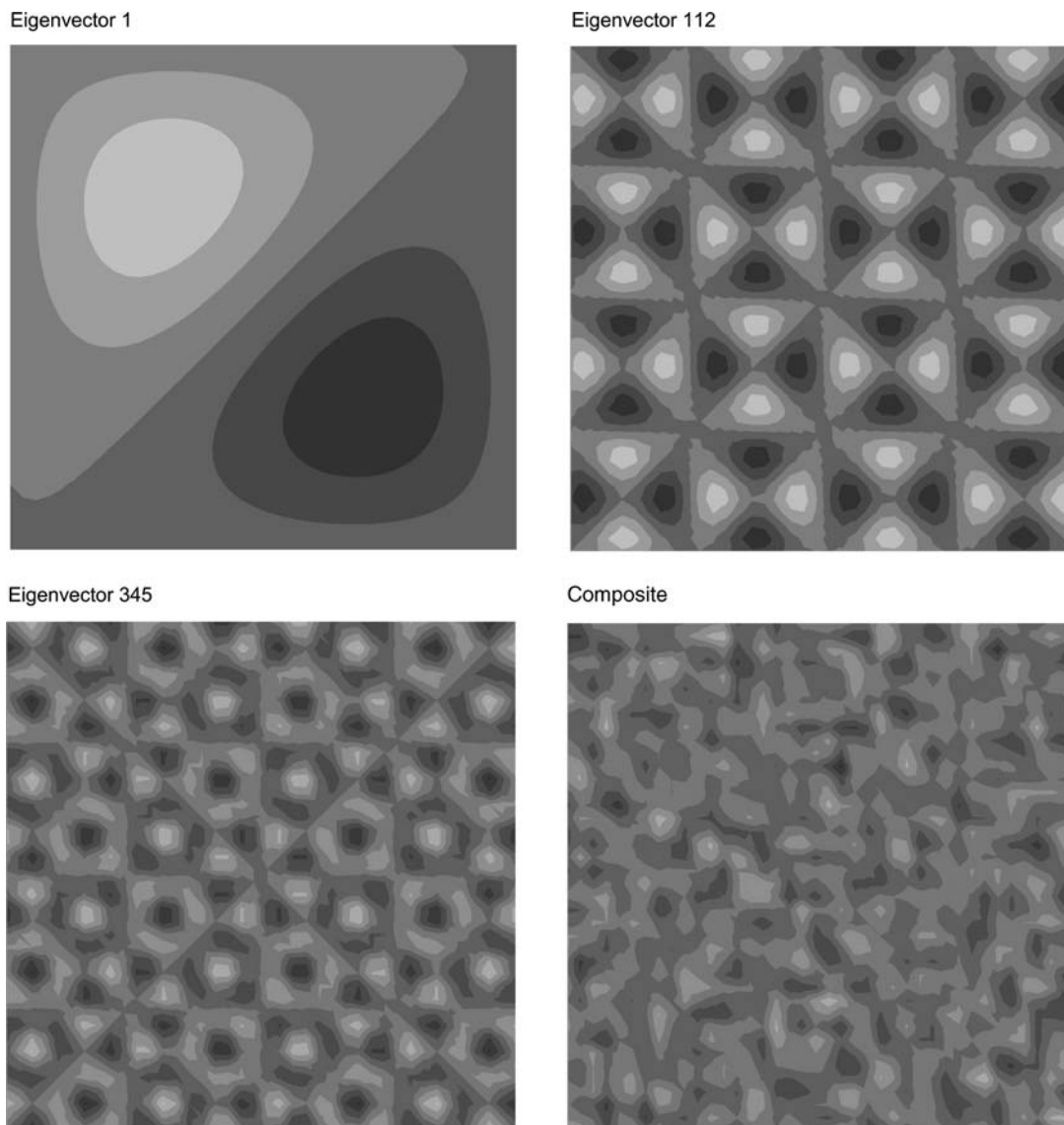


FIG. 2. Examples of spatial patterns depicted by eigenvector maps. Three distance-based eigenvectors (numbers 1, 112, and 345, based upon a descending ordering of the accompanying eigenvalues) are presented. These were extracted from a spatial structure matrix for a 40×40 regular grid. The composite map (bottom right panel) represents a linear combination of eigenvector maps (see *Eigenfunction-based spatial-filtering methodologies*). Eigenvector maps whose associated eigenvalues are the largest have large Moran coefficient (MC) values and represent coarse spatial scales, whereas axes with eigenvalues closer to zero correspond to finer spatial scales.

Gallagher (2001) report that this transformation might provide better resolutions for species, by making them behave more like a Gaussian distribution, than the commonly used chi-square metric employed in CCA (ter Braak 1986). In addition, Peres-Neto et al. (2006) showed that this transformation provides appropriate estimates of the redundancy statistic $R^2_{Y|X}$ (i.e., the RDA equivalent of the regression coefficient of determination, R^2). Next, species-transformed values were detrended to remove the linear component of variation. Detrended values represent regression residuals of the

Hellinger-transformed data using geographic coordinates as covariates.

We began by comparing the sets of multivariate scores for the first two canonical axes of two separate RDAs based on the two sets of spatial predictors (Fig. 3). Canonical scores are simply based on a principal components analysis of the covariance matrix of predicted values between species, and are routinely used to construct bi-plots to identify environmental and spatial gradients regarding species distributions (Legendre and Legendre 1998). The plots of scores vs. their

coordinates show that both sets of spatial predictors provide high concordance in terms of their predicted values. Next, we conducted a variance partitioning (Borcard et al. 1992, Peres-Neto et al. 2006; see Appendix C for a complete description of the method) based on the two sets of spatial predictors and the environmental predictors. In this approach, the total percentage of variation explained by the model ($R^2_{Y|X}$) is partitioned into unique and common contributions of the sets of environmental and spatial predictors. Results of variance partitioning based on adjusted fractions of variance (Peres-Neto et al. 2006; see Appendix C) are presented in Fig. 4, and they indicate that differences between the two methods are not substantial with regard to the amount of variation solely explained by environment and space. We also conducted a variance-partitioning exercise using the two sets of spatial predictors, with results indicating that most of the explained variation in species distribution is shared between the topology- and distance-based eigenvectors. The unique fractions of variation related to the topology-based and distance-based eigenvectors are 0.093 and 0.080, respectively, whereas the common fraction accounts for the largest amount of variation (0.166).

In order to assess the success of both types of spatial predictors in removing the effects of spatial dependence in regression residuals, for each oribatid species we calculated a global Moran's coefficient I (MC) (see Appendix A for calculations and tests of significance) for the environmental model residuals (Table 1). Regression analysis results for 12 and 11 species reveal the presence of significant spatial dependence according to the topology- and distance-based neighborhood matrices, respectively. Here a forward selection procedure based on including variables that decrease the amount of positive spatial autocorrelation in residuals was implemented (see Appendix B for details). In this approach, eigenvectors are added to a model until the spatial autocorrelation measured by I in the residuals is no longer significant according to a preestablished alpha level. Most species that display significant levels of positive autocorrelation for one set of spatial predictors also do so for the other set. Both sets of predictors are equally successful in removing the spatial trends in the model residuals. In addition, we also tested the environmental contribution after spatial predictors were entered into a model (see Legendre and Legendre 1998:608–612). For both sets of predictors, unique percentages of contribution and probability levels were comparable across all species regarding the influence of environmental factors driving their distributions after spatial dependence was removed. In Appendix D, we compare results based on the eigenvector filtering methods with those obtained by estimating a simultaneous autoregressive (SAR) spatial model, and we show that the SAR model does not completely remove the autocorrelation among regression residuals for all species. This is a restricted comparison in the sense that we only consider

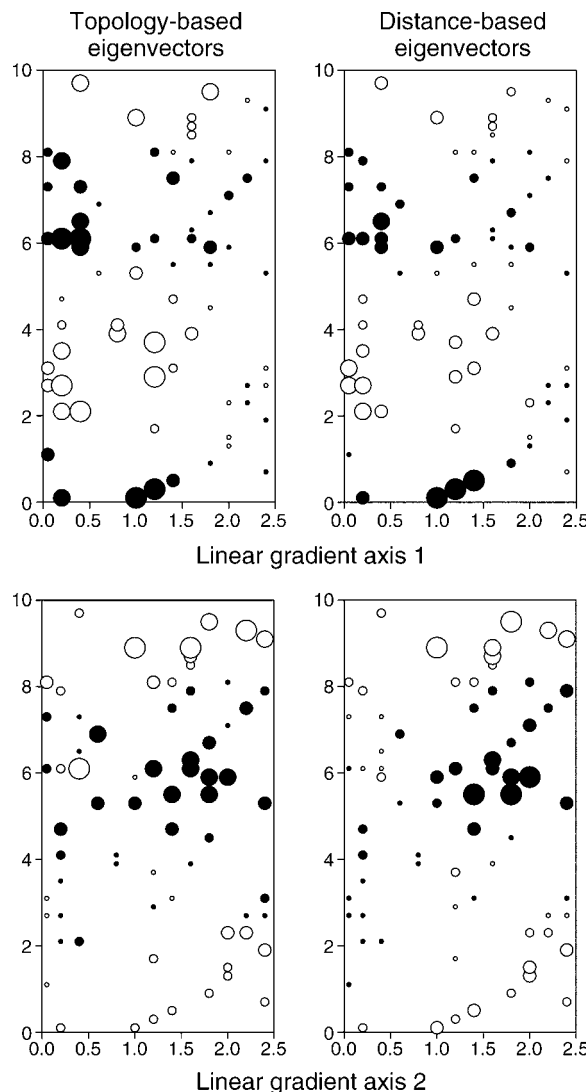


FIG. 3. The first two canonical axes of an RDA of the Hellinger-transformed oribatid mite species data explained by topology- and eigenvector-based spatial predictors.

the species studied here. However, it points out that, given the selection procedure implemented here, the eigenvector method is designed to assure that residuals are not spatially autocorrelated, which is not the case for the autoregressive family based on maximum likelihood estimates (Tiefelsdorf and Griffith 2006). Indeed, autoregressive models might not be completely successful in removing autocorrelation in model residuals (e.g., see Lichstein et al. (2002: Fig. 2) regarding the CAR model, and Tognelli and Kelt (2004: Fig. 4) regarding the SAR and CAR models).

DISCUSSION

Our goal here is to expose ecologists to a new class of spatial predictors based on eigenfunction analysis and to compare the two implementation methods that are

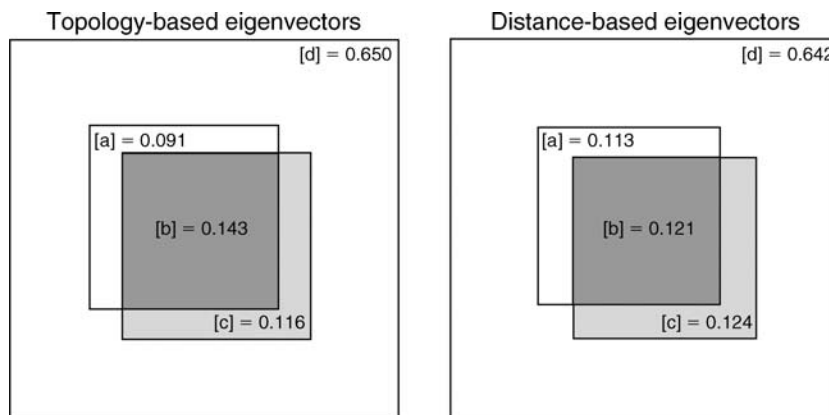


FIG. 4. Variance partitioning Venn diagrams representing percentages of unique contribution of environmental [a] and spatial [c] components to the oribatid mite distribution based on topology- and distance-based eigenvectors. In the figure, [b] represents the amount of shared variation between the environmental and spatial components, and [d] represents the residual fraction (i.e., unexplained variation).

available in this relatively new and emerging spatial approach. In addition, we compare the two methods based on multi- and single-species analysis, showing the flexibility of this class of methods to both types of analysis. The modeling approaches (redundancy analysis [RDA] and multiple regression) used here are based on ordinary least-squares regression. Although the issues of normality here may appear as a concern, and if so can be addressed as in Griffith (2002) and Griffith (2004), the Hellinger-transformation provides reliable estimates in variance partitioning (Peres-Neto et al. 2006) and representation in canonical space (Legendre and Gallagher 2001). In addition, all conducted statistical tests reported here are based on randomization procedures to avoid potential problems with the normality assumptions associated with inference. However, the two classes of eigenvector methods can be easily extended to generalized linear models (GLMs; e.g., logistic/binomial and Poisson regressions), whereas the autoregressive approach becomes cumbersome and difficult to estimate when extended to GLMs. The principal problem with autoregression is the normalizing constant, which is the Jacobian term in the case of a linear model and requires Markov chain Monte Carlo estimation in the GLM case. In addition, there is nothing akin to the autoregressive model for modeling communities (i.e., multiple species). The methods presented here can be easily incorporated in a GLM framework where binary, percentage, and count data are to be analyzed using environmental and eigenvector predictors, with selection continuing to be based on choosing eigenvectors that reduce autocorrelation in model residuals. In Supplement 3, we make available software that performs the eigenvector selection procedure described here for OLS and GLM (logistic and Poisson) estimation.

Another important advantage of the spatial eigenvectors is that a forward selection method based on the idea of reducing the degree of dependence between residuals

can be performed, since we can estimate the degree of global autocorrelation in these residuals. This approach is important when assessing the contribution of environmental predictors because, if all spatial eigenvectors (28 and 22 for the topology- and distance-based, respectively, in the example data used here) are kept in a model jointly with environmental predictors, one may decrease statistical power due to the loss of degrees of freedom. This would be the case either when assessing the unique contribution of environment in a variance-partitioning scheme or when interpreting individual slopes for each environmental predictor in single-species models. However, if the interest is to estimate the importance of space rather than control for it, then we suggest that all predictors are used and contributions are adjusted in accordance with the number of predictors in a model. Based upon simulation experiments, Peres-Neto et al. (2006) found that adjusted fractions in variance partitioning are better estimated when all regressors are used rather than just a reduced set identified with forward model selection techniques. Another point to consider is that ecologists seem to often apply correlograms to judge whether applying a spatial model successfully removes spatial dependence among residuals. Although we have considered global measures of autocorrelation, one could easily modify the eigenvector selection procedure in order to reduce the autocorrelation between particular classes of distances.

Although both eigenfunction-based methods we compare are quite similar in terms of their results regarding the example employed here, this outcome may not always be the case, and choice might depend on the specific success of each method in removing spatial dependence in regression residuals. In addition, both methods are flexible enough, given that they can be further modified. For instance, we could apply the surface partitioning used in generating the topology-based eigenvector maps to the distance-based method

TABLE 1. Moran's I for regression residuals and associated probabilities for models considering only environment (E), and environment and topology- or distance-based eigenvector spatial predictors (E + S) for the oribatid mite data set.

Species	Topology-based eigenvectors						Distance-based eigenvectors					
	Moran's I				Environmental contribution		Moran's I				Environmental contribution	
	E	P	E + S	P	Percentage	P	E	P	E + S	P	Percentage	P
Brachy	-0.081	0.619	0.149	0.037	-0.042	0.401	0.149	0.036
PHTH	-0.102	0.734	0.063	0.203	-0.027	0.312	0.063	0.191
HPAV	-0.244	0.998	0.124	0.058	-0.038	0.374	0.124	0.062
RARD	0.255	0.001	0.012	0.063	0.297	0.001	0.272	0.001	-0.029	0.158	0.186	0.005
SSTR	0.085	0.028	0.000	0.129	0.270	0.001	0.129	0.005	0.037	0.046	0.360	0.001
Protopl	0.179	0.001	0.036	0.068	-0.050	0.720	0.085	0.014	-0.001	0.125	-0.034	0.595
MEGR	-0.020	0.294	0.061	0.202	-0.117	0.852	0.061	0.184
MPRO	0.137	0.007	0.006	0.116	0.011	0.369	0.102	0.011	0.006	0.135	-0.034	0.653
TVIE	-0.196	0.975	0.028	0.328	-0.024	0.310	0.028	0.322
HMIN	-0.069	0.546	0.331	0.002	-0.016	0.258	0.331	0.002
HMIN2	-0.002	0.219	0.102	0.106	0.074	0.024	0.013	0.092	0.148	0.054
NPRA	0.105	0.017	0.004	0.123	0.117	0.058	-0.002	0.190	0.062	0.161
TVEL	0.254	0.001	0.004	0.082	0.175	0.001	0.199	0.003	0.000	0.085	0.194	0.001
ONOV	-0.045	0.437	0.408	0.001	-0.048	0.493	0.408	0.002
SUCT	0.112	0.013	0.003	0.113	0.302	0.001	0.144	0.002	0.009	0.071	0.196	0.002
LCIL	0.035	0.091	0.337	0.001	0.074	0.027	0.000	0.117	0.386	0.001
Oribatul1	0.056	0.077	0.064	0.197	0.006	0.157	0.064	0.197
Ceratoz1	-0.034	0.347	0.025	0.299	-0.032	0.361	0.025	0.306
HRUF	-0.075	0.575	-0.056	0.706	-0.007	0.208	-0.056	0.684
Trhypochth1	0.185	0.001	0.025	0.088	0.612	0.001	0.128	0.009	0.015	0.094	0.471	0.001
PPEL	-0.038	0.379	-0.084	0.817	-0.033	0.363	-0.084	0.827
NCOR	-0.065	0.517	0.171	0.016	-0.026	0.303	0.171	0.018
SLAT	0.093	0.018	0.001	0.148	0.004	0.437	0.040	0.069	-0.026	0.571
FSET	0.009	0.171	0.231	0.005	-0.037	0.388	0.231	0.005
Lepidozetes	-0.005	0.218	0.057	0.216	-0.081	0.686	0.057	0.208
Eupelops	0.003	0.204	0.138	0.081	-0.038	0.360	0.138	0.081
Minigalumna	-0.038	0.379	0.054	0.243	0.032	0.079	0.054	0.229
LRUG	0.218	0.002	0.005	0.100	0.352	0.001	0.136	0.006	-0.001	0.124	0.243	0.001
PLAG2	0.104	0.027	0.000	0.130	-0.026	0.620	0.051	0.054	-0.054	0.727
Ceratoz3	-0.104	0.718	0.086	0.115	-0.020	0.285	0.086	0.133
Oppia	0.029	0.122	0.207	0.006	0.046	0.064	0.207	0.014
Trimalaco2	0.114	0.014	0.000	0.181	0.030	0.257	0.133	0.005	-0.003	0.130	-0.047	0.881

Notes: Environmental contribution refers to the amount of contribution to the model. In cases where Moran's I is not significant, this percentage is simply the adjusted R^2 from regression models considering only spatial predictors, whereas if Moran's I is significant, then the percentage is the unique adjusted contribution of environment after the influence of spatial predictors is factored out via variance partitioning. Abbreviations: Brachy, *Brachychochthonius* sp.; PHTH, *Phthiracarus* sp.; HPAV, *Hoplophthiracarus pavidus*; RARD, *Rhysotritia ardua*; SSTR, *Steganacarus striculus*; Protopl, *Protophthiracarus* sp.; MEGR, *Malacothonrus cf. egregius*; MPRO, *Malacothonrus cf. processus*; TVIE, *Trimalaconothrus cf. vietsi*; HMIN, *Hypothoniella minutissima*; HMIN2, *Hypothoniella* sp.; NPRA, *Nothrus pratensis*; TVEL, *Tectocepheus velatus*; ONOV, *Oppiella nova*; SUCT, *Suctobelba* spp.; LCIL, *Limozetes cf. ciliatus*; Orbatul1, *Oribatulidae* sp. 1; Ceratoz1, *Ceratozetes* sp. 1; HRUF, *Hypochthonius rufulus*; Trhypochth1, *Trhypochthonius* sp. 1; PPEL, *Platynothrus peltifer*; NCOR, *Nanhermannia coronata*; SLAT, *Scheloribates latipes*; FSET, *Fuscozetes setosus*; Lepidozetes, *Lepidozetes* sp.; Eupelops, *Eupelops* sp.; Minigalumna, *Galumnidae* sp.; LRUG, *Limozetes cf. rugosus*; PLAG2, *Protoribates* sp.; Ceratoz3, *Ceratozetes* sp. 2; Oppia, *Oppia minus* (Paoli); Trimalaco2, *Trimalaconothrus* sp.

where the 1's in the connectivity matrix (Fig. 1) could be replaced by geographic distances. Conversely, the truncated connectivity matrix used in the distance-based method could have its distances replaced by 1's. There are many possible avenues for applying and expanding the applications of eigenvector maps, and our study is an illustration of two methods that can be unified under the class of spatial eigenfunction maps, a relatively new and flexible technique for spatial analysis.

ACKNOWLEDGMENTS

Both authors contributed equally to this work. An early version of this paper was presented at the Spatial Ecology Workshop, University of Miami, Coral Gables, 7–10 January 2005. We thank Pierre Legendre and Stéphane Dray for lengthy discussions on the distance-based eigenvector maps. This

manuscript was greatly improved by the comments provided by two anonymous reviewers. This research was supported in part by an NSERC grant to P. R. Peres-Neto.

LITERATURE CITED

- Bavaud, F. 1998. Models for spatial weights: a systematic look. *Geographical Analysis* **30**:152–171.
- Borcard, D., and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* **153**:51–68.
- Borcard, D., P. Legendre, C. Avois-Jacquet, and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. *Ecology* **85**:1826–1832.
- Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. *Ecology* **73**: 1045–1055.

- Clifford, P., S. Richardson, and D. Hémon. 1988. Assessing the significance of the correlation between two spatial processes. *Biometrics* **45**:123–134.
- Cormack, R., and J. Ord. 1979. Spatial and temporal analysis in ecology. International Co-operative Publishing House, Fairland, Maryland, USA.
- Cressie, N. 1991. Statistics for spatial data. Wiley series in probability and mathematical statistics. John Wiley and Sons, New York, New York, USA.
- Dale, M. R. T., and M. J. Fortin. 2002. Spatial autocorrelation and statistical tests in ecology. *Écoscience* **9**:162–167.
- Dray, S., P. Legendre, and P. R. Peres-Neto. 2006. Spatial modeling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). *Ecological Modelling*, in press.
- Dutilleul, P. 1993. Modifying the *t* test for assessing the correlation between two spatial processes. *Biometrics* **49**:305–314.
- Getis, A. 1990. Screening for spatial dependence in regression analysis. *Papers of the Regional Science Association* **69**:69–81.
- Getis, A. 1995. Spatial filtering in a regression framework: experiments on regional inequality, government expenditures, and urban crime. Pages 172–188 in L. Anselin and R. Florax, editors. *New directions in spatial econometrics*. Springer, Berlin, Germany.
- Getis, A., and D. Griffith. 2002. Comparative spatial filtering in regression analysis. *Geographical Analysis* **34**:130–140.
- Getis, A., and J. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* **24**:189–206.
- Griffith, D. 2000a. A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems* **2**:141–156.
- Griffith, D. 2000b. Eigenfunction properties and approximations of selected incidence matrices employed in spatial. *Linear Algebra and its Applications* **321**:95–112.
- Griffith, D. 2002. A spatial filtering specification for the auto-Poisson model. *Statistics and Probability Letters* **58**:245–251.
- Griffith, D. 2003. Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer-Verlag, Berlin, Germany.
- Griffith, D. 2004. A spatial filtering specification for the autologistic model. *Environment and Planning A* **36**:1791–1811.
- He, F., J. Zhou, and H. T. Zhu. 2003. Autologistic regression model for the distribution of vegetation. *Journal of Agricultural, Biological, and Environmental Statistics* **8**:205–222.
- Kaiser, M., and N. Cressie. 1997. Modeling Poisson variables with positive spatial dependence. *Statistics and Probability Letters* **35**:423–432.
- Keitt, T. H., O. N. Bjørnstad, P. M. Dixon, and S. Citron-Pousty. 2002. Accounting for spatial pattern when modeling organism–environment interactions. *Ecography* **25**:616–625.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**:1659–1673.
- Legendre, P., M. R. T. Dale, M.-J. Fortin, J. Gurevitch, M. Hohn, and D. Myers. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* **25**:601–615.
- Legendre, P., and E. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**:271–280.
- Legendre, P., and L. Legendre. 1998. Numerical ecology. Second English edition. Elsevier Science, Amsterdam, The Netherlands.
- Lichstein, J. W., T. R. Simons, S. A. Shriner, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* **72**:445–463.
- Norcliffe, G. B. 1969. On the use and limitations of trend surface models. *Canadian Geographer* **13**:338–348.
- Peres-Neto, P. R. 2004. Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. *Oecologia* **140**:352–360.
- Peres-Neto, P. R., P. Legendre, S. Dray, and D. Borcard. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, in press.
- Rao, C. R. 1964. The use and interpretation of principal component analysis in applied research. *Sankhyā Series A* **26**:329–358.
- Richardson, S. 1990. Some remarks on the testing of association between spatial processes. Pages 277–309 in D. Griffith, editor. *Spatial statistics: past, present, and future*. Institute of Mathematical Geography, Ann Arbor, Michigan, USA.
- Ripley, B. 1977. Modeling spatial patterns. *Journal of the Royal Statistical Society B* **39**:172–212.
- Ripley, B. 1988. Statistical inference for spatial processes. Cambridge University Press, Cambridge, UK.
- Scarlett, M. J. 1972. Problems of analysis of spatial distribution. Pages 928–931 in W. P. Adams and F. M. Helleiner, editors. *Congrès international de géographie*. University of Toronto Press, Montreal, Canada.
- Selmi, S., and T. Boulinier. 2001. Ecological biogeography of Southern Ocean islands: the importance of considering spatial issues. *American Naturalist* **158**:426–437.
- Smith, E. 2002. Ecological statistics. Pages 589–602 in A. El-Shaarawi and W. Piergorsch, editors. *Encyclopedia of environmental statistics*. Wiley, Chichester, UK.
- ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**:1167–1179.
- Tiefelsdorf, M., and D. Griffith. 2006. Semi-parametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A*, in press.
- Tognelli, M. F., and D. A. Kelt. 2004. Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. *Ecography* **27**:427–436.
- Wartenberg, D. 1985. Canonical trend surface analysis: a method for describing geographic pattern. *Systematic Zoology* **34**:259–279.

APPENDIX A

Detailed description for calculating a global Moran's *I* and significance test based on a permutation procedure for environmental model residuals based on topology-based and distance-based eigenvector maps (*Ecological Archives* E087-157-A1).

APPENDIX B

Forward selection procedure for eigenvector maps (*Ecological Archives* E087-157-A2).

APPENDIX C

Adjusted redundancy statistic and variation partitioning in redundancy analysis (*Ecological Archives* E087-157-A3).

APPENDIX D

A comparison between eigenvector maps and simultaneous autoregressive models (SAR) using the oribatid mite data set for controlling spatial autocorrelation in regression residuals (*Ecological Archives* E087-157-A4).

SUPPLEMENT 1

Matlab functions for generating topology-based and distance-based eigenvector maps (*Ecological Archives* E087-157-S1).

SUPPLEMENT 2

Matlab function for performing model selection of eigenvector maps (*Ecological Archives* E087-157-S2).

SUPPLEMENT 3

Matlab compiled software to perform the eigenvector selection procedure for generalized linear models (normal, logistic, and Poisson) (*Ecological Archives* E087-157-S3).