

Partitioning variation in ecological communities: do the numbers add up?

Benjamin Gilbert^{1*†} and Joseph R. Bennett^{2†}

¹Marine Sciences Building, Ecology, Evolution & Marine Biology, University of California, Santa Barbara, CA 93106-9610, USA; and ²Centre for Applied Conservation Research, University of British Columbia, Vancouver, BC V6T 1Z4, USA

Summary

1. Statistical tests partitioning community variation into environmental and spatial components have been widely used to test ecological theories and explore the determinants of community structure for applied conservation questions. Despite the wide use of these tests, there is considerable debate about their relative effectiveness.

2. We used simulated communities to evaluate the most commonly employed tests that partition community variation: regression on distance matrices and canonical ordination using a third-order polynomial, principal components of neighbour matrices (PCNM) or Moran's eigenvector maps (MEM) to model spatial components. Each test was evaluated under a variety of realistic sampling scenarios.

3. All tests failed to correctly model spatial and environmental components of variation, and in some cases produced biased estimates of the relative importance of components. Regression on distance matrices under-fit the spatial component, and ordination models consistently under-fit the environmental component. The PCNM and MEM approaches often produced inflated R^2 statistics, apparently as a result of statistical artefacts involving selection of superfluous axes. This problem occurred regardless of the forward-selection technique used.

4. Both sample configuration and the underlying linear model used to analyse species–environment relationships also revealed strong potential to bias results.

5. *Synthesis and applications.* Several common applications of variation partitioning in ecology now appear inappropriate. These potentially include decisions for community conservation based on inferred relative strengths of niche and dispersal processes, inferred community responses to climate change, and numerous additional analyses that depend on precise results from multivariate variation-partitioning techniques. We clarify the appropriate uses of these analyses in research programmes, and outline potential steps to improve them.

Key-words: beta diversity, canonical ordination, Moran's eigenvector maps, principal coordinates of neighbour matrices, redundancy analysis, regression on distance matrices, spatial variation, trend surface analysis, variation partitioning

Introduction

A fundamental challenge in ecology is to understand the determinants of community composition. Traditionally, ecologists relied on environmental niches to explain community variation (e.g. Hutchinson 1957), but the role of spatial dynamics, such as dispersal differences amongst species, has increasingly been recognized (e.g. MacArthur & Wilson 1967; Hanski 1991).

More recently, neutral models of biodiversity have abandoned the role of niches altogether, instead emphasizing the ability of dispersal limitation alone to produce realistic spatial distributions of competitively equivalent species (Hubbell 2001; Chave & Leigh 2002). Parallel to these theoretical advances, statistical tests have been designed to determine the relative importance of environmental heterogeneity and dispersal limitation in structuring communities (Borcard, Legendre & Drapeau 1992; Legendre & Legendre 1998; Borcard & Legendre 2002).

Developments in theory and statistical tests have precipitated dozens of comparative analyses of the influences of environmental heterogeneity and spatial dynamics on species

*Correspondence author. E-mail: bgilbert@msi.ucsb.edu

†Both authors contributed equally to this work.

distributions. For example, Gilbert & Lechowicz (2004) used a sampling regime that removed spatial autocorrelation in the environment sampled, and used species' spatial and environmental correlations to show that the sampled community was inconsistent with neutral predictions. Cottenie (2005) developed a framework in which the relative importance of spatial and environmental correlations was used to infer a range of processes, from neutral to source-sink to environmental sorting. Several studies have taken similar approaches both for testing theory (e.g. Tuomisto, Ruokolainen & Yli-Halla 2003), and for a range of applied ecological questions, from elucidating scales of isolation in metapopulations (Yamanaka *et al.* 2009) to developing predictive models for species distributions under climate scenarios (e.g. Heikkinen *et al.* 2006).

There is, however, much controversy about the relative merits of the statistical options for partitioning spatial and environmental components of variation (e.g. Legendre, Borcard & Peres-Neto 2005, 2008; Tuomisto & Ruokolainen 2006, 2008; Laliberté 2008; Pélissier, Couteron & Dray 2008). This controversy over appropriate statistical methods and their interpretation is important to both theoretical and applied ecology. Indeed, understanding the theoretical underpinnings of meta-communities (Cottenie 2005) and the impacts of environmental changes such as climate warming, land-use change and eutrophication on biodiversity (e.g. McLachlan, Hellmann & Schwartz 2007), depends on correctly identifying the processes that structure species distributions.

In this study, we test the effectiveness of the most commonly used multivariate partitioning techniques, with a particular focus on a recently developed and increasingly used approach: redundancy analysis (RDA) with a principal coordinates of neighbour matrices (PCNM) spatial matrix. While recent papers evaluating these techniques (Laliberté 2008; Legendre, Borcard & Peres-Neto 2008; Pélissier, Couteron & Dray 2008; Tuomisto & Ruokolainen 2008) have focused almost exclusively on modelling spatial patterns to infer dispersal limitation in neutral communities, we took a broader approach by evaluating how well each method performs from modelling both spatial and environmental processes. Our goal was not to further develop or test the mechanics of these methods, but rather to assess how well they represent known causes of species distributions in simple yet realistic communities.

We partitioned variation amongst simulated communities using multiple regression on distance matrices (MRDM) and raw-data approaches (RDA) that differ in both environmental models (linear and eigenvector) and spatial models [trend surface, PCNM and Moran's eigenvector maps (MEM)]. Although previous studies have argued the relative merits of these different methods (e.g. Pelletier, Fyles & Dutilleul 1999; Legendre, Borcard & Peres-Neto 2008; Tuomisto & Ruokolainen 2008; Peres-Neto & Legendre 2010), they are all widely used by ecologists and have yet to be systematically compared. We began by simulating ecological communities with known levels of environmental and spatial control of species distributions, using levels that encompassed *c.* 90% of studies reported in the most recent meta-analysis of environment–space partitioning papers (Cottenie 2005). Species distributions in our

simulations were generated through three distinct processes: response to a spatially autocorrelated environmental gradient, response to a spatially random environmental gradient, and source-sink dispersal. We then used each statistical method to partition the variation explained by environment and space, and compared this to the known fractions explained. Through these analyses, we address the following questions: (i) how accurate is each method at determining both absolute and relative importance of spatial and environmental drivers; (ii) does this accuracy change as the relative importance of each driver changes; and (iii) how sensitive are the statistical methods to the spatial configuration of sampling regime?

Materials and methods

SIMULATION DESIGN

We designed three sets of communities, with each community constructed to represent passive dispersers with distinct levels of environmental control and dispersal effects: one was influenced purely by environment; one had a large environmental influence and small dispersal influence; and one had large dispersal and environmental influences. Each community was created by initially generating a 129×129 cell lattice grid, and assigning values from two environmental gradients to each cell. Each cell, which represents one potential sampling location, was environmentally homogenous internally. The two environmental gradients differed in that one was spatially autocorrelated (hereafter *E_s*) and the other was randomly distributed in space (*E_r*). We began by generating *E_s* to be similar to spatially autocorrelated gradients observed in natural communities (With & King 1997; Gergel & Turner 2002). We employed a two-dimensional random fractal model using the midpoint displacement method (Saupe 1988), keeping the environmental autocorrelation consistent across simulations (fractal dimension = 2.8), and scaled the simulations so that *E_s* was evenly distributed with values ranging from 0 to 1. The scaled environment maintained the patchy pattern typical of spatially autocorrelated environmental gradients (Gergel & Turner 2002; Fig. S1, Supporting information). The second environmental gradient, *E_r*, was also uniformly distributed with values ranging from 0–1, but was randomly distributed in space.

We used the environment to structure species' initial distributions. The initial abundance of each species was modelled using a unimodal distribution along each environmental gradient. In particular, we used the difference between the environmental optima of species *k* (*E_{s_k}*) and the environment in cell *j* (*E_{s_j}*):

$$f(E_s) = e^{-\frac{(E_{s_k} - E_{s_j})^2}{0.02}} \quad \text{eqn 1}$$

The same function was used for *E_r* (replace *E_s* with *E_r*). The initial abundance of species *k* in cell *j* was determined as:

$$\text{Initial}_{kj} = 0.75f(E_s) + 0.25f(E_r) \quad \text{eqn 2}$$

As can be seen in eqn 2, the effect of the two environmental gradients was additive and the gradients were weighted differently. This weighting is arbitrary, but was used to reflect commonly observed differences in the explanatory power of different environmental variables (e.g. Gilbert & Lechowicz 2004). Once the initial abundances of species were determined, we added a dispersal component. We did this by treating initial abundances as fixed, and weighted all cells within the community by a dispersal kernel. The dispersal contribution for species *k* at cell *j* was generated with the following formula:

$$D_{kj} = \sum_{i \neq j} \text{initial}_{ki} \times f(\text{distance}_{ij}), \quad \text{eqn 3}$$

where initial is from eqn 2, and distance is the Euclidean distance between sites i and j . The distance function (f) varied amongst sets of communities, with the first community having no dispersal component, while the second had $f \propto 1/\text{distance}^2$, and the third had $f \propto e^{-\text{distance}/0.5}$. To avoid edge effects, a cell's initial value (eqn 2) only weighted cells that were within 32 units, and only the central 64×64 cells of each simulated community were retained after D was determined.

The D term was then scaled to give a dispersal component (D') that had, on average, the same magnitude as the environmental component, so that $D'_{kj} = D_{kj} \times \sum \text{Initial}_{kj} / \sum D_{kj}$. This dispersal component is qualitatively similar to distribution patterns in source-sink communities of passive dispersers, with the initial environmentally determined abundance dictating the strength of the source, and the distance function (eqn 3) determining the spread of sink populations.

The relative importance of the dispersal component in each community was varied by giving the dispersal component and the environmental component different predictive weights in final species abundances. Final abundances for each species in each cell were Poisson random variates with mean abundances equal to $30(w_s D'_{kj} + w_e \text{Initial}_{kj})$. In addition to the distance functions above, the weighting was varied to generate the three types of communities: one with no dispersal component ($w_s = 0$, $w_e = 1$); one with a minor dispersal component ($w_s = 0.25$, $f \propto 1/\text{distance}^2$, $w_e = 0.75$); and one with a relatively large dispersal component ($w_s = 0.5$, $f \propto e^{-\text{distance}/0.5}$, $w_e = 0.5$). Graphical examples of the dispersal components after weighting are shown in Fig. S2, Supporting information. We then used the known environmental and dispersal components to partition the variation in each community. To determine the variation explained by each variable, we used the standard multivariate analysis (eqn 1 in Peres-Neto *et al.* 2006) with the programmed components (weighted values from eqns 2 and 3) used as the predicted Y values. These programmed components represent the true deterministic trends in the data.

SAMPLING DESIGN AND STATISTICAL ANALYSES

All statistical methods tested use a common framework of partial analyses to partition the variation explained by spatial and environmental components. For detailed descriptions of partial canonical analyses, see Legendre & Legendre (1998). In summary, three matrices are required: the species matrix, the environmental matrix and the spatial matrix. Three analyses are conducted to partition the variation in the species matrix between the environmental and spatial matrices: the first uses both the environmental and spatial matrices as predictors of the species matrix, and represents the total variation explained. The total variation explained is made up of three components: the environment (E), space (S) and the covariation between the environment and space (ES). The second analysis uses only the environmental matrix as a predictor to determine the conditional (non-independent) environmental variation. When the conditional environmental variation is subtracted from the total variation explained, the remaining variation is the independent space effect (S). The third analysis subtracts the conditional variation explained by the spatial matrix from the total variation explained to determine the independent environment effect (E).

All statistical analyses were performed using R v. 2.8.0 (R development core team 2008). For canonical ordinations, we used the 'packfor' library forward-selection procedure designed by Blanchet,

Legendre & Borcard (2008) to reduce Type I error, with each explanatory matrix only subjected to forward selection if a test of the entire matrix was first found to be significant. Species abundances were transformed using the Hellinger transformation (Legendre & Gallagher 2001) (R, vegan library). We also used the 'varpart' function (R, vegan library) to partition variation; this function computes the adjusted canonical R^2 , analogous to the adjusted R^2 in multiple regression (Peres-Neto *et al.* 2006). Environmental variables were not transformed unless otherwise stated. Statistical tests were assessed for their accuracy by comparing the variation explained in each component with the true variation programmed into the simulated communities. Specific methods are explained in detail below.

We evaluated MRDM using transformed matrices by creating a Bray–Curtis distance matrix for each environmental predictor and a Euclidean distance matrix for geographic distances. All predictor matrices were log-transformed as necessary. We also performed standard Mantel tests (Mantel 1967), which are closely related to MRDM, but differ in that environmental variables have identical weightings in a single predictive matrix. Because of this difference, we only report the Mantel results in the Supporting information. Significance of both MRDM and Mantel tests was evaluated by permutation (Legendre & Legendre 1998).

Polynomial Trend Surface Analysis of community data uses polynomial terms of spatial coordinates as variables to model spatial variation in communities within an RDA. The method was first proposed by Legendre (1990), but relies on previous trend surface modelling using polynomial terms (reviewed in Legendre & Legendre 1998; Borcard, Legendre & Drapeau 1992). It involves generating a spatial matrix using the centred geographic coordinates of the sampled area. The spatial matrix consists of nine terms, using a third-order polynomial of the x and y coordinates: $x, y, xy, x^2, y^2, x^2y, xy^2, x^3, y^3$.

The PCNM technique was designed by Borcard & Legendre (2002), Borcard *et al.* (2004) to provide a more flexible method for explaining spatial relationships with species composition than the polynomial method, and as an attempt to account in a systematic way for different scales of spatial dependence. The technique represents the spatial configuration of sample points using principal coordinates of a truncated distance matrix amongst points. The resulting PCNM axes with positive eigenvalues are used as spatial components in variation partitioning, with each axis potentially modelling species clustering at different distances amongst sampling units. We used the R package QuickPCNM v. 7.7-1 (available at <http://www.bio.umontreal.ca/legendre/>) to generate the spatial PCNM axes for this analysis.

The MEM technique was introduced by Dray, Legendre & Peres-Neto (2006) as an improvement to the PCNM method. A full description of the method is given in Dray, Legendre & Peres-Neto (2006). For our analysis, we used the same method to generate axes as for the PCNM, but only retained those axes that were strongly and positively spatially structured based on Moran's I (using the spacemakeR library in R; Dray, Legendre & Peres-Neto 2006).

It should be noted that there is debate over how to deal with degrees of freedom when using eigenvector approaches such as PCNM or MEM. One issue is how to penalize the inclusion of multiple eigenvectors, when several eigenvectors may be modelling a single spatial process. Our approach with the PCNM and MEM methods was to count each eigenvector as a single predictor – this approach is the most conservative in its penalization of degrees of freedom and adjusted R^2 statistics.

In addition to the various spatial models used in RDA, we used one transformation that included environmental variables that we term PCNMe. Species in our simulated communities, and often in

ecological communities generally, have unimodal relationships along environmental gradients. We suspected that the linear regression approach used to fit environment–species relationships in RDA might be a poor model of this phenomenon. While canonical correspondence analysis (CCA) should in theory represent unimodal species/environment relationships well (ter Braak 1985), CCA suffers from poor performance in simulations (Økland 1999) and problems with its underlying distance metric (Legendre & Gallagher 2001). Other options are to transform environmental variables to linearize relationships, or examine more complex relationships directly through testing various transformations at once (e.g. Jones *et al.* 2008). We explored the latter concept through an analysis of PCNM axes from both spatial and environmental variables. In other words, we generated PCNM axes for each of the environmental gradients, using the same criteria that we used for spatial variables. This is similar to other transformations of the environmental gradient, but we felt it may better model unimodal trends than other approaches, such as quadratic transformations, based on results reported in Borcard & Legendre (2002). The steps involved in the analysis are the same as the PCNM analysis, with three resulting matrices (one spatial and one for each environmental gradient).

EFFECT OF SAMPLING REGIME

While it is recognized that both scale and sampling effort may influence results in spatial analyses, especially with nonlinear decay in similarity (Fortin & Dale 2005), there has been no systematic comparison of the relative importance of these influences amongst community variation-partitioning tests. We sampled 500 simulated communities for each community type, using three sampling configurations (contiguous sampling, random sampling and uniform sampling) at two different levels of sampling effort: 64 cells and 256 cells (Supporting information, Fig. S3) for a total of 9000 analyses per statistical method.

ADDITIONAL STATISTICAL ANALYSES

Given a current trend towards using PCNM or MEM as the preferred means of accounting for spatial variation in community data, and in light of results reported below, we decided to further investigate the details of data fitting in these methods. This further exploration was conducted on a single simulated species distributed unimodally along a linear transect, with simulations run to determine the effect of species abundance and unexplained variation on forward selection and overfitting. The basic approach used was similar to that described in Borcard & Legendre (2002) for detecting Gaussian curves, with a deterministic abundance structure programmed into the species distribution and variable levels of unexplained variation also included. Full details of these simulations are given in the Supporting information.

We also tested a new forward-selection procedure proposed by Peres-Neto & Legendre (2010). This method involves testing the significance of each eigenvector for each species if the spatial matrix is globally significant. Because this technique is not yet widely used and was found to exacerbate problems with selecting eigenvectors, we present the methods and results in the Supporting information.

Results

The simulated communities were constructed with known levels of environmental and dispersal effects (from w_{es} , eqns 1 and 2 and D'), which allowed us to separate the actual variation explained by the dispersal component alone, the environmental gradients alone, and the covariation between the environment and dispersal (Table 1). However, unlike our calculations of the true variation explained by each component, the statistical methods used may also model part of the independent environment as ES. This is a recognized behaviour of these statistics (Legendre & Legendre 1998), and occurs because one environmental variable is spatially autocorrelated. To compare between the known variation explained and that found in the models, we therefore compare the S portion from the statistical models with our independent dispersal effect. We also compare the sum of E and ES from the statistical models with the summed contribution of the environmental gradients and the environment-dispersal covariation (Fig. 1, left panel). For these tests, we first use the 256 contiguous sample configuration to compare across methods and communities (Fig. 1). In addition, we assess the accuracy of the tests by comparing the relative contribution of S [i.e. $S/(S + E + ES)$] across models (Fig. 1, right panel).

All techniques exhibited unexpectedly large differences in explained variation from the true variation in the data set (Figs 1 and 2, Supporting information, Fig. S6). In most cases, the amount of explained variation was considerably lower than that programmed into the data set, supporting the conclusions of Økland (1999) and others who have noted that high amounts of unexplained variability may mislead researchers into thinking their results are not meaningful. However, deviations from the true variation explained were not consistent. Instead, the different techniques were often biased in over-representing one component of variation relative to another (Fig. 1).

In addition to differences amongst methods, all analyses showed sensitivity to sampling configuration (Fig. 2). Regardless of the statistical method used, communities sampled using

Table 1. Levels of explained variation in simulated community types

Community type	Number of simulated communities	Independent environment	Independent dispersal	Environment-dispersal covariation
No dispersal component	500	89	0	0
Small dispersal component	500	77	1	7
Large dispersal component	500	45	13	22

Note: The simulated communities were constructed with known levels of environmental determinism (from w_{es} , eqns 1 and 2), and a known scaled dispersal component (D'), allowing us to separate the independent effects used to test variation partitioning techniques.

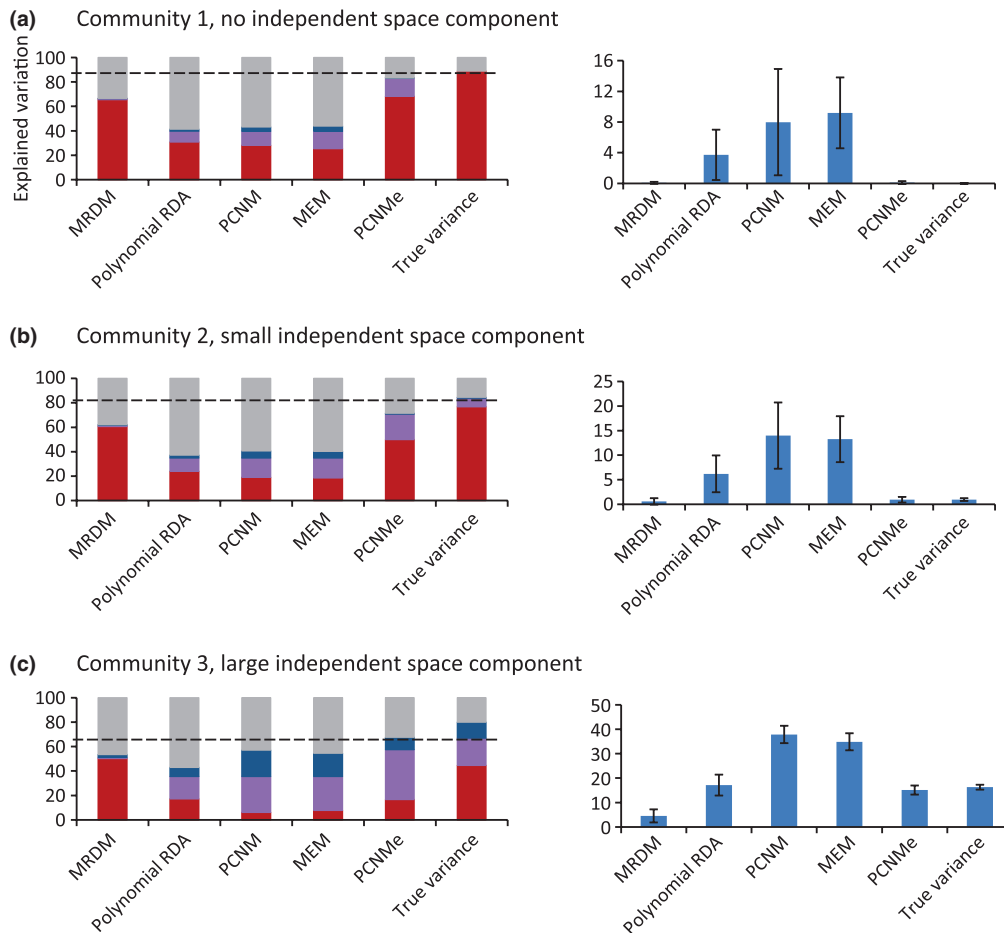


Fig. 1. Comparison of all methods in the 256 contiguous sampling regime (16×16 cells), the regime at which the maximum spatial signal could be detected. The far right bar in all graphs shows the actual variation explained by each component. Red represents the independent environment signal (E); blue represents the independent spatial signal (S); purple represents space–environment covariation (ES); grey represents unexplained variation. Left hand graphs present absolute variation explained. Dashed horizontal lines mark the known combined E and ES. The accuracy of each method is measured by how well the estimates of these two components match the known E and ES components (bar on far right) and also how close in size the S component is to its true value. Right hand graphs present proportional variation explained by S [i.e. $S/(E + S + ES)$]. Error bars show one standard deviation.

contiguous plots consistently produced a higher S component than those sampled with larger distances between cells when there was an independent dispersal effect (Fig. 2). The configuration of spaced plots (uniform vs. random) had little effect, but sampling effort and spacing amongst plots had a large effect, due both to changes in statistical power and to dispersal patterns occurring mainly at relatively fine scales (Supporting information, Fig. S2). For example, in uniform and random sampling regimes, the PCNM and MEM methods modelled the relative importance of space as doubling between samples with 64 cells and those with 256 (Fig. 2). Below, we outline in detail the results for each statistical method tested.

The MRDM technique represented the environmental signal well, and was in fact better at modelling this signal than the linear model used in the RDA techniques (Fig. 1, A5). However, the MRDM was poor at modelling spatial patterns (Fig. 1, S4).

Redundancy analysis using polynomial trend surfaces to model the spatial signal explained less variation than was

programmed into the simulated communities, although there was no consistent bias towards over- or under-representing the relative importance of components (Fig. 1). Instead, the polynomial approach tended to over-represent S when this component was actually low, and under-represent it when it was high. In particular, the polynomial trend surface modelled an independent spatial signal in the community where none existed (Fig. 1), and overfit the independent spatial signal in the community with a small spatial component by nearly double (Fig. 1). This overfitting of the spatial component may not result from a problem with the spatial model per se, but rather a poor fit by the environmental model. In particular, the linear model that is used as a default in RDA provides a very poor fit for most species (Fig. 3). When there is a poor fit between the environment and species matrices, the polynomial trend surface could model species distributions along the spatially structured environment better than linear regression with the environment, and then attribute some of this variation to the inde-

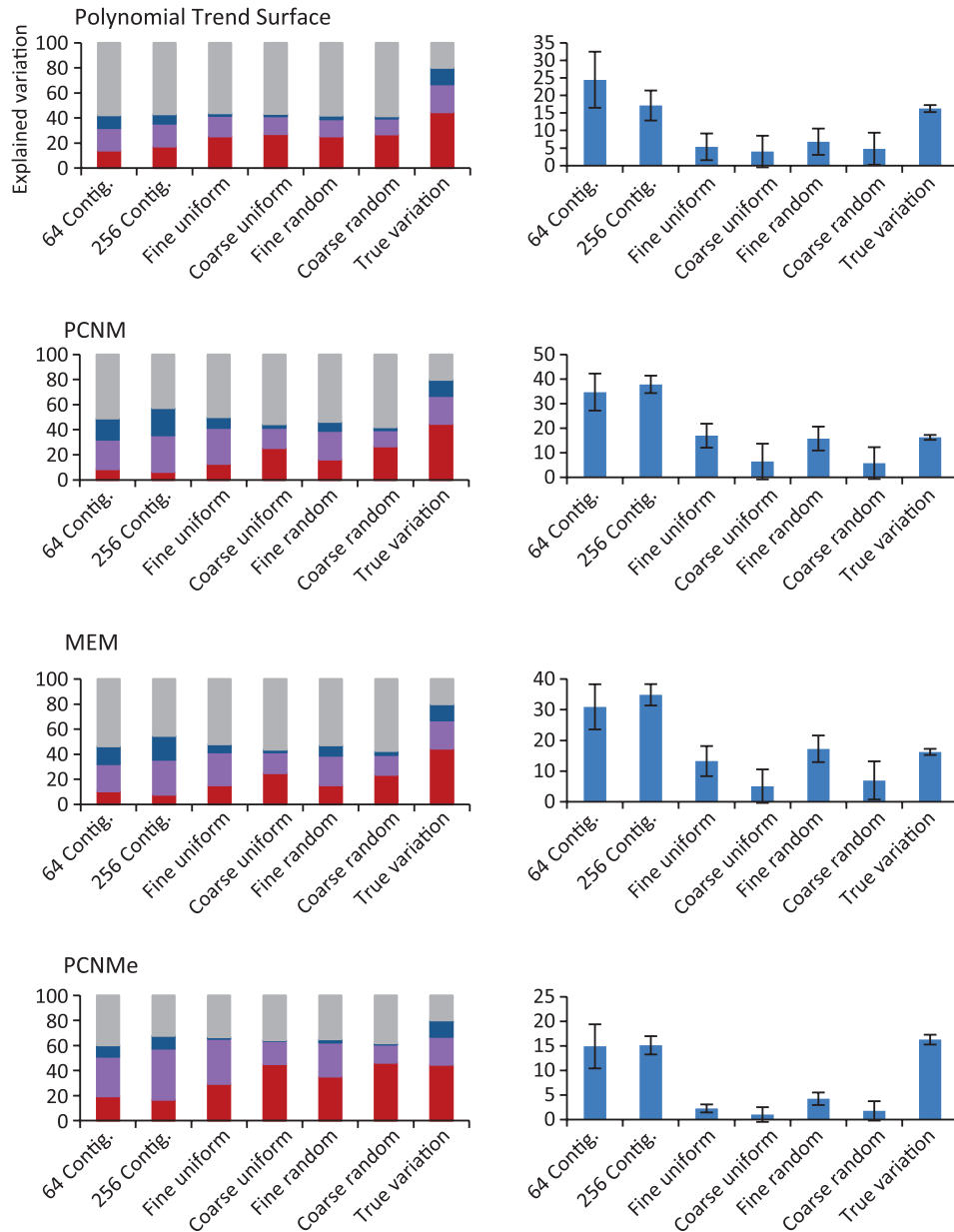


Fig. 2. The effect of the different sampling methods on each of the raw-data (RDA) analyses for Community 3 (large independent spatial signal); segment colours as per Fig. 1. Left hand graphs present absolute variation explained, while right hand graphs present the independent spatial signal as a proportion of explained variation: $S/(E + S + ES)$. For uniform and random configurations, 'fine' refers to analyses with 256 lattice cells sampled and 'coarse' refers to those with 64 lattice cells sampled. 'Contig' refers to contiguously sampled cells (Supporting information, Fig. S3). Distance-based tests were consistently far inferior at modelling spatial patterns, and are not presented here (but are presented in Supporting information, Figs S4 and S5). Error bars show one standard deviation.

pendent spatial component rather than space/environment covariation.

The eigenvector techniques (PCNM and MEM) produced results that were virtually indistinguishable. In the contiguous sample configurations, they predicted greater S than was truly present in all simulated community types (Fig. 1), and also produced biased results for the relative importance of space [i.e. $S/(S + E + ES)$]. This bias favoured the S component, representing it as larger than what was modelled into the simulations (Fig. 1), with PCNM producing slightly larger estimates of S than MEM. The bias in the eigenvector techniques

probably resulted from two processes: a poor modelling of the E component using the linear model (Fig. 3) and overfitting the S component.

Further analysis revealed that using PCNM axes as predictor variables can indeed greatly inflate the variation explained, even when an adjusted R^2 is used (Figs 3 and 4). To understand the accuracy of the PCNM method, we considered its performance using species distributions along two gradients that were uncorrelated with each other. In the simulated communities with no pure spatial component, the environmental gradients (E_s and E_r) were uncorrelated, while in the other sim-

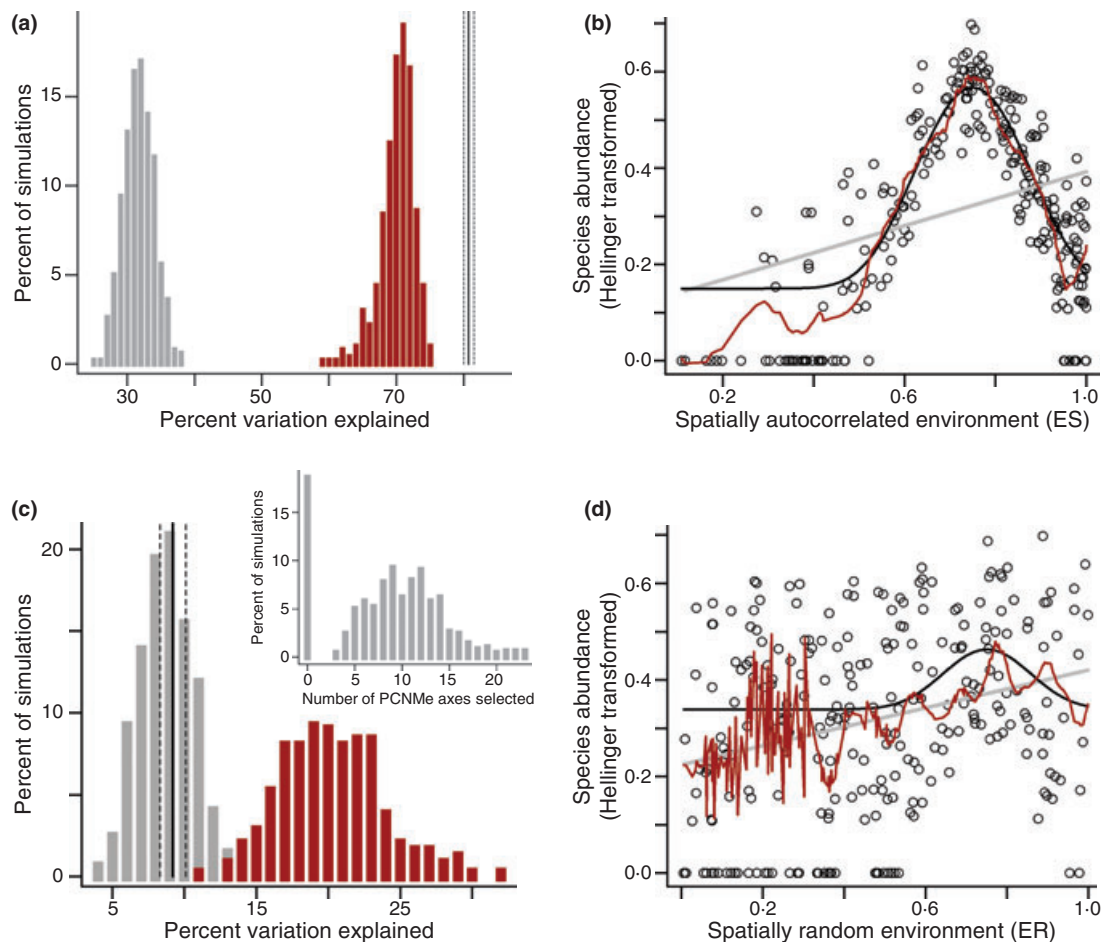


Fig. 3. Variation explained by linear and PCNME models in simulations with no independent space. (a) Percentage variation explained by the spatially autocorrelated environment (ES) using the linear model (grey), vs. the PCNME (red). The true variation explained by the deterministic model is given in black (dashed lines are ± 2 standard deviations). (b) An example of model fits along the Es axis for one species (species 13) in one simulation. Points show the species abundance after Hellinger transformation. The solid black line shows the deterministic trend programmed into the simulations (in this case $R^2 = 0.76$). The grey line is the best fit from the linear model ($R^2 = 0.10$), and the red line is the best fit from the PCNME model for the Es gradient ($R^2 = 0.77$). (c) Percentage variation explained by the spatially random environment (Er) using the linear model (grey), PCNME (red), and the true variation explained (black ± 2 SD). The inset histogram shows the per cent of simulations that selected specific numbers of PCNME axes. (d) Model fits for the same species as in b, but along the Er axis. Points show the species abundance after Hellinger transformation. The solid black line shows the deterministic trend ($R^2 = 0.09$). The grey line is the best fit from the linear model ($R^2 = 0.08$), and the red line is the best fit from the PCNME model ($R^2 = 0.15$).

ulations, only Er was uncorrelated with the other drivers (the spatial function and ES are correlated by definition). Thus, Es and Er were examined for the PCNME in simulations with no pure spatial component (Figs 3 and 4). Examination of the variation explained by these PCNME models (Figs 3b and 4c) illustrated the risk of inflation of variation explained in the PCNM method. In simulations with no pure space component, the PCNME for the ES gradient did a reasonable job at fitting trends; however, the PCNME for Er explained over two times its actual variation explained when PCNME axes were selected. The over-inflation appeared to be partly due to an anomaly that occurs in the selection of both PCNM and MEM axes (Fig. 3c, Supporting information, Fig. S13). The distribution of the number of PCNM and MEM axes selected was bimodal, with one mode at zero (no axes selected) and the other close to 10. The structure of the eigenvectors appears to

make selection of subsequent axes more likely once one has already been selected. Similar bimodal distributions of selected axes were in fact found for all gradients (spatial and environmental) with relatively low explanatory power ($< 10\%$ true variation explained) across all simulations, and occurred despite use of the forward-selection procedure designed by Blanchet, Legendre & Borcard (2008) to correct for Type 1 error. Implementation of a newer forward-selection procedure (Peres-Neto & Legendre 2010) exacerbated this problem (see Supporting information), and we therefore focused our subsequent tests on the Blanchet, Legendre & Borcard method.

Additional analyses using a single simulated species along a linear transect (Fig. 4; Supporting information, Figs S7–S11) indicated that regression using PCNM and MEM axes can cause a positive feedback whereby selecting some axes increases the chance of selecting additional axes, perhaps due

to patterns created in residuals after fitting initial axes (Supporting information, Fig. S9). The increased number of axes selected can in turn inflate the explained variation by large amounts (Fig. 4; Supporting information, Figs S10 and S11). Greater spread of species abundance along a transect and a greater amount of background noise in the data both result in a greater problem with overfit (Fig. 4; Supporting information, Figs S10 and S11). Our simulation results found that the eigenvector approaches (MEM and PCNM) behaved similarly, with the MEM being only slightly more conservative (Supporting information, Fig. S11). These results are consistent with the community simulations (Figs 1 and 2), and suggest that eigenvector analyses are prone to statistical artefacts that inflate the variation explained.

Discussion

Despite recent and intense debates on the relative merits of the methods tested here (Legendre, Borcard & Peres-Neto 2005, 2008; Tuomisto & Ruokolainen 2006, 2008; Laliberté 2008; Pélissier, Couteron & Dray 2008), we found that all methods failed to correctly model the relative importance of environmental and dispersal drivers of community composition (Figs 1 and 2). Indeed, each analysis performed best under limited circumstances that could not be known *a priori* in real ecosystems, and therefore no single technique that we tested can be recommended over the others. From a statistical viewpoint, this failure resulted from improper models of both spatial and environmental signals (Figs 1 and 3). In addition, the challenges associated with selecting an appropriate sampling scale may be equally important as the influence of the statistical method (Fig. 2).

Although we have identified several pitfalls of using incorrect sampling or statistical methods, the questions addressed by studies that partition environmental and spatial controls of community composition are important to both basic and applied ecology (e.g. Jones *et al.* 2006; Legendre *et al.* 2009). Moreover, these questions are likely to increase in importance as anthropogenic environmental change proceeds. The challenge for ecologists is therefore twofold: to determine when and where the current statistical methods may yield reasonable results, and to develop better methods for partitioning spatial and environmental correlates.

STATISTICAL CONSIDERATIONS

Although our results indicated that no method was consistently good at modelling the relative importance of spatial and environmental drivers of community composition, some methods performed better than others. For example, MRDM failed to adequately model spatial patterns in our simulations, which produced strongly biased estimates of the S and ES portion of species distributions in simulations that had a dispersal component. Mantel-type tests have previously been shown to produce lower correlation coefficients than RDA (Dutilleul *et al.* 2000; Legendre, Borcard & Peres-Neto 2005), and variation partitioning using distance matrices has been criticized for failing to represent

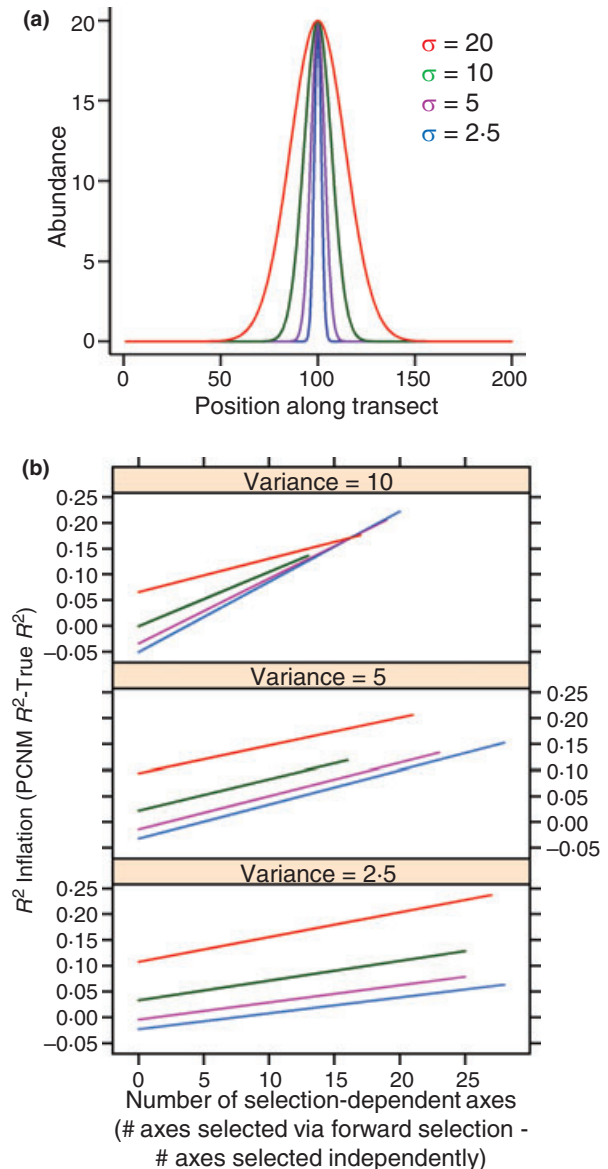


Fig. 4. Overfitting by the PCNM as a function of the amount of unexplained variation in simulations, the abundance of the species and the number of selection-dependent PCNM axes included through the forward-selection procedure. (a) The spread of data points along a linear transect for four different simulation conditions. (b) Exacerbation of R^2 inflation by the PCNM technique as the number of selection-dependent axes (i.e. those selected via forward selection, but not significant when tested independently), increases. The colours show the spread of the species (from a), and each panel shows simulation results from a specific level of unexplained variation in the background noise. Full results with all data points are presented in the Supporting information.

the true variation in species distributions (Legendre, Borcard & Peres-Neto 2005, 2008). Our findings support previous criticisms of Mantel-type methods (Legendre, Borcard & Peres-Neto 2005, 2008). Regardless of the philosophical merits of distance-based or raw-data based methods for testing beta diversity (Legendre, Borcard & Peres-Neto 2005; Tuomisto & Ruokolainen 2006), it is clear that correlations based on distance matrices are inferior to

RDA for modelling spatial patterns. Given that several approaches to species distribution models use similar distance-based approaches to model spatial dependency in species distributions (e.g. Allouche *et al.* 2008), this concern could extend to a range of modelling applications.

While MRDM fails to adequately model spatial patterns, the PCNM and MEM approaches suffer the opposite problem. Indeed, despite the recent popularity of these methods for detecting spatial patterns (e.g. Brind'Amour *et al.* 2005; Soinen & Weckstrom 2009), our results indicate that eigenvector methods can inflate the variation explained by a given causal process (Fig. 1). The over-estimation of variation explained was consistent even in simulations of a single species along a single spatial axis (Fig. 4, Supporting information), and resulted in R^2_{adj} values that were up to 0.5 higher than actual R^2 values, with average R^2_{adj} values inflated by 0–0.25 (Fig. 4, Supporting information). Visual inspection of PCNMe axis fitting indicated that some of this overfitting is due to eigenvector axes accounting for random noise (Fig. 3d). We also found that there is an increased likelihood of selecting eigenvector axes once one axis has been selected, even though the conservative forward-selection procedure developed by Blanchet, Legendre & Borcard (2008) was used. Not surprisingly, the inflated R^2 statistics often occurred in conjunction with additional axes being selected (Fig. 4). The inflation of R^2 statistics and the irregularities in the forward selection of eigenvectors indicate that the PCNM and MEM methods are unstable and vulnerable to statistical artefacts. These issues must be addressed before the results of these analyses can be considered trustworthy.

Despite the problems encountered with the eigenvector techniques, they have a number of desirable aspects (Borcard *et al.* 2004), including decomposition of spatial scales for analysing scale-dependent community structuring. The future challenge for statisticians will therefore be to incorporate these strengths within a more robust approach to generation and selection of spatial variables directly relevant to the ecological questions being posed.

In addition to issues with the spatial models used in multivariate partitioning, our analyses also indicate that the default linear models are insufficient to model community–environment relationships (Fig. 3). Although this problem has been recognized for some time (ter Braak 1986), recent developments in multivariate partitioning have focused almost exclusively on increasing the complexity of the spatial model used (e.g. Borcard & Legendre 2002; Borcard *et al.* 2004; Dray, Legendre & Peres-Neto 2006; Peres-Neto & Legendre 2010). In the application of these methods, complex spatial relationships are almost always tested against single linear terms for environmental variables. This approach is analogous to comparing the R^2 of a linear regression with that of a general additive model. However, even standard transformations of the environmental data may not be sufficient to represent the importance of environmental processes relative to the potentially inflated spatial signal generated by eigenvector analyses. For example, we reran a subset of the analyses using linear and quadratic terms for the environmental variables and eigenvectors for the spatial model (results not shown). While this did

improve the accuracy of the results somewhat, we nonetheless found a 99% ‘false positive’ rate, where simulated communities with no programmed independent spatial signal erroneously showed a statistically significant independent spatial signal.

The problem of finding appropriate transformations for environmental data is often even more critical than is suggested by our simulated communities, which contain simple unimodal responses along two environmental gradients. There is ample literature to indicate that complex environmental influences, including interactions amongst variables, often occur. However, including numerous, complex environmental predictors without *a priori* justification may also lead to problems of overfitting, especially if interactions amongst all variables are considered. More flexible ordination techniques that fit response curves based on maximum likelihood solutions (e.g. Yee 2004, 2006) may offer new avenues for more balanced approaches to modelling the environmental component in variation-partitioning analyses, although these models also pose challenges in terms of model convergence and incorporating complex spatial and environmental models into a single framework. Regardless of the approach taken, we suggest that environmental and spatial models be developed at carefully considered and logically defensible levels of complexity.

CONSIDERATIONS FOR EMPIRICAL STUDIES

Ecologists have used multivariate partitioning of environmental and spatial correlates extensively, with goals that range from contrasting theoretical hypotheses (e.g. Cottenie 2005), to modelling species distributions, (e.g. Karst, Gilbert & Lechowicz 2005; Spiesman & Cumming 2008), to informing management decisions (e.g. Kohler *et al.* 2006; Urban *et al.* 2006). Our results indicate potential problems with such analyses. Comparison of studies using different sampling or statistical techniques is problematic, as the methods we tested exhibit quite different behaviour and can lead to different conclusions, even when analysing the same community. Interpretations of results in many individual studies may also be suspect, especially when ecologists have relied on precise estimates from partitioning analyses. Despite these downfalls, partitioning analyses are nonetheless useful for a number of applications. While we cannot address every possible use of partitioning analyses, our study provides insight into the appropriateness of several applications.

Testing ecological theory directly with environment–space partitioning analyses can be particularly difficult using the methods we critique, because many theories incorporate both a species–environment relationship and a species–space relationship (through species interactions or dispersal, Leibold *et al.* 2004; Snyder & Chesson 2004). However, specific models, such as a spatial neutral model (Chave & Leigh 2002; Condit *et al.* 2002) can be tested directly if other confounding factors are controlled (e.g. space–environment correlation, Gilbert & Lechowicz 2004). Other models that explicitly predict the role of spatial processes, such as neighbour competition, can be tested directly without the need for partitioning analyses (e.g. Fajardo & McIntire 2007).

Several authors have suggested that studies can discriminate amongst theoretical hypotheses based on the relative strength of the environmental and spatial components of a partitioning analysis (e.g. Cottenie 2005; Soininen & Weckstrom 2009). Apart from the concern that many theoretical hypotheses have no specific link to the spatial models used (e.g. PCNM axes and neutral theory; Tuomisto & Ruokolainen 2008), our results indicate that the observed relative importance of different components is sensitive to both sampling and statistical methodology. For example, Cottenie (2005) tested a specific theoretical framework by partitioning variation in numerous data sets that utilized a variety of sampling configurations. Although Cottenie's framework was valuable for advancing theory, our results indicate that both the technique used (polynomial RDA) and the scale of sampling may have led to false conclusions. For example, the theoretical framework tested relied critically on whether an independent spatial component (S) was present; in our simulated communities the same type of analysis found a significant S component 89% of the time even when there was no true dispersal component (Fig. 1a).

Such fundamental problems with partitioning spatial vs. environmental control of communities also have profound implications for applied ecologists. Variation partitioning of community data has been used in numerous and diverse applications. For example, variation partitioning using PCNM to represent spatial signals was used by Kohler *et al.* (2006) to model the type and intensity of cattle effects at different scales, and by Urban *et al.* (2006) to infer a high spatial signal in stream communities, and thus recommend the maintenance of corridors through urban areas. All such applications make the crucial assumption that the sample configuration and analytical technique are adequate to address the issue at hand.

In addition, the theoretical frameworks that have been extensively evaluated using partitioning techniques have well-documented and divergent management ramifications. For example, management according to Neutral Theory implies that distinct characteristics of habitats on a landscape are unimportant relative to habitat area, and that communities may be restored if an adequate number of propagules, rather than specific habitats, are preserved (Daleo, Alberti & Iribarne 2009). Management approaches that account for species-specific niches require that amenable extrinsic conditions are also preserved or restored (Tuomisto, Ruokolainen & Yli-Halla 2003). Ecologists relying on flawed estimates of the relative importance of habitat conditions risk making suboptimal decisions for restoration or maintenance of biodiversity.

Based on these concerns, we suggest that ecologists use two approaches when testing the spatial and environmental drivers of species distributions. First, the appropriate scale of sampling must be determined. Just as our results indicate that certain spatial processes are hard to detect when the sampling design is too coarse (Fig. 2), other spatial and environmental processes are sensitive to the extent of the area sampled (Fortin & Dale 2005). A first step is to consult the considerable literature on spatial sampling that explores trade-offs amongst sampling designs (e.g. Fortin & Dale 2005; Schlup & Wagner 2008), and likewise to examine previous studies on the organisms of con-

cern to determine if they demonstrate sampling-dependent results. Where previous tests are not available, process-based models can be used to predict how sampling designs can capture patterns that would emerge from different ecological processes, thus informing sampling procedures which then test these models.

Second, we suggest that the results from specific studies be tested with independent supporting analyses or experimental studies. For example, Moore & Elmendorf (2006) performed a seed addition experiment to test sampling-generated hypotheses about the degree of microsite and propagule limitation in grassland plants. Similarly, Karst, Gilbert & Lechowicz (2005) used independent surveys to test for consistency in species' environmental niches.

The key message is that studies should not rely on precise differences in variation explained by spatial and environmental correlates to infer the exact proportional strength of each effect. The initial partitioning is a first step, and exploring species distributions with more specific models or experimental tests is critical.

Conclusions

When tested against realistic simulated communities, the most common variation-partitioning models fail to accurately represent environmental and spatial components of community variation. This failure results from poor environmental models, poor spatial models, or both. Despite these problems, there is a growing need for accurate models, both for basic and applied ecology. We suggest that statisticians continue to develop more robust spatial and environmental models of community composition. Multivariate methods for inferring environmental and spatial controls of species distributions are in their infancy, and new developments are likely to improve accuracy. Until such models are developed and tested, we suggest that empirical ecologists use multivariate partitioning analyses mainly as exploratory tools to develop hypotheses about the environmental and spatial determinants of species distributions.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), through a Postdoctoral Fellowship (PDF) to B.G. and a Canada Graduate Scholarship (CGS) to J.R.B. Mark Drever, Justine Karst, Andrew MacDougall, Pedro Peres-Neto and an anonymous referee provided valuable comments.

References

- Allouche, O., Steinitz, O., Rotem, D., Rosenfeld, A. & Kadmon, R. (2008) Incorporating distance constraints into species distribution models. *Journal of Applied Ecology*, **45**, 599–609.
- Blanchet, F.G., Legendre, P. & Borcard, D. (2008) Forward selection of explanatory variables. *Ecology*, **89**, 2623–2632.
- Borcard, D. & Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.
- Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.

- Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H. (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, **85**, 1826–1832.
- ter Braak, C.J.F. (1985) Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics*, **41**, 859–873.
- ter Braak, C.J.F. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**, 1167–1179.
- Brind'Amour, A., Boisclair, D., Legendre, P. & Borcard, D. (2005) Multiscale spatial distribution of a littoral fish community in relation to environmental variables. *Limnology and Oceanography*, **50**, 465–479.
- Chave, J. & Leigh, E.G. (2002) A spatially explicit neutral model of β -diversity in tropical forests. *Theoretical and Population Biology*, **62**, 153–168.
- Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B., Núñez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H.C., Losos, E. & Hubbell, S.P. (2002) Beta-diversity in tropical forest trees. *Science*, **295**, 666–669.
- Cottenie, K. (2005) Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters*, **8**, 1175–1182.
- Daleo, P., Alberti, J. & Iribarne, O. (2009) Biological invasions and the neutral theory. *Diversity and Distributions*, **15**, 547–553.
- Dray, S., Legendre, P. & Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, **196**, 483–493.
- Dutilleul, P., Stockwell, J.D., Frigon, D. & Legendre, P. (2000) The Mantel-Pearson paradox: statistical considerations and ecological implications. *Journal of Agricultural, Biological and Environmental Statistics*, **5**, 131–150.
- Fajardo, A. & McIntire, E.J.B. (2007) Distinguishing microsite and competition processes in tree growth dynamics: an a priori spatial modeling approach. *American Naturalist*, **169**, 647–661.
- Fortin, M. & Dale, M.R.T. (2005) *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press, Cambridge, UK.
- Gergel, S.E. & Turner, M.G. eds. (2002) *Learning Landscape Ecology: A Practical Guide to Concepts & Techniques*. Springer-Verlag, New York, NY, USA.
- Gilbert, B. & 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.
- Hanski, I. (1991) Metapopulation dynamics: brief history and conceptual domain. *Biological Journal of the Linnean Society*, **42**, 3–16.
- Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W. & Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **30**, 1–27.
- Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ, USA.
- Hutchinson, G.E. (1957) Concluding remarks. *Quantitative Biology*, **22**, 415–427.
- Jones, M.M., Tuomisto, H., Borcard, D., Legendre, P., Clark, D.B. & Olivas, P.C. (2008) Explaining variation in tropical plant community composition: influence of environmental and spatial data quality. *Oecologia*, **155**, 593–604.
- Jones, M.M., Tuomisto, H., Clark, D.B. & Olivas, P. (2006) Effects of mesoscale environmental heterogeneity and dispersal limitation on floristic variation in rain forest ferns. *Journal of Ecology*, **94**, 181–195.
- Karst, J., Gilbert, B. & Lechowicz, M.J. (2005) Fern community assembly: the roles of chance and the environment at local and intermediate scales. *Ecology*, **86**, 2473–2486.
- Kohler, F., Gillet, F., Reust, S., Wagner, H.H., Gadallah, F., Gobat, J.-M. & Buttler, A. (2006) Spatial and seasonal patterns of cattle habitat use in a mountain wooded pasture. *Landscape Ecology*, **21**, 281–295.
- Laliberté, E. (2008) Analyzing or explaining beta diversity? Comment. *Ecology*, **89**, 3232–3237.
- Legendre, P. (1990) Quantitative methods & biogeographic analysis. *Evolutionary Biogeography of the Marine Algae of the North Atlantic* (eds D.J. Garbary & R.R. South), pp. 9–34. Springer-Verlag, Berlin.
- Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecology*, **75**, 435–450.
- Legendre, P., Borcard, D. & Peres-Neto, P.R. (2008) Analyzing or explaining beta diversity? Comment. *Ecology*, **89**, 3238–3244.
- Legendre, P. & Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271–280.
- Legendre, P. & Legendre, L. (1998) *Numerical Ecology*, 2nd edn. Elsevier Science, Amsterdam.
- Legendre, P., Mi, X., Ren, H., Ma, K., Yu, M., Sun, I. & He, F. (2009) Partitioning beta diversity in a subtropical broad-leaved forest of China. *Ecology*, **90**, 663–674.
- Leibold, M.A. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601–613.
- Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt, R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M. & Gonzalez, A. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601–613.
- MacArthur, R.H. & Wilson, E.O. (1967) *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ.
- Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.
- McLachlan, J.S., Hellmann, J.J. & Schwartz, M.W. (2007) A framework for debate of assisted migration in an era of climate change. *Conservation Biology*, **21**, 297–302.
- Moore, K.A. & Elmendorf, S.C. (2006) Propagule vs. niche limitation: untangling the mechanisms behind plant species' distributions. *Ecology Letters*, **9**, 797–804.
- Økland, R.H. (1999) On the variation explained by ordination and constrained ordination axes. *Journal of Vegetation Science*, **10**, 131–136.
- Pélissier, R., Couteron, P. & Dray, S. (2008) Analyzing or explaining beta diversity? Comment. *Ecology*, **89**, 3227–3232.
- Pelletier, B., Fyles, J.W. & Dutilleul, P. (1999) Tree species control and spatial structure of forest floor properties in a mixed-species stand. *Écoscience*, **6**, 79–91.
- Peres-Neto, P.R. & Legendre, P. (2010) Estimating and controlling for spatial structure in the study of ecological communities. *Global Ecology and Biogeography*, **19**, 174–184.
- Peres-Neto, P.R., Legendre, P., Dray, S. & Borcard, D. (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, **87**, 2614–2625.
- R Development Core Team. (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Saupe, D. (1988) Algorithms for random fractals. *The Science of Fractal Images* (eds H.O. Peitgen & D. Saupe), pp. 71–113. Springer-Verlag, New York.
- Schlup, B.M. & Wagner, H.H. (2008) Effects of study design and analysis on the spatial community structure detected by multiscale ordination. *Journal of Vegetation Science*, **19**, 621–632.
- Snyder, R. & Chesson, P. (2004) How the spatial scales of dispersal, competition, and environmental heterogeneity interact to affect coexistence. *American Naturalist*, **164**, 633–650.
- Soininen, J. & Weckstrom, J. (2009) Diatom community structure along environmental and spatial gradients in lakes and streams. *Fundamental and Applied Limnology*, **174**, 205–213.
- Spiesman, B.J. & Cumming, G.S. (2008) Communities in context: the influences of multiscale environmental variation on local ant community structure. *Landscape Ecology*, **23**, 313–325.
- Tuomisto, H. & Ruokolainen, K. (2006) Analyzing or explaining beta diversity? Understanding the targets of different methods of analysis. *Ecology*, **87**, 2697–2708.
- Tuomisto, H. & Ruokolainen, K. (2008) Analyzing or explaining beta diversity? Reply. *Ecology*, **89**, 3244–3256.
- Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. (2003) Dispersal, environment and floristic variation of western Amazonian forests. *Science*, **299**, 241–244.
- Urban, M.C., Skelly, D.K., Burchsted, D., Price, W. & Lowry, S. (2006) Stream communities across a rural–urban landscape gradient. *Diversity and Distributions*, **12**, 337–350.
- With, K.A. & King, A.W. (1997) The use and misuse of landscape models in ecology. *Oikos*, **79**, 219–229.
- Yamanaka, T., Tanaka, K., Hamasaki, K., Nakatani, Y., Iwasaki, N., Sprague, D.S. & Bjørnstad, O.N. (2009) Evaluating the relative importance of patch quality and connectivity in a damselfly metapopulation from a one-season survey. *Oikos*, **118**, 67–76.
- Yee, T.W. (2004) A new technique for maximum-likelihood canonical gaussian ordination. *Ecological Monographs*, **74**, 685–701.
- Yee, T.W. (2006) Constrained additive ordination. *Ecology*, **87**, 203–213.

Received 29 April 2009; accepted 20 July 2010
Handling Editor: Marc Cadotte

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Fig. S1. An example of one species' distribution from a simulation of a community with a high space component.

Fig. S2. The effect of the dispersal kernel after weighting.

Fig. S3. Uniform and random sampling configurations used for analyses.

Fig. S4. Variation explained by the basic Mantel test.

Fig. S5. The effect of the different sampling methods on Mantel-type analyses.

Fig. S6. The effect of the different sampling methods on all analyses for the simulated community with no true independent spatial signal.

Fig. S7. Simple linear transect simulation to explore PCNM forward selection.

Fig. S8. Predicted fits for the three independently significant PCNM axes for simulated individual species.

Fig. S9. Distribution of residuals after the first three independently significant PCNM axes had been fit to an individual simulated species.

Fig. S10. Inflated R^2 from PCNM axes ($\text{PCNM } R^2 - \text{True } R^2$) for an individual simulated species.

Fig. S11. Inflated R^2 from MEM axes ($\text{MEM } R^2 - \text{True } R^2$) for an individual simulated species.

Fig. S12. Comparison across community types of MEM method using species-by-species forward selection vs. standard forward selection.

Fig. S13. Histograms showing examples of bimodal selection of MEM spatial eigenvectors using species-by-species forward selection.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.